



Age-related differences in sensitivity to facial trustworthiness: Perceptual representation and the role of emotional development

Journal:	<i>Child Development</i>
Manuscript ID	2019-0092.R2
Wiley - Manuscript type:	Empirical Article
Subject Area:	Representation < Cognitive Processes, Face perception < Perceptual, Sensory, Motor, Social skills < Interpersonal Relations, Developmental trajectories < Methods, History, Theory, Social cognition < Social, Emotional, Moral, Personality
Keywords:	trustworthiness, perceptual discrimination, face representational space
Abstract:	The ability to discriminate social signals from faces is a fundamental component of human social interactions whose developmental origins are still debated. In this study, 5-year-old (N = 29) and 7-year-old children (N = 31) and adults (N= 34) made perceptual similarity and trustworthiness judgements of a set of female faces varying in level of expressed trustworthiness. All groups represented perceived similarity of the faces as a function of trustworthiness intensity, but such representation becomes more fine-grained with development. Moreover, 5-year-olds' accuracy in choosing the more trustworthy face in a pair varied as a function of children's score at the Test of Emotion Comprehension, suggesting that the ability to perform face-to-trait inferences is related to the development of emotion understanding.

Running Head: Children's sensitivity to facial trustworthiness

Age-related differences in sensitivity to facial trustworthiness: Perceptual representation and the role of emotional development

Date of submission: July 29th, 2019

For Review Only

Children's sensitivity to facial trustworthiness

Age-related differences in sensitivity to facial trustworthiness: Perceptual representation and the role of emotional development

Abstract

The ability to discriminate social signals from faces is a fundamental component of human social interactions whose developmental origins are still debated. In this study, 5-year-old (N = 29) and 7-year-old children (N = 31) and adults (N= 34) made perceptual similarity and trustworthiness judgements of a set of female faces varying in level of expressed trustworthiness. All groups represented perceived similarity of the faces as a function of trustworthiness intensity, but such representation becomes more fine-grained with development. Moreover, 5-year-olds' accuracy in choosing the more trustworthy face in a pair varied as a function of children's score at the Test of Emotion Comprehension, suggesting that the ability to perform face-to-trait inferences is related to the development of emotion understanding.

Keywords: face traits; trustworthiness; development; childhood; perceptual discrimination; face representational space; emotional development.

Children's sensitivity to facial trustworthiness

**Age-related differences in children's sensitivity to facial cues to trustworthiness:
Perceptual representation and the role of emotional development**

As humans, we display fine-grained sensitivity to non-verbal cues from faces, which we use to derive a great amount of social information about the person we are approaching. This competence has its origins in the very early stages of development. For example, by the age of 7 months infants are able to discriminate between different emotion categories, such as happiness and fear, based on facial expressions (Grossmann, 2010; Nelson, 1987), and use facial expression of emotional states to regulate their behaviour (Cohn & Tronick, 1983; Hirshberg, 1990).

One important piece of social information we automatically derive from faces is whether an individual represents safety or threat, i.e. whether he/she could be trusted and approached or better not trusted and avoided. Humans seem to be very prone to decode the facial information that drives social perception of trustworthiness: adults can identify whether a stranger represents a threat solely based on his/her facial characteristics after only 39 ms (Bar, Neta, & Linz, 2006), and use subtle differences between facial characteristics to generate explicit judgments of trustworthiness (Ames, Fiske, & Todorov, 2011; Todorov, Said, Engell, & Oosterhof, 2008). Trustworthiness judgements appear to be vehiculated by specific facial features (i.e., Action Units pattern for trustworthiness, see Jack & Schyns, 2015), such as up/downturned eyebrows, upward/downturned curving mouth and a wrinkling nose, which are also involved in emotion perception. Indeed, a possible mechanism through which these features induce social perception of trustworthiness is an overgeneralization of responses to facial configurations resembling emotional expressions (Said, Sebe, & Todorov, 2009; Zebrowitz, Fellous, Mignault, & Andreoletti, 2003). According to the emotion overgeneralization hypothesis, cues like lowered eyebrows, which, if stressed, might signal

Children's sensitivity to facial trustworthiness

anger, would induce the perceiver to make a dispositional assumption about an otherwise emotionally neutral face, resulting in social perception of unfriendliness (Ames et al., 2011). Accordingly, in both adults (e.g., Oosterhof & Todorov, 2008) and children (Caulfield, Ewing, Bank, & Rhodes, 2016) perceived trustworthiness is robustly associated to the attribution of emotional states, as overt angry expressions lead to subjective judgments of untrustworthiness, and overt happy expressions lead to subjective judgments of trustworthiness. Nevertheless, because social perception of trustworthiness and untrustworthiness also occurs for faces that are perceived as emotionally neutral (e.g., Lischke, Junge, Hamm, & Weymar, 2018), a thorough consideration of how the specific face information that elicits the perception of trustworthiness differ from that subtending the perception of emotional states is at stake (e.g., see review by Jack & Schyns, 2015). For example, Gill and colleagues (Gill, Garrod, Jack, & Schyns, 2014) showed that transient facial movements vehiculating trustworthiness judgements represent a unique Action Units configuration that differ from, and therefore should not be reduced to, Action Units configurations of emotional expressions.

Although much is known about the neurocognitive mechanisms underlying adults' sensitivity to facial cues to trustworthiness and/or their proneness to make trustworthiness inferences from faces, how this ability develops over time has been scarcely investigated. In particular, the question of whether and how this ability relates to the development of emotion understanding remains unexplored.

Available evidence suggests that the ability to perform explicit face-based trait judgments develops throughout childhood. At the age of 3 years, children are capable of making explicit judgements about how 'mean' or 'nice' a person appears to be, and by 6 years of age these attributions are at adult levels of consistency (Cogsdill, Todorov, Spelke, & Banaji, 2014). When specifically asked to judge whether a series of faces were trustworthy by rating each of them on a 3-point scale, 8-, 10-, and 12-year-old children all showed within-age agreement in

Children's sensitivity to facial trustworthiness

their judgments, although these judgments were more consistent and more similar to those of the adults in the older children than in the younger ones (Ma, Xu, & Luo, 2016). Ewing and colleagues (Ewing, Caulfield, Read, & Rhodes, 2015) also showed that face-based trustworthiness judgments influence children's behaviour, as 5- and 10-year-olds were more likely to place their trust in partners who looked trustworthy than in those who looked untrustworthy while playing an economic trust game. Overall, these findings led some authors to conclude that the ability to derive trait inferences from faces appears rather early in development, and do not require prolonged social experience as it builds on adaptive mechanisms developed to actively respond to threat (Cogsdill, Todorov, Spelke, & Banaji, 2014; LoBue, 2009).

Consistent with this hypothesis, two recent studies by Jessen & Grossmann (2016, 2017) suggest that even preverbal infants are sensitive to the face information that, in older children and adults, convey trust perception. In these studies, 7-months-old infants showed neural discrimination between neutral faces and those rated high or low on trustworthiness (Jessen & Grossmann, 2016), even when faces were presented subliminally (Jessen & Grossmann, 2017). Moreover, infants preferentially oriented their attention towards faces judged as trustworthy by adults rather than those judged as untrustworthy. These findings are in line with those showing that infants in the first year of life prefer prosocial individuals to antisocial others (Hamlin & Wynn, 2011; Van de Vondervoort & Hamlin, 2017), and preferably approach a stranger that their mother approached positively (Fein, 1975), just like older children (6- to 11-year-olds) tend to trust those who help others (Fu, Heyman, Chen, Liu, & Lee, 2015). Overall, this evidence suggests that humans are sensitive to other people's approachability from very early in the development.

Notwithstanding the relevance of these studies, they leave open a few questions, which the current study aimed to address. The first question is how perceptual sensitivity to fine-grained

Children's sensitivity to facial trustworthiness

differences in facial information subtending social perception of trustworthiness develops in time.

Previous studies investigated infants' (Jessen & Grossmann, 2016, 2017) and children's (e.g., Cogsdill et al., 2014; Cogsdill & Banaji, 2015; Ma et al., 2016) responses to computer-generated faces obtained from data-driven modelling (but see Cogsdill & Banaji, 2015) lying at the opposites of the trustworthiness continuum. On the one hand, this might have inflated participants' performance in the tasks due to high distinctiveness of the trustworthiness opposites. On the other hand, asking children to distinguish between very trustworthy and very untrustworthy faces may limit our understanding of their sensitivity to social signals from faces in real-life situations. Indeed, during our everyday social interactions we constantly decode subtle facial information and discriminate the slightest variations in other people's facial expressions. Finally, although artificial faces allow for a strictly controlled manipulation of the selected features, they may not fully reflect participants' expertise at face processing, including perceptual discrimination (e.g., see Crookes et al., 2015).

In the aim to overcome these limitations, in the current study we used as stimulus material a set of seven parametrically-manipulated variations of one real female face identity, slightly varying in level of perceived trustworthiness. This allowed us to trace developmental differences in children's perceptual sensitivity to subtle variations in physical cues to trustworthiness (see Method for a description of embedded in an exemplar of a face category (i.e., female faces) that is highly familiar to young children (see Quinn, Yahr, Kuhn, Slater, & Pascalis, 2002; Ramsey-Rennels & Langlois, 2006). Perceptual sensitivity was inferred from 5-year-olds', 7-year-olds', and adults' performance in an *Oddmanout task* (e.g., see Nishimura, Maurer, & Gao, 2009), where participants were simply asked to choose the face that appeared to be most different from the rest in a set of three simultaneously presented faces selected from our trustworthiness continuum. Children's and adults' perception of similarity/dissimilarity

Children's sensitivity to facial trustworthiness

among all the seven faces composing our trustworthiness continuum allowed us to compute pairwise dissimilarity scores, which provided a measure of their ability to discriminate the facial information that varied along the continuum.

By collecting dissimilarity scores, we were also able to address **a second important issue left open in the literature, which is whether developmental differences exist between childhood and adulthood in the representation of physical cues to trustworthiness.**

According to the influential face-space model proposed by Valentine (1991; also see review by Valentine, Lewis, & Hills, 2016), throughout the life span we build a representational model that maps the information that our visual system extracts from the faces we encounter in our social environment into a multidimensional perceptual space. Faces are organized in this space according to their perceived similarities along different dimensions, each of which represents a critical cue that is important for discriminating among individual faces (Valentine, 1991). Therefore, face-space organization is dependent on the amount and quality of one's own experience with faces, and becomes more fine-grained across development (Gao, Maurer, & Nishimura, 2010; Humphreys & Johnson, 2007; Rodger, Vizioli, Ouyang, & Caldara, 2015). Indeed, available evidence indicates that, although children represent faces in a multidimensional face space that has some adult-like characteristics at least from the age of 4 years (Jeffery et al., 2010), considerable refinement of this representation occurs throughout childhood. For example, separable representations of faces belonging to different categories defined by race, gender, and age emerge between 5 and 8 years of age (e.g., Short, Hatry, & Mondloch, 2011; Short, Lee, Fu, & Mondloch, 2014), and the representation of changeable facial traits like emotional expressions is also subject to critical changes across this same age range (Rodger et al., 2015). Other studies investigating identity discriminations have shown that, from the age of 7 years onwards, there are no differences between children and adults in the number of dimensions along which they represent faces, but children rely more heavily on

Children's sensitivity to facial trustworthiness

one single dimension (e.g., hair cues or eye colour) when making similarity judgments, whereas adults use all dimensions equally (Nishimura et al., 2009; Pedelty, Levine, & Shevell, 1985).

In the current study, we explored how face information subtending social perception of trustworthiness is represented in younger children's, older children's, and adults' memory by using dissimilarity scores derived from the *Oddmanout task* to build a Representational Dissimilarity Matrix (RDM) and perform a cluster analysis for each of the three age groups. This allowed us to describe the organization of face representation by unveiling how participants grouped together faces that varied in the level of expressed trustworthiness based on their perceived similarity (Sireci & Geisinger, 1992). It has been recently argued that perceptual, social and cultural experience all play a role in establishing and tuning face-trait mappings across development, by influencing both the perceptual representation of faces and the conceptual representation of others' personality traits (Over & Cook, 2018). In the current study, we focused on developmental changes in the perceptual representation of face information subtending trustworthiness perception. In light of earlier demonstrations that children from the age of 3 years can make explicit face-trait judgments on very distinctive trustworthiness opposites (Cogsdill et al., 2014), we hypothesized that the organization of younger children's representation of extreme physical cues to trustworthiness would be comparable to that of adults, but representation of the intermediate levels of the trustworthiness continuum would become more fine-grained with increasing age.

A third issue left unexplored by current developmental research relates to the impact of emotional development on children's social perception of trustworthiness. Recent evidence has shown that individual differences in personality and social behavior have an impact on how facial traits to trustworthiness are detected and utilized to perform social trait inferences (Baccolo & Macchi Cassia, 2019; Meconi, Luria, & Sessa, 2014; Young, Slepian, & Sacco, 2015). Most notably, impairments in social cognition and mentalizing abilities (i.e., attributing

Children's sensitivity to facial trustworthiness

mental states to other people) are known to be associated to diminished abilities to discriminate socially relevant characteristics of faces. For example, adults with ASD show abnormal face-based judgments of trustworthiness in association with atypical emotion recognition abilities (Adolphs, Sears, & Piven, 2001; Forgeot d'Arc et al., 2016). Despite this evidence supports the alleged relationship between emotion recognition skills and social judgements from facial cues, to the best of our knowledge, no studies have explored whether inter-individual variations in emotion comprehension abilities are reflected in corresponding variations in children's perceptual sensitivity to physical cues to trustworthiness as well as in their proneness to use such cues to generate trustworthiness judgements. Social cognition undergoes dramatic developmental changes in the first five years of life, and this is especially true for the so called *Theory of Mind*, which is the ability to understand other people's mental states, like desires, motives, emotions, and thoughts, and to respond to them adequately (Astington & Dack, 2008). By 3 to 4 years of age children can grasp other people's emotional states based on their facial expressions, but it is by the age of 5 years that they first develop critical components of emotion understanding (e.g., the situational causes of the outward expression of emotion), with other important components (i.e., the relationship between one's beliefs and his/her emotional states) appearing during school years, and reaching adult-like levels of performance in early adolescence (Pons, Harris, & de Rosnay, 2004).

In light of this evidence, in the current study, we explored the relation between children's emotional development and their perception of trustworthiness traits from faces by focusing on the 5- to 7-year age range. Specifically, we investigated whether 5- and 7-year-old children's ability to comprehend other people's emotions affected their perceptual representation of facial cues to trustworthiness and/or their judgements of perceived trustworthiness. To this end, we correlated children's score in the Test of Emotion Comprehension (TEC; Pons & Harris, 2000) to their performance in the *Oddmanout task* as well as in a second task – i.e., the *Pairwise*

Children's sensitivity to facial trustworthiness

Preference task – which allowed us to acquire explicit trustworthiness judgements on the seven faces of the continuum using a child-friendly procedure. Previous studies with children measured explicit judgements of trustworthiness by using rating scales (e.g., Cogsdill et al., 2014; Cogsdill & Banaji, 2015; Ma et al., 2016), which, however, could prove challenging for young children as they require reference to an internal rating scale and memory of the values assigned to previous faces, possibly resulting in inconsistent use of the scale across trials. Unlike these previous studies, in the *Pairwise Preference task* children were asked to indicate which face they would trust more within a pair randomly selected from our trustworthiness continuum, and the participant's response was used to compute a trustworthiness score for each face.

To sum up, the current study had three main aims: (1) to investigate whether perceptual sensitivity to face information subtending social perception of trustworthiness changes across childhood and into adulthood, (2) to explore the presence of age-related differences in the structure of the mental representation of facial cues to trustworthiness, and (3) to investigate whether children's emotion understanding skills affect their social perception of trustworthiness from faces.

Five- and 7-year-old children were selected as target age groups because we wanted our data to be comparable with those obtained by previous studies exploring children's trustworthiness judgments from faces, which targeted this same age range (Caulfield et al., 2016; Ewing et al., 2015). Moreover, the 5- to 8-years age range is also critical for the development of face representation, including the representation of changeable facial traits such as emotional expressions (e.g., Rodger et al., 2015), with the age of 7 marking the time when the structure of children's face representational space becomes adult-like (e.g., Nishimura et al., 2009). Finally, emotional intelligence and emotion comprehension show important improvements across the 5- to 7-year age range, when children become able not only to distinguish between

Children's sensitivity to facial trustworthiness

facial expressions of emotions and understand situational causes, but also to understand the mentalistic nature of emotions, such as the connection to desires and beliefs, and the distinction between expressed and felt emotion (Pons et al., 2004).

Materials and methods

Participants

Sample size was based on a Power Analysis for a univariate ANOVA with three groups (5-year-olds, 7-year-olds and adults), which revealed that about 64 participants should lead to an 80% chance to observe a significant effect with an alpha level of .05 and a large effect size. Data collection took place between October 2017 and May 2018. Analyses were performed on a total of 94 subjects: 29 5-year-old children (14 females; mean age = 5 years 5 months, range = 4 years 11 month - 5 years 11 months), 31 7-year-olds (12 females; mean age = 7 years 8 months, range = 7 years 1 month - 7 years 12 months), and 34 young adults (25 females; mean age = 23.03 years, range = 19 - 28 years). All children were recruited from preschools and schools within a major city area, and were attending preschool or primary school full-time at the time of testing. They all came from middle-class Caucasian families (except one Hispanic) and lived in a racially homogeneous neighbourhood. Adults were either undergraduate or graduate university students from middle-class families receiving course credits or recruited from the community by word of mouth on a voluntary basis. An additional 17 children (10 5-year-olds) were excluded from the final sample due to being distracted during the test. All procedures used in the current study complied with the Ethics Standards outlined by the Declaration of Helsinki (BMJ 1991; 302: 1194) and were approved by the Ethics Committee of the University of (omitted for blind reviewing). Adult participants signed an informed consent before testing; all participants' parents gave informed written consent prior to commencement of the study, and children gave their verbal assent before testing.

Children's sensitivity to facial trustworthiness

Stimuli

Stimuli were seven variations of one female facial identity reflecting a continuum of trustworthiness that ranged from 1 (very untrustworthy) to 7 (very trustworthy), interleaved by a neutral face (see Figure 1). The 7-steps continuum was created by morphing an averaged neutral face towards an averaged untrustworthy and an averaged trustworthy face using WebMorph (DeBruine, 2017), an online program for image transformation, specifically designed to perform face morphing and transforming. All the averaged faces were created by averaging three different face identities selected from the Chicago Face Database (Ma, Correll, & Wittenbrink, 2015). This Database provides a wide range of photographs of female and male identities rated on different face dimensions, including face trustworthiness. The averaged neutral face was created by averaging three different face identities that were rated as neutral on the trustworthiness dimension (i.e. not trustworthy nor untrustworthy), while the averaged faces used as references for morphing the neutral face towards the untrustworthy/trustworthy extremes were created by averaging the three faces rated as the most untrustworthy and trustworthy in the Database. We morphed the averaged neutral face 3 steps (30%, 60% and 100%) towards the very untrustworthy averaged face and 3 steps (30%, 60% and 100%) towards the very trustworthy averaged face, thus obtaining a 7-step trustworthiness continuum which included the neutral face. Figure 1 shows the output of an image difference analysis describing the physical variations among the seven faces included in the continuum. The analysis was carried out with the MATLAB (Mathworks Inc.) function *imshowpair* and the *diff* method to create a difference image between the most untrustworthy face (i.e., face 1) and each of the other faces. The images show that the physical aspects of the face that change the most along the continuum relate to the eyes area (i.e., eyes opening and eyebrows curvature), the

Children's sensitivity to facial trustworthiness

corners of the mouth (i.e., downturned in untrustworthy faces and upturned in trustworthy faces), and the nostrils (i.e, nose wrinkling).

----- Insert Figure 1 about here -----

Stimuli validation. In order to validate the stimuli (i.e., to ensure that the faces actually reflected a continuum of expressed trustworthiness), we asked an independent sample of 42 adults (34 females; mean age = 23.36 years; range = 18 - 35) to rate each step of the trustworthiness continuum on a scale ranging from 1 ("I wouldn't trust this person at all) to 9 ("I would definitely trust this person"). Adults' ratings were entered into a repeated-measures ANOVA with trustworthiness intensity as the within-subject factor, which attained statistical significance, $F(6,246) = 20.295$, $p < .001$, $p\eta^2 = 0.331$. A test of within-subjects contrasts revealed a significant linear trend, $F(1,41) = 58.760$, $p < .001$, $p\eta^2 = 0.589$, meaning that participants' trustworthiness judgements varied monotonically as a function of the faces' position along the trustworthiness continuum. Moreover, we explored whether our face stimuli elicited explicit judgements on other dimensions. Indeed, the abovementioned emotion overgeneralisation hypothesis (Said et al., 2009; Zebrowitz et al., 2003) suggests that trustworthiness judgements arise from an overgeneralisation of spontaneous responses to emotional expressions. Similarly, it has been suggested that face-to-trait trustworthiness judgements might be dependent on facial dimensions such as typicality (with atypical faces being perceived as less trustworthy; Sofer, Dotsch, Wigboldus, & Todorov, 2015; Todorov, Mende-Siedlecki, & Dotsch, 2013) or attractiveness (with attractive faces being perceived as more trustworthy; Hu, Abbasi, Zhang, & Chen, 2018; Oosterhof & Todorov, 2008; Schmidt, Leventsten, & Ambadar, 2012). Therefore, we asked a second group of young adults (N = 46, 25 females; mean age = 23.98 years; range = 19 - 35) to evaluate each face of the trustworthiness continuum on perceived emotion, typicality and attractiveness. For each face,

Children's sensitivity to facial trustworthiness

participants were asked if they perceived it to be emotional, typical and attractive, with the three questions presented in a random order. If participants responded the face to be emotional, they were asked to select the emotion they perceived (happy, angry, sad, scared or other) and rate the intensity of the selected emotion on a scale ranging from 1 ("This face is hardly happy/angry/sad/scared/other") to 9 ("This face is quite happy/angry/sad/scared/other"). Similarly, if participants responded the face to be typical or attractive, they were asked to rate the level of typicality/attractiveness on a scale ranging from 1 ("This face is hardly typical/attractive") to 9 ("This face is quite typical/attractive"). For all dimensions, negative answers (i.e., the face is not emotional, not typical or not attractive) were treated as zero values. To compare the average intensity of the trustworthiness judgements elicited by the seven faces of the continuum to that of the emotional, typicality and attractiveness judgements, we performed six independent *t*-tests, one for each of the four listed emotions (i.e., happiness, anger, sadness and fear), and one for each of the typicality and the attractiveness judgements. On average, our face stimuli were judged as more trustworthy than emotional (all $ps < .001$), whereas the intensity of perceived trustworthiness did not differ from that of perceived typicality, $t(12) = .845$, $p = .414$, and attractiveness, $t(12) = 1.043$, $p = .318$. These data suggest that, on average, trustworthiness judgments are poorly influenced by emotional cues: in fact, all faces from the trustworthiness continuum are, overall, scarcely perceived as emotional. Instead, trustworthiness judgements are not discernible from those on typicality and attractiveness in terms of overall intensity. Correlational analyses performed on intensity judgements elicited by each of the seven faces along the continuum revealed that increase in perceived trustworthiness was related to a corresponding increase in perceived attractiveness, $r = .965$, $p < .001$, 95% confidence interval (CI) [0.775, 0.995], and a decrease in perceived anger, $r = -.875$, $p = .01$, 95% confidence interval (CI) [-0.981, -0.356] (Figure 1). Therefore, although our face stimuli were scarcely perceived as emotional, in accord with the

Children's sensitivity to facial trustworthiness

literature (Said, Sebe, et al., 2009; Zebrowitz et al., 2003) the more untrustworthy faces in the continuum were hardly perceived as expressing negatively valenced emotional expressions. Moreover, and still in accord with the literature (Ma et al., 2016), the more trustworthy faces were perceived as more attractive, as attractiveness has been found to act as an heuristic property in trustworthiness judgements.

Apparatus and Procedure

Participants were tested individually in a quiet room (for the 5-year-old and the 7-year-old children the room was located in the school where the testing took place). All of them performed the *Oddmanout task* first, followed by the *Pairwise Preference task*. Both tasks were administered while participants seated 60 cm from a 17.3-inch touch-screen monitor with a resolution of 1080p. Stimulus presentation and response collection were controlled by ASF (Schwarzbach, 2011) and MATLAB Psychtoolbox for Windows (Brainard, 1997). For the 5-years-old and the 7-years-old children the experimenter manually controlled trial presentation by starting the trial as soon as the child fixated the monitor. Moreover, for the children the two tasks were interleaved by the administration of the *Test of Emotion Comprehension* (TEC) (Pons & Harris, 2000).

Oddmanout task

The *Oddmanout task* was designed to acquire measures of perceived dissimilarity among faces varying in their level of expressed trustworthiness. Participants observed three simultaneously presented faces appearing on the computer screen. On each trial, the three faces were all different, and randomly selected from the 7 trustworthiness intensities composing the continuum. Adults were asked to select the one they judged to be more different from the others by using the touch-screen interface of the computer. Children also provided they response by

Children's sensitivity to facial trustworthiness

using the touch-screen interface after being told the story of a young princess who was imprisoned in a castle tower by a witch, and hoped to be set free by the experimenter, who needed the child's help to succeed. The experimenter was able to steal the tower keys and reach the princess' room but the witch had created two avatars of the real princess in order to disguise him. Therefore, the experimenter asked the child "Can you help me to find out which one is the real princess? To find out, you should look carefully at the three faces and choose the one that looks more different from the others, as that one for sure is the real one!". To ensure that children thoroughly understood the instructions, we asked each child to repeat to the experimenter what they were meant to do. In addition, all children were given 5 practice trials prior to test trials so as to familiarize them with the task. Since we aimed at recording participants' responses for all possible triplet combinations to compute pairwise dissimilarity scores, participants viewed a total of 35 trials (the binomial coefficient obtained by selecting 3 faces out of a total of 7 faces, without considering triplets repetition or order). This way, each trustworthiness intensity appeared for a total of 15 times, and each pairwise comparison of the same two trustworthiness intensities appeared for a total of 5 times; the positions of the faces on the screen were randomized across trials. Each trial started with a central fixation cross (Figure 2), that remained on the screen for 1000 ms for the adult participants, or until the experimenter turned on the stimuli for the children. The stimuli remained on the screen until a response was made.

Pairwise Preference task

The *Pairwise Preference task* was designed to acquire explicit trustworthiness judgements on the seven faces of the continuum. On each trial, two faces randomly selected from the trustworthiness continuum simultaneously appeared on the computer screen, and participants were asked to select the face they trusted more by using the touch-screen interface. Children

Children's sensitivity to facial trustworthiness

were told a second story in which, after saving the princess, the experimenter got lost in a supermarket. While searching for the exit door, he bumped into two identical girls who pretend to know where the exit door is, one being the real princess he/she had saved earlier ("the good one"), and the other being one of the princess' avatars created by the witch ("the mean one"). In order to actually reach the exit door and find the way home, the experimenter asked the child "Can you help me to find out which one is the real princess? The princess is good, she wants to help me as we are friends now, she is someone I can trust. The other girl, instead, is mean, she wants to disguise me, and I should better not trust her. Can you help me to find out which one is the real princess who will help me to find the way home?". In order to ensure that task instructions were fully understood, before the task commenced each child was asked to explain back to the experimenter what he/she was meant to do. Since we aimed at recording participants' responses for all possible pairwise combinations of the seven faces of the continuum, the task was composed of 21 trials (the binomial coefficient obtained by selecting 2 faces out of a total of 7 faces, without considering pairs repetition or order). Each trustworthiness intensity was compared to all other intensities for a total of 6 times; the position of the faces on the screen was randomized across trials. Each trial started with a central fixation cross (Figure 2), that remained on the screen for 1000 ms for the adult participants, or until the experimenter turned on the stimuli for the children. The stimuli remained on the screen until a response was made.

Test of Emotion Comprehension

In between the *Oddmanout task* and the *Pairwise Preference task*, children were all administered the Italian version of the *Test of Emotion Comprehension* (TEC) (Albanese & Molina, 2008), which assesses nine dimensions of emotion understanding, namely the recognition of emotions based on facial expressions, the understanding of external emotional

Children's sensitivity to facial trustworthiness

causes, impact of desire on emotions, emotions based on beliefs, memory influence on emotion, emotion regulation, the ability to hide emotional states, understanding of mixed emotions and the relationship between morality and emotional experiences (Pons & Harris, 2000). The test is suitable for use with children aged 3 to 11 years, and consists of a booklet of illustrations divided into a set of stories, each depicting a different situation with four possible outcomes represented by different emotional facial expressions (happy, sad, angry, afraid). Children are asked to assign an emotion to the situation by selecting the corresponding facial expression; responses could be provided either verbally or by pointing to the drawing. The administration of the test took approximately 20 to 30 minutes. Each child obtained a score between 0 and 9 resulting from the sum of the partial scores attributed to the nine dimensions of emotion understanding; these scores were converted into *z scores*.

----- Insert Figure 2 about here -----

Results

Oddmanout task: Perceived dissimilarity between face pairs

In order to confirm that both the younger and the older children correctly understood the task, we performed two one-sample *t*-tests, one for each age group, on percent accuracy on a subset of trials ($N = 2$) in which one of the three faces was maximally distinguishable from the other two (i.e., trials in which faces 7-6-1 or 1-2-7 were shown). Performance accuracy in these trials was significantly above the chance level (50%) for both the 5-year-olds ($M = 76\%$), $t(28) = 5.477$, $p < .001$, and the 7-year-olds (92%), $t(30) = 12.490$, $p < .001$, as well as for the adults ($M = 87\%$), $t(33) = 9.574$, $p < .001$.

Representational Dissimilarity Matrices of Pairwise Dissimilarity Scores. To obtain a measure of perceived similarity/dissimilarity among all the seven faces composing the trustworthiness continuum, participant's response on each trial was used to compute three pairwise

Children's sensitivity to facial trustworthiness

dissimilarity scores, one for each face pair within the triplet. Once the subject had selected the most different face of the triad, the face pair composed of the non-selected faces was given a distance score of 0 (minimum dissimilarity), while the face pairs composed of the selected face and the non-selected ones were given a score of 1 (maximum dissimilarity). For each subject, the sum of the dissimilarity scores obtained for each face pair was scaled to 0-1 by dividing it by the number of trials in which that specific face pair appeared ($N = 5$), and used to build a 7x7 Representational Dissimilarity Matrix (RDM). The obtained RDM shows the level of perceived dissimilarity between face pairs: each column and row represent the dissimilarity scores of one trustworthiness intensity against all other trustworthiness intensities. Since the diagonal represents the dissimilarity of each trustworthiness level with itself, it contains only zero values, and each RDM is specular along the diagonal.

----- Insert Figure 3 about here -----

Within each age group, individual RDMs were averaged across subjects to obtain an RDM for each age group (see Figure 3). In order to explore age-related differences in the level of perceived dissimilarity among the faces composing the trustworthiness continuum, Pearson correlation analyses were performed on age-specific RDMs to assess whether the matrices resulting from the acquired dissimilarity scores had similar configurations for the three age groups. All three RDMs proved to highly correlate (5-year-olds and 7-year-olds: $r = 0.92$, $p < 0.001$, 95% confidence interval (CI) [0.799, 0.965]; 5-year-olds and adults: $r = 0.879$, $p < 0.001$, 95% confidence interval (CI) [0.721, 0.95]; 7-year-olds and adults: $r = 0.9334$, $p < 0.001$, 95% confidence interval (CI) [0.840, 0.973]), meaning that the pattern of perceived dissimilarities across the seven faces composing the trustworthiness continuum was similar for all groups.

Cluster Analysis on Pairwise Dissimilarity Scores. In order to investigate age-related differences in how facial cues to trustworthiness are perceptually represented in participants'

Children's sensitivity to facial trustworthiness

memory, pairwise dissimilarity scores were used to perform separate agglomerative hierarchical cluster analyses for each age group (see Everitt, 2011). Cluster analyses provided a description of how participants from each age group aggregate the faces from the trustworthiness continuum based on their perceptual similarities. The analyses were performed using the *average linkage* method in MATLAB (Mathworks Inc.), which groups items by creating a multilevel hierarchy to form a hierarchical tree based on average distance between items. Figure 4 represents the dendrograms resulting from the cluster analyses performed on the pairwise dissimilarity scores for the 5-year-olds, the 7-year-olds, and the adults. The X-axis represents the logical order derived from the dissimilarity judgements; the Y-axis represents the degree of perceived difference between faces, which is the distance that the function *linkage* computes between couple of items. Clusters are depicted in grayscales when their linkage is lower than 70% of the maximum linkage.

----- Insert Figure 4 about here -----

Intra-group consistency of Pairwise Dissimilarity Scores. In order to further explore the presence of age-related differences in sensitivity to variations in facial cues to trustworthiness, we compared intra-group consistency in participants' perceived dissimilarity across the three age groups. To this end, for each single age group we used RDMS of single participants to compute cosine distances of pairwise dissimilarity scores within each age group. Cosine distance can be defined as one minus the angle cosine of two vectors of an inner product space. A cosine distance of 0 is found whenever two vectors have the same orientation, while two perpendicular vectors have a cosine distance of 1. Cosine distance can therefore range between 0 (lowest distance) and 1 (greatest distance). We computed the cosine distance between all possible pairwise combinations of vectorized RDMS (upper triangular part of the matrices) of single subjects, separately for each of the three age groups. A univariate ANOVA on cosine

Children's sensitivity to facial trustworthiness

distances with age as between-subjects factor was conducted, which proved significant, $F(2,1429) = 136.006, p < 0.001, p\eta^2 = 0.160$. Pairwise dissimilarity scores were more consistent among adults ($M_{\text{cosine distance}} = 0.065, SD = 0.020$) than both the 5-year-olds ($M_{\text{cosine distance}} = 0.093, SD = 0.027$), $p < .001$, and the 7-year-olds ($M_{\text{cosine distance}} = 0.081, SD = 0.032$), $p < .001$, while the 7-year-olds were more consistent than the 5-year-olds, $p < .001$. Figure 5 shows the empirical cumulative distribution of cosine distances for the three age groups.

----- Insert Figure 5 about here -----

In order to investigate whether dissimilarity scores within each age group were equally consistent across participants for all trustworthiness intensities, we computed cosine distances between dissimilarity scores for each trustworthiness intensity. Indeed, each row (or column) of the RDM represents the pairwise dissimilarity scores between one trustworthiness intensity and each of the other trustworthiness intensities in the continuum. For each age group, we calculated the cosine distances between each row of the RDM of each single participant and the corresponding row of RDMs of all other participants. The obtained values were entered into a repeated-measures ANOVA with trustworthiness intensity as the within-subjects factor and age group as the between-subjects factor. Both main effects were significant (trustworthiness intensity: $F(6,8574) = 251.291, p < .001, p\eta^2 = 0.150, \text{power} = 1$; age group: $F(2,1429) = 110.788, p < .001, p\eta^2 = 0.134, \text{power} = 1$), as was the interaction between the two factors, $F(12,2850) = 17.596, p < .001, p\eta^2 = 0.024, \text{power} = 1$. For all age groups, a test of within-subjects contrasts revealed a significant quadratic trend for trustworthiness intensity, $ps < .001$ (see Figure 4).

Pairwise Preference task: explicit trustworthiness judgements

Response accuracy. Participant's response on each trial of the *Pairwise Preference task* was used to compute percent accuracy, i.e. the percentage of trials in which the subject selected the

Children's sensitivity to facial trustworthiness

face with higher trustworthiness intensity. To examine whether participants in all age groups systematically selected the face displaying more intense cues to trustworthiness, we performed three one-sample *t*-tests, one for each age group, which all proved significant, indicating that accuracy was significantly above chance for all age groups (5-year-olds: $t(28) = 7.177$, $p < .001$; 7-year-olds: $t(30) = 11.049$, $p < .001$; adults: $t(33) = 12.575$, $p < .001$). However, a univariate ANOVA with age group as the between-subjects factor also proved significant, $F(2,91) = 4.704$, $p = .011$, $\eta^2 = 0.094$, power = 0.776, showing age-related differences in the accuracy with which participants selected the more trustworthy face in the pair. Post-hoc comparisons (Bonferroni corrected) revealed that the 5-year-old children ($M = 72.25\%$, $SD = 16.69$) performed more poorly than the adults ($M = 84.17\%$, $SD = 15.85$), $p = .014$, and showed a marginal trend towards being also less accurate than the 7-year-olds ($M = 82.18\%$, $SD = 16.22$), $p = .06$ (see Figure 6).

----- Insert Figure 6 about here -----

Trustworthiness Scores. To obtain a measure of how consistently each face from the trustworthiness continuum was selected as the more trustworthy in a pair, participant's response on each trial of the *Pairwise Preference task* was used to compute a trustworthiness score for each face. Once the subject judged a face to be more trustworthy than the other, the selected face was given a score of 1, and the non-selected face was given a score of 0. For each subject, trustworthiness scores for each face of the continuum were summed across trials. Since each face appeared for a total of 6 times, the related trustworthiness score could range from 0 (if never selected as most trustworthy) to 6 (if always selected as more trustworthy). In order to investigate whether, for all age groups, participants' accuracy in selecting the more trustworthy face in a pair varied as a function of the intensity of the physical cues to trustworthiness displayed by the face to be selected (i.e., the higher the trustworthiness intensity, the higher the trustworthiness score), we performed a repeated-measures ANOVA

Children's sensitivity to facial trustworthiness

with trustworthiness intensity as the within-subjects factor and age group as the between-subjects factor. The analysis revealed a main effect of trustworthiness intensity, $F(6,546) = 119.327$, $p < .001$, $\eta^2 = 0.576$, power = 1, and a Trustworthiness Intensity x Age interaction, $F(12,174) = 2.821$, $p < 0.01$, $\eta^2 = 0.06$, power = 0.98 (see Figure 6). Post-hoc analysis (Bonferroni corrected) revealed that, for faces 1 and 2, the 5-year-olds provided judgements of higher trustworthiness compared to the adults ($p = .004$ and $p = .006$), whereas for face 5 the adults provided judgements of higher trustworthiness compared to the 7-year-olds ($p = .017$). A within-subjects contrast analysis revealed the presence of a significant linear trend in the trustworthiness scores for all age groups (all $ps < .001$).

Test of Emotion Comprehension (TEC)

In order to check for the existence of the expected developmental differences in children's performance at TEC (see Pons et al., 2004), we performed an independent-samples t -test on the z -transformed TEC scores of the 5-year-old and the 7-year-old age groups. The comparison proved significant, $t(58) = 5.45$, $p < 0.001$, revealing that the older children ($M = 8.03$, $SD = 0.91$) scored higher than the younger ones ($M = 6.17$, $SD = 1.65$) in the test. In light of this age-related overall difference in children's emotion comprehension abilities, we explored whether individual variability in these abilities was associated with differences in the perceptual representation of facial cues to trustworthiness and in using these cues to generate trustworthiness judgements by running correlational analyses between children's TEC scores and their performance in the *Oddmanout task* as well as in the *Pairwise Preference task*, separately for the two age groups.

TEC and Pairwise Dissimilarity scores. To investigate whether children's perceptual representation of facial cues to trustworthiness varied as a function of their TEC scores, we computed the intragroup cosine distance between the vectorized RDMs (i.e., dissimilarity

Children's sensitivity to facial trustworthiness

judgements between all face pairs) and the euclidean distance between the TEC scores separately for 5-year-olds and 7-year-olds. The euclidean distance between two points is the length of the linear segment that connects them. The Euclidean distance can be used as an index of similarity between couple of items. We therefore calculated the Euclidean distance between all possible pairwise combinations of TEC scores in order to have a measure of similarity between all TEC scores within each age group. We conjectured that, if mental representation of facial cues to trustworthiness varies as a function of the ability to understand other people's emotions, greater interindividual differences in TEC scores (measured as Euclidean distances) within a given age group should be related to greater interindividual differences in the RDMs (measured as cosine distances). For the 5-year-olds, the correlational analysis led to a significant, though very weak, positive correlation, $r = 0.053$, $p < .001$, 95% confidence interval (CI) [-0.045, 0.149]. For the 7-year-olds, no significant correlation was found, $r = -0.018$, $p = .685$, 95% confidence interval (CI) [-0.106, 0.0698].

TEC and Pairwise Preference accuracy. To investigate whether individual differences in the ability to attribute explicit trustworthiness judgements are related to individual differences in emotion comprehension, we correlated participants' response accuracy in the *Pairwise Preference task* with z-transformed TEC scores, separately for the 5-year-old and the 7-year-old children. A positive correlation was found for the younger children, $r = 0.44$, $p = .017$, 95% confidence interval (CI) [0.087, 0.694], but not for the older ones, $r = 0.02$, $p = .921$, 95% confidence interval (CI) [-0.338, 0.371].

Discussion

In the current study, we aimed at investigating the development of perceptual sensitivity to facial cues to trustworthiness and their representation in long term memory, taking into account the role of individual differences in emotion comprehension. To investigate perceptual

Children's sensitivity to facial trustworthiness

sensitivity to physical cues to trustworthiness, we tested younger and older children and adults in an *Oddmanout task* designed to provide measures of perceived similarity/dissimilarity among faces slightly varying in their level of expressed trustworthiness. Analysis of participants' performance revealed that, already at the age of 5 years, children represent faces as a function of the level of the trustworthiness they express. In fact, average RDMs describing perceived dissimilarity between face pairs were highly correlated across age groups, meaning that there were no age-related qualitative differences in participants' sensitivity to variations in physical cues to trustworthiness. However, our findings also showed that such sensitivity becomes increasingly fine-grained with development, as intra-group consistency in dissimilarity judgements increased with age. Indeed, adults' judgements were more consistent than those provided by the 5-year-old and the 7-year-old children, and 7-year-olds' judgements were, in turn, more consistent than those provided by the 5-year-olds. The *Oddmanout task* also allowed us to examine age-related differences in the consistency among participants' dissimilarity scores for each trustworthiness intensity. The adults' scores were more consistent than those of the 5-year-olds for all trustworthiness intensities, and for almost all trustworthiness intensities adults were also more consistent than the 7-year-olds. This is in line with earlier evidence that within-age consistency in explicit judgements of trustworthiness increases with age (Ma et al., 2016).

The finding that participants' proficiency at attributing dissimilarity judgements for faces that only slightly vary in the level of expressed trustworthiness improves with age is also evident from the results of the cluster analysis performed on dissimilarity scores from the *Oddmanout task*. As already noted, both adults and children represented faces as a function of the intensity of the physical cues to trustworthiness they express. However, the structure of the hierarchical clustering becomes more differentiated with increasing age. Five-year-old children's dissimilarity judgments formed one cluster composed of the two most untrustworthy

Children's sensitivity to facial trustworthiness

faces (faces 1 and 2) and another cluster composed of two trustworthy faces (faces 5 and 6). Proceeding along the hierarchy, these two clusters enlarged including all the three untrustworthy faces (faces 1, 2, 3) on one side, and all the three trustworthy faces (faces 5, 6, 7) on the other, with the neutral face associated to this latter cluster. The structure of children's representation becomes more differentiated at 7 years of age, when already at the bottom of the hierarchy the three untrustworthy faces (faces 1, 2, 3) form one cluster, and the two moderately trustworthy faces (faces 5 and 6) compose another cluster, which also includes the neutral face (face 4), and to which the most trustworthy face (face 7) is hierarchically associated. The hierarchical clustering of dissimilarity judgements reaches the greatest differentiation in adults, who show three different clusters including, respectively, the two most untrustworthy faces (faces 1 and 2), the neutral face and the one next to the neutral (faces 3 and 4), and the three trustworthy faces (5, 6, 7). Proceeding along the hierarchy, the clusters including the more untrustworthy faces (faces 1 and 2) and the more neutral faces (faces 3 and 4) are then incorporated into a larger supra-cluster. In keeping with Valentine's theory (Valentine et al., 2016), these results corroborate the idea that the representation of facial characteristics that are relevant for discriminating among individual faces and/or face types, becomes more fine-grained and differentiated across development. To the best of our knowledge, our results provide the first evidence of how the perceptual space which maps the physical face information that subtend perception of a social trait changes across childhood and into adulthood.

In fact, in addition to age-related differences in the structure of the mental representation of facial cues to trustworthiness, we also observed consistency in performance across age groups. Indeed, analyses on intra-group cosine distances of dissimilarity scores revealed that participants' agreement in attributing dissimilarity judgements showed similar variations across trustworthiness levels for all age groups. Consistency of dissimilarity scores of both

Children's sensitivity to facial trustworthiness

younger and older children, as well as adults, showed a significant quadratic trend, with less consistent scores for the central hub of the trustworthiness continuum (around the neutral face) and most consistent scores for the continuum extremes (very trustworthy and very untrustworthy faces). This finding replicates earlier evidence that, in adults, facial cues to trustworthiness that yield to more extreme trustworthiness judgments (i.e., very trustworthy or very untrustworthy) are easier to discriminate than those yielding to less extreme judgments, independently of their valence (i.e., whether the face is very trustworthy or very untrustworthy) (Baccolo & Macchi Cassia, 2019). Accordingly, neuroimaging studies with adults reported a similar valence-independent sensitivity of the amygdala to trustworthiness cues (Said, Baron, & Todorov, 2009; Said, Dotsch, & Todorov, 2010), and electrophysiological studies with infants reported neural discrimination between neutral faces and both very trustworthy (+3 SD) and very untrustworthy (-3 SD) faces, but not between trustworthy and untrustworthy faces (Jessen & Grossmann, 2016).

The finding that faces including more intense physical cues to social traits enjoy a processing advantage over those including less intense cues, irrespective of the valence of such cues, might be due to the fact that responding to socially connoted faces is of greater adaptive value than responding to neutral ones, just like responding to emotional faces is easier and faster than responding to neutral ones, even when only little face information conveying emotional intensity is available (Roesch, Sander, Mumenthaler, Kerzel, & Scherer, 2010). Another possibility is that participants in the current study agreed more on attributing dissimilarity judgments for faces at the extreme opposites of the trustworthiness continuum than for faces at the centre of the continuum because the former were perceived as less prototypical than the latter. Indeed, previous studies showed that deviations in typicality explain the amygdala response better than valence (Said et al., 2010). Nonetheless, we view this possibility as unlikely in light of the results of our stimulus validation procedure, which

Children's sensitivity to facial trustworthiness

revealed that judgements of perceived trustworthiness for each of the seven faces of the continuum were not discernible from judgements of perceived typicality in terms of overall intensity.

An age-related trend similar to that observed for perceptual sensitivity to physical cues to trustworthiness was present in our data also for explicit trustworthiness judgements recorded during the *Pairwise Preference task*. Just like adults, both younger and older children performed above chance when asked to select the more trustworthy face in a pair, and for all age groups the pairwise preference scores varied linearly, showing that participants made explicit judgements of trustworthiness intensity for each of the seven faces as a function of the stimulus position along the trustworthiness continuum. At the same time, though, data also showed that the adults were more accurate than the younger children, but not the older ones, suggesting that, by age 7 years, children's performance has become adult-like. This is in accord with earlier demonstration that by 7 years of age children are as sensitive as adults in attributing explicit trustworthiness judgements (Cogsdill et al., 2014; Cogsdill & Banaji, 2015). Of note, in addition to being less accurate and less consistent in their explicit judgements of trustworthiness, the 5-year-old children overestimated the trustworthiness intensity of the faces at the untrustworthy extreme of the continuum (faces 1 and 2), for which their judgements overrode those of the adults. This finding is consistent with earlier reports of inflated trustworthiness ratings for untrustworthy faces in 5-year-old children (Caulfield, Ewing, Bank, & Rhodes, 2016), and further suggests that sensitivity to this social cue from faces is not fully developed at this young age.

In fact, our findings showed that, at 5 years but not at 7 years, children differed in their ability to make trustworthiness judgements from faces as a function of their emotion understanding skills, as measured through the *Test of Emotion Comprehension*: higher scores at TEC were related to higher performance at successfully detecting the more trustworthy face

Children's sensitivity to facial trustworthiness

in a pair in the *Pairwise Preference task*. In contrast, we found no evidence that children's emotion comprehension skills modulated how they represented facial cues to trustworthiness, as we observed no significant correlations between TEC scores and children's intragroup cosine distance between the vectorized RDMs built on pairwise dissimilarity scores obtained from the *Oddmanout task*.

The finding that the association between emotion comprehension skills and social perception of trustworthiness in the *Pairwise Preference task* was confined to the 5-year-old further suggests that emotional development impacts the perception of social traits from faces. Indeed, although performance of the 5-year-old children in the *Pairwise Preference task* was only slightly lower than that of the older children, and similarly variable (5-year-olds: SD = 16.69; 7-year-olds: SD = 16.22), in line with previous evidence (Pons et al., 2004) the 5-year-olds scored significantly lower than the older children at TEC, and showed higher variability in their responses to the questionnaire (5-year-olds: SD = 1.65; 7-year-olds: SD = 0.91). This indicates that emotion understanding abilities are not evenly distributed in our 5-year-old sample, and is in accord with earlier reports of 5 years being the first developmental time in which critical components of emotion understanding (i.e., understanding of the outward expression of emotion and their situational causes) emerge, with other important components (i.e., the understanding of the mentalistic nature of emotions) appearing only later in development (Pons et al., 2004). The positive correlation between 5-year-olds' performance at TEC and the *Pairwise Preference task* indicates that children's early emotion understanding abilities are associated to their proficiency at inferring trustworthiness traits from faces. Overall these findings suggest that it is right in the earlier critical stages of the development of emotion understanding that the ability to use face information to infer trustworthiness traits builds on the ability to consistently use transient facial cues to infer internal emotional states.

Children's sensitivity to facial trustworthiness

As the present study is the first to explore the relation between emotional development and perception of social traits from faces, it has a number of limitations that could be addressed in future research. First, additional measures could be used to examine children's inferences about trustworthiness from faces. We assessed this ability by measuring children's responses in a paired-preference task in which they selected the face they could trust more in a pair. Although this is a viable way to evaluate young children's use of face information to infer social traits without straining on their limited cognitive resources, it will be important to develop new measures to explore how such inferences affect children's approach/avoidance behaviour (e.g., see Ewing et al., 2015), and how trust behaviour is associated with emotional development.

Second, more work is needed to examine how the findings from the present study generalize across culture, and across face ethnicity. It has been shown that Caucasian and Asian adults rely on similar facial cues to judge trustworthiness (Xu et al., 2012), and trust perception generalizes across face ethnicity (Birkás, Dzhelyova, Lábadi, Bereczkei, & Perrett, 2014). However, this may not be the case for children, especially in light of the fact that, starting from preschool years, children increasingly draw on racial information when making social judgments (e.g., Bigler & Liben, 2007; Killen & Stangor, 2001). Moreover, cultural differences may emerge in how trust perception is related to emotional understanding, as cultural norms and practice affect display rules.

Third, the use of averaged identities in this study reduced the influence of idiosyncratic facial features on perceptual sensitivity to physical cues of trustworthiness and explicit judgements of trustworthiness; **still, the use of one single average face identity runs the risk of obtaining non-generalizable findings. Future** studies shall put the generalizability of the current results under scrutiny by using a new set of averaged stimuli created starting from a different pool of face identities.

Children's sensitivity to facial trustworthiness

Finally, future studies shall investigate whether the relation between trust perception and individual variability in emotional development generalizes across age, for example by testing whether individual differences in temperament and self-regulation abilities affect sensitivity to facial cues to trustworthiness in preverbal infants, who have been shown to discriminate between faces based on those cues (Jessen & Grossmann, 2017).

To sum up, the present study shows that, already at the age of 5 years, the mental representation of perceived differences between facial cues associated to trustworthiness is organized along a continuum of trustworthiness intensity. Nevertheless, with increasing age, this representation becomes more fine-grained. Similarly, although the 5-year-olds were overall less accurate than the adults in their explicit judgements of trustworthiness and showed a positivity bias for the very untrustworthy faces, those judgements were linearly distributed along the trustworthiness continuum for both children and adults. Most importantly, at 5 years of age more accurate judgements of trustworthiness were associated with more accurate emotional understanding. Overall, implicit measures of perceptual sensitivity to physical cues of trustworthiness and explicit judgements of trustworthiness intensity converged in showing that the ability to discriminate facial cues associated to trustworthiness and to use such cues to make trustworthiness attributions is apparent at the age of 5 years but becomes more adult-like by the age of 7 years, and its development is related to the development of emotion comprehension.

Children's sensitivity to facial trustworthiness

References

- Albanese, O., & Molina, P. (2008). *Lo sviluppo della comprensione delle emozioni e la sua valutazione: la standardizzazione italiana del Test di comprensione delle emozioni (TEC)*. Milano : Unicopli. 2008.
- Ames, D. L., Fiske, S. T., & Todorov, A. T. (2011). *Impression Formation: A Focus on Others' Intents*. *The Oxford Handbook of Social Neuroscience*. Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780195342161.013.0028>
- Astington, J. W., & Dack, L. A. (2008). Theory of mind. In M. M. Haith & J. B. Benson (Eds.), *Encyclopedia of infant and early childhood development* (pp. 343–356). San Diego, CA: Academic Press.
- Astington, J. W., & Hughes, C. (2013). Theory of mind: Self-reflection and social understanding. In P. D. Zelazo (Ed.), *Oxford Handbook of Developmental Psychology*. New York: Oxford University Press.
- Baccolo, E., & Macchi Cassia, V. (2019). Individual differences in perceptual sensitivity and representation of facial signals of trustworthiness. *Journal of Experimental Psychology: Human Perception and Performance*. <https://doi.org/10.1037/xhp0000601>
- Bar, M., Neta, M., & Linz, H. (2006). Very first impressions. *Emotion (Washington, D.C.)*, 6(2), 269–278. <https://doi.org/10.1037/1528-3542.6.2.269>
- Bigler, R. S., & Liben, L. S. (2007). Developmental Intergroup Theory: Explaining and Reducing Children's Social Stereotyping and Prejudice. *Current Directions in Psychological Science*, 16(3), 162–166. <https://doi.org/10.04.87/j.1467-8721.2007.00496.x>
- Birkás, B., Dzhelyova, M., Lábadi, B., Bereczkei, T., & Perrett, D. I. (2014). Cross-cultural perception of trustworthiness: The effect of ethnicity features on evaluation of faces' observed trustworthiness across four samples. *Personality and Individual Differences*, 69,

Children's sensitivity to facial trustworthiness

56–61. <https://doi.org/10.1016/j.paid.2014.05.012>

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*, 433–436.

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* (London, England: 1953), *107*(3), 503–518. <https://doi.org/10.1111/bjop.12156>

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/http://10.0.3.248/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. <https://doi.org/10.1177/0956797614523297>

Cohn, J. F., & Tronick, E. Z. (1983). Three-month-old infants' reaction to simulated maternal depression. *Child Development*, *54*(1), 185–193. <https://doi.org/10.2307/1129876>

Cohn, J. F., & Tronick, E. Z. (1987). Mother-Infant Face-to-Face Interaction: The Sequence of Dyadic States at 3, 6, and 9 Months. *Developmental Psychology*, *23*(1), 68–77. <https://doi.org/10.1037/0012-1649.23.1.68>

Crookes, K., Ewing, L., JD, G., Kloth, N., WG, H., Oxner, M., ... Rhodes, G. (2015). How Well Do Computer-Generated Faces Tap Face Expertise? *Plos One*. <https://doi.org/10.1371/journal.pone.0141353>

DeBruine, L. (2017). Webmorph (Version v0.0.0.9001). *Zenodo*. <https://doi.org/10.5281/zenodo.1073696>

Everitt, B. (2011). *Cluster Analysis* (Vol. 5th ed). Chichester, West Sussex: Wiley.

Ewing, L., Caulfield, F., Read, A., & Rhodes, G. (2015). Perceived trustworthiness of faces drives trust behaviour in children. *Developmental Science*, *18*(2), 327–334.

Children's sensitivity to facial trustworthiness

<https://doi.org/10.1111/desc.12218>

Fein, G. G. (1975). Children's sensitivity to social contexts at 18 months of age. *Developmental Psychology*, *11*(6), 853–854. <https://doi.org/10.1037/0012-1649.11.6.853>

Forgeot d'Arc, B., Ramus, F., Lefebvre, A., Brottier, D., Zalla, T., Moukawane, S., ... Delorme, R. (2016). Atypical Social Judgment and Sensitivity to Perceptual Cues in Autism Spectrum Disorders. *Journal of Autism & Developmental Disorders*, *46*(5), 1574–1581. <https://doi.org/10.1007/s10803-014-2208-5>

Fu, G., Heyman, G. D., Chen, G., Liu, P., & Lee, K. (2015). Children trust people who lie to benefit others. *Journal of Experimental Child Psychology*, *129*, 127–139. <https://doi.org/http://10.03.248/j.jecp.2014.09.006>

Gao, X., Maurer, D., & Nishimura, M. (2010). Similarities and Differences in the Perceptual Structure of Facial Expressions of Children and Adults. *Journal of Experimental Child Psychology*, *105*(1), 98–115. <https://doi.org/http://dx.doi.org/10.1016/j.jecp.2009.09.001>

Gill, D., Garrod, O. G. B., Jack, R. E., & Schyns, P. G. (2014). Facial Movements Strategically Camouflage Involuntary Social Signals of Face Morphology. *Psychological Science* (0956-7976), *25*(5), 1079–1086. <https://doi.org/10.1177/0956797614522274>

Grossmann, T. (2010). The development of emotion perception in face and voice during infancy. *Restorative Neurology and Neuroscience*, *28*, 219–236. <https://doi.org/10.3233/RNN-2010-0499>

Hamlin, J. K., & Wynn, K. (2011). Young infants prefer prosocial to antisocial others. *Cognitive Development*, *26*, 30–39. <https://doi.org/10.03.248/j.cogdev.2010.09.001>

Hirshberg, L. (1990). When Infants Look to Their Parents: II. Twelve-Month-Olds' Response to Conflicting Parental Emotional Signals. *Child Development*, *61*(4), 1187. <https://doi.org/10.2307/1130886>

Hu, Y., Abbasi, N. U. H., Zhang, Y., & Chen, H. (2018). The Effect of Target Sex, Sexual

Children's sensitivity to facial trustworthiness

- Dimorphism, and Facial Attractiveness on Perceptions of Target Attractiveness and Trustworthiness. *Frontiers In Psychology*, 9, 942. <https://doi.org/10.3389/fpsyg.2018.00942>
- Humphreys, K., & Johnson, M. H. (2007). The development of “face-space” in infancy. *Visual Cognition*, 15(5), 578–598. <https://doi.org/10.1080/13506280600943518>
- Jack, R. E., & Schyns, P. G. (2015). Review: The Human Face as a Dynamic Tool for Social Communication. *Current Biology*, 25(14), R621. <https://doi.org/10.1016/j.cub.2015.05.052>
- Jeffery, L., Haynes, R., Pellicano, E., Rhodes, G., McKone, E., & Firth, E. (2010). Four-to-six-year-old children use norm-based coding in face-space. *Journal of Vision*, 10(5). <https://doi.org/10.1167/10.5.18>
- Jessen, S., & Grossmann, T. (2016). Neural and Behavioral Evidence for Infants' Sensitivity to the Trustworthiness of Faces. *Journal Of Cognitive Neuroscience*, 28(11), 1728–1736. https://doi.org/10.1162/jocn_a_00999
- Jessen, S., & Grossmann, T. (2017). Neural evidence for the subliminal processing of facial trustworthiness in infancy. *Neuropsychologia*. <https://doi.org/10.1016/j.neuropsychologia.2017.04.025>
- Killen, M., & Stangor, C. (2001). Children's Social Reasoning about Inclusion and Exclusion in Gender and Race Peer Group Contexts. *Child Development*, 72(1), 174. <https://doi.org/10.1111/1467-8624.00272>
- Lischke, A., Junge, M., Hamm, A. O., & Weymar, M. (2018). Enhanced processing of untrustworthiness in natural faces with neutral expressions. *Emotion (Washington, D.C.)*, 18(2), 181–189. <https://doi.org/10.1037/emo0000318>
- LoBue, V. (2009). More than just another face in the crowd: superior detection of threatening facial expressions in children and adults. *Developmental Science*, 12(2), 305–313.

Children's sensitivity to facial trustworthiness

<https://doi.org/10.1111/j.1467-7687.2008.00767.x>

Ma, D. S., Correll, J., & Wittenbrink, B. (2015). The Chicago face database: A free stimulus set of faces and norming data. *Behavior Research Methods*.

<https://doi.org/10.3758/s13428-014-0532-5>

Ma, F., Xu, F., & Luo, X. (2016). Children's Facial Trustworthiness Judgments: Agreement and Relationship with Facial Attractiveness. *Frontiers In Psychology*, 7, 499.

<https://doi.org/10.3389/fpsyg.2016.00499>

Meconi, F., Luria, R., & Sessa, P. (2014). Individual differences in anxiety predict neural measures of visual working memory for untrustworthy faces. *Social Cognitive And Affective Neuroscience*, 9(12), 1872–1879. <https://doi.org/10.1093/scan/nst189>

Nelson, C. A. (1987). The recognition of facial expressions in the first two years of life: mechanisms of development. *Child Development*, 58(4), 889–909.

<https://doi.org/10.2307/1130530>

Nelson, C. A., & De Haan, M. (1996). Neural Correlates of Infants' Visual Responsiveness to Facial Expressions of Emotion. *Developmental Psychobiology*, 29(7), 577.

<https://doi.org/10.1002/9780470147658.chpsy0201>

Nishimura, M., Maurer, D., & Gao, X. (2009). Exploring children's face-space: A multidimensional scaling analysis of the mental representation of facial identity. *Journal of Experimental Child Psychology*, 103, 355–375.

<https://doi.org/http://10.03.248/j.jecp.2009.02.005>

Oosterhof, N. N., & Todorov, A. (2008). The Functional Basis of Face Evaluation. *Proceedings of the National Academy of Sciences of the United States of America*, 105(32), 11087.

<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. <https://doi.org/10.1016/j.cognition.2017.10.002>

<https://doi.org/10.1016/j.cognition.2017.10.002>

Children's sensitivity to facial trustworthiness

- Pedelty, L., Levine, S. C., & Shevell, S. K. (1985). Developmental changes in face processing: Results from multidimensional scaling. *Journal of Experimental Child Psychology*, *39*(3), 421–436. [https://doi.org/10.1016/0022-0965\(85\)90050-5](https://doi.org/10.1016/0022-0965(85)90050-5)
- Pons, F., Harris, P., & de Rosnay, M. (2004). Emotion comprehension between 3 and 11 years: Developmental periods and hierarchical organization. *European Journal of Developmental Psychology*, *1*(2), 127–152. <https://doi.org/http://10.0.4.56/17405620344000022>
- Pons, F., & Harris, P. L. (2000). *TEC (Test of Emotion Comprehension)*. Oxford University Press.
- Quinn, P. C., Yahr, J., Kuhn, A., Slater, A. M., & Pascalis, O. (2002). Representation of the gender of human faces by infants: a preference for female. *Perception*, *31*(9), 1109–1121. <https://doi.org/10.1068/p3331>
- Ramsey-Rennels, J. L., & Langlois, J. H. (2006). Infants' Differential Processing of Female and Male Faces. *Current Directions in Psychological Science*, *15*(2), 59–62. <https://doi.org/10.0.4.87/j.0963-7214.2006.00407.x>
- Rodger, H., Vizioli, L., Ouyang, X., & Caldara, R. (2015). Mapping the development of facial expression recognition. *Developmental Science*, *18*(6), 926–939. <https://doi.org/10.1111/desc.12281>
- Roesch, E. B., Sander, D., Mumenthaler, C., Kerzel, D., & Scherer, K. R. (2010). Psychophysics of emotion: The QUEST for Emotional Attention. *Journal of Vision*. <https://doi.org/10.1167/10.3.4>
- Said, C. P., Baron, S. G., & Todorov, A. (2009). Nonlinear amygdala response to face trustworthiness: contributions of high and low spatial frequency information. *Journal Of Cognitive Neuroscience*, *21*(3), 519–528. <https://doi.org/10.1162/jocn.2009.21041>
- Said, C. P., Dotsch, R., & Todorov, A. (2010). The amygdala and FFA track both social and

Children's sensitivity to facial trustworthiness

- non-social face dimensions. *Neuropsychologia*, 48(12), 3596–3605.
<https://doi.org/10.1016/j.neuropsychologia.2010.08.009>
- Said, C. P., Sebe, N., & Todorov, A. (2009). Structural resemblance to emotional expressions predicts evaluation of emotionally neutral faces. *Emotion (Washington, D.C.)*, 9(2), 260–264. <https://doi.org/10.1037/a0014681>
- Schmidt, K., Leventsten, R., & Ambadar, Z. (2012). Intensity of Smiling and Attractiveness as Facial Signals of Trustworthiness in Women. *Perceptual & Motor Skills*, 114(3), 964–978. <https://doi.org/10.0.9.162/07.09.21.PMS.114.3.964-978>
- Schwarzbach, J. (2011). A simple framework (ASF) for behavioral and neuroimaging experiments based on the psychophysics toolbox for MATLAB. *Behavior Research Methods*. <https://doi.org/10.3758/s13428-011-0106-8>
- Short, L. A., Hatry, A. J., & Mondloch, C. J. (2011). The development of norm-based coding and race-specific face prototypes: An examination of 5- and 8-year-olds' face space. *Journal of Experimental Child Psychology*, 108, 338–357.
<https://doi.org/10.0.3.248/j.jecp.2010.07.007>
- Short, L. A., Lee, K., Fu, G., & Mondloch, C. J. (2014). Category-specific face prototypes are emerging, but not yet mature, in 5-year-old children. *Journal of Experimental Child Psychology*, 126, 161–177. <https://doi.org/10.1016/j.jecp.2014.04.004>
- Sireci, S., & Geisinger, K. (1992). Analyzing Test Content Using Cluster Analysis and Multidimensional Scaling. *Applied Psychological Measurement*, 16(1), 17.
<https://doi.org/https://doi.org/10.1177/014662169201600102>
- Sofer, C., Dotsch, R., Wigboldus, D. H. J., & Todorov, A. (2015). What is typical is good: the influence of face typicality on perceived trustworthiness. *Psychological Science*, 26(1), 39–47. <https://doi.org/10.1177/0956797614554955>
- Todorov, A., Mende-Siedlecki, P., & Dotsch, R. (2013). Social judgments from faces. *Current*

Children's sensitivity to facial trustworthiness

Opinion in Neurobiology, 23(Social and emotional neuroscience), 373–380.

<https://doi.org/10.1016/j.conb.2012.12.010>

Todorov, A., Said, C. P., Engell, A. D., & Oosterhof, N. N. (2008). Understanding evaluation of faces on social dimensions. *Trends In Cognitive Sciences*, 12(12), 455–460.

<https://doi.org/10.1016/j.tics.2008.10.001>

Valentine, T. (1991). A unified account of the effects of distinctiveness, inversion, and race in face recognition. *The Quarterly Journal Of Experimental Psychology. A, Human Experimental Psychology*, 43(2), 161–204. <https://doi.org/10.1080/14640749108400966>

Valentine, T., Lewis, M. B., & Hills, P. J. (2016). Face-space: A unifying concept in face recognition research. *Quarterly Journal of Experimental Psychology*, 69(10), 1996–2019.

<https://doi.org/10.1080/17470218.2014.990392>

Van de Vondervoort, J. W., & Hamlin, J. K. (2017). The early emergence of sociomoral evaluation: infants prefer prosocial others. *Current Opinion In Psychology*, 20, 77–81.

<https://doi.org/10.1016/j.copsyc.2017.08.014>

Xu, F., Wu, D., Toriyama, R., Ma, F., Itakura, S., & Lee, K. (2012). Similarities and differences in Chinese and Caucasian adults' use of facial cues for trustworthiness judgments. *Plos One*. <https://doi.org/10.1371/journal.pone.0034859>

Young, S. G., Slepian, M. L., & Sacco, D. F. (2015). Sensitivity to Perceived Facial Trustworthiness Is Increased by Activating Self-Protection Motives. *Social Psychological and Personality Science*, 6(6), 607. <https://doi.org/10.1177/1948550615573329>

Zebrowitz, L. A., Fellous, J.-M., Mignault, A., & Andreoletti, C. (2003). Trait impressions as overgeneralized responses to adaptively significant facial qualities: evidence from connectionist modeling. *Personality And Social Psychology Review: An Official Journal Of The Society For Personality And Social Psychology, Inc*, 7(3), 194–215.

https://doi.org/10.1207/S15327957PSPR0703_01

Children's sensitivity to facial trustworthiness

Figure Captions

Figure 1. The seven variations of the female face identity used as stimulus material, representing a trustworthiness continuum ranging from 1 (very untrustworthy) to 7 (very trustworthy) (a). Difference images between the most untrustworthy face (i.e., face 1) and each of the other six faces included in the continuum, as resulted from an image difference analysis describing the physical variations among the stimuli (b).

Figure 2. Example of experimental trials from the Oddmanout task (a). After the offset of the fixation cross, three faces randomly selected from the trustworthiness continuum were simultaneously presented, which remained on the screen until the participant responded. Example of experimental trials from the Pairwise Preference task (b). After the offset of the fixation cross, two faces randomly selected from the trustworthiness continuum were presented simultaneously and remained on the screen until the participant responded.

Figure 3. Average representational dissimilarity matrices resulting from the pairwise dissimilarity scores derived from the Oddmanout task for each age group. Yellow [light grey] represents maximum dissimilarity, while blue [dark grey] represents minimum dissimilarity. The matrix is symmetrical across the diagonal. See the online article for the colour version of this figure.

Figure 4. Hierarchical plots describing the results of the cluster analyses performed on the pairwise dissimilarity scores derived from the Oddmanout task for the 5-year-olds (a), the 7-year-olds (b), and the adults (c). Different clusters are depicted in different shades of grey.

Figure 5. Empirical cumulative distributions of intra-group cosine distances of dissimilarity scores (a). Intra-group cosine distances of dissimilarity scores recorded for the three age groups for each of the seven trustworthiness intensities (b). ** $p < .01$ *** $p < .001$

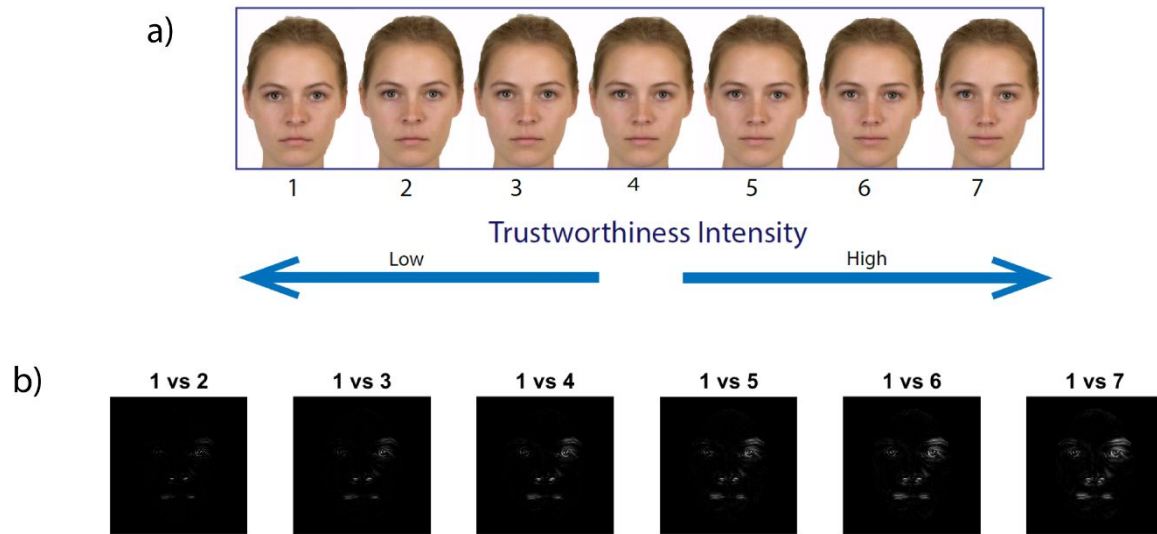
Children's sensitivity to facial trustworthiness

Figure 6. Mean percent accuracy obtained in the Pairwise Preference task by children in the three age groups (a). Mean trustworthiness scores obtained from the Pairwise Preference task for each of the seven trustworthiness intensities (b). $*p < .05$ $***p < .001$

For Review Only

Children's sensitivity to facial trustworthiness

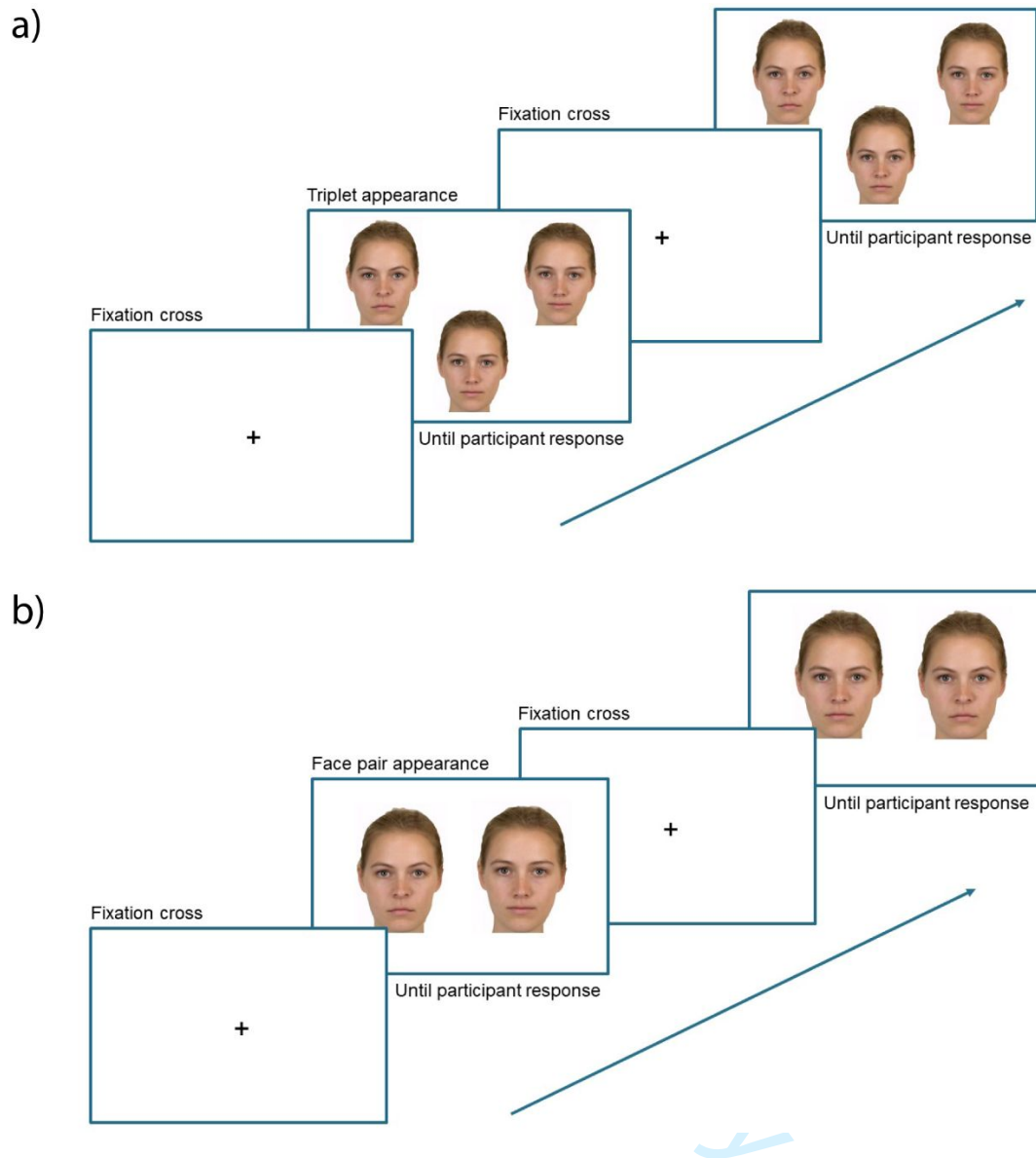
Figure 1



Review Only

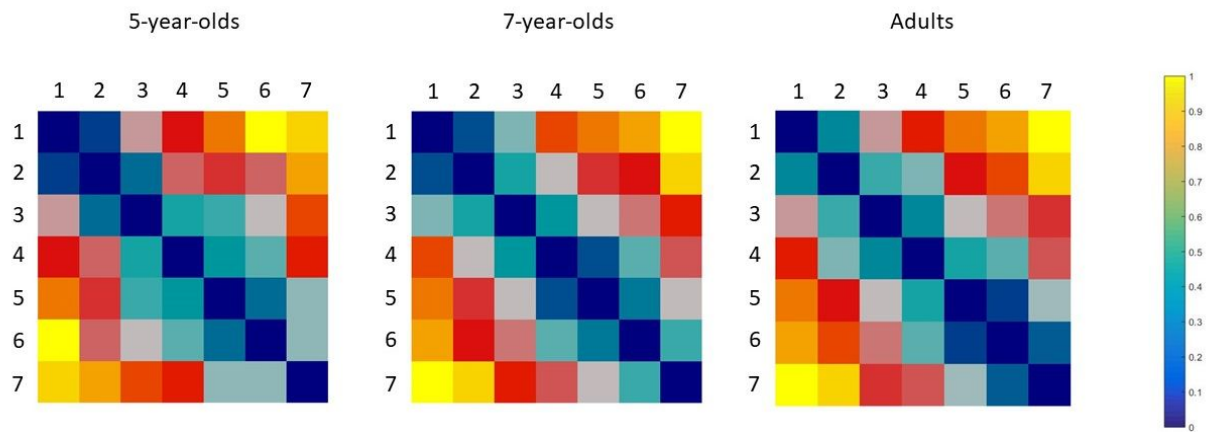
Children's sensitivity to facial trustworthiness

Figure 2



Children’s sensitivity to facial trustworthiness

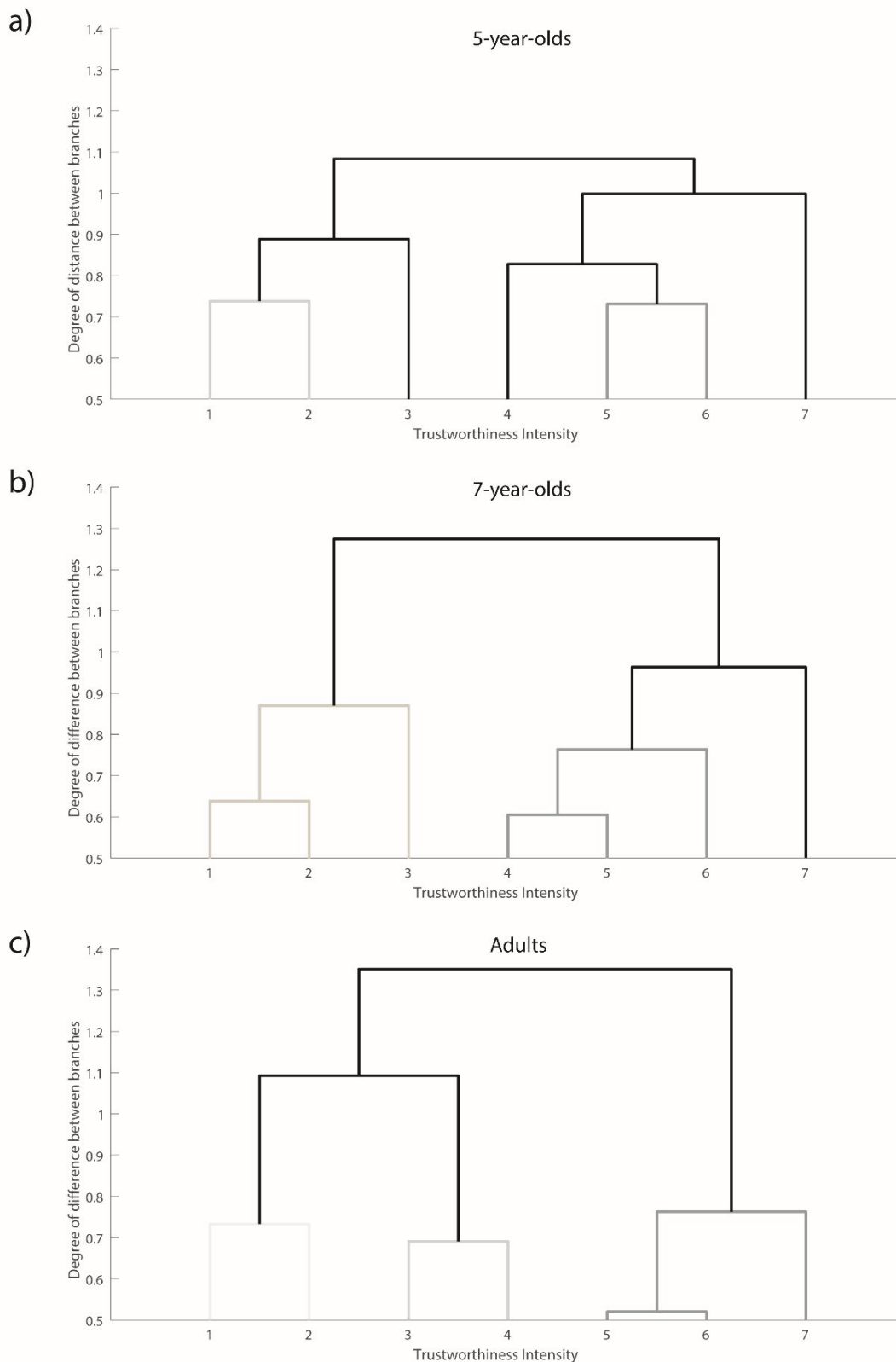
Figure 3



For Review Only

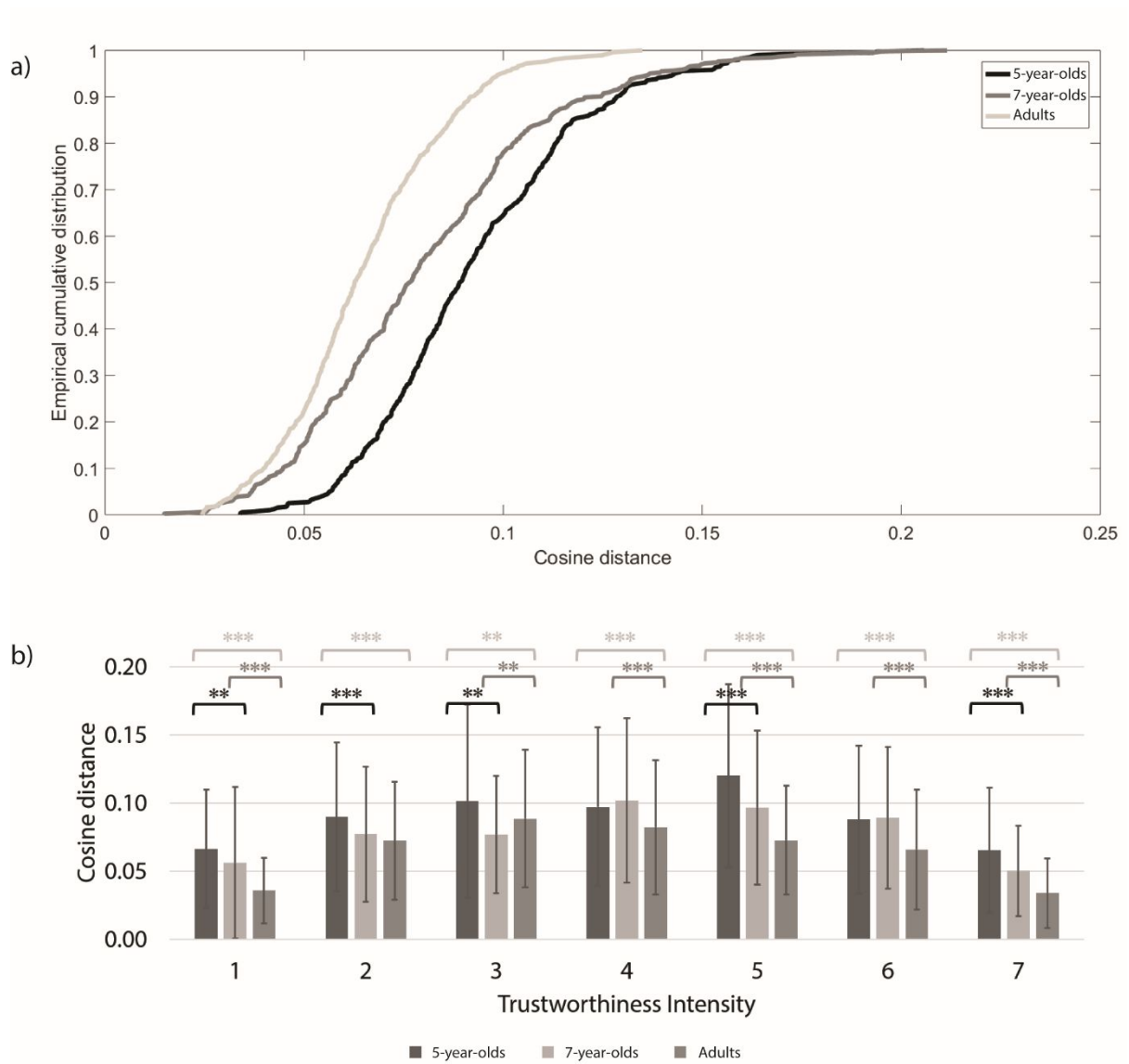
Children's sensitivity to facial trustworthiness

Figure 4



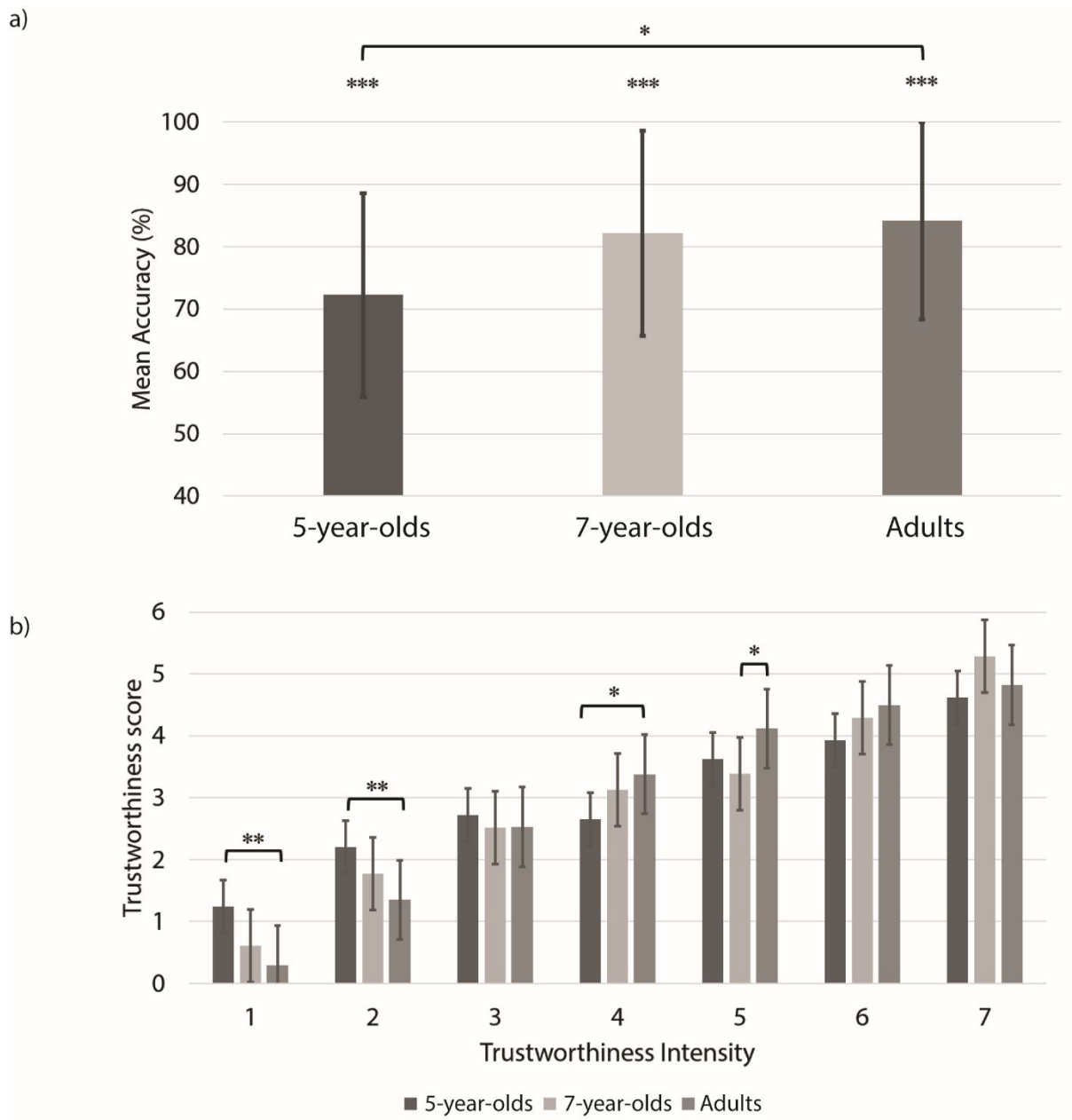
Children's sensitivity to facial trustworthiness

Figure 5



Children's sensitivity to facial trustworthiness

Figure 6



Dear Editor,

Please find with this letter the revised version of our manuscript entitled “Age-related differences in sensitivity to facial trustworthiness: Perceptual representation and the role of emotional development”.

We would like to thank you for your editorial decision on our manuscript, and both you and Reviewer 2 for your thoughtful suggestions on how to further improve the quality of the paper.

Below you will find our response to each comment raised by Reviewer 2 with a reference to where, in the text, the changes were made. To ensure clarity, the changes are also highlighted in red in the text.

We would like to thank you once again for your interest in our work. We hope that, with these requested changes, the manuscript will now be suitable for publication in Child Development.

Best regards,

Elisa Baccolo

For Review Only

Response to Reviewer 2

Q1. Main Comment

I wish the participants had been tested on more faces. Basing the study on a single face runs the risk of idiosyncratic findings.

Response.

We thank the Reviewer for prompting us to further stress this limitation of the study. To this end, we rephrased our comment on the matter as follows (p. 29): *“Third, the use of averaged identities in this study reduced the influence of idiosyncratic facial features on perceptual sensitivity to physical cues of trustworthiness and explicit judgements of trustworthiness; still, the use of one single average identity runs the risk of obtaining non-generalizable findings. Future studies shall put the generalizability of the current results under scrutiny by using a new set of averaged stimuli created starting from a different pool of face identities”.*

Q2. Main Comment

The authors mention three goals of the study. Goal#2 is to document differences in the representation of cues to trustworthiness. While the study indicates that there may be developmental differences, the exact nature of differences in the underlying representations is not clear. Perhaps the authors might wish to rephrase to indicate that their goal was to detect the presence of developmental differences rather than to identify the exact nature of these differences.

Response.

We thank the Reviewer for this suggestion. We rephrased the description of Goal #2 as follows (p. 6): *“By collecting dissimilarity scores, we were also able to address a second important issue left open in the literature, which is whether developmental differences exist between childhood and adulthood in the representation of physical cues to trustworthiness”.*

Q3. Main Comment

The framing and discussion of the outcome of goal # 3 (to examine the relation between emotion understanding and trustworthiness judgment in childhood) could also use some tweaking. While the paper is written to emphasize that there was a correlation at 5 years but not at 7, the rebuttal letter indicates that there was a correlation if all children’s data were combined. This brings up the potential interpretive problem of whether there are developmental differences. Was there any a priori reason to expect 5-year-olds to differ from 7-year-olds in the relation between emotion understanding and trustworthiness judgement? If so, that might justify differentiating between the two age groups. Otherwise, it may be best to avoid making strong claims about differences between 5-year-olds and 7-year-olds.

Response.

We thank the Reviewer for pushing us to make even more explicit the rationale behind our choice to include the 5- and the 7-year-old age groups in our sample in order to explore the impact of emotional development on children’s social perception of trustworthiness. Indeed, previous studies have pointed to 5 years as a relevant developmental time when subtle components of emotion

understanding first emerge, with other important components appearing during school years (Pons et al., 2004). This is why we opted for focusing on the 5-year-old and the 7-year-old groups in our attempt to explore the relation between children's emotional understanding abilities and their perception of trustworthiness traits from faces.

Given these important developmental differences in emotion understanding between the 5-year-old and the 7-year-old children, which were also confirmed by the significant independent-samples *t*-test on the z-transformed TEC scores from the two age groups, we believe that reporting the correlational analyses separately per each age group would reflect more accurately the data.

It is indeed true that we found a significant correlation if all children's data were combined, but it is also true that this overall correlation was driven by the 5-year-olds' data, as the correlation for the 7-year-old children was almost non-existent ($r = 0.02$) and far from being significant ($p = .921$).

Overall, prompted by the Reviewer's comment, we realized that we had failed to provide a convincing discussion of the rationale behind our choice of testing the two separate age groups in the Introduction. This has now been fixed, and the following paragraph has been amended (p. 8): *"By 3 to 4 years of age children can grasp other people's emotional states based on their facial expressions, but it is by the age of 5 years that they first develop critical components of emotion understanding (e.g., the situational causes of the outward expression of emotion), with other important components (i.e., the relationship between one's beliefs and his/her emotional states) appearing during school years, and reaching adult-like levels of performance in early adolescence. In light of this evidence, in the current study, we explored the relation between children's emotional development and their perception of trustworthiness traits from faces by focusing on the 5- to 7-year age range. Specifically, we investigated..."*

Running Head: Children's sensitivity to facial trustworthiness

**Age-related differences in sensitivity to facial trustworthiness: Perceptual representation
and the role of emotional development**

Date of submission: July 29th, 2019

For Review Only

Children's sensitivity to facial trustworthiness

Age-related differences in sensitivity to facial trustworthiness: Perceptual representation and the role of emotional development

Abstract

The ability to discriminate social signals from faces is a fundamental component of human social interactions whose developmental origins are still debated. In this study, 5-year-old ($N = 29$) and 7-year-old children ($N = 31$) and adults ($N = 34$) made perceptual similarity and trustworthiness judgements of a set of female faces varying in level of expressed trustworthiness. All groups represented perceived similarity of the faces as a function of trustworthiness intensity, but such representation becomes more fine-grained with development. Moreover, 5-year-olds' accuracy in choosing the more trustworthy face in a pair varied as a function of children's score at the Test of Emotion Comprehension, suggesting that the ability to perform face-to-trait inferences is related to the development of emotion understanding.

Keywords: face traits; trustworthiness; development; childhood; perceptual discrimination; face representational space; emotional development.

Children's sensitivity to facial trustworthiness

**Age-related differences in children's sensitivity to facial cues to trustworthiness:
Perceptual representation and the role of emotional development**

As humans, we display fine-grained sensitivity to non-verbal cues from faces, which we use to derive a great amount of social information about the person we are approaching. This competence has its origins in the very early stages of development. For example, by the age of 7 months infants are able to discriminate between different emotion categories, such as happiness and fear, based on facial expressions (Grossmann, 2010; Nelson, 1987), and use facial expression of emotional states to regulate their behaviour (Cohn & Tronick, 1983; Hirshberg, 1990).

One important piece of social information we automatically derive from faces is whether an individual represents safety or threat, i.e. whether he/she could be trusted and approached or better not trusted and avoided. Humans seem to be very prone to decode the facial information that drives social perception of trustworthiness: adults can identify whether a stranger represents a threat solely based on his/her facial characteristics after only 39 ms (Bar, Neta, & Linz, 2006), and use subtle differences between facial characteristics to generate explicit judgments of trustworthiness (Ames, Fiske, & Todorov, 2011; Todorov, Said, Engell, & Oosterhof, 2008). Trustworthiness judgements appear to be vehiculated by specific facial features (i.e., Action Units pattern for trustworthiness, see Jack & Schyns, 2015), such as up/downturned eyebrows, upward/downturned curving mouth and a wrinkling nose, which are also involved in emotion perception. Indeed, a possible mechanism through which these features induce social perception of trustworthiness is an overgeneralization of responses to facial configurations resembling emotional expressions (Said, Sebe, & Todorov, 2009; Zebrowitz, Fellous, Mignault, & Andreoletti, 2003). According to the emotion overgeneralization hypothesis, cues like lowered eyebrows, which, if stressed, might signal

Children's sensitivity to facial trustworthiness

anger, would induce the perceiver to make a dispositional assumption about an otherwise emotionally neutral face, resulting in social perception of unfriendliness (Ames et al., 2011). Accordingly, in both adults (e.g., Oosterhof & Todorov, 2008) and children (Caulfield, Ewing, Bank, & Rhodes, 2016) perceived trustworthiness is robustly associated to the attribution of emotional states, as overt angry expressions lead to subjective judgments of untrustworthiness, and overt happy expressions lead to subjective judgments of trustworthiness. Nevertheless, because social perception of trustworthiness and untrustworthiness also occurs for faces that are perceived as emotionally neutral (e.g., Lischke, Junge, Hamm, & Weymar, 2018), a thorough consideration of how the specific face information that elicits the perception of trustworthiness differ from that subtending the perception of emotional states is at stake (e.g., see review by Jack & Schyns, 2015). For example, Gill and colleagues (Gill, Garrod, Jack, & Schyns, 2014) showed that transient facial movements vehiculating trustworthiness judgements represent a unique Action Units configuration that differ from, and therefore should not be reduced to, Action Units configurations of emotional expressions.

Although much is known about the neurocognitive mechanisms underlying adults' sensitivity to facial cues to trustworthiness and/or their proneness to make trustworthiness inferences from faces, how this ability develops over time has been scarcely investigated. In particular, the question of whether and how this ability relates to the development of emotion understanding remains unexplored.

Available evidence suggests that the ability to perform explicit face-based trait judgments develops throughout childhood. At the age of 3 years, children are capable of making explicit judgements about how 'mean' or 'nice' a person appears to be, and by 6 years of age these attributions are at adult levels of consistency (Cogsdill, Todorov, Spelke, & Banaji, 2014). When specifically asked to judge whether a series of faces were trustworthy by rating each of them on a 3-point scale, 8-, 10-, and 12-year-old children all showed within-age agreement in

Children's sensitivity to facial trustworthiness

their judgments, although these judgments were more consistent and more similar to those of the adults in the older children than in the younger ones (Ma, Xu, & Luo, 2016). Ewing and colleagues (Ewing, Caulfield, Read, & Rhodes, 2015) also showed that face-based trustworthiness judgments influence children's behaviour, as 5- and 10-year-olds were more likely to place their trust in partners who looked trustworthy than in those who looked untrustworthy while playing an economic trust game. Overall, these findings led some authors to conclude that the ability to derive trait inferences from faces appears rather early in development, and do not require prolonged social experience as it builds on adaptive mechanisms developed to actively respond to threat (Cogsdill, Todorov, Spelke, & Banaji, 2014; LoBue, 2009).

Consistent with this hypothesis, two recent studies by Jessen & Grossmann (2016, 2017) suggest that even preverbal infants are sensitive to the face information that, in older children and adults, convey trust perception. In these studies, 7-months-old infants showed neural discrimination between neutral faces and those rated high or low on trustworthiness (Jessen & Grossmann, 2016), even when faces were presented subliminally (Jessen & Grossmann, 2017). Moreover, infants preferentially oriented their attention towards faces judged as trustworthy by adults rather than those judged as untrustworthy. These findings are in line with those showing that infants in the first year of life prefer prosocial individuals to antisocial others (Hamlin & Wynn, 2011; Van de Vondervoort & Hamlin, 2017), and preferably approach a stranger that their mother approached positively (Fein, 1975), just like older children (6- to 11-year-olds) tend to trust those who help others (Fu, Heyman, Chen, Liu, & Lee, 2015). Overall, this evidence suggests that humans are sensitive to other people's approachability from very early in the development.

Notwithstanding the relevance of these studies, they leave open a few questions, which the current study aimed to address. The first question is how perceptual sensitivity to fine-grained

Children's sensitivity to facial trustworthiness

differences in facial information subtending social perception of trustworthiness develops in time.

Previous studies investigated infants' (Jessen & Grossmann, 2016, 2017) and children's (e.g., Cogsdill et al., 2014; Cogsdill & Banaji, 2015; Ma et al., 2016) responses to computer-generated faces obtained from data-driven modelling (but see Cogsdill & Banaji, 2015) lying at the opposites of the trustworthiness continuum. On the one hand, this might have inflated participants' performance in the tasks due to high distinctiveness of the trustworthiness opposites. On the other hand, asking children to distinguish between very trustworthy and very untrustworthy faces may limit our understanding of their sensitivity to social signals from faces in real-life situations. Indeed, during our everyday social interactions we constantly decode subtle facial information and discriminate the slightest variations in other people's facial expressions. Finally, although artificial faces allow for a strictly controlled manipulation of the selected features, they may not fully reflect participants' expertise at face processing, including perceptual discrimination (e.g., see Crookes et al., 2015).

In the aim to overcome these limitations, in the current study we used as stimulus material a set of seven parametrically-manipulated variations of one real female face identity, slightly varying in level of perceived trustworthiness. This allowed us to trace developmental differences in children's perceptual sensitivity to subtle variations in physical cues to trustworthiness (see Method for a description of embedded in an exemplar of a face category (i.e., female faces) that is highly familiar to young children (see Quinn, Yahr, Kuhn, Slater, & Pascalis, 2002; Ramsey-Rennels & Langlois, 2006). Perceptual sensitivity was inferred from 5-year-olds', 7-year-olds', and adults' performance in an *Oddmanout task* (e.g., see Nishimura, Maurer, & Gao, 2009), where participants were simply asked to choose the face that appeared to be most different from the rest in a set of three simultaneously presented faces selected from our trustworthiness continuum. Children's and adults' perception of similarity/dissimilarity

Children's sensitivity to facial trustworthiness

among all the seven faces composing our trustworthiness continuum allowed us to compute pairwise dissimilarity scores, which provided a measure of their ability to discriminate the facial information that varied along the continuum.

By collecting dissimilarity scores, we were also able to address a second important issue left open in the literature, which is whether developmental differences exist between childhood and adulthood in the representation of physical cues to trustworthiness.

According to the influential face-space model proposed by Valentine (1991; also see review by Valentine, Lewis, & Hills, 2016), throughout the life span we build a representational model that maps the information that our visual system extracts from the faces we encounter in our social environment into a multidimensional perceptual space. Faces are organized in this space according to their perceived similarities along different dimensions, each of which represents a critical cue that is important for discriminating among individual faces (Valentine, 1991). Therefore, face-space organization is dependent on the amount and quality of one's own experience with faces, and becomes more fine-grained across development (Gao, Maurer, & Nishimura, 2010; Humphreys & Johnson, 2007; Rodger, Vizioli, Ouyang, & Caldara, 2015). Indeed, available evidence indicates that, although children represent faces in a multidimensional face space that has some adult-like characteristics at least from the age of 4 years (Jeffery et al., 2010), considerable refinement of this representation occurs throughout childhood. For example, separable representations of faces belonging to different categories defined by race, gender, and age emerge between 5 and 8 years of age (e.g., Short, Hatry, & Mondloch, 2011; Short, Lee, Fu, & Mondloch, 2014), and the representation of changeable facial traits like emotional expressions is also subject to critical changes across this same age range (Rodger et al., 2015). Other studies investigating identity discriminations have shown that, from the age of 7 years onwards, there are no differences between children and adults in the number of dimensions along which they represent faces, but children rely more heavily on

Children's sensitivity to facial trustworthiness

one single dimension (e.g., hair cues or eye colour) when making similarity judgments, whereas adults use all dimensions equally (Nishimura et al., 2009; Pedelty, Levine, & Shevell, 1985).

In the current study, we explored how face information subtending social perception of trustworthiness is represented in younger children's, older children's, and adults' memory by using dissimilarity scores derived from the *Oddmanout task* to build a Representational Dissimilarity Matrix (RDM) and perform a cluster analysis for each of the three age groups. This allowed us to describe the organization of face representation by unveiling how participants grouped together faces that varied in the level of expressed trustworthiness based on their perceived similarity (Sireci & Geisinger, 1992). It has been recently argued that perceptual, social and cultural experience all play a role in establishing and tuning face-trait mappings across development, by influencing both the perceptual representation of faces and the conceptual representation of others' personality traits (Over & Cook, 2018). In the current study, we focused on developmental changes in the perceptual representation of face information subtending trustworthiness perception. In light of earlier demonstrations that children from the age of 3 years can make explicit face-trait judgments on very distinctive trustworthiness opposites (Cogsdill et al., 2014), we hypothesized that the organization of younger children's representation of extreme physical cues to trustworthiness would be comparable to that of adults, but representation of the intermediate levels of the trustworthiness continuum would become more fine-grained with increasing age.

A third issue left unexplored by current developmental research relates to the impact of emotional development on children's social perception of trustworthiness. Recent evidence has shown that individual differences in personality and social behavior have an impact on how facial traits to trustworthiness are detected and utilized to perform social trait inferences (Baccolo & Macchi Cassia, 2019; Meconi, Luria, & Sessa, 2014; Young, Slepian, & Sacco, 2015). Most notably, impairments in social cognition and mentalizing abilities (i.e., attributing

Children's sensitivity to facial trustworthiness

mental states to other people) are known to be associated to diminished abilities to discriminate socially relevant characteristics of faces. For example, adults with ASD show abnormal face-based judgments of trustworthiness in association with atypical emotion recognition abilities (Adolphs, Sears, & Piven, 2001; Forgeot d'Arc et al., 2016). Despite this evidence supports the alleged relationship between emotion recognition skills and social judgements from facial cues, to the best of our knowledge, no studies have explored whether inter-individual variations in emotion comprehension abilities are reflected in corresponding variations in children's perceptual sensitivity to physical cues to trustworthiness as well as in their proneness to use such cues to generate trustworthiness judgements. Social cognition undergoes dramatic developmental changes in the first five years of life, and this is especially true for the so called *Theory of Mind*, which is the ability to understand other people's mental states, like desires, motives, emotions, and thoughts, and to respond to them adequately (Astington & Dack, 2008). By 3 to 4 years of age children can grasp other people's emotional states based on their facial expressions, but it is by the age of 5 years that they first develop critical components of emotion understanding (e.g., the situational causes of the outward expression of emotion), with other important components (i.e., the relationship between one's beliefs and his/her emotional states) appearing during school years, and reaching adult-like levels of performance in early adolescence (Pons, Harris, & de Rosnay, 2004).

In light of this evidence, in the current study, we explored the relation between children's emotional development and their perception of trustworthiness traits from faces by focusing on the 5- to 7-year age range. Specifically, we investigated whether 5- and 7-year-old children's ability to comprehend other people's emotions affected their perceptual representation of facial cues to trustworthiness and/or their judgements of perceived trustworthiness. To this end, we correlated children's score in the Test of Emotion Comprehension (TEC; Pons & Harris, 2000) to their performance in the *Oddmanout task* as well as in a second task – i.e., the *Pairwise*

Children's sensitivity to facial trustworthiness

Preference task – which allowed us to acquire explicit trustworthiness judgements on the seven faces of the continuum using a child-friendly procedure. Previous studies with children measured explicit judgements of trustworthiness by using rating scales (e.g., Cogsdill et al., 2014; Cogsdill & Banaji, 2015; Ma et al., 2016), which, however, could prove challenging for young children as they require reference to an internal rating scale and memory of the values assigned to previous faces, possibly resulting in inconsistent use of the scale across trials. Unlike these previous studies, in the *Pairwise Preference task* children were asked to indicate which face they would trust more within a pair randomly selected from our trustworthiness continuum, and the participant's response was used to compute a trustworthiness score for each face.

To sum up, the current study had three main aims: (1) to investigate whether perceptual sensitivity to face information subtending social perception of trustworthiness changes across childhood and into adulthood, (2) to explore the presence of age-related differences in the structure of the mental representation of facial cues to trustworthiness, and (3) to investigate whether children's emotion understanding skills affect their social perception of trustworthiness from faces.

Five- and 7-year-old children were selected as target age groups because we wanted our data to be comparable with those obtained by previous studies exploring children's trustworthiness judgments from faces, which targeted this same age range (Caulfield et al., 2016; Ewing et al., 2015). Moreover, the 5- to 8-years age range is also critical for the development of face representation, including the representation of changeable facial traits such as emotional expressions (e.g., Rodger et al., 2015), with the age of 7 marking the time when the structure of children's face representational space becomes adult-like (e.g., Nishimura et al., 2009). Finally, emotional intelligence and emotion comprehension show important improvements across the 5- to 7-year age range, when children become able not only to distinguish between

Children's sensitivity to facial trustworthiness

facial expressions of emotions and understand situational causes, but also to understand the mentalistic nature of emotions, such as the connection to desires and beliefs, and the distinction between expressed and felt emotion (Pons et al., 2004).

Materials and methods

Participants

Sample size was based on a Power Analysis for a univariate ANOVA with three groups (5-year-olds, 7-year-olds and adults), which revealed that about 64 participants should lead to an 80% chance to observe a significant effect with an alpha level of .05 and a large effect size. Data collection took place between October 2017 and May 2018. Analyses were performed on a total of 94 subjects: 29 5-year-old children (14 females; mean age = 5 years 5 months, range = 4 years 11 month - 5 years 11 months), 31 7-year-olds (12 females; mean age = 7 years 8 months, range = 7 years 1 month - 7 years 12 months), and 34 young adults (25 females; mean age = 23.03 years, range = 19 - 28 years). All children were recruited from preschools and schools within a major city area, and were attending preschool or primary school full-time at the time of testing. They all came from middle-class Caucasian families (except one Hispanic) and lived in a racially homogeneous neighbourhood. Adults were either undergraduate or graduate university students from middle-class families receiving course credits or recruited from the community by word of mouth on a voluntary basis. An additional 17 children (10 5-year-olds) were excluded from the final sample due to being distracted during the test. All procedures used in the current study complied with the Ethics Standards outlined by the Declaration of Helsinki (BMJ 1991; 302: 1194) and were approved by the Ethics Committee of the University of (omitted for blind reviewing). Adult participants signed an informed consent before testing; all participants' parents gave informed written consent prior to commencement of the study, and children gave their verbal assent before testing.

Children's sensitivity to facial trustworthiness

Stimuli

Stimuli were seven variations of one female facial identity reflecting a continuum of trustworthiness that ranged from 1 (very untrustworthy) to 7 (very trustworthy), interleaved by a neutral face (see Figure 1). The 7-steps continuum was created by morphing an averaged neutral face towards an averaged untrustworthy and an averaged trustworthy face using WebMorph (DeBruine, 2017), an online program for image transformation, specifically designed to perform face morphing and transforming. All the averaged faces were created by averaging three different face identities selected from the Chicago Face Database (Ma, Correll, & Wittenbrink, 2015). This Database provides a wide range of photographs of female and male identities rated on different face dimensions, including face trustworthiness. The averaged neutral face was created by averaging three different face identities that were rated as neutral on the trustworthiness dimension (i.e. not trustworthy nor untrustworthy), while the averaged faces used as references for morphing the neutral face towards the untrustworthy/trustworthy extremes were created by averaging the three faces rated as the most untrustworthy and trustworthy in the Database. We morphed the averaged neutral face 3 steps (30%, 60% and 100%) towards the very untrustworthy averaged face and 3 steps (30%, 60% and 100%) towards the very trustworthy averaged face, thus obtaining a 7-step trustworthiness continuum which included the neutral face. Figure 1 shows the output of an image difference analysis describing the physical variations among the seven faces included in the continuum. The analysis was carried out with the MATLAB (Mathworks Inc.) function *imshowpair* and the *diff* method to create a difference image between the most untrustworthy face (i.e., face 1) and each of the other faces. The images show that the physical aspects of the face that change the most along the continuum relate to the eyes area (i.e., eyes opening and eyebrows curvature), the

Children's sensitivity to facial trustworthiness

corners of the mouth (i.e., downturned in untrustworthy faces and upturned in trustworthy faces), and the nostrils (i.e, nose wrinkling).

----- Insert Figure 1 about here -----

Stimuli validation. In order to validate the stimuli (i.e., to ensure that the faces actually reflected a continuum of expressed trustworthiness), we asked an independent sample of 42 adults (34 females; mean age = 23.36 years; range = 18 - 35) to rate each step of the trustworthiness continuum on a scale ranging from 1 ("I wouldn't trust this person at all) to 9 ("I would definitely trust this person"). Adults' ratings were entered into a repeated-measures ANOVA with trustworthiness intensity as the within-subject factor, which attained statistical significance, $F(6,246) = 20.295$, $p < .001$, $p\eta^2 = 0.331$. A test of within-subjects contrasts revealed a significant linear trend, $F(1,41) = 58.760$, $p < .001$, $p\eta^2 = 0.589$, meaning that participants' trustworthiness judgements varied monotonically as a function of the faces' position along the trustworthiness continuum. Moreover, we explored whether our face stimuli elicited explicit judgements on other dimensions. Indeed, the abovementioned emotion overgeneralisation hypothesis (Said et al., 2009; Zebrowitz et al., 2003) suggests that trustworthiness judgements arise from an overgeneralisation of spontaneous responses to emotional expressions. Similarly, it has been suggested that face-to-trait trustworthiness judgements might be dependent on facial dimensions such as typicality (with atypical faces being perceived as less trustworthy; Sofer, Dotsch, Wigboldus, & Todorov, 2015; Todorov, Mende-Siedlecki, & Dotsch, 2013) or attractiveness (with attractive faces being perceived as more trustworthy; Hu, Abbasi, Zhang, & Chen, 2018; Oosterhof & Todorov, 2008; Schmidt, Leventsten, & Ambadar, 2012). Therefore, we asked a second group of young adults ($N = 46$, 25 females; mean age = 23.98 years; range = 19 - 35) to evaluate each face of the trustworthiness continuum on perceived emotion, typicality and attractiveness. For each face,

Children's sensitivity to facial trustworthiness

participants were asked if they perceived it to be emotional, typical and attractive, with the three questions presented in a random order. If participants responded the face to be emotional, they were asked to select the emotion they perceived (happy, angry, sad, scared or other) and rate the intensity of the selected emotion on a scale ranging from 1 ("This face is hardly happy/angry/sad/scared/other") to 9 ("This face is quite happy/angry/sad/scared/other"). Similarly, if participants responded the face to be typical or attractive, they were asked to rate the level of typicality/attractiveness on a scale ranging from 1 ("This face is hardly typical/attractive") to 9 ("This face is quite typical/attractive"). For all dimensions, negative answers (i.e., the face is not emotional, not typical or not attractive) were treated as zero values. To compare the average intensity of the trustworthiness judgements elicited by the seven faces of the continuum to that of the emotional, typicality and attractiveness judgements, we performed six independent *t*-tests, one for each of the four listed emotions (i.e., happiness, anger, sadness and fear), and one for each of the typicality and the attractiveness judgements. On average, our face stimuli were judged as more trustworthy than emotional (all $ps < .001$), whereas the intensity of perceived trustworthiness did not differ from that of perceived typicality, $t(12) = .845$, $p = .414$, and attractiveness, $t(12) = 1.043$, $p = .318$. These data suggest that, on average, trustworthiness judgments are poorly influenced by emotional cues: in fact, all faces from the trustworthiness continuum are, overall, scarcely perceived as emotional. Instead, trustworthiness judgements are not discernible from those on typicality and attractiveness in terms of overall intensity. Correlational analyses performed on intensity judgements elicited by each of the seven faces along the continuum revealed that increase in perceived trustworthiness was related to a corresponding increase in perceived attractiveness, $r = .965$, $p < .001$, 95% confidence interval (CI) [0.775, 0.995], and a decrease in perceived anger, $r = -.875$, $p = .01$, 95% confidence interval (CI) [-0.981, -0.356] (Figure 1). Therefore, although our face stimuli were scarcely perceived as emotional, in accord with the

Children's sensitivity to facial trustworthiness

literature (Said, Sebe, et al., 2009; Zebrowitz et al., 2003) the more untrustworthy faces in the continuum were hardly perceived as expressing negatively valenced emotional expressions. Moreover, and still in accord with the literature (Ma et al., 2016), the more trustworthy faces were perceived as more attractive, as attractiveness has been found to act as an heuristic property in trustworthiness judgements.

Apparatus and Procedure

Participants were tested individually in a quiet room (for the 5-year-old and the 7-year-old children the room was located in the school where the testing took place). All of them performed the *Oddmanout task* first, followed by the *Pairwise Preference task*. Both tasks were administered while participants seated 60 cm from a 17.3-inch touch-screen monitor with a resolution of 1080p. Stimulus presentation and response collection were controlled by ASF (Schwarzbach, 2011) and MATLAB Psychtoolbox for Windows (Brainard, 1997). For the 5-years-old and the 7-years-old children the experimenter manually controlled trial presentation by starting the trial as soon as the child fixated the monitor. Moreover, for the children the two tasks were interleaved by the administration of the *Test of Emotion Comprehension* (TEC) (Pons & Harris, 2000).

Oddmanout task

The *Oddmanout task* was designed to acquire measures of perceived dissimilarity among faces varying in their level of expressed trustworthiness. Participants observed three simultaneously presented faces appearing on the computer screen. On each trial, the three faces were all different, and randomly selected from the 7 trustworthiness intensities composing the continuum. Adults were asked to select the one they judged to be more different from the others by using the touch-screen interface of the computer. Children also provided they response by

Children's sensitivity to facial trustworthiness

using the touch-screen interface after being told the story of a young princess who was imprisoned in a castle tower by a witch, and hoped to be set free by the experimenter, who needed the child's help to succeed. The experimenter was able to steal the tower keys and reach the princess' room but the witch had created two avatars of the real princess in order to disguise him. Therefore, the experimenter asked the child "Can you help me to find out which one is the real princess? To find out, you should look carefully at the three faces and choose the one that looks more different from the others, as that one for sure is the real one!". To ensure that children thoroughly understood the instructions, we asked each child to repeat to the experimenter what they were meant to do. In addition, all children were given 5 practice trials prior to test trials so as to familiarize them with the task. Since we aimed at recording participants' responses for all possible triplet combinations to compute pairwise dissimilarity scores, participants viewed a total of 35 trials (the binomial coefficient obtained by selecting 3 faces out of a total of 7 faces, without considering triplets repetition or order). This way, each trustworthiness intensity appeared for a total of 15 times, and each pairwise comparison of the same two trustworthiness intensities appeared for a total of 5 times; the positions of the faces on the screen were randomized across trials. Each trial started with a central fixation cross (Figure 2), that remained on the screen for 1000 ms for the adult participants, or until the experimenter turned on the stimuli for the children. The stimuli remained on the screen until a response was made.

Pairwise Preference task

The *Pairwise Preference task* was designed to acquire explicit trustworthiness judgements on the seven faces of the continuum. On each trial, two faces randomly selected from the trustworthiness continuum simultaneously appeared on the computer screen, and participants were asked to select the face they trusted more by using the touch-screen interface. Children

Children's sensitivity to facial trustworthiness

were told a second story in which, after saving the princess, the experimenter got lost in a supermarket. While searching for the exit door, he bumped into two identical girls who pretend to know where the exit door is, one being the real princess he/she had saved earlier ("the good one"), and the other being one of the princess' avatars created by the witch ("the mean one"). In order to actually reach the exit door and find the way home, the experimenter asked the child "Can you help me to find out which one is the real princess? The princess is good, she wants to help me as we are friends now, she is someone I can trust. The other girl, instead, is mean, she wants to disguise me, and I should better not trust her. Can you help me to find out which one is the real princess who will help me to find the way home?". In order to ensure that task instructions were fully understood, before the task commenced each child was asked to explain back to the experimenter what he/she was meant to do. Since we aimed at recording participants' responses for all possible pairwise combinations of the seven faces of the continuum, the task was composed of 21 trials (the binomial coefficient obtained by selecting 2 faces out of a total of 7 faces, without considering pairs repetition or order). Each trustworthiness intensity was compared to all other intensities for a total of 6 times; the position of the faces on the screen was randomized across trials. Each trial started with a central fixation cross (Figure 2), that remained on the screen for 1000 ms for the adult participants, or until the experimenter turned on the stimuli for the children. The stimuli remained on the screen until a response was made.

Test of Emotion Comprehension

In between the *Oddmanout task* and the *Pairwise Preference task*, children were all administered the Italian version of the *Test of Emotion Comprehension* (TEC) (Albanese & Molina, 2008), which assesses nine dimensions of emotion understanding, namely the recognition of emotions based on facial expressions, the understanding of external emotional

Children's sensitivity to facial trustworthiness

causes, impact of desire on emotions, emotions based on beliefs, memory influence on emotion, emotion regulation, the ability to hide emotional states, understanding of mixed emotions and the relationship between morality and emotional experiences (Pons & Harris, 2000). The test is suitable for use with children aged 3 to 11 years, and consists of a booklet of illustrations divided into a set of stories, each depicting a different situation with four possible outcomes represented by different emotional facial expressions (happy, sad, angry, afraid). Children are asked to assign an emotion to the situation by selecting the corresponding facial expression; responses could be provided either verbally or by pointing to the drawing. The administration of the test took approximately 20 to 30 minutes. Each child obtained a score between 0 and 9 resulting from the sum of the partial scores attributed to the nine dimensions of emotion understanding; these scores were converted into *z scores*.

----- Insert Figure 2 about here -----

Results

Oddmanout task: Perceived dissimilarity between face pairs

In order to confirm that both the younger and the older children correctly understood the task, we performed two one-sample *t*-tests, one for each age group, on percent accuracy on a subset of trials ($N = 2$) in which one of the three faces was maximally distinguishable from the other two (i.e., trials in which faces 7-6-1 or 1-2-7 were shown). Performance accuracy in these trials was significantly above the chance level (50%) for both the 5-year-olds ($M = 76\%$), $t(28) = 5.477$, $p < .001$, and the 7-year-olds (92%), $t(30) = 12.490$, $p < .001$, as well as for the adults ($M = 87\%$), $t(33) = 9.574$, $p < .001$.

Representational Dissimilarity Matrices of Pairwise Dissimilarity Scores. To obtain a measure of perceived similarity/dissimilarity among all the seven faces composing the trustworthiness continuum, participant's response on each trial was used to compute three pairwise

Children's sensitivity to facial trustworthiness

dissimilarity scores, one for each face pair within the triplet. Once the subject had selected the most different face of the triad, the face pair composed of the non-selected faces was given a distance score of 0 (minimum dissimilarity), while the face pairs composed of the selected face and the non-selected ones were given a score of 1 (maximum dissimilarity). For each subject, the sum of the dissimilarity scores obtained for each face pair was scaled to 0-1 by dividing it by the number of trials in which that specific face pair appeared ($N = 5$), and used to build a 7x7 Representational Dissimilarity Matrix (RDM). The obtained RDM shows the level of perceived dissimilarity between face pairs: each column and row represent the dissimilarity scores of one trustworthiness intensity against all other trustworthiness intensities. Since the diagonal represents the dissimilarity of each trustworthiness level with itself, it contains only zero values, and each RDM is specular along the diagonal.

----- Insert Figure 3 about here -----

Within each age group, individual RDMs were averaged across subjects to obtain an RDM for each age group (see Figure 3). In order to explore age-related differences in the level of perceived dissimilarity among the faces composing the trustworthiness continuum, Pearson correlation analyses were performed on age-specific RDMs to assess whether the matrices resulting from the acquired dissimilarity scores had similar configurations for the three age groups. All three RDMs proved to highly correlate (5-year-olds and 7-year-olds: $r = 0.92$, $p < 0.001$, 95% confidence interval (CI) [0.799, 0.965]; 5-year-olds and adults: $r = 0.879$, $p < 0.001$, 95% confidence interval (CI) [0.721, 0.95]; 7-year-olds and adults: $r = 0.9334$, $p < 0.001$, 95% confidence interval (CI) [0.840, 0.973]), meaning that the pattern of perceived dissimilarities across the seven faces composing the trustworthiness continuum was similar for all groups.

Cluster Analysis on Pairwise Dissimilarity Scores. In order to investigate age-related differences in how facial cues to trustworthiness are perceptually represented in participants'

Children's sensitivity to facial trustworthiness

memory, pairwise dissimilarity scores were used to perform separate agglomerative hierarchical cluster analyses for each age group (see Everitt, 2011). Cluster analyses provided a description of how participants from each age group aggregate the faces from the trustworthiness continuum based on their perceptual similarities. The analyses were performed using the *average linkage* method in MATLAB (Mathworks Inc.), which groups items by creating a multilevel hierarchy to form a hierarchical tree based on average distance between items. Figure 4 represents the dendrograms resulting from the cluster analyses performed on the pairwise dissimilarity scores for the 5-year-olds, the 7-year-olds, and the adults. The X-axis represents the logical order derived from the dissimilarity judgements; the Y-axis represents the degree of perceived difference between faces, which is the distance that the function *linkage* computes between couple of items. Clusters are depicted in grayscales when their linkage is lower than 70% of the maximum linkage.

----- Insert Figure 4 about here -----

Intra-group consistency of Pairwise Dissimilarity Scores. In order to further explore the presence of age-related differences in sensitivity to variations in facial cues to trustworthiness, we compared intra-group consistency in participants' perceived dissimilarity across the three age groups. To this end, for each single age group we used RDMS of single participants to compute cosine distances of pairwise dissimilarity scores within each age group. Cosine distance can be defined as one minus the angle cosine of two vectors of an inner product space. A cosine distance of 0 is found whenever two vectors have the same orientation, while two perpendicular vectors have a cosine distance of 1. Cosine distance can therefore range between 0 (lowest distance) and 1 (greatest distance). We computed the cosine distance between all possible pairwise combinations of vectorized RDMS (upper triangular part of the matrices) of single subjects, separately for each of the three age groups. A univariate ANOVA on cosine

Children's sensitivity to facial trustworthiness

distances with age as between-subjects factor was conducted, which proved significant, $F(2,1429) = 136.006, p < 0.001, p\eta^2 = 0.160$. Pairwise dissimilarity scores were more consistent among adults ($M_{\text{cosine distance}} = 0.065, SD = 0.020$) than both the 5-year-olds ($M_{\text{cosine distance}} = 0.093, SD = 0.027$), $p < .001$, and the 7-year-olds ($M_{\text{cosine distance}} = 0.081, SD = 0.032$), $p < .001$, while the 7-year-olds were more consistent than the 5-year-olds, $p < .001$. Figure 5 shows the empirical cumulative distribution of cosine distances for the three age groups.

----- Insert Figure 5 about here -----

In order to investigate whether dissimilarity scores within each age group were equally consistent across participants for all trustworthiness intensities, we computed cosine distances between dissimilarity scores for each trustworthiness intensity. Indeed, each row (or column) of the RDM represents the pairwise dissimilarity scores between one trustworthiness intensity and each of the other trustworthiness intensities in the continuum. For each age group, we calculated the cosine distances between each row of the RDM of each single participant and the corresponding row of RDMs of all other participants. The obtained values were entered into a repeated-measures ANOVA with trustworthiness intensity as the within-subjects factor and age group as the between-subjects factor. Both main effects were significant (trustworthiness intensity: $F(6,8574) = 251.291, p < .001, p\eta^2 = 0.150, \text{power} = 1$; age group: $F(2,1429) = 110.788, p < .001, p\eta^2 = 0.134, \text{power} = 1$), as was the interaction between the two factors, $F(12,2850) = 17.596, p < .001, p\eta^2 = 0.024, \text{power} = 1$. For all age groups, a test of within-subjects contrasts revealed a significant quadratic trend for trustworthiness intensity, $ps < .001$ (see Figure 4).

Pairwise Preference task: explicit trustworthiness judgements

Response accuracy. Participant's response on each trial of the *Pairwise Preference task* was used to compute percent accuracy, i.e. the percentage of trials in which the subject selected the

Children's sensitivity to facial trustworthiness

face with higher trustworthiness intensity. To examine whether participants in all age groups systematically selected the face displaying more intense cues to trustworthiness, we performed three one-sample *t*-tests, one for each age group, which all proved significant, indicating that accuracy was significantly above chance for all age groups (5-year-olds: $t(28) = 7.177$, $p < .001$; 7-year-olds: $t(30) = 11.049$, $p < .001$; adults: $t(33) = 12.575$, $p < .001$). However, a univariate ANOVA with age group as the between-subjects factor also proved significant, $F(2,91) = 4.704$, $p = .011$, $\eta^2 = 0.094$, power = 0.776, showing age-related differences in the accuracy with which participants selected the more trustworthy face in the pair. Post-hoc comparisons (Bonferroni corrected) revealed that the 5-year-old children ($M = 72.25\%$, $SD = 16.69$) performed more poorly than the adults ($M = 84.17\%$, $SD = 15.85$), $p = .014$, and showed a marginal trend towards being also less accurate than the 7-year-olds ($M = 82.18\%$, $SD = 16.22$), $p = .06$ (see Figure 6).

----- Insert Figure 6 about here -----

Trustworthiness Scores. To obtain a measure of how consistently each face from the trustworthiness continuum was selected as the more trustworthy in a pair, participant's response on each trial of the *Pairwise Preference task* was used to compute a trustworthiness score for each face. Once the subject judged a face to be more trustworthy than the other, the selected face was given a score of 1, and the non-selected face was given a score of 0. For each subject, trustworthiness scores for each face of the continuum were summed across trials. Since each face appeared for a total of 6 times, the related trustworthiness score could range from 0 (if never selected as most trustworthy) to 6 (if always selected as more trustworthy). In order to investigate whether, for all age groups, participants' accuracy in selecting the more trustworthy face in a pair varied as a function of the intensity of the physical cues to trustworthiness displayed by the face to be selected (i.e., the higher the trustworthiness intensity, the higher the trustworthiness score), we performed a repeated-measures ANOVA

Children's sensitivity to facial trustworthiness

with trustworthiness intensity as the within-subjects factor and age group as the between-subjects factor. The analysis revealed a main effect of trustworthiness intensity, $F(6,546) = 119.327$, $p < .001$, $\eta^2 = 0.576$, power = 1, and a Trustworthiness Intensity x Age interaction, $F(12,174) = 2.821$, $p < 0.01$, $\eta^2 = 0.06$, power = 0.98 (see Figure 6). Post-hoc analysis (Bonferroni corrected) revealed that, for faces 1 and 2, the 5-year-olds provided judgements of higher trustworthiness compared to the adults ($p = .004$ and $p = .006$), whereas for face 5 the adults provided judgements of higher trustworthiness compared to the 7-year-olds ($p = .017$). A within-subjects contrast analysis revealed the presence of a significant linear trend in the trustworthiness scores for all age groups (all $ps < .001$).

Test of Emotion Comprehension (TEC)

In order to check for the existence of the expected developmental differences in children's performance at TEC (see Pons et al., 2004), we performed an independent-samples t -test on the z -transformed TEC scores of the 5-year-old and the 7-year-old age groups. The comparison proved significant, $t(58) = 5.45$, $p < 0.001$, revealing that the older children ($M = 8.03$, $SD = 0.91$) scored higher than the younger ones ($M = 6.17$, $SD = 1.65$) in the test. In light of this age-related overall difference in children's emotion comprehension abilities, we explored whether individual variability in these abilities was associated with differences in the perceptual representation of facial cues to trustworthiness and in using these cues to generate trustworthiness judgements by running correlational analyses between children's TEC scores and their performance in the *Oddmanout task* as well as in the *Pairwise Preference task*, separately for the two age groups.

TEC and Pairwise Dissimilarity scores. To investigate whether children's perceptual representation of facial cues to trustworthiness varied as a function of their TEC scores, we computed the intragroup cosine distance between the vectorized RDMs (i.e., dissimilarity

Children's sensitivity to facial trustworthiness

judgements between all face pairs) and the euclidean distance between the TEC scores separately for 5-year-olds and 7-year-olds. The euclidean distance between two points is the length of the linear segment that connects them. The Euclidean distance can be used as an index of similarity between couple of items. We therefore calculated the Euclidean distance between all possible pairwise combinations of TEC scores in order to have a measure of similarity between all TEC scores within each age group. We conjectured that, if mental representation of facial cues to trustworthiness varies as a function of the ability to understand other people's emotions, greater interindividual differences in TEC scores (measured as Euclidean distances) within a given age group should be related to greater interindividual differences in the RDMs (measured as cosine distances). For the 5-year-olds, the correlational analysis led to a significant, though very weak, positive correlation, $r = 0.053$, $p < .001$, 95% confidence interval (CI) [-0.045, 0.149]. For the 7-year-olds, no significant correlation was found, $r = -0.018$, $p = .685$, 95% confidence interval (CI) [-0.106, 0.0698].

TEC and Pairwise Preference accuracy. To investigate whether individual differences in the ability to attribute explicit trustworthiness judgements are related to individual differences in emotion comprehension, we correlated participants' response accuracy in the *Pairwise Preference task* with z-transformed TEC scores, separately for the 5-year-old and the 7-year-old children. A positive correlation was found for the younger children, $r = 0.44$, $p = .017$, 95% confidence interval (CI) [0.087, 0.694], but not for the older ones, $r = 0.02$, $p = .921$, 95% confidence interval (CI) [-0.338, 0.371].

Discussion

In the current study, we aimed at investigating the development of perceptual sensitivity to facial cues to trustworthiness and their representation in long term memory, taking into account the role of individual differences in emotion comprehension. To investigate perceptual

Children's sensitivity to facial trustworthiness

sensitivity to physical cues to trustworthiness, we tested younger and older children and adults in an *Oddmanout task* designed to provide measures of perceived similarity/dissimilarity among faces slightly varying in their level of expressed trustworthiness. Analysis of participants' performance revealed that, already at the age of 5 years, children represent faces as a function of the level of the trustworthiness they express. In fact, average RDMs describing perceived dissimilarity between face pairs were highly correlated across age groups, meaning that there were no age-related qualitative differences in participants' sensitivity to variations in physical cues to trustworthiness. However, our findings also showed that such sensitivity becomes increasingly fine-grained with development, as intra-group consistency in dissimilarity judgements increased with age. Indeed, adults' judgements were more consistent than those provided by the 5-year-old and the 7-year-old children, and 7-year-olds' judgements were, in turn, more consistent than those provided by the 5-year-olds. The *Oddmanout task* also allowed us to examine age-related differences in the consistency among participants' dissimilarity scores for each trustworthiness intensity. The adults' scores were more consistent than those of the 5-year-olds for all trustworthiness intensities, and for almost all trustworthiness intensities adults were also more consistent than the 7-year-olds. This is in line with earlier evidence that within-age consistency in explicit judgements of trustworthiness increases with age (Ma et al., 2016).

The finding that participants' proficiency at attributing dissimilarity judgements for faces that only slightly vary in the level of expressed trustworthiness improves with age is also evident from the results of the cluster analysis performed on dissimilarity scores from the *Oddmanout task*. As already noted, both adults and children represented faces as a function of the intensity of the physical cues to trustworthiness they express. However, the structure of the hierarchical clustering becomes more differentiated with increasing age. Five-year-old children's dissimilarity judgments formed one cluster composed of the two most untrustworthy

Children's sensitivity to facial trustworthiness

faces (faces 1 and 2) and another cluster composed of two trustworthy faces (faces 5 and 6). Proceeding along the hierarchy, these two clusters enlarged including all the three untrustworthy faces (faces 1, 2, 3) on one side, and all the three trustworthy faces (faces 5, 6, 7) on the other, with the neutral face associated to this latter cluster. The structure of children's representation becomes more differentiated at 7 years of age, when already at the bottom of the hierarchy the three untrustworthy faces (faces 1, 2, 3) form one cluster, and the two moderately trustworthy faces (faces 5 and 6) compose another cluster, which also includes the neutral face (face 4), and to which the most trustworthy face (face 7) is hierarchically associated. The hierarchical clustering of dissimilarity judgements reaches the greatest differentiation in adults, who show three different clusters including, respectively, the two most untrustworthy faces (faces 1 and 2), the neutral face and the one next to the neutral (faces 3 and 4), and the three trustworthy faces (5, 6, 7). Proceeding along the hierarchy, the clusters including the more untrustworthy faces (faces 1 and 2) and the more neutral faces (faces 3 and 4) are then incorporated into a larger supra-cluster. In keeping with Valentine's theory (Valentine et al., 2016), these results corroborate the idea that the representation of facial characteristics that are relevant for discriminating among individual faces and/or face types, becomes more fine-grained and differentiated across development. To the best of our knowledge, our results provide the first evidence of how the perceptual space which maps the physical face information that subtend perception of a social trait changes across childhood and into adulthood.

In fact, in addition to age-related differences in the structure of the mental representation of facial cues to trustworthiness, we also observed consistency in performance across age groups. Indeed, analyses on intra-group cosine distances of dissimilarity scores revealed that participants' agreement in attributing dissimilarity judgements showed similar variations across trustworthiness levels for all age groups. Consistency of dissimilarity scores of both

Children's sensitivity to facial trustworthiness

younger and older children, as well as adults, showed a significant quadratic trend, with less consistent scores for the central hub of the trustworthiness continuum (around the neutral face) and most consistent scores for the continuum extremes (very trustworthy and very untrustworthy faces). This finding replicates earlier evidence that, in adults, facial cues to trustworthiness that yield to more extreme trustworthiness judgments (i.e., very trustworthy or very untrustworthy) are easier to discriminate than those yielding to less extreme judgments, independently of their valence (i.e., whether the face is very trustworthy or very untrustworthy) (Baccolo & Macchi Cassia, 2019). Accordingly, neuroimaging studies with adults reported a similar valence-independent sensitivity of the amygdala to trustworthiness cues (Said, Baron, & Todorov, 2009; Said, Dotsch, & Todorov, 2010), and electrophysiological studies with infants reported neural discrimination between neutral faces and both very trustworthy (+3 SD) and very untrustworthy (-3 SD) faces, but not between trustworthy and untrustworthy faces (Jessen & Grossmann, 2016).

The finding that faces including more intense physical cues to social traits enjoy a processing advantage over those including less intense cues, irrespective of the valence of such cues, might be due to the fact that responding to socially connoted faces is of greater adaptive value than responding to neutral ones, just like responding to emotional faces is easier and faster than responding to neutral ones, even when only little face information conveying emotional intensity is available (Roesch, Sander, Mumenthaler, Kerzel, & Scherer, 2010). Another possibility is that participants in the current study agreed more on attributing dissimilarity judgments for faces at the extreme opposites of the trustworthiness continuum than for faces at the centre of the continuum because the former were perceived as less prototypical than the latter. Indeed, previous studies showed that deviations in typicality explain the amygdala response better than valence (Said et al., 2010). Nonetheless, we view this possibility as unlikely in light of the results of our stimulus validation procedure, which

Children's sensitivity to facial trustworthiness

revealed that judgements of perceived trustworthiness for each of the seven faces of the continuum were not discernible from judgements of perceived typicality in terms of overall intensity.

An age-related trend similar to that observed for perceptual sensitivity to physical cues to trustworthiness was present in our data also for explicit trustworthiness judgements recorded during the *Pairwise Preference task*. Just like adults, both younger and older children performed above chance when asked to select the more trustworthy face in a pair, and for all age groups the pairwise preference scores varied linearly, showing that participants made explicit judgements of trustworthiness intensity for each of the seven faces as a function of the stimulus position along the trustworthiness continuum. At the same time, though, data also showed that the adults were more accurate than the younger children, but not the older ones, suggesting that, by age 7 years, children's performance has become adult-like. This is in accord with earlier demonstration that by 7 years of age children are as sensitive as adults in attributing explicit trustworthiness judgements (Cogsdill et al., 2014; Cogsdill & Banaji, 2015). Of note, in addition to being less accurate and less consistent in their explicit judgements of trustworthiness, the 5-year-old children overestimated the trustworthiness intensity of the faces at the untrustworthy extreme of the continuum (faces 1 and 2), for which their judgements overrode those of the adults. This finding is consistent with earlier reports of inflated trustworthiness ratings for untrustworthy faces in 5-year-old children (Caulfield, Ewing, Bank, & Rhodes, 2016), and further suggests that sensitivity to this social cue from faces is not fully developed at this young age.

In fact, our findings showed that, at 5 years but not at 7 years, children differed in their ability to make trustworthiness judgements from faces as a function of their emotion understanding skills, as measured through the *Test of Emotion Comprehension*: higher scores at TEC were related to higher performance at successfully detecting the more trustworthy face

Children's sensitivity to facial trustworthiness

in a pair in the *Pairwise Preference task*. In contrast, we found no evidence that children's emotion comprehension skills modulated how they represented facial cues to trustworthiness, as we observed no significant correlations between TEC scores and children's intragroup cosine distance between the vectorized RDMs built on pairwise dissimilarity scores obtained from the *Oddmanout task*.

The finding that the association between emotion comprehension skills and social perception of trustworthiness in the *Pairwise Preference task* was confined to the 5-year-old further suggests that emotional development impacts the perception of social traits from faces. Indeed, although performance of the 5-year-old children in the *Pairwise Preference task* was only slightly lower than that of the older children, and similarly variable (5-year-olds: $SD = 16.69$; 7-year-olds: $SD = 16.22$), in line with previous evidence (Pons et al., 2004) the 5-year-olds scored significantly lower than the older children at TEC, and showed higher variability in their responses to the questionnaire (5-year-olds: $SD = 1.65$; 7-year-olds: $SD = 0.91$). This indicates that emotion understanding abilities are not evenly distributed in our 5-year-old sample, and is in accord with earlier reports of 5 years being the first developmental time in which critical components of emotion understanding (i.e., understanding of the outward expression of emotion and their situational causes) emerge, with other important components (i.e., the understanding of the mentalistic nature of emotions) appearing only later in development (Pons et al., 2004). The positive correlation between 5-year-olds' performance at TEC and the *Pairwise Preference task* indicates that children's early emotion understanding abilities are associated to their proficiency at inferring trustworthiness traits from faces. Overall these findings suggest that it is right in the earlier critical stages of the development of emotion understanding that the ability to use face information to infer trustworthiness traits builds on the ability to consistently use transient facial cues to infer internal emotional states.

Children's sensitivity to facial trustworthiness

As the present study is the first to explore the relation between emotional development and perception of social traits from faces, it has a number of limitations that could be addressed in future research. First, additional measures could be used to examine children's inferences about trustworthiness from faces. We assessed this ability by measuring children's responses in a paired-preference task in which they selected the face they could trust more in a pair. Although this is a viable way to evaluate young children's use of face information to infer social traits without straining on their limited cognitive resources, it will be important to develop new measures to explore how such inferences affect children's approach/avoidance behaviour (e.g., see Ewing et al., 2015), and how trust behaviour is associated with emotional development.

Second, more work is needed to examine how the findings from the present study generalize across culture, and across face ethnicity. It has been shown that Caucasian and Asian adults rely on similar facial cues to judge trustworthiness (Xu et al., 2012), and trust perception generalizes across face ethnicity (Birkás, Dzhelyova, Lábadi, Bereczkei, & Perrett, 2014). However, this may not be the case for children, especially in light of the fact that, starting from preschool years, children increasingly draw on racial information when making social judgments (e.g., Bigler & Liben, 2007; Killen & Stangor, 2001). Moreover, cultural differences may emerge in how trust perception is related to emotional understanding, as cultural norms and practice affect display rules.

Third, the use of averaged identities in this study reduced the influence of idiosyncratic facial features on perceptual sensitivity to physical cues of trustworthiness and explicit judgements of trustworthiness; still, the use of one single average face identity runs the risk of obtaining non-generalizable findings. Future studies shall put the generalizability of the current results under scrutiny by using a new set of averaged stimuli created starting from a different pool of face identities.

Children's sensitivity to facial trustworthiness

Finally, future studies shall investigate whether the relation between trust perception and individual variability in emotional development generalizes across age, for example by testing whether individual differences in temperament and self-regulation abilities affect sensitivity to facial cues to trustworthiness in preverbal infants, who have been shown to discriminate between faces based on those cues (Jessen & Grossmann, 2017).

To sum up, the present study shows that, already at the age of 5 years, the mental representation of perceived differences between facial cues associated to trustworthiness is organized along a continuum of trustworthiness intensity. Nevertheless, with increasing age, this representation becomes more fine-grained. Similarly, although the 5-year-olds were overall less accurate than the adults in their explicit judgements of trustworthiness and showed a positivity bias for the very untrustworthy faces, those judgements were linearly distributed along the trustworthiness continuum for both children and adults. Most importantly, at 5 years of age more accurate judgements of trustworthiness were associated with more accurate emotional understanding. Overall, implicit measures of perceptual sensitivity to physical cues of trustworthiness and explicit judgements of trustworthiness intensity converged in showing that the ability to discriminate facial cues associated to trustworthiness and to use such cues to make trustworthiness attributions is apparent at the age of 5 years but becomes more adult-like by the age of 7 years, and its development is related to the development of emotion comprehension.

Children's sensitivity to facial trustworthiness

References

- Albanese, O., & Molina, P. (2008). *Lo sviluppo della comprensione delle emozioni e la sua valutazione: la standardizzazione italiana del Test di comprensione delle emozioni (TEC)*. Milano : Unicopli. 2008.
- Ames, D. L., Fiske, S. T., & Todorov, A. T. (2011). *Impression Formation: A Focus on Others' Intents*. *The Oxford Handbook of Social Neuroscience*. Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780195342161.013.0028>
- Astington, J. W., & Dack, L. A. (2008). Theory of mind. In M. M. Haith & J. B. Benson (Eds.), *Encyclopedia of infant and early childhood development* (pp. 343–356). San Diego, CA: Academic Press.
- Astington, J. W., & Hughes, C. (2013). Theory of mind: Self-reflection and social understanding. In P. D. Zelazo (Ed.), *Oxford Handbook of Developmental Psychology*. New York: Oxford University Press.
- Baccolo, E., & Macchi Cassia, V. (2019). Individual differences in perceptual sensitivity and representation of facial signals of trustworthiness. *Journal of Experimental Psychology: Human Perception and Performance*. <https://doi.org/10.1037/xhp0000601>
- Bar, M., Neta, M., & Linz, H. (2006). Very first impressions. *Emotion (Washington, D.C.)*, 6(2), 269–278. <https://doi.org/10.1037/1528-3542.6.2.269>
- Bigler, R. S., & Liben, L. S. (2007). Developmental Intergroup Theory: Explaining and Reducing Children's Social Stereotyping and Prejudice. *Current Directions in Psychological Science*, 16(3), 162–166. <https://doi.org/10.04.87/j.1467-8721.2007.00496.x>
- Birkás, B., Dzhelyova, M., Lábadi, B., Bereczkei, T., & Perrett, D. I. (2014). Cross-cultural perception of trustworthiness: The effect of ethnicity features on evaluation of faces' observed trustworthiness across four samples. *Personality and Individual Differences*, 69,

Children's sensitivity to facial trustworthiness

56–61. <https://doi.org/10.1016/j.paid.2014.05.012>

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*, 433–436.

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* (London, England: 1953), *107*(3), 503–518. <https://doi.org/10.1111/bjop.12156>

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/http://10.0.3.248/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. <https://doi.org/10.1177/0956797614523297>

Cohn, J. F., & Tronick, E. Z. (1983). Three-month-old infants' reaction to simulated maternal depression. *Child Development*, *54*(1), 185–193. <https://doi.org/10.2307/1129876>

Cohn, J. F., & Tronick, E. Z. (1987). Mother-Infant Face-to-Face Interaction: The Sequence of Dyadic States at 3, 6, and 9 Months. *Developmental Psychology*, *23*(1), 68–77. <https://doi.org/10.1037/0012-1649.23.1.68>

Crookes, K., Ewing, L., JD, G., Kloth, N., WG, H., Oxner, M., ... Rhodes, G. (2015). How Well Do Computer-Generated Faces Tap Face Expertise? *Plos One*. <https://doi.org/10.1371/journal.pone.0141353>

DeBruine, L. (2017). Webmorph (Version v0.0.0.9001). *Zenodo*. <https://doi.org/10.5281/zenodo.1073696>

Everitt, B. (2011). *Cluster Analysis* (Vol. 5th ed). Chichester, West Sussex: Wiley.

Ewing, L., Caulfield, F., Read, A., & Rhodes, G. (2015). Perceived trustworthiness of faces drives trust behaviour in children. *Developmental Science*, *18*(2), 327–334.

Children's sensitivity to facial trustworthiness

<https://doi.org/10.1111/desc.12218>

- Fein, G. G. (1975). Children's sensitivity to social contexts at 18 months of age. *Developmental Psychology*, *11*(6), 853–854. <https://doi.org/10.1037/0012-1649.11.6.853>
- Forgeot d'Arc, B., Ramus, F., Lefebvre, A., Brottier, D., Zalla, T., Moukawane, S., ... Delorme, R. (2016). Atypical Social Judgment and Sensitivity to Perceptual Cues in Autism Spectrum Disorders. *Journal of Autism & Developmental Disorders*, *46*(5), 1574–1581. <https://doi.org/10.1007/s10803-014-2208-5>
- Fu, G., Heyman, G. D., Chen, G., Liu, P., & Lee, K. (2015). Children trust people who lie to benefit others. *Journal of Experimental Child Psychology*, *129*, 127–139. <https://doi.org/http://10.03.248/j.jecp.2014.09.006>
- Gao, X., Maurer, D., & Nishimura, M. (2010). Similarities and Differences in the Perceptual Structure of Facial Expressions of Children and Adults. *Journal of Experimental Child Psychology*, *105*(1), 98–115. <https://doi.org/http://dx.doi.org/10.1016/j.jecp.2009.09.001>
- Gill, D., Garrod, O. G. B., Jack, R. E., & Schyns, P. G. (2014). Facial Movements Strategically Camouflage Involuntary Social Signals of Face Morphology. *Psychological Science* (0956-7976), *25*(5), 1079–1086. <https://doi.org/10.1177/0956797614522274>
- Grossmann, T. (2010). The development of emotion perception in face and voice during infancy. *Restorative Neurology and Neuroscience*, *28*, 219–236. <https://doi.org/10.3233/RNN-2010-0499>
- Hamlin, J. K., & Wynn, K. (2011). Young infants prefer prosocial to antisocial others. *Cognitive Development*, *26*, 30–39. <https://doi.org/10.03.248/j.cogdev.2010.09.001>
- Hirshberg, L. (1990). When Infants Look to Their Parents: II. Twelve-Month-Olds' Response to Conflicting Parental Emotional Signals. *Child Development*, *61*(4), 1187. <https://doi.org/10.2307/1130886>
- Hu, Y., Abbasi, N. U. H., Zhang, Y., & Chen, H. (2018). The Effect of Target Sex, Sexual

Children's sensitivity to facial trustworthiness

- Dimorphism, and Facial Attractiveness on Perceptions of Target Attractiveness and Trustworthiness. *Frontiers In Psychology*, 9, 942. <https://doi.org/10.3389/fpsyg.2018.00942>
- Humphreys, K., & Johnson, M. H. (2007). The development of “face-space” in infancy. *Visual Cognition*, 15(5), 578–598. <https://doi.org/10.1080/13506280600943518>
- Jack, R. E., & Schyns, P. G. (2015). Review: The Human Face as a Dynamic Tool for Social Communication. *Current Biology*, 25(14), R621. <https://doi.org/10.1016/j.cub.2015.05.052>
- Jeffery, L., Haynes, R., Pellicano, E., Rhodes, G., McKone, E., & Firth, E. (2010). Four-to-six-year-old children use norm-based coding in face-space. *Journal of Vision*, 10(5). <https://doi.org/10.1167/10.5.18>
- Jessen, S., & Grossmann, T. (2016). Neural and Behavioral Evidence for Infants' Sensitivity to the Trustworthiness of Faces. *Journal Of Cognitive Neuroscience*, 28(11), 1728–1736. https://doi.org/10.1162/jocn_a_00999
- Jessen, S., & Grossmann, T. (2017). Neural evidence for the subliminal processing of facial trustworthiness in infancy. *Neuropsychologia*. <https://doi.org/10.1016/j.neuropsychologia.2017.04.025>
- Killen, M., & Stangor, C. (2001). Children's Social Reasoning about Inclusion and Exclusion in Gender and Race Peer Group Contexts. *Child Development*, 72(1), 174. <https://doi.org/10.1111/1467-8624.00272>
- Lischke, A., Junge, M., Hamm, A. O., & Weymar, M. (2018). Enhanced processing of untrustworthiness in natural faces with neutral expressions. *Emotion (Washington, D.C.)*, 18(2), 181–189. <https://doi.org/10.1037/emo0000318>
- LoBue, V. (2009). More than just another face in the crowd: superior detection of threatening facial expressions in children and adults. *Developmental Science*, 12(2), 305–313.

Children's sensitivity to facial trustworthiness

<https://doi.org/10.1111/j.1467-7687.2008.00767.x>

Ma, D. S., Correll, J., & Wittenbrink, B. (2015). The Chicago face database: A free stimulus set of faces and norming data. *Behavior Research Methods*.

<https://doi.org/10.3758/s13428-014-0532-5>

Ma, F., Xu, F., & Luo, X. (2016). Children's Facial Trustworthiness Judgments: Agreement and Relationship with Facial Attractiveness. *Frontiers In Psychology*, 7, 499.

<https://doi.org/10.3389/fpsyg.2016.00499>

Meconi, F., Luria, R., & Sessa, P. (2014). Individual differences in anxiety predict neural measures of visual working memory for untrustworthy faces. *Social Cognitive And Affective Neuroscience*, 9(12), 1872–1879. <https://doi.org/10.1093/scan/nst189>

Nelson, C. A. (1987). The recognition of facial expressions in the first two years of life: mechanisms of development. *Child Development*, 58(4), 889–909.

<https://doi.org/10.2307/1130530>

Nelson, C. A., & De Haan, M. (1996). Neural Correlates of Infants' Visual Responsiveness to Facial Expressions of Emotion. *Developmental Psychobiology*, 29(7), 577.

<https://doi.org/10.1002/9780470147658.chpsy0201>

Nishimura, M., Maurer, D., & Gao, X. (2009). Exploring children's face-space: A multidimensional scaling analysis of the mental representation of facial identity. *Journal of Experimental Child Psychology*, 103, 355–375.

<https://doi.org/http://10.03.248/j.jecp.2009.02.005>

Oosterhof, N. N., & Todorov, A. (2008). The Functional Basis of Face Evaluation. *Proceedings of the National Academy of Sciences of the United States of America*, 105(32), 11087.

<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. <https://doi.org/10.1016/j.cognition.2017.10.002>

<https://doi.org/10.1016/j.cognition.2017.10.002>

Children's sensitivity to facial trustworthiness

- Pedelty, L., Levine, S. C., & Shevell, S. K. (1985). Developmental changes in face processing: Results from multidimensional scaling. *Journal of Experimental Child Psychology*, *39*(3), 421–436. [https://doi.org/10.1016/0022-0965\(85\)90050-5](https://doi.org/10.1016/0022-0965(85)90050-5)
- Pons, F., Harris, P., & de Rosnay, M. (2004). Emotion comprehension between 3 and 11 years: Developmental periods and hierarchical organization. *European Journal of Developmental Psychology*, *1*(2), 127–152. <https://doi.org/http://10.0.4.56/17405620344000022>
- Pons, F., & Harris, P. L. (2000). *TEC (Test of Emotion Comprehension)*. Oxford University Press.
- Quinn, P. C., Yahr, J., Kuhn, A., Slater, A. M., & Pascalis, O. (2002). Representation of the gender of human faces by infants: a preference for female. *Perception*, *31*(9), 1109–1121. <https://doi.org/10.1068/p3331>
- Ramsey-Rennels, J. L., & Langlois, J. H. (2006). Infants' Differential Processing of Female and Male Faces. *Current Directions in Psychological Science*, *15*(2), 59–62. <https://doi.org/10.0.4.87/j.0963-7214.2006.00407.x>
- Rodger, H., Vizioli, L., Ouyang, X., & Caldara, R. (2015). Mapping the development of facial expression recognition. *Developmental Science*, *18*(6), 926–939. <https://doi.org/10.1111/desc.12281>
- Roesch, E. B., Sander, D., Mumenthaler, C., Kerzel, D., & Scherer, K. R. (2010). Psychophysics of emotion: The QUEST for Emotional Attention. *Journal of Vision*. <https://doi.org/10.1167/10.3.4>
- Said, C. P., Baron, S. G., & Todorov, A. (2009). Nonlinear amygdala response to face trustworthiness: contributions of high and low spatial frequency information. *Journal Of Cognitive Neuroscience*, *21*(3), 519–528. <https://doi.org/10.1162/jocn.2009.21041>
- Said, C. P., Dotsch, R., & Todorov, A. (2010). The amygdala and FFA track both social and

Children's sensitivity to facial trustworthiness

- non-social face dimensions. *Neuropsychologia*, 48(12), 3596–3605.
<https://doi.org/10.1016/j.neuropsychologia.2010.08.009>
- Said, C. P., Sebe, N., & Todorov, A. (2009). Structural resemblance to emotional expressions predicts evaluation of emotionally neutral faces. *Emotion (Washington, D.C.)*, 9(2), 260–264. <https://doi.org/10.1037/a0014681>
- Schmidt, K., Leventsten, R., & Ambadar, Z. (2012). Intensity of Smiling and Attractiveness as Facial Signals of Trustworthiness in Women. *Perceptual & Motor Skills*, 114(3), 964–978. <https://doi.org/10.0.9.162/07.09.21.PMS.114.3.964-978>
- Schwarzbach, J. (2011). A simple framework (ASF) for behavioral and neuroimaging experiments based on the psychophysics toolbox for MATLAB. *Behavior Research Methods*. <https://doi.org/10.3758/s13428-011-0106-8>
- Short, L. A., Hatry, A. J., & Mondloch, C. J. (2011). The development of norm-based coding and race-specific face prototypes: An examination of 5- and 8-year-olds' face space. *Journal of Experimental Child Psychology*, 108, 338–357.
<https://doi.org/10.0.3.248/j.jecp.2010.07.007>
- Short, L. A., Lee, K., Fu, G., & Mondloch, C. J. (2014). Category-specific face prototypes are emerging, but not yet mature, in 5-year-old children. *Journal of Experimental Child Psychology*, 126, 161–177. <https://doi.org/10.1016/j.jecp.2014.04.004>
- Sireci, S., & Geisinger, K. (1992). Analyzing Test Content Using Cluster Analysis and Multidimensional Scaling. *Applied Psychological Measurement*, 16(1), 17.
<https://doi.org/https://doi.org/10.1177/014662169201600102>
- Sofer, C., Dotsch, R., Wigboldus, D. H. J., & Todorov, A. (2015). What is typical is good: the influence of face typicality on perceived trustworthiness. *Psychological Science*, 26(1), 39–47. <https://doi.org/10.1177/0956797614554955>
- Todorov, A., Mende-Siedlecki, P., & Dotsch, R. (2013). Social judgments from faces. *Current*

Children's sensitivity to facial trustworthiness

Opinion in Neurobiology, 23(Social and emotional neuroscience), 373–380.

<https://doi.org/10.1016/j.conb.2012.12.010>

Todorov, A., Said, C. P., Engell, A. D., & Oosterhof, N. N. (2008). Understanding evaluation of faces on social dimensions. *Trends In Cognitive Sciences*, 12(12), 455–460.

<https://doi.org/10.1016/j.tics.2008.10.001>

Valentine, T. (1991). A unified account of the effects of distinctiveness, inversion, and race in face recognition. *The Quarterly Journal Of Experimental Psychology. A, Human Experimental Psychology*, 43(2), 161–204. <https://doi.org/10.1080/14640749108400966>

Valentine, T., Lewis, M. B., & Hills, P. J. (2016). Face-space: A unifying concept in face recognition research. *Quarterly Journal of Experimental Psychology*, 69(10), 1996–2019.

<https://doi.org/10.1080/17470218.2014.990392>

Van de Vondervoort, J. W., & Hamlin, J. K. (2017). The early emergence of sociomoral evaluation: infants prefer prosocial others. *Current Opinion In Psychology*, 20, 77–81.

<https://doi.org/10.1016/j.copsyc.2017.08.014>

Xu, F., Wu, D., Toriyama, R., Ma, F., Itakura, S., & Lee, K. (2012). Similarities and differences in Chinese and Caucasian adults' use of facial cues for trustworthiness judgments. *Plos One*. <https://doi.org/10.1371/journal.pone.0034859>

Young, S. G., Slepian, M. L., & Sacco, D. F. (2015). Sensitivity to Perceived Facial Trustworthiness Is Increased by Activating Self-Protection Motives. *Social Psychological and Personality Science*, 6(6), 607. <https://doi.org/10.1177/1948550615573329>

Zebrowitz, L. A., Fellous, J.-M., Mignault, A., & Andreoletti, C. (2003). Trait impressions as overgeneralized responses to adaptively significant facial qualities: evidence from connectionist modeling. *Personality And Social Psychology Review: An Official Journal Of The Society For Personality And Social Psychology, Inc*, 7(3), 194–215.

https://doi.org/10.1207/S15327957PSPR0703_01

Children's sensitivity to facial trustworthiness

Figure Captions

Figure 1. The seven variations of the female face identity used as stimulus material, representing a trustworthiness continuum ranging from 1 (very untrustworthy) to 7 (very trustworthy) (a). Difference images between the most untrustworthy face (i.e., face 1) and each of the other six faces included in the continuum, as resulted from an image difference analysis describing the physical variations among the stimuli (b).

Figure 2. Example of experimental trials from the Oddmanout task (a). After the offset of the fixation cross, three faces randomly selected from the trustworthiness continuum were simultaneously presented, which remained on the screen until the participant responded. Example of experimental trials from the Pairwise Preference task (b). After the offset of the fixation cross, two faces randomly selected from the trustworthiness continuum were presented simultaneously and remained on the screen until the participant responded.

Figure 3. Average representational dissimilarity matrices resulting from the pairwise dissimilarity scores derived from the Oddmanout task for each age group. Yellow [light grey] represents maximum dissimilarity, while blue [dark grey] represents minimum dissimilarity. The matrix is symmetrical across the diagonal. See the online article for the colour version of this figure.

Figure 4. Hierarchical plots describing the results of the cluster analyses performed on the pairwise dissimilarity scores derived from the Oddmanout task for the 5-year-olds (a), the 7-year-olds (b), and the adults (c). Different clusters are depicted in different shades of grey.

Figure 5. Empirical cumulative distributions of intra-group cosine distances of dissimilarity scores (a). Intra-group cosine distances of dissimilarity scores recorded for the three age groups for each of the seven trustworthiness intensities (b). ** $p < .01$ *** $p < .001$

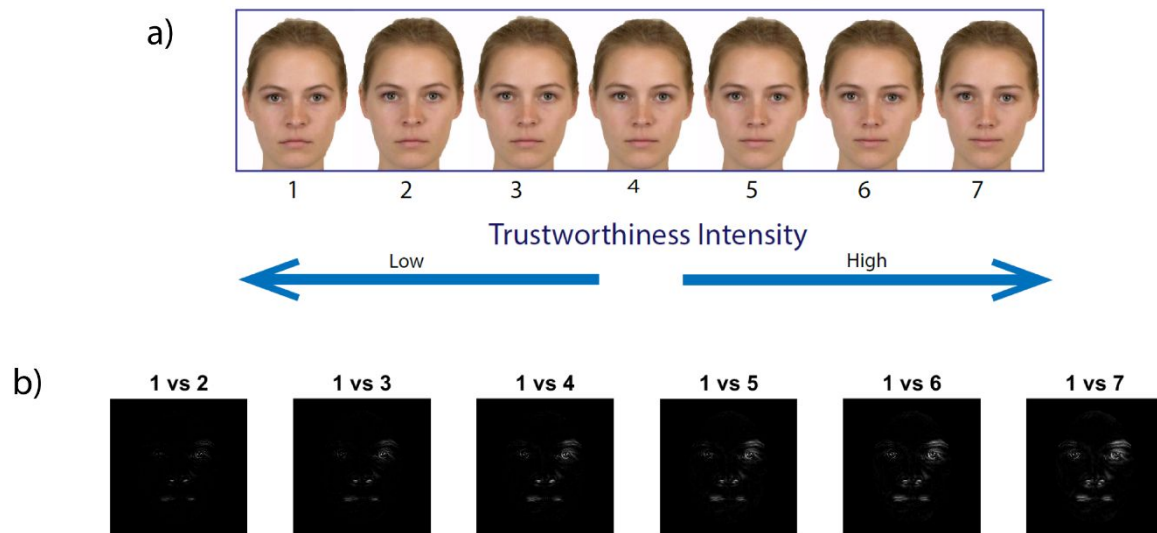
Children's sensitivity to facial trustworthiness

Figure 6. Mean percent accuracy obtained in the Pairwise Preference task by children in the three age groups (a). Mean trustworthiness scores obtained from the Pairwise Preference task for each of the seven trustworthiness intensities (b). $*p < .05$ $***p < .001$

For Review Only

Children's sensitivity to facial trustworthiness

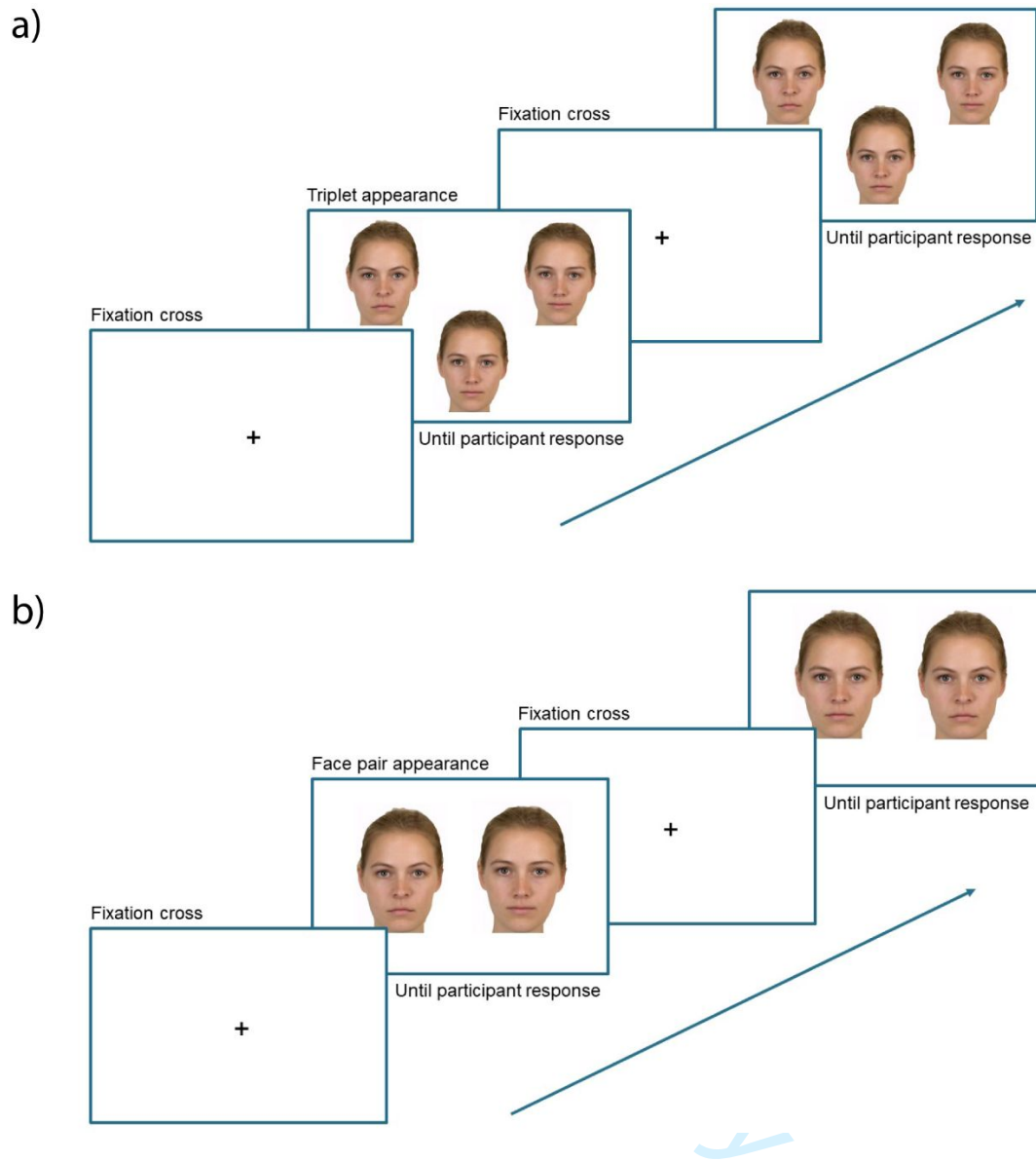
Figure 1



Review Only

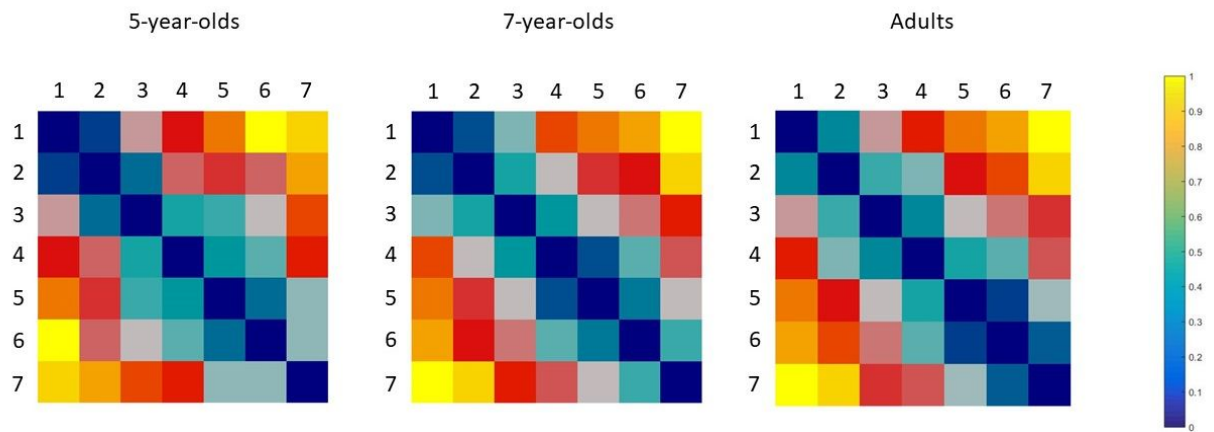
Children's sensitivity to facial trustworthiness

Figure 2



Children's sensitivity to facial trustworthiness

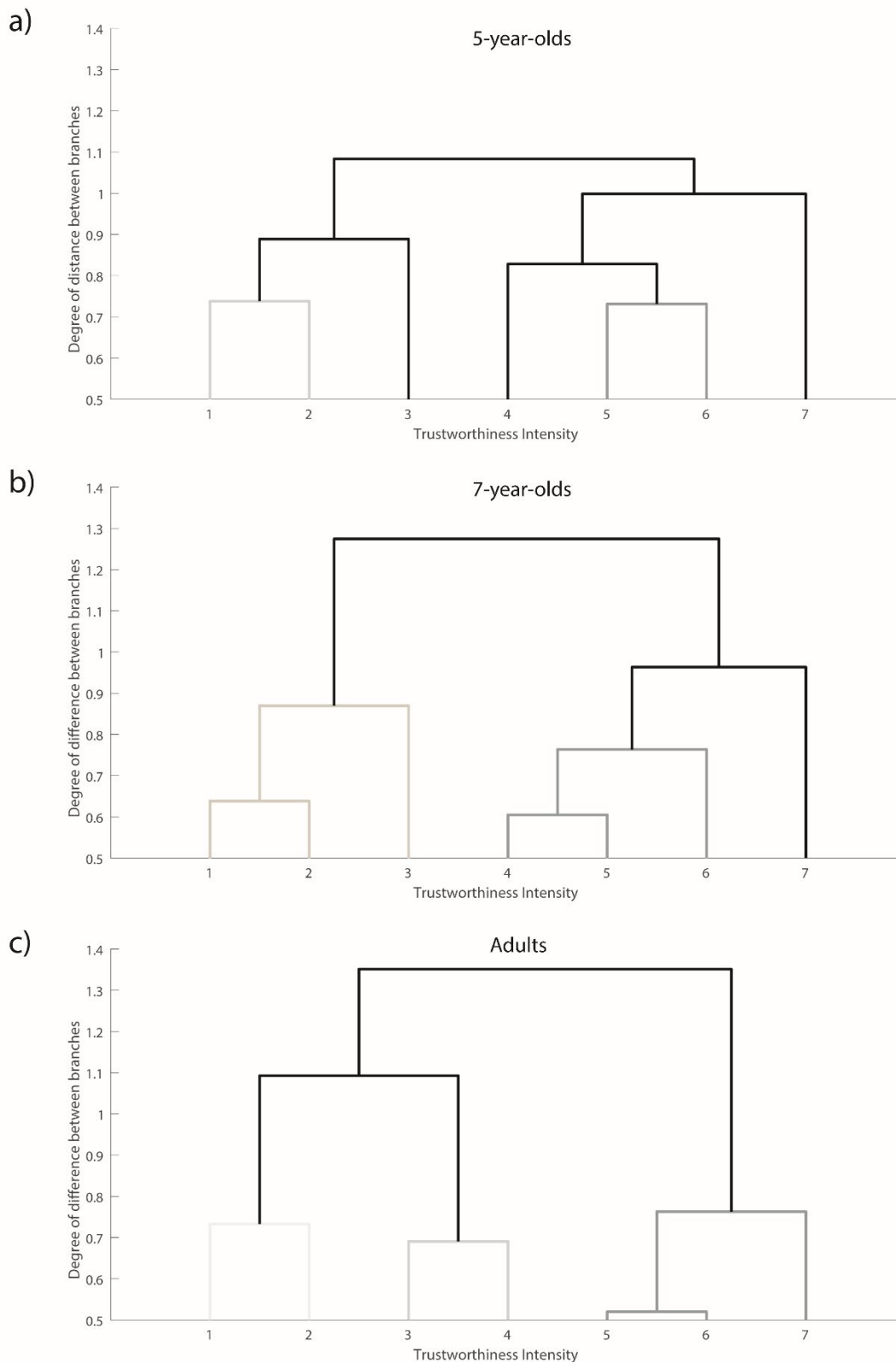
Figure 3



For Review Only

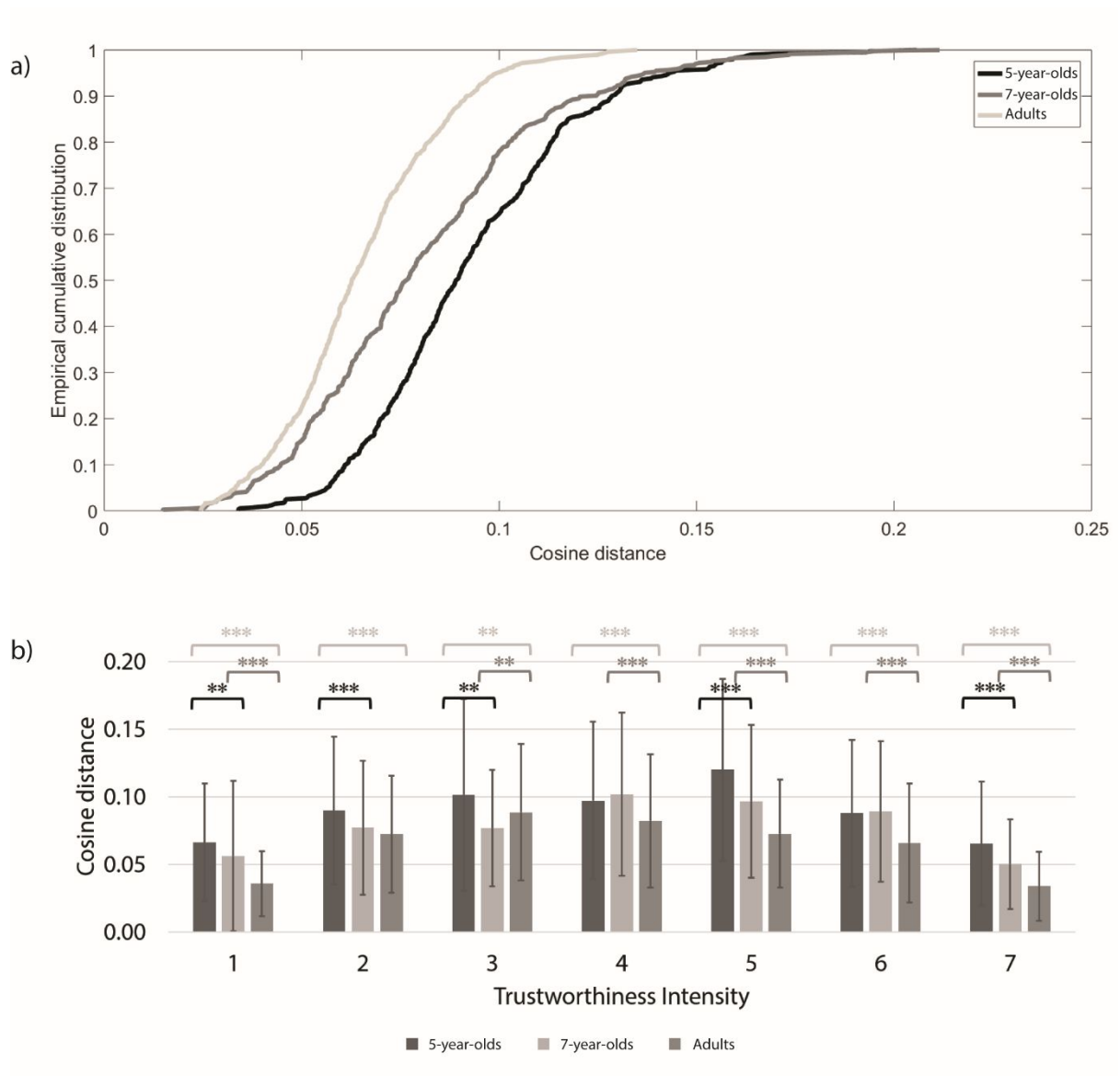
Children's sensitivity to facial trustworthiness

Figure 4



Children's sensitivity to facial trustworthiness

Figure 5



Children’s sensitivity to facial trustworthiness

Figure 6

