Children's Facial Trustworthiness Judgments: Agreement and Relationship with Facial Attractiveness

Fengling Ma[1], Fen Xu[1]* and Xianming Luo[2]*

[1] Department of Psychology, Zhejiang Sci-Tech University, Hangzhou, China, [2] School of Environmental Science and Public Health, Wenzhou Medical University, Wenzhou, China

This study examined developmental changes in children's abilities to make trustworthiness judgments based on faces and the relationship between a child's perception of trustworthiness and facial attractiveness. One hundred and one 8-, 10-, and 12-year-olds, along with 37 undergraduates, were asked to judge the trustworthiness of 200 faces. Next, they issued facial attractiveness judgments. The results indicated that children made consistent trustworthiness and attractiveness judgments based on facial appearance, but with-adult and within-age agreement levels of facial judgments increased with age. Additionally, the agreement levels of judgments made by girls were higher than those by boys. Furthermore, the relationship between trustworthiness and attractiveness judgments increased with age, and the relationship between two judgments made by girls was closer than those by boys. These findings suggest that face-based trait judgment ability develops throughout childhood and that, like adults, children may use facial attractiveness as a heuristic cue that signals a stranger's trustworthiness.

Keywords: trustworthiness judgment, attractiveness, face, agreement, school-aged children

INTRODUCTION

People often form first impressions about others based on their physical appearance—they judge the book by its cover. This strategy serves as an efficient and effective strategy to make interpersonal decisions during social interactions when other information is lacking (Todorov et al., 2005; Bar et al., 2006; Willis and Todorov, 2006; Ballew and Todorov, 2007). One commonly made social decision is a trustworthiness judgment.

The literature has indicated that adults use each other's facial appearances to make trustworthiness judgments just 50 ms after seeing another person's face (Willis and Todorov, 2006; Engell et al., 2007; Oosterhof and Todorov, 2008; Todorov et al., 2009), and that amygdala activity increases with the degree of facial untrustworthiness (Winston et al., 2002; Oosterhof and Todorov, 2008; Todorov and Engell, 2008). Furthermore, face-based trustworthiness judgments are high in agreement among adults (Oosterhof and Todorov, 2008; Todorov et al., 2008). This agreement occurs even when the faces in question belong to a race with which adults are entirely unfamiliar (Xu et al., 2012; Birkás et al., 2014). A recent study reports that older and younger adults displayed similar levels of within-age agreement in their impressions of trustworthiness (Zebrowitz et al., 2013). These findings suggest that adults may rely on certain common facial properties to make

trustworthiness judgments (Todorov et al., 2008). Existing work has consistently supported the close relationship between facial trustworthiness and attractiveness judgments (Willis and Todorov, 2006; Oosterhof and Todorov, 2008; Xu et al., 2012). Moreover, Todorov et al. (2009) found that adults indeed perceived more attractive faces to be more trustworthy.

Although a growing body of evidence suggests that adults have evolved into highly sophisticated decoders of trustworthiness based on facial properties alone (Winston et al., 2002; Engell et al., 2007), little research has focused on the development of facial trustworthiness judgment during childhood. Nevertheless, the evidence has demonstrated that children can make social trait judgments (e.g., trustworthiness, dominance, competence, aggressiveness, and attractiveness) based on facial appearances (Keating and Bai, 1986; Antonakis and Dalgas, 2009; Vannatta et al., 2009; Short et al., 2012; Caulfield et al., 2014; Cogsdill and Banaji, 2015; Ewing et al., 2015; Ma et al., 2015). For example, Cogsdill and Banaji (2015) find that 3- to 12-year-old children selected "nice/mean" novel faces in a manner similar to adults. Using the Token Quest paradigm, Ewing et al. (2015) investigate the influence of facial appearances on trust behavior across development. They observed that, like adults, 5- to 10-year-old children selectively invested tokens in trustworthy-looking partners rather than untrustworthy-looking partners. In developmental psychology, most facial trait judgments concentrate on attractiveness. Extensive evidence suggests that from an early age, children exhibit clear preferences for adult-judged attractive vs. unattractive faces (Langlois et al., 1987; Slater et al., 1998; Ramsey et al., 2004). Children also use perceptions of physical attractiveness to make decisions in their social interactions (Cross and Cross, 1971; Dion, 1973; Langlois et al., 1990; Slater et al., 1998; Vannatta et al., 2009; Bascandziev and Harris, 2014).

The ability to infer social traits is a crucial component of social functioning and development. Previous studies have focused primarily on the inference of social traits from facial appearances in infancy and early childhood; little is known about school-aged children. Therefore, we do not know whether the ability to infer social traits based on face improves continuously during childhood. Furthermore, compared with preschoolers, school-aged children are gradually developing independence from their parents and have more opportunities for independent participation in social activities. In fact, we also believed that school-aged children may have some experience similar to judging strangers' trustworthiness from faces. For example, when children get lost, they choose whom to ask the way. The ability to infer trustworthiness based on facial appearances during initial interactions is important for self-protection, social adjustment, and the preservation of wellbeing (Rotenberg et al., 2004, 2005).

Thus, this study aimed to explore the developmental changes in 8- to 12-year-old children's trustworthiness judgments of novel faces. Specifically, we were interested in two questions. The first concerned whether children's trustworthiness judgments based on novel faces were consistent with those of adults (with-adult agreement) and with those of individuals their own age (within-age agreement), and their age-related differences. If children could make consistent judgments, then the second

question further explored whether like adults, children used facial attractiveness as a shortcut or heuristics cue for trustworthiness judgments and their age-related differences.

For the first question, we expected that children's trustworthiness judgments of novel faces were similar to those by adults and by individuals their own age. Existing evidence proves that preschoolers can distinguish trustworthy-looking faces from untrustworthy-looking faces in the same pattern as adults (Caulfield et al., 2014; Cogsdill et al., 2014; Ewing et al., 2015). However, beyond the similar preferences between two extreme faces, we expected age-related differences in the agreement levels of facial judgment during childhood. Consistent with this expectation, evidence showed that although children could make consensus nice or mean judgments in the same manner as adults, this consensus (with-adult agreement) increased with age during ages 3 to 12 (Cogsdill and Banaji, 2015). Similarly, previous studies report that there are developmental changes in the perception of facial attractiveness after infancy and even until puberty (Geldart et al., 1999; Cooper et al., 2006). One possibility is that due to limited face perception ability, there are age variations in discrimination between subtle facial differences during development (Bruce et al., 2000; Mondloch et al., 2003, 2010). Another possibility is that, according to an experience-based explanation (Geldart et al., 1999; Cooper et al., 2006), due to limited face experience, children could not form adult-like trustworthy-looking prototypes. Young children's trust originates from close caregivers, and different individuals have different (and different-looking) significant others, friends, and foes. Therefore, one's face experience most likely influences his/her judgments of novel faces (Smith and DeCoster, 2000; DeBruine, 2002). Given the above evidence, we assumed that agreement levels in children might be weaker than those in adults.

Regarding the second question, existing research finds that trustworthiness judgment is closely associated with facial attractiveness (Buckingham et al., 2006; Willis and Todorov, 2006; Oosterhof and Todorov, 2008; Todorov et al., 2009; Xu et al., 2012). If children's judgment patterns were similar to those of adults, then we expected that children's trustworthiness judgments of novel faces were also related to attractiveness. The "beauty is good" stereotype provides other possible support for this hypothesis. The beauty halo effect leads to systematic human perceptual biases, and considerable research has supported that people with more attractive faces are judged more positively on a host of personality traits (Eagly et al., 1991; Langlois et al., 2000; Zebrowitz and Montepare, 2008). Furthermore, this stereotype emerges early in life (Dion et al., 1972; Langlois and Stephan, 1977; Eagly et al., 1991; Langlois et al., 2000; Ramsey et al., 2004; Bascandziev and Harris, 2014). Dion reported that preschoolers are drawn to attractive children as potential friends and that they exhibit a corresponding dislike for unattractive children (Dion, 1973). Bascandziev and Harris also reported that preschoolers preferred to seek and accept information from more attractive people (Bascandziev and Harris, 2014). During initial interaction, we have little information and time to learn about a novel individual's ability, honesty, and benevolence to make a reliable decision about his/her trustworthiness. Thus, we assumed that

like adults, children may also use facial attractiveness—the readily accessible and useful indicator of social dispositions—as a heuristic cue for signaling a stranger's trustworthiness.

In this study, we used a data-driven statistical model of 3D faces to generate 200 adult male faces with neutral expressions. Children ages 8 to 12 and a comparison group of adults were asked to judge the faces' trustworthiness and attractiveness. The with-adult and within-age agreement of facial judgments and their relationships to facial attractiveness were analyzed.

METHODS

Participants

The total valid sample consisted of 138 participants, including 34 8-year-olds ($M = 8$ years, $SD = 4$ months; 17 boys), 34 10-year-olds ($M = 10$ years, $SD = 5$ months; 17 boys), and 33 12-year-olds ($M = 12$ years, $SD = 4$ months; 17 boys), 37 undergraduates were recruited as a comparison group ($M = 20$ years, $SD = 19$ months; 16 males). Additional 21 participants were not included in the final analyses for the following reasons: nine participants (including three 8-year-olds, two 10-year-olds, and four12-year-olds) did not understand the task or were not serious; 12 participants (including four 8-year-olds, four 10-year-olds, two 12-year-olds, and two adult) did not perform the attractiveness judgment in the second test. The participants were all Chinese, and most of the sample (92%) was Han Chinese. They were from families of mixed socioeconomic backgrounds.

This study was approved by the local ethics committees [Institutional Review Board (IRB)] of the Zhejiang Sci-Tech University. Written informed consent was obtained from the parents and teachers of all the children included in this study.

Face Stimuli

FaceGen Modeler 3.1 (http://facegen.com) was used to generate emotionally neutral faces with direct gazes. In this study, the 200 randomly generated a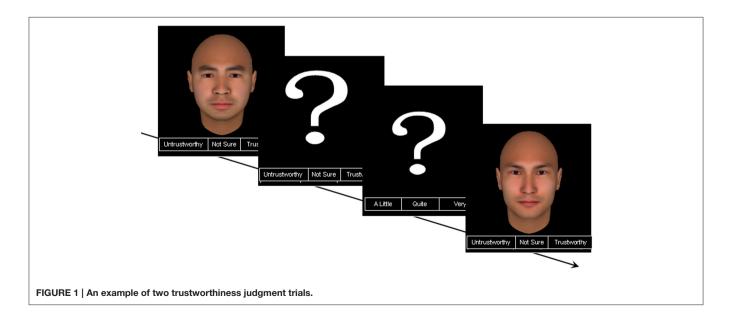dult male faces were all East Asian to avoid cross-race effects. The faces were set to appear to be between 20 and 30 years of age. To avoid the influence of symmetry on trustworthiness and attractiveness judgments, all the faces were set to be symmetrical. These procedures resulted in 200 bitmap face images with a resolution of 400 × 400 pixels (see **Figure 1** for an example).

Procedure

The participants were seated in comfortable chairs, and each faced a 14-inch computer screen. E-Prime software was used to individually display the faces in the middle of the screen, which had a black background. The participants were told that they would be shown a series of faces and that their task was to judge how trustworthy (Session 1) and attractive (Session 2) each face was according to their first impressions. They were told that they would make their decision using a button box. The presentation order of the faces was completely randomized. To avoid participant fatigue, four rest intervals were implemented during the testing. The length of the rest intervals was controlled by the children themselves. After resting well, the children pressed the space key on the keyboard to continue the test procedure.

Session 1: Facial Trustworthiness Judgment

Given the limitations of children's cognitive abilities, the rating procedure was simplified as follows: First, the participants were asked to judge whether the face was trustworthy using a 3-point scale labeled "*untrustworthy / not sure / trustworthy*" (or "*trustworthy / not sure / untrustworthy*," counterbalanced among participants, and the "*not sure*" label was interpreted as "*neither untrustworthy nor trustworthy*"). If the participants rated the face "*untrustworthy*" or "*trustworthy*," then they were asked to rate the degree of untrustworthiness or trustworthiness on a different 3-point scale labeled "*a little /quite / very*" (or "*very / quite /a little*," counterbalanced among participants). Thus, the rating scale

FIGURE 1 | An example of two trustworthiness judgment trials.

ranged from –3 ("*very untrustworthy*") to 3 ("*very trustworthy*"), with 0 indicating "*neither untrustworthy nor trustworthy.*"

Each face was presented in the middle of the screen for 3000 ms (see **Figure 1**). The rating scale appeared under the face image. There was no time limit for the trials; a trial was terminated when the participant made a choice by pressing the preset buttons on the keyboard. When a face was initially rated "*untrustworthy*" or "*trustworthy*," a second screen with a question mark appeared, and the participants rated the degree of the face's trustworthiness or untrustworthiness. However, if the participants were "*not sure*" about the face's trustworthiness (or lack thereof), the trial ended immediately, and the participants judged the next face.

First, six practice trials were conducted with novel faces that were not shown during the formal test. All the participants responded systematically during the practice trials and demonstrated that they understood the task. Then, the experimental trials began. The entire rating process lasted approximately 20 min.

Session 2: Facial Attractiveness Judgment

The procedure for the attractiveness judgment was similar to that used for the trustworthiness judgment. The participants were instructed to judge facial attractiveness using a 3-point scale with "*unattractive / not sure / attractive*" (or "*attractive / not sure / unattractive*," counterbalanced among participants, and the label "*not sure*" was interpreted as "*neither unattractive nor attractive*"); then, the participants continued to rate the degree of unattractiveness or attractiveness on a different 3-point scale with "*a little / quite / very*" (or "*very / quite / a little*," counterbalanced among participants). Thus, combining the two refined scales into a single overall rating scale created a score that ranged from –3 ("*very unattractive*") to 3 ("*very attractive*"), with 0 indicating "*neither unattractive nor attractive.*"

All the participants completed the trustworthiness judgments in the first session and the attractiveness judgments in the second session to avoid the influence of the "beauty-is-good" stereotype. In addition, to avoid the possible influence of trustworthiness judgments on attractiveness judgments, all the participants completed the two sessions at intervals of 1 month or more.

RESULTS

Within each age group, we computed the inter-face reliabilities for trustworthiness and attractiveness judgments for children and adults separately. *Cronbach's αlphas* ranged from 0.94 to 0.97, indicating that both children and adults showed high reliability in their judgments.

Agreement in Facial Judgments
With-Adult Agreement in Facial Judgments

To test whether children's trustworthiness and attractiveness judgments were similar to those of adults, we separately correlated each child's ratings of trustworthiness and attractiveness with the mean of all adults' ratings. The correlations for each child were Fisher Z-transformed to normalize them (Franklin and Adams, 2009; Zebrowitz et al.,

2013). These correlations provided indices of with-adult agreement for each participant (see **Table 1**). *T*-tests (comparing means with zero) indicated that all with-adult agreement means were significantly greater than zero ($p < 0.001$), suggesting that overall patterns of trustworthiness and attractiveness judgments among child participants were similar to those among adults. To examine the with-adult agreement differences in the facial judgment type, age, and gender, a 2 (judgment type: trustworthiness, attractiveness) × 2 (gender: boy, girl) × 3 (age group: 8-year-olds, 10-year-olds, 12-year-olds) repeated-measures ANOVA was performed, with judgment type as within-subject variable. The results showed a main effect of judgment type, $F_{(1, 95)} = 15.54$, $p < 0.001$, partial $\eta^2 = 0.14$, suggesting with-adult agreement of attractiveness judgment being significantly higher than that of trustworthiness judgment (for trustworthiness, $M = 0.18$, $SD = 0.13$; for attractiveness, $M = 0.24$, $SD = 0.15$). The main effect of gender was significant, $F_{(1, 95)} = 9.94$, $p < 0.01$, partial $\eta^2 = 0.10$, suggesting that with-adult agreement of girls was significantly higher than that of boys (for girls, $M = 0.26$, $SD = 0.16$; for boys, $M = 0.18$, $SD = 0.13$). Additionally, the age-related differences were significant, $F_{(2, 95)} = 5.67$, $p < 0.01$, partial $\eta^2 = 0.11$. *Post hoc* (LSD) tests indicated that the agreement for facial judgment in the 8-year-old group was significantly lower than that in the 10- ($p < 0.05$) and 12-year-old groups ($p < 0.01$; for the 8-year-old group, $M = 0.16$, $SD = 0.13$; for the 10-year-old group, $M = 0.22$, $SD = 0.14$; for the 12-year-old group, $M = 0.25$, $SD = 0.16$). No significant interactions were found.

Within-Age Agreement in Facial Judgments

To examine the children's and adults' agreement with participants in the same age group in judging trustworthiness and attractiveness, we correlated each participant's ratings with the mean of all the other individuals in his/her age group, excluding the participant. The correlations for each child were Fisher Z-transformed to normalize them (Franklin and Adams, 2009; Zebrowitz et al., 2013, **Table 2**). These correlations provided indices of within-age agreement for each participant. *T*-tests (comparing means with zero) indicated that all within-age agreement means were significantly greater than zero ($p < 0.001$), suggesting that the overall pattern of judgment in each participant was similar to that of his/her age-mates. To examine the within-age agreement differences in the facial judgment type, age, and gender, a 2 (judgment type: trustworthiness, attractiveness) × 2 (gender: boy, girl) × 3 (age group: 8-year-olds, 10-year-olds, 12-year-olds) repeated-measures ANOVA was performed, with judgment type as

TABLE 1 | Descriptive statistics for the with-adult agreements (Z-score) for facial trustworthiness and attractiveness judgments in each group of children ($M \pm SD$)

	8-year-olds ($n = 34$)	10-year-olds ($n = 34$)	12-year-olds ($n = 33$)
Trustworthiness judgment	0.14 ± 0.12	0.22 ± 0.14	0.22 ± 0.14
Attractiveness judgment	0.19 ± 0.14	0.25 ± 0.15	0.30 ± 0.17

TABLE 2 | Descriptive statistics for the within-age agreements (Z-score) for facial judgments in each age group (M ± SD)

	8-year-olds (n = 34)	10-year-olds (n = 34)	12-year-olds (n = 33)	Adults (n = 37)
Trustworthiness judgment	0.12 ± 0.09	0.20 ± 0.12	0.19 ± 0.12	0.44 ± 0.15
Attractiveness judgment	0.19 ± 0.13	0.22 ± 0.14	0.27 ± 0.16	0.60 ± 0.15

TABLE 3 | The Zr transformed from Pearson Correlation (r) between trustworthiness and attractiveness judgments in each age group for boys and girls.

	8-year-olds	10-year-olds	12-year-olds	Adults
Boys	0.27	0.34	0.42	0.95
Girls	0.56	0.66	0.63	1.10
Z (two-tailed test)	2.89*	3.18*	2.08*	1.49

*$p < 0.05$ (two-tailed test).

TABLE 4 | The results of Z-tests between any two age groups separately for boys and girls on the correlation of trustworthiness and attractiveness judgments.

	Boys	Girls
Differences between 8-year-olds and adults	6.75*	5.40*
Differences between 10-year-olds and adults	6.05*	4.37*
Differences between 12-year-olds and adults	5.26*	4.66*
Differences between 8- and 10-year-olds	0.69	0.99
Differences between 8 and 12-year-olds	1.49	0.69
Differences between 10-year-olds and 12-year-olds	0.79	0.30

*$p < 0.05$ (two-tailed test).

within-subject variable. The results showed a main effect of judgment type, $F_{(1, 130)} = 42.21$, $p < 0.001$, partial $\eta^2 = 0.25$, suggesting that within-age agreement of attractiveness judgment was significantly higher than that of trustworthiness judgment (for trustworthiness, $M = 0.24$, $SD = 0.17$; for attractiveness, $M = 0.33$, $SD = 0.22$). The main effect of gender was significant, $F_{(1, 130)} = 16.05$, $p < 0.001$, partial $\eta^2 = 0.11$, suggesting that within-age agreement of girls was significantly higher than that of boys (for girls, $M = 0.33$, $SD = 0.22$; for boys, $M = 0.24$, $SD = 0.18$). The age-related differences were significant, $F_{(3, 130)} = 78.59$, $p < 0.001$, partial $\eta^2 = 0.66$. *Post hoc* (LSD) tests revealed that the within-age agreements in the three child groups were significantly lower than those of the adult group ($p < 0.001$), and within-age agreement in the 8-year-old group was significantly lower than those of the 10- and 12-year-old groups ($p < 0.05$). Additionally, there was a significant two-way interaction between judgment type and age group, $F_{(3, 130)} = 5.11$, $p < 0.01$, partial $\eta^2 = 0.11$. Simple effect analysis revealed that only in the 10-year-old group was there no significant difference in the within-age agreements between trustworthiness and attractiveness judgments; in the other three age groups, the within-age agreement of attractiveness was significantly higher than that of trustworthiness.

The Relationships between Facial Trustworthiness and Attractiveness Judgments

Based on the gender effect in the agreement of facial judgments, we separately analyzed the relationship between the two judgments in boys and girls. We averaged the facial judgments of boys and girls to obtain the mean trustworthiness and attractiveness score for each face in each age group. Paired *T*-tests showed that the trustworthiness and attractiveness judgments were significantly different from each other in each age group ($p < 0.001$).

We conducted Pearson correlation analyses to examine whether the facial trustworthiness scores were correlated with the attractiveness scores, with the adults as the comparison group. The results revealed that these two scores were highly correlated with each other in each age group for both boys and girls ($p < 0.001$). We used the Fisher transformation to transform r to Z-scores (**Table 3**) and then subjected the Z-scores to Z-tests. Four Z-tests were performed to test the gender effect in each age group. The results indicated that the relationships between the two judgments of girls were significantly closer than those of boys in all the child age groups; however, there was no significant gender

effect in the adult group. Then, we computed six Z-tests between any two age groups to test the age effect. The results showed that the relationships between trustworthiness and attractiveness in all the child age groups were significantly different from that of the adult group; however, the differences between any two child age groups were not significant (**Table 4**).The above differences were similar in the boys and girls.

DISCUSSION

In this study, we investigated the development of facial trustworthiness judgments of made by 8- to 12-year-old children. The participants were asked to rate trustworthiness and attractiveness based on facial appearances. In general, the findings revealed that school-aged children showed consensus in trustworthiness and attractiveness impressions within their own age groups, and they rated the two impressions in a manner similar to that of adults. Some age-related differences were found: as expected, with-adult and within-age agreements of facial judgments increased with age. We also found that facial attractiveness judgments were more consistent than trustworthiness judgments across all age groups. Additionally, we found that girls made facial judgments more consistently with adults and with individuals their own age than boys. Further exploration of the relationships between facial trustworthiness and attractiveness judgments showed that, like adults, close relationships existed between the two facial judgments during childhood, especially for girls; additionally, some developmental changes existed.

Using the same rating procedure as that used for adults, children aged 8–12 were asked to rate randomly generated novel faces on a revised 7-point Likert scale. Overall, children were

able to form similar trustworthiness judgments based on facial appearances across individuals. This finding was consistent with existing evidence (Caulfield et al., 2014; Cogsdill et al., 2014; Cogsdill and Banaji, 2015; Ewing et al., 2015). Prior research that used a forced-choice paradigm found that, as early as preschool age, children could distinguish traits (e.g., trustworthiness or untrustworthiness, competence or incompetence, dominance, or submissiveness) from novel faces in the same way that adults do (Keating and Bai, 1986; Antonakis and Dalgas, 2009; Cogsdill et al., 2014).

As expected, our work provides some evidence that facial trustworthiness judgment abilities gradually improve during childhood. Increasing with age, children's trustworthiness judgments were more consistent with those of adults; additionally, the same developmental difference was found—that the consistency of children's judgments with those of individuals their own age increased with age during childhood; even 12-year-old children did not reach an adult level. One explanation for the developmental changes in trustworthiness judgments may be children's immature face-processing abilities. The research suggests that adult-like expertise in face-based judgments is slow to develop from late childhood through adulthood (Bruce et al., 2000; Mondloch et al., 2003). Furthermore, related brain areas that aid in facial trustworthiness judgments, such as the amygdala, continue to develop throughout late childhood and adolescence and show corresponding functional differences (Giedd et al., 1996; Thomas et al., 2007; Todorov et al., 2008; Baron et al., 2011). Another plausible explanation is based on facial experience. Young children's facial experience comes from close caregivers, and different individuals have different (and different-looking) significant others, friends, and foes. This face environment can influence one's trustworthiness judgments of novel faces (Jones and Hill, 1993; Smith and DeCoster, 2000; Elfenbein and Ambady, 2003; Bronstad and Russell, 2007; Zebrowitz et al., 2012).

With respect to facial attractiveness judgments, similar developmental patterns were found. Children could form consensual attractiveness judgments from novel faces; however, the agreement levels of judgments continued to increase with age. These results were consistent with existing findings, suggesting that children made attractiveness judgments similar to those of adults (Dion, 1973). Our findings provide a more comprehensive overview of the development of attractiveness judgments during childhood.

More importantly, children's agreement levels for attractiveness judgments were higher than those for their trustworthiness judgments. These findings confirmed that the ability to perform facial attractiveness judgments appears early in life, and it can shape children's social decision making. As expected, children's trustworthiness judgments were closely related to facial attractiveness judgments. The results were consistent with existing evidence finding that facial cues signaling trustworthiness overlapped with features that drive attractiveness ratings (Willis and Todorov, 2006; Oosterhof and Todorov, 2008; Todorov et al., 2009; Xu et al., 2012; Ma et al., 2015). For example, Xu et al. (2012) reported that, regardless of their facial experience, Caucasian and Chinese participants

used brow ridge (high or low), cheekbones (shallow or pronounced), and face (heavy or light) for facial trustworthiness and attractiveness judgments. Similarly, Ma et al. (2015) also found that children aged 8–12 used the same facial features (such as the shape of the brow ridge, chin and nose) for both trustworthiness and attractiveness judgments in a similar pattern to adults. Another possibility for the close relationship may be that a more attractive face biases people to rate a person as more positive along many dimensions (such as trustworthiness). That is the "beauty is good" stereotype, which appears at a very early age. Even newborn babies display preferences for adult-judged attractive faces, and 1-year-old babies attribute positive behaviors and traits to attractive people and select attractive individuals more than unattractive ones as playmates (Dion, 1973; Langlois et al., 1990; Slater et al., 1998). When encountering unfamiliar individuals, attractiveness may first act as a simple, efficient facial cue for trustworthiness judgments and then induce subsequent approach-or-avoid actions.

We also found that the relationships between trustworthiness and attractiveness judgments in children groups were significantly weaker than that of the adult group. One explanation for this age effect may be that, facial experience and social experience might serve to refine this link between facial attractiveness and trustworthiness. Unlike adults, children participants are more likely to use unique standards (such as faces that resemble their own or the "look" of an important person) to judge facial trustworthiness rather than the shared standards used across other raters. As facial experience increases, common facial cues may play a substantial role in the trustworthiness and attractiveness judgments, as in adults (Cooper et al., 2006). Additionally, consistent with the social learning mechanism, the "beauty is good" impression may gradually develop through children's daily experiences in witnessing the association between attractive individuals and trustworthy behaviors. As Smith and DeCoster proposed, learning about this association takes a long time and requires ample experience (Smith and DeCoster, 2000). Another compounding factor for this age effect may be memory ability differences between children and adults. During the same intervals between facial trustworthiness and attractiveness judgments, adults may have more memory of these stimuli faces, so the degree of correlation between the two judgments was higher than those of children. Further work is needed to verify these findings.

Additional interesting finding was gender effects. Girls made trustworthiness and attractiveness judgments more consistently across individuals in both child and adult groups. Additionally, the relationships between two judgments made by girls were closer than those by boys. These results suggested that girls were more sophisticated at decoding facial trustworthiness and more dependent on facial attractiveness as a heuristic cue to decide who was trustworthy. To our knowledge, no research has reported the stable gender effects on the facial trustworthiness judgment. One plausible explanation was facial trait inference created by variations in facial expression (Zebrowitz and Montepare, 2008), and women have a significant advantage over men in interpreting facial expression information (Judith, 1978). Existing literature

has examined the differences in the attractiveness ratings of female and male faces by female and male raters; however, the results were mixed (Wernick and Manaster, 1984; Furnham et al., 2004; Foos and Clark, 2011). Some research proposed that the gender differences in trait judgments may be unstable (Foos and Clark, 2011). In the present study, one possible explanation for the female's advantage may be that all the stimuli were male faces, which might naturally elicit stronger attractiveness judgments from heterosexual females. These gender effects need to be more carefully examined in further research.

Several important and interesting questions were further explored, as follows: First, prior work focused primarily on the consensus or agreement among trustworthiness judgments based on novel faces; little is known about the sources of idiosyncratic variation in facial judgments, particularly in children. For example, adults perceive self-resembling faces as more trustworthy and show more trusting behaviors in economic games with self-resembling patterns (DeBruine, 2002; Krupp et al., 2008). Willis and colleagues have reported that individual differences in trait anxiety are associated with facial trustworthiness judgments (Willis et al., 2013), and facial expressions play an important role in trustworthiness judgments (Willis et al., 2011a,b). An important question for future research concerns the influential factors of trustworthiness judgments based on novel faces during childhood. Second, the gender effect on the trait judgments is another interesting question to explore further. In the present work, we found that women were better at decoding the trustworthy trait based on novel faces than men. From an evolutionary perspective, in social interaction, men may be more likely to value others' competence, while women may be more likely to focus on others' warmth. Then, in the next step, we will investigate the gender effect on other trait judgments (e.g., dominance, competence) based on faces.

This study addressed a developmental question concerning trustworthiness judgments and their relationship with facial attractiveness. Like adults, children as young as 8 years old can make consistent trustworthiness judgments and attractiveness judgments based on facial appearances. Our results clearly demonstrated that age-related differences exist in the trait judgments of faces; that is, the with-adult and within-age agreement levels of facial trustworthiness and facial attractiveness increased with age during childhood. Furthermore, children's trustworthiness judgments were closely related to facial attractiveness, and this relationship increased with age. We also found that girls were more sophisticated at facial judgments than boys. Overall, our findings provide a more comprehensive overview of the development of trustworthiness judgment and add to the recently growing body of work claiming that attractiveness is a universal facial cue for trustworthiness judgments during childhood.

AUTHOR CONTRIBUTIONS

FM and FX conceived and designed the experiments; FM and XL performed the experiments; FM and XL analyzed the data; and FM, FX, and XL wrote the paper. All the listed authors have approved the final version.

FUNDING

ACKNOWLEDGMENTS

REFERENCES

Antonakis, J., and Dalgas, O. (2009). Predicting elections: child's play! *Science* 323, 1183. doi: 10.1126/science.1167748

Ballew, C. C., and Todorov, A. (2007). Predicting political elections from rapid and unreflective face judgements. *Proc. Natl. Acad. Sci. U.S.A.* 104, 17948–17953. doi: 10.1073/pnas.0705435104

Bar, M., Neta, M., and Linz, H. (2006). Very first *impressions. Emotion* 6, 269–278. doi: 10.1037/1528-3542.6.2.269

Baron, S. G., Gobbini, M. I., Engell, A. D., and Todorov, A. (2011). Amygdala and dorsomedial prefrontal cortex responses to appearance-based and behavior-based person impressions. *Soc. Cogn. Affect. Neurosci.* 6, 572–581. doi: 10.1093/scan/nsq086

Bascandziev, I., and Harris, P. L. (2014). In beauty we trust: children prefer information from more attractive informants. *Br. J. Dev. Psychol.* 32, 94–99. doi: 10.1111/bjdp.12022

Birkás, B., Dzhelyova, M., Lábadi, B., Bereczkei, T., and Perrett, D. I. (2014). Cross-cultural perception of trustworthiness: the effect of ethnicity features on evaluation of faces' observed trustworthiness across four samples. *Pers. Individ. Dif.* 69, 56–61. doi: 10.1016/j.paid.2014.05.012

Bronstad, P. M., and Russell, R. (2007). Beauty is in the "we" of the beholder: greater agreement on facial attractiveness among close relations. *Perception* 36, 1674–1681. doi: 10.1068/p5793

Bruce, V., Campbell, R. N., Doherty-Sneddon, G., Langton, S., McAuley, S., and Wright, R. (2000). Testing face processing skills in children. *Br. J. Dev. Psychol.* 18, 319–333. doi: 10.1348/026151000165715

Buckingham, G., DeBruine, L. M., Little, A. C., Welling, L. L. M., Conway, C. A., Tiddeman, B. P., et al. (2006). Visual adaptation to masculine and feminine faces influences generalized preferences and perceptions of trustworthiness. *Evol. Hum. Behav.* 27, 381–389. doi: 10.1016/j.evolhumbehav.2006.03.001

Caulfield, F., Ewing, L., Burton, N., Avard, E., and Rhodes, G. (2014). Facial trustworthiness judgements in children with ASD are modulated by happy and angry emotional cues. *PLoS ONE* 9:e97644. doi: 10.1371/journal.pone.0097644

Cogsdill, E. J., and Banaji, M. R. (2015). Face-trait inferences show robust child-adult agreement: evidence from three types of faces. *J. Exp. Soc. Psychol.* 60, 150–156. doi: 10.1016/j.jesp.2015.05.007

Cogsdill, E. J., Todorov, A. T., Spelke, E. S., and Banaji, M. R. (2014). Inferring character from faces: a developmental study. *Psychol. Sci.* 25, 1132–1139. doi: 10.1177/0956797614523297

Cooper, P. A., Geldart, S. S., Mondloch, C. J., and Maurer, D. (2006). Developmental changes in perceptions of attractiveness: a role of experience? *Dev. Sci.* 9, 530–543. doi: 10.1111/j.1467-7687.2006.00520.x

Cross, J. F., and Cross, J. (1971). Age, sex, race, and the perception of facial beauty. *Dev. Psychol.* 5, 433–439. doi: 10.1037/h0031591

DeBruine, L. M. (2002). Facial resemblance enhances trust. *Proc. Biol. Sci.* 269, 1307–1312. doi: 10.1098/rspb.2002.2034

Dion, K., Berscheid, E., and Walster, E. (1972). What is beautiful is good. *J. Pers. Soc. Psychol.* 24, 285–290. doi: 10.1037/h0033731

Dion, K. K. (1973). Young children's stereotyping of facial attractiveness. *Dev. Psychol.* 9, 183–188. doi: 10.1037/h0035083

Eagly, A. H., Ashmore, R. D., Makhijani, M. G., and Longo, L. C. (1991). What is beautiful is good, but…: a meta-analytic review of research on the physical attractiveness stereotype. *Psychol. Bull.* 110, 109–128. doi: 10.1037/0033-2909.110.1.109

Elfenbein, H. A., and Ambady, N. (2003). When familiarity breeds accuracy: cultural exposure and facial emotion recognition. *J. Pers. Soc. Psychol.* 85, 276–290. doi: 10.1037/0022-3514.85.2.276

Engell, A. D., Haxby, J. V., and Todorov, A. (2007). Implicit trustworthiness decisions: automatic coding of face properties in the human amygdala. *J. Cogn. Neurosci.* 19, 1508–1519. doi: 10.1162/jocn.2007.19.9.1508

Ewing, L., Caulfield, F., Read, A., and Rhodes, G. (2015). Perceived trustworthiness of faces drives trust behaviour in children. *Dev. Sci.* 18, 327–334. doi: 10.1111/desc.12218

Foos, P. W., and Clark, M. C. (2011). Adult age and gender differences in perceptions of facial attractiveness: beauty is in the eye of the older beholder. *J. Genet. Psychol.* 172, 162–175. doi: 10.1080/00221325.2010.526154

Franklin, R. G. Jr., and Adams, R. B. Jr. (2009). A dual-process account of female facial attractiveness preferences: sexual and nonsexual routes. *J. Exp. Soc. Psychol.* 45, 1156–1159. doi: 10.1016/j.jesp.2009.06.014

Furnham, A., Mistry, D., and McClelland, A. (2004). The influence of age of the face and the waist to hip ratio on judgements of female attractiveness and traits. *Pers. Individ. Dif.* 36, 1171–1185. doi: 10.1016/S0191-8869(03)00209-5

Geldart, S., Maurer, D., and Henderson, H. (1999). Effects of the height of the internal features of faces on adults' aesthetic ratings and 5-month-olds' looking times. *Perception* 28, 839–850. doi: 10.1068/p2943

Giedd, J. N., Vaituzis, A. C., Hamburger, S. D., Lange, N., Rajapakse, J. C., Kaysen, D., et al. (1996). Quantitative MRI of the temporal lobe, amygdala, and hippocampus in normal human development. Ages 4–18 years. *J. Comp. Neurol.* 366, 223–230.

Jones, D., and Hill, K. (1993). Criteria of facial attractiveness in five populations. *Hum. Nat.* 4, 271–296. doi: 10.1007/BF02692202

Judith, H. A. (1978). Women gender effects in decoding nonverbal cues. *Psychol. Bull.* 85, 845–857. doi: 10.1037/0033-2909.85.4.845

Keating, C. F., and Bai, D. L. (1986). Children's attributions of social dominance from facial cues. *Child. Dev.* 57, 1269–1276. doi: 10.2307/1130449

Krupp, D., DeBruine, L., and Barclay, P. (2008). A cue of kinship promotes cooperation for the public good. *Evol. Hum. Behav.* 29, 49–55. doi: 10.1016/j.evolhumbehav.2007.08.002

Langlois, J. H., Kalakanis, L., Rubenstein, A. J., Larson, A., Hallam, M., and Smoot, M. (2000). Maxims or myths of beauty? A meta-analytic and theoretical review. *Psychol. Bull.* 126, 390–423. doi: 10.1037/0033-2909.126.3.390

Langlois, J. H., Roggman, L. A., Casey, R. J., Ritter, J. M., Rieser-Danner, L. A., and Jenkins, V. Y. (1987). Infant preferences for attractive faces: rudiments of a stereotype? *Dev. Psychol.* 23, 363–369. doi: 10.1037/0012-1649.23.3.363

Langlois, J. H., Roggman, L. A., and Rieser-Danner, L. A. (1990). Infants' differential social responses to attractive and unattractive faces. *Dev. Psychol.* 26, 153–159. doi: 10.1037/0012-1649.26.1.153

Langlois, J. H., and Stephan, C. (1977). The effects of physical attractiveness and ethnicity on children's behavioral attributions and peer preferences. *Child Dev.* 48, 1694–1698. doi: 10.1111/j.1467-8624.1977.tb03985.x

Ma, F., Xu, F., and Luo, X. (2015). Children's and adults' judgments of facial trustworthiness: the relationship to facial attractiveness. *Percept. Mot. Skills* 121, 179–198. doi: 10.2466/27.22.PMS.121c10x1

Mondloch, C. J., Geldart, S., Maurer, D., and Le Grand, R. (2003). Developmental changes in face processing skills. *J. Exp. Child Psychol.* 86, 67–84. doi: 10.1016/S0022-0965(03)00102-4

Mondloch, C. J., Le Grand, R., and Maurer, D. (2010). "Development of expertise in face recognition," in *Perceptual Expertise: Bridging Brain and Behavior,* eds I. Gauthier, M. J. Tarr, and D. Bub (New York, NY: Oxford University Press), 67–106.

Oosterhof, N. N., and Todorov, A. (2008). The functional basis of face evaluation. *Proc. Natl. Acad. Sci. U.S.A.* 105, 11087–11092. doi: 10.1073/pnas.0805664105

Ramsey, J. L., Langlois, J. H., Hoss, R. A., Rubenstein, A. J., and Griffin, A. M. (2004). Origins of a stereotype: categorization of facial attractiveness by 6-month-old infants. *Dev. Sci.* 7, 201–211. doi: 10.1111/j.1467-7687.2004.00339.x

Rotenberg, K. J., Boulton, M. J., and Fox, C. L. (2005). Cross-sectional and longitudinal relations among children's trust beliefs, psychological maladjustment and social relationships: are very high as well as very low trusting children at risk? *J. Abnorm. Child Psychol.* 33, 595–610. doi: 10.1007/s10802-005-6740-9

Rotenberg, K. J., McDougall, P., Boulton, M. J., Vaillancourt, T., Fox, C., and Hymel, S. (2004). Cross-sectional and longitudinal relations among peer-reported trustworthiness, social relationships, and psychological adjustment in children and early adolescents from the United Kingdom and Canada. *J. Exp. Child Psychol.* 88, 46–67. doi: 10.1016/j.jecp.2004.01.005

Short, L. A., Mondloch, C. J., McCormick, C. M., Carré, J. M., Ma, R., Fu, G., et al. (2012). Detection of propensity for aggression based on facial structure irrespective of face race. *Evol. Hum. Behav.* 33, 121–129. doi: 10.1016/j.evolhumbehav.2011.07.002

Slater, A., Von der Schulenburg, C., Brown, E., Badenoch, M., Butterworth, G., Parsons, S., et al. (1998). Newborn infants prefer attractive faces. *Infant Behav. Dev.* 21, 345–354. doi: 10.1016/S0163-6383(98)90011-X

Smith, E. R., and DeCoster, J. (2000). Dual-process models in social and cognitive psychology: conceptual integration and links to underlying memory systems. *Pers. Soc. Psychol. Rev.* 4, 108–131. doi: 10.1207/S15327957PSPR0402_01

Thomas, L. A., De Bellis, M. D., Graham, R., and LaBar, K. S. (2007).Development of emotional facial recognition in late childhood and adolescence. *Dev. Sci.* 10, 547–558. doi: 10.1111/j.1467-7687.2007.00614.x

Todorov, A., Baron, S. G., and Oosterhof, N. N. (2008). Evaluating face trustworthiness: a model based approach. *Soc. Cogn. Affect. Neurosci.* 3, 119–127. doi: 10.1093/scan/nsn009

Todorov, A., and Engell, A. D. (2008). The role of the amygdala in implicit evaluation of emotionally neutral faces. *Soc. Cogn. Affect. Neurosci.* 3, 303–312. doi: 10.1093/scan/nsn033

Todorov, A., Mandisodza, A. N., Goren, A., and Hall, C. C. (2005). Inferences of competence from faces predict election outcomes. *Science* 308, 1623–1626. doi: 10.1126/science.1110589

Todorov, A., Pakrashi, M., and Oosterhof, N. N. (2009). Evaluating faces on trustworthiness after minimal time exposure. *Soc. Cogn.* 27, 813–833. doi: 10.1521/soco.2009.27.6.813

Vannatta, K., Gartstein, M. A., Zeller, M., and Noll, R. B. (2009). Peer acceptance and social behavior during childhood and adolescence: how important are appearance, athleticism, and academic competence? *Int. J. Behav. Dev.* 33, 303–311. doi: 10.1177/0165025408101275

Wernick, M., and Manaster, G. J. (1984). Age and the perception of age and attractiveness. *Gerontologist* 24, 408–414. doi: 10.1093/geront/24.4.408

Willis, J., and Todorov, A. (2006). First impressions: making up your mind after a 100-ms exposure to a face. *Psychol. Sci.* 17, 592–598. doi: 10.1111/j.1467-9280.2006.01750.x

Willis, M. L., Dodd, H. F., and Palermo, R. (2013). The relationship between anxiety and the social judgements of approachability and trustworthiness. *PLoS ONE* 8:e76825. doi: 10.1371/journal.pone.0076825

Willis, M. L., Palermo, R., and Burke, D. (2011a). Social judgements are influenced by both facial expression and direction of eye gaze. *Soc. Cogn.* 29, 415–429. doi: 10.1521/soco.2011.29.4.415

Willis, M. L., Palermo, R., and Burke, D. (2011b). Judging approachability on the face of it: the influence of face and body expressions on the perception of approachability. *Emotion* 11, 514–523. doi: 10.1037/a0022571

Winston, J. S., Strange, B. A., O'Doherty, J., and Dolan, R. J. (2002). Automatic and intentional brain responses during evaluation of trustworthiness of faces. *Nat. Neurosci.* 5, 277–283. doi: 10.1038/nn816

Xu, F., Wu, D., Toriyama, R., Ma, F., Itakura, S., and Lee, K. (2012). Similarities and differences in Chinese and Caucasian adults' use of facial cues for trustworthiness judgements. *PLoS ONE* 7:e34859. doi: 10.1371/journal.pone.0034859

Zebrowitz, L. A., Franklin, R. G. Jr., Hillman, S., and Boc, H. (2013). Older and younger adults' first impressions from faces: similar in agreement but different in positivity. *Psychol. Aging* 28, 202–212. doi: 10.1037/a0 030927

Zebrowitz, L. A., and Montepare, J. M. (2008). Social psychological face perception: why appearance matters. *Soc. Pers. Psychol. Compass* 2, 1497–1517. doi: 10.1111/j.1751-9004.2008.00109.x

Zebrowitz, L. A., Wang, R., Bronstad, P. M., Eisenberg, D., Undurraga, E., and Reyes-Garcia, V. (2012). First impressions from faces among U.S. *and culturally isolated Tsimane' people in the Bolivian rainforest. J. Cross Cult. Psychol.* 43, 119–134. doi: 10.1177/0022022111411386