SCUOLA DI DOTTORATO
UNIVERSITÀ DEGLI STUDI DI MILANO-BICOCCA

Department of Psychology

PhD program in Psychology, Linguistics and Cognitive Neuroscience
Cycle XXXII

Curriculum: The Mind, Brain and Behavior

# IT'S WRITTEN ALL OVER YOUR FACE

The ontogeny of sensitivity to facial cues to trustworthiness

**Surname:** Baccolo

**Name:** Elisa

**Registration number:** 823009

**Tutor:** Viola Macchi Cassia

**Coordinator:** Marco Perugini

**ACADEMIC YEAR** 2018/2019

# Table of contents

# Abstract

Human beings are hypersensitive to those facial properties that convey social signals. Their ability to attribute trustworthiness judgements, i.e., whether a person can be safely approached or better avoided, based on facial cues is known to be fast, automatic and based on very little information.

A large part of the literature focused on disentangling the cognitive and neural processes underlying this ability in adults, but little effort has been dedicated to investigating whether and how sensitivity to fine-grained differences in facial cues to trustworthiness changes across development, and how this sensitivity is modulated by experience and individual differences related to temperamental and personality traits.

This doctoral dissertation includes 5 studies aimed at investigating (1) whether and how individual variations in social personality traits affect adults' perceptual sensitivity to subtle variations in physical cues to trustworthiness and the ability to use these variations to make trustworthiness judgments (Study 1), (2) whether and how this sensitivity changes across development from preschool years to adulthood and whether emotional development can modulate these changes (Study 2), (3) whether neural and behavioral signatures of the ability to discriminate facial cues to trustworthiness are apparent as early as the first year of life (Study 3a, Study 3b, Study 4), and (4) whether this ability is a universal phenomenon or it is modulated by culture and/or perceptual expertise with faces of the ethnicity that is more frequently represented in the beholder's social environment.

Specifically, Study 1 investigated whether individuals' variability in personality traits influences perception of fine-grained differences in facial cues to trustworthiness and the way these are represented in the long-term memory. Results showed that individual differences along the introversion-extraversion dimension affect the time required to successfully disentangle fine-grained variations in trustworthiness intensity expressed by the faces. The results are discussed within a developmental framework where individual differences related to social attitude constrain the amount and quality of facial experience, modulating perceptual attunement to social cues from faces across the life span.

Study 2 showed that sensitivity to facial cues to trustworthiness and the ability to use these cues to generate trustworthiness judgements is present in preschool years, but the representation of these cues in long-term memory becomes adult-like at the age of 7 years. Moreover, the ability make trustworthiness judgements from faces is related to the development of emotion understanding abilities.

Study 3 and 4 used two different electrophysiological paradigms to measure neural sensitivity to facial cues to trustworthiness in 6-month-old infants, and whether this ability is related to individual differences in temperamental traits. Results extended earlier evidence by demonstrating that the infants' brain discriminated high trustworthy faces from neutral faces and low trustworthy faces, and that temperamental differences in Surgency and Negative Affect modulate electrophysiological responses.

Study 5 describes the results of the validation of a novel stimulus set that will be used to investigate cross-cultural differences in the perception of face trustworthiness in young children.

Overall, the results of the presented studies agree in suggesting that the ability to discriminate the physical cues that are used to generate trustworthiness judgments appears early in development. The sensitivity to and perceptual representation of these cues refines over the

course of development as a result of the quantity and quality of facial experience provided by the social environment to each individual, which may be also affected by individual temperamental and personality traits and emotional development trajectories.

Keywords: facial cues; trustworthiness; development; perceptual sensitivity; face representational space; neural correlates; social judgments; perceptual expertise; EEG, fast-periodic-visual stimulation, event-related potentials.

*"The most entertaining surface*

*on earth is the human face."*

*(Georg Christoph Lichtenberg, 1742-1799)*

# Introduction

### The human face in social communication

As humans, we are fine-tuned to other people social signals: we easily detect even the slightest switch in their emotional status, mental state or intention, and use this information to consequently adapt our social behavior. This sensitivity is so fundamental for human social interactions that an impairment in successfully displaying or detecting social signals can entail dysfunctional social relationships.

Much of the social information about other people is derived from faces, a visual stimulus to which we reserve special attention. Adults can identify faces much faster than other environmental stimuli, even when they appear in the visual periphery (Hershler, Golan, Bentin, & Hochstein, 2010; Hershler & Hochstein, 2005) or are presented subconsciously (Hoshiyama, Kakigi, Watanabe, Miki, & Takeshima, 2003). This drive to attend to other people's faces has precocious roots. Since very early in our development, in fact, we manifest a peculiar attentional bias for face-like patterns, which allows us to extrapolate information about facial structure, and facilitates the recognition of other people's identities (Morton & Johnson, 1991; Dziurawiec, Ellis, & Morton, 1991). Perception of invariant aspects of faces is as much important as perception of more changeable aspects of faces, such as expressions, which are particularly important for social communication (Haxby et al., 2011). Indeed, humans' ability to perform and decode dynamic and complex facial expressions stands above that of all other animals (Adams, Albohn, & Kveraga, 2017). By four weeks of age, infants can mimic facial expressions initiated by a stranger (Meltzoff & Moore, 1977) and, by the age of 7 months, they can discriminate between different kinds of facial expressions signalling emotional states, such as happiness and fear (Grossmann, 2010; Nelson, 1987), and use this information to consequently adapt their behaviour (Cohn & Tronick, 1983; Hirshberg, 1990).

Notwithstanding the effortlessness through which we extrapolate social pieces of information from other people's facial cues, faces are a complex, high-dimensional and dynamic stimulus category, and so is the process of social communication through which social information transmission and decoding takes place (Jack & Schyns, 2015). In order to fully comprehend the mechanisms underlying our sensitivity to social signals from faces, Jack & Schyns (2015) proposed that we should first put a focus on how the process of social communication takes place at all. In their review, the authors start from a definition of *communication* as a dynamic act of information transfer, i.e. a situation where a person - "the sender" - transmits an information relative to his/her mental state, intention or disposition, which is able to influence the conduct of another person - "the receiver". For example, through a specific facial configuration, e.g. a sad facial expression, that serves as a communication channel, a sender (voluntarily or involuntarily) transmits the information relative to his/her mental state to a receiver, who decodes the information to form an interpretation about the other person's mental state (i.e., he/she is sad). To successfully decode the information, receivers must own a suitable computational system that allow them to extrapolate the information from the communication channel and consequently respond to that (McCullough & Reed, 2016). Moreover, individual characteristics of the receiver (prior experience, personality traits, mental state) filter the incoming information, influencing its processing (Hehman, Flake, Xie, Stolier, & Freeman, 2019). Therefore, when studying the face as a communication tool, we first need to understand **which distinct face information** of the

**sender** can evoke the perception of a **distinct social message in the perceiver**, taking into account how **individual differences related to perceiver**'s character or expertise might shape his/her perception of the social signal itself.

**Social signals from faces: the construct of trustworthiness**

When it comes to the topic of facial cues in social communication, it is important to notice that a vast part of the literature focused on a small amount of information that an highly dynamic stimulus such as the face can vehiculate, namely the six universal facial expressions of emotion (Ekman & Oster, 1979; Ekman, Sorenson, & Friesen, 1969). Nevertheless, we are able to express and extrapolate a wide variety of cues from faces, including more subtle and complex ones. For example, when presented with an array of identical female faces with neutral expressions, created as composite pictures of women scoring high or low on different health (physical and mental) and personality (agreeableness, extraversion, neuroticism and intellect) dimensions, subjects are able to tell the difference between those faces scoring high or low on those dimensions, solely based on the tiny, fine-grained pieces of information contained in their internal facial features (Kramer & Ward, 2010).

Another kind of cue we rapidly and intuitively extract from faces relates to the information about other people's intentions, that is whether they are likely to approach us friendly or hostilely. This type of cue is defined in the literature as face *trustworthiness* (Todorov, Said, Engell, & Oosterhof, 2008), and the ability to infer this disposition is known to be fast, automatic, based on very little information and performed with high consensus. Indeed, people can assess whether an emotionally neutral face belongs to someone who is approaching them hostilely or friendly solely based on facial perceptual information, and after a very short exposure time (Bar, Neta, & Linz, 2006; Todorov, Pakrashi, & Oosterhof, 2009; Willis & Todorov, 2006). Moreover, even tiny differences between facial characteristics are used to generate explicit judgments of trustworthiness (Ames, Fiske, & Todorov, 2011; Todorov et al., 2008).

Even though non-necessarily accurate, these perceptually based attributions have important real-word consequences. For example, players of economic games tend to trust less an individual with untrustworthy-looking face (Chang, Doll, van't Wout, Frank, & Sanfey, 2010; Rezlescu, Duchaine, Olivola, & Chater, 2012). Judgements of untrustworthiness also predict judgements of guilt (Porter, ten Brinke, & Gustaw, 2010; Wilson & Rule, 2015) and the outcome of political elections (Ballew & Todorov, 2007). Interestingly, trustworthiness judgements on children facial appearance were found to predict their real-world behaviour, with children judged as untrustworthy by strangers solely based on their facial appearance being the least liked ones by their own classmates (Li, Heyman, Mei, & Lee, 2017).

But which distinct facial pieces of information of the sender can evoke the perception of trustworthiness in the perceiver? Even though this is still to be further investigated, Dotsch and Todorov (2012) observed that, for what concerns inner facial features, people rely on information found in the mouth, eye and eyebrows regions. More specifically, up/downturned eyebrows, upward/downturned curving mouth and a more or less wrinkling nose are often reported as those facial cues that might be used to infer whether a person is likely to approach us safely or hostilely (Gill, Garrod, Jack, & Schyns, 2014; Hehman, Flake, & Freeman, 2015; Oosterhof & Todorov, 2008). Jack and Schyns (Gill et al., 2014; Jack & Schyns, 2015) propose that these can be thought of as trustworthiness' Action Units (AUs), namely facial patterns of movement or movement configurations that consistently communicate specific social messages. In the case of trustworthiness, AUs can be either morphological (a face type that embodies configural patterns of trustworthiness/untrustworthiness as static traits) or dynamic (a face configuration that changes within an individual when he/she intends to signal hostility or friendlessness).

According to the Emotion Overgeneralization hypothesis (Said, Sebe, & Todorov, 2009; Zebrowitz, Fellous, Mignault, & Andreoletti, 2003), trustworthiness judgements should be interpreted as overgeneralized responses to facial cues resembling emotional expressions. Indeed, faces with different emotional valence (angry, happy) carry information about others' approachability (Marsh, Ambady, & Kleck, 2005); for example, approaching angry people lead to higher chances of verbal or physical attacks (Lischke, Junge, Hamm, & Weymar, 2018). Congruently, both typically developing children (Caulfield, Ewing, Bank, & Rhodes, 2016) and children in the autistic spectrum (Caulfield, Ewing, Burton, Avard, & Rhodes, 2014) are influenced by the intensity of emotional cues expressed by faces when judging trustworthiness, such that overt angry expressions prompt judgements of untrustworthiness, and overt happy expressions prompt judgements of trustworthiness. Nevertheless, facial cues such as low eyebrows (that, when stressed, communicate anger) are used by the perceiver to make assumptions about the other person's disposition (unfriendlessness) even if the other person's face is not perceived as expressing an overt emotion (i.e., angriness; Ames et al., 2011). Therefore, precisely because trustworthiness is a trait we attribute also to neutral, non-emotional faces (Lischke et al., 2018; Oosterhof & Todorov, 2008; Todorov, Dotsch, Porter, Oosterhof, & Falvello, 2013), there's still need for a systematic investigation of the extent to which trustworthiness AUs differ from the AUs that vehiculate emotional states. A first attempt in this direction is represented by a study by Gill and colleagues (2014), which openly questioned the Emotion Overgeneralization theory by proving that dynamic social masks, i.e. transient facial movements vehiculating trustworthiness (thanks to muscles such as the brow raiser, the lip corner, the nose wrinkler, etc), represent a unique sets of facial movement configurations that differ from, and therefore should not be reduced to, the classic six emotional facial expressions.

## The origins of the sensitivity to facial cues to trustworthiness: phylogenetic and ontogenetic accounts

Sensitivity to facial cues to trustworthiness is thought to be rooted in our phylogeny. Indeed, evolutionary theories claim that some aspects of human social cognition might have evolved because they are adaptive, in the sense that they guarantee the species' survival through human sociality. In this respect, being able to distinguish between the friendly or hostile intentions of conspecifics has strong impact on an individual's chances of survival (Zebrowitz et al., 2003). This evolutionary pressure potentially explains why, as previously reported, humans are spontaneous, incredibly fast, need minimal cognitive effort and generally agree when assessing face trustworthiness (Bar et al., 2006; Todorov et al., 2009; Willis & Todorov, 2006). Indeed, we might have learned throughout evolution to pay attention to those facial cues signalling potential harm, thus becoming extremely hypersensitive to such cues as a result (Schaller & Skowronski, 2012). As a matter of fact, evolution might have wired the human brain system to detect the presence of relevant stimuli in the environment (e.g. stimuli that represent threat) and to allow them a preferential access to awareness (Leppänen & Nelson, 2009). Not by chance, humans manifest a precocious sense for threat, which affects their behavior and attentional responses. Between 4 and 6 months of age, infants show approach behaviours when other people display happy facial expressions, and avoidance behaviours towards those displaying anger or negative expressions (Serrano, Iglesias, & Loeches, 1995). Moreover infants' attention disengagement and looking duration are modulated by fearful expressions, which signal a threat in the environment (Peltola, Leppanen, Palokangas, & Hietanen, 2008). Congruently, emotional facial expressions signalling potential aggressive behaviours (i.e. potential danger), such as angriness, are more effective in catching and maintaining our attention (Fox et al., 2000; Schupp et al., 2004), even when subliminally presented (van Honk et al., 1998; Williams et al., 2004). These fast and automatic responses are mediated by a fast, sub-cortical route in the adult brain, enabling even hemianoptic patients to process fearful faces in their blind field (de Gelder, Vroomen, Pourtois, & Weiskrantz, 1999; Johnson, Senju, & Tomalski, 2015).

Other arguments supporting the phylogenetic origins of sensitivity to facial cues to trustworthiness come from the the small set of cross-cultural studies conducted on the topic. For example, Xu and colleagues (Xu et al., 2012) found that Chinese and Caucasian adults use the same facial physical information to generate explicit judgments of trustworthiness from Caucasian faces, suggesting that cues to trustworthiness are somewhat universal. Analogously, Birkás et al. (Birkás, Dzhelyova, Lábadi, Bereczkei, & Perrett, 2014) observed that participants of Caucasian and Asian origins used the same facial cues to detect the level of a target's trustworthiness, irrespective of ethnical characteristics (Caucasian, African, South Asian and East Asian) and therefore unrelated to an individual's expertise. Sakuta and colleagues (Sakuta, Kanazawa, & Yamaguchi, 2018) even found a preference for trustworthy versus untrustworthy Caucasian faces in preverbal Japanese infants. These studies altogether suggest that perception of trustworthiness might be generalized across ethnicity, possibly as a result of evolutionary pressures.

Notwithstanding the relevance of adaptive mechanisms, ontogenetic mechanisms likely play an important role as well, in driving the emergence of sensitivity to trustworthiness facial cues in the human species. Indeed, the expression of genes into phenotypes is known to be influenced by multiple environmental variables, originating from both inside and outside the individual (Karmiloff-Smith, Casey, Massand, Tomalski, & Thomas, 2014). In the wake of these considerations, Over & Cook (2018) proposed that sensitivity to facial cues to trustworthiness might also be under environmental control, and therefore shaped by experience. Inspired by the Face Space model proposed by Valentine (Valentine, Lewis, & Hills, 2016), the authors propose that trait inferences from faces might be the result of a face-trait mapping where the visual system extrapolates the variations among the faces encountered in the individual's everyday life by representing them in a multidimensional space, called *face space* (Valentine et al., 2016). After having associated a certain face to a certain social trait as a result of social interactions and experience, all the faces that are close to that face in the represented face space are mapped onto the same trait, which is also represented in a *trait space* that is built upon the individual's experience within the social environment. Therefore, experience across development would have a critical role in shaping and refining the structuring of this face-trait mapping. Congruently, Keefe and colleagues (Keefe, Dzhelyova, Perrett, & Barraclough, 2013) adopted a forced-choice paradigm where participants judged which face was more untrustworthy after adaptation to different levels of trustworthiness, and observed that adaptation to a certain level of facial trustworthiness enhanced discrimination of the same trustworthiness level compared to a condition without adaptation or adaptation to another level of trustworthiness, suggesting that our visual experience plays an important role in shaping and refining our sensitivity to facial trustworthiness.

To conclude, sensitivity to the physical facial cues that drive trustworthiness judgements possibly results from the interaction between biological constraints evolved from evolutionary pressure to successfully distinguish between friendly or hostile conspecifics, and experience-based mechanisms that shape and refine our perceptual sensitivity through social experience with conspecifics. By tracing the developmental trajectory of such sensitivity, developmental research may provide critical contribution to the understanding of the way biological and environmental mechanisms intertwine in producing human's perception of trustworthiness from faces.

**The development of sensitivity to social cues from faces**

Although much of the literature focused on adults' sensitivity to facial cues to trustworthiness, its developmental trajectories are still scarcely investigated.

The existing literature seems to point out that children gain understanding of the construct of trustworthiness by the age of 4 years, when they use verbal and behavioral hints to make inferences about whether or not they can trust somebody (Harris, 2007; Heyman, 2008). The ability to make explicit trustworthiness judgements based on other people's faces also emerges quite early in development: children as young as 3 years of age can judge whether a person appears to be 'mean' or 'nice' based on his/her facial look, and these judgements reach adult level of consistency by the age of 6 (Cogsdill, Todorov, Spelke, & Banaji, 2014). Congruently, children aged 8, 10, and 12 years of age show within-age agreement when asked to judge the level of trustworthiness expressed by a face, and their performance becomes increasingly consistent with that of the adults with increasing age (Ma, Xu, & Luo, 2016). Furthermore, judgements of trustworthiness based on facial appearance also influence children's behaviors: while playing an economic trust game, 5- and 10- years-olds were shown to trust more trustworthy-looking partners than untrustworthy-looking ones (Ewing, Caulfield, Read, & Rhodes, 2015).

Recent studies suggest that sensitivity to face characteristics that drive trustworthiness judgements in adults and children might emerge even earlier than preschool-years, during infancy. Jessen & Grossmann (2016, 2017, 2019) recorded 7-month-old infants' brain activity while looking at very trustworthy and untrustworthy faces, and found evidence of significant discrimination both when the faces were presented supraliminally (i.e., above infants' faces visibility threshold; Sarah Jessen & Grossmann, 2016) and subliminally (i.e., below infants' visibility threshold; Sarah Jessen & Grossmann, 2017). In addition, infants' attention is caught more by trustworthy faces than untrustworthy ones, as they look more at the former face category than the latter in a preferential looking task (Jessen & Grossmann, 2016; Sakuta, Kanazawa, & Yamaguchi, 2018), and allocate enhanced attention to objects that are attended to by trustworthy faces in a gaze-cueing paradigm (Jessen & Grossmann, 2019). These findings are in line with those showing that five to nine-month-old infants preferentially attend to prosocial individuals than to antisocial ones (Hamlin & Wynn, 2011; Van de Vondervoort & Hamlin, 2017), and to a stranger who has been positively approached as compared to avoided by their mother (Fein, 1975), just like 6- to 11-years-old children preferentially place their trust in those who help others (Fu, Heyman, Chen, Liu, & Lee, 2015). Altogether, these findings suggest that humans are sensitive to other people's hostile/friendly attitude from very early in development; nonetheless, the developmental trajectories of humans' sensitivity to the physical facial cues that drive trustworthiness inferences still need to be fully understood.

**Clearing the conceptual ground**

The term trustworthiness is a multifaceted construct that has been widely used in the literature. In a recent review, Wilson and Rule (2017) called for a greater precision when referring to this concept, observing that ambiguity about what trustworthiness represents might hinder the nature and quality of experimental questions. Indeed, this term assumes a variety of meanings that go well beyond the definition I've previously offered (i.e., a judgement on other people's approachability), such as in studies on lie detection (Charles, Berry, & Omar, 1994; Bond, Charles, & DePaulo, 2006), criminal behaviour (Porter et al., 2010; Wilson & Rule, 2015), "second-hand" impression formation or impression formation via interaction (Delgado, Frank, & Phelps, 2005; Ferrari et al., 2016; King-Casas et al., 2005), where the term is used to signify cooperation, lack of aggression, faithfulness, honesty and other constructs related to human behaviour (Wilson & Rule, 2017).

The ground for a broader interpretation of what trustworthiness represents can be found in the etymology of the word itself - *worthy of trust* -, which evokes a wide range of meanings connected to other people's integrity, support, alliance, reliability, and so on. Therefore, lingering over taxonomy seems advisable before discussing the studies described in the current dissertation.

*Trustworthiness* is not the only term used in the literature to designate other people's approachability. In a review, Fiske and colleagues (Fiske, Cuddy, & Glick, 2007) proposed another expression that might describe humans' ability to judge another person's intent for good or ill: *warmth* judgements. The authors support the idea that warmth is a dimension of social cognition which promotes survival, as it answers a basic survival question: *do other people intent to hurt me/help me?* Therefore, the word would describe humans' sensitivity to perceived intents, and more specifically to approach-avoidance tendencies (i.e., other people's friendliness vs threat/aggressiveness). This dimensions is claimed to be universal, as it would account for 82% of variance of impressions from everyday social behaviours (Wojciszke, Bazinska, & Jaworski, 1998). Moreover, warmth judgements are described as primary, as they stand above all other social judgements, such as competence (e.g., after exposure of 100 ms, people judge face trustworthiness more reliably than face competence, Willis & Todorov, 2006), and elicit fast behavioural reactions. Finally, warmth is described as a dimension that predicts the valence of social judgement (i.e., whether the judgements is positive or negative).

Even though Fiske and colleagues (Fiske et al., 2007) seem to intend warmth as a synonym for trustworthiness, we could argue that, from an etymological perspective, *warmth* (i.e., being cold or warm) recalls the *perceptual* impression of another person's approachability, while *trustworthiness* (i.e., being worthy of trust) relates more to the explicit act of judging the other person's approachability, as a consequence of that perceptual act. This relationship between perception and social evaluation is well described by Fenske and colleagues (Fenske, Raymond, Kessler, Westoby, & Tipper, 2005) when they state that faces whose perceptual impression is associated to an approach response are judged as trustworthy, while faces whose perceptual impression is associated to an avoidance response are judged as untrustworthy.

In light of these considerations, this thesis intends to address humans' sensitivity to facial cues to trustworthiness, meant as those facial features that affect our first impression of a stranger by inducing us to believe he/she is approaching us warmly (positive attitude) or rather represents a potential threat (hostile attitude) before getting to access any second-hand information (based on behaviour or description) about the person in question. Moreover, we are interested in investigating the sensitivity to those internal facial features whose configuration can vehiculate trustworthiness judgements, rather than which type of face identity (and the related morphology) is perceived as more or less trustworthy. In fact, the studies presented in the following chapters are framed in the line of research that does not consider trustworthiness as a stable feature of a single face identity, but it is rather interested in investigating which perceptual cues of the face, independently of identity, are crucial in influencing perception of trustworthiness and trustworthiness judgement. Indeed, we argue that trustworthiness judgements are attributed to specific face identities as a result of an extension of automatic and spontaneous perceptually-cued attributions to face morphologies that statically incorporate the action units (Jack & Schyns, 2015) that are dynamically recruited to express willingness to approach/avoid someone (e.g. upturning/downturning lips and eyebrows).

**An introduction to the studies**

This thesis aims at addressing three main questions: (1) Is sensitivity to facial cues to trustworthiness modulated by individual variations in social personality characteristics? (Study 1); (2) what is the developmental trajectory of this sensitivity? (Study 2, 3, 4); (3) is sensitivity to subtle variations in facial cues to trustworthiness a universal phenomenon or is it modulated by culture and/or face ethnicity? (Study 5 – in progress).

Chapter 1 addresses the question of whether individual differences in fine-grained perceptual sensitivity and mental representation of facial features related to trustworthiness judgements are associated with individual differences in social motivation. Indeed, although much effort has been put in investigating the cognitive mechanisms underlying the ability to make trustworthiness judgements based on facial appearance in the average adult population (see Paragraph 1.2), still little is known about whether individual differences related to personality and social behaviour relate to how individuals differ in their sensitivity to facial cues that drive those judgements. As claimed by the sender and receiver model proposed by Jack and Schyns (2015; see Paragraph 1.1), it is important to take into consideration the receiver's characteristics and knowledge to fully understand the dynamics behind perception and social communication. Indeed, plenty of studies showed how perception is modulated by the characteristics of the perceiver. For example, Rhodes and colleagues (Rhodes, Jeffery, Watson, Clifford, & Nakayama, 2003) showed how individual experience with a specific type of faces (e.g., related to everyday routines or massmedia exposure) can alter the individual's perception of typicality and attractiveness. Similarly, psychological characteristics of the perceiver modulate neural responses to faces and alter social impressions: individual with high or low levels of extraversion show different pattern of face-specific electrophysiological responses (Cheung, Rutherford, Mayes, & McPartland, 2010), and individuals with high levels of anxiety traits tend to attribute more negative judgements to faces (Dimberg & Thunberg, 2007; Kim, Yoon, Shin, Lee, & Kim, 2016; Schofield, Coles, & Gibb, 2007).

In Study 1, we aimed to investigate whether individuals' variability on the extraversion dimension influenced perception of fine-grained differences in facial cues to trustworthiness. The motivation behind the focus on the extraversion dimension is embedded within a developmental perspective that highlights the role of proactive learning from the species-specific environment as well as the individual environment in the emergence of neurocognitive specialization. Specifically, we reasoned that individual differences related to one's social attitude and motivation may affect the amount and quality of facial experience, thus constraining perceptual attunement to social cues from faces across the life span (see Li et al., 2010). Accordingly, we hypothesized that individuals with higher extraversion scores would be more efficient in assessing perceptual similarity between faces that slightly varied in the level of trustworthiness they expressed. In order to test this hypothesis, subjects performed a similarity task where they had to identify among two probe faces the one they perceived as more similar to a target face. The experimental stimuli were computer-generated faces that only slightly varied in the level of perceived trustworthiness, composing a 7-steps trustworthiness continuum (Oosterhof & Todorov, 2008). We were interested in using fine-grained variations in trustworthiness intensity to tackle individual differences in perceptual sensitivity as we conjectured that the more sensitive a person is, the tiniest is the information needed to evaluate face trustworthiness.

The studies presented in Chapter 2 (Study 2, 3 and 4) further explore the hypothesis that differential ontogenetic experience in the social domain can affect perceptual sensitivity to face trustworthiness by focusing on development. More specifically, Study 2 investigated how perceptual sensitivity to and mental representation of fine-grained differences in facial information subtending social perception of trustworthiness develop in time, taking into account individual differences in emotional development. Indeed, if a large part of the literature questioned the cognitive and neural processes behind adults' sensitivity to face trustworthiness, still little is known about how this sensitivity develops in time. As outlined in Paragraph 1.4, previous studies proved that by the age of 3 years children make explicit evaluations on how mean or nice a stranger looks, reaching adults level of consistency by the age of 6 years (e.g., Cogsdill et al., 2014). Nevertheless, these studies used as stimulus material computer-generated faces (except for Cogsdill & Banaji, 2015) placed at the extremes of the trustworthiness continuum, which might have made the task quite easy to perform. Therefore, in Study 2 we used as experimental stimuli a set of seven variations of one female identity that was parametrically manipulated to only slightly vary in the level of expressed trustworthiness. Stimuli were created by using a data-driven approach similar to the one used by Todorov and colleagues (Oosterhof & Todorov, 2008), which prevented us to specify a-priori which facial AUs should vary.

Children aged 5 and 7 years, together with adults, were presented with a perceptual sensitivity task where they were asked to select the face they judged to be the most different among three simultaneously presented faces, all selected randomly from the trustworthiness continuum. This allowed us to collect dissimilarity scores, which were used to explore the structure of mental representation of facial cues to trustworthiness in subjects' long term memory, and how this changes across time. In addition, similarly to Study 1, Study 2 focused on individual differences in social perception by exploring whether and how individual variations in the development of emotion understanding affect children's perception of facial cues to trustworthiness.

In light of the evidence provided by Study 2 that pre-schoolers are sensitive to facial cues associated to trustworthiness, Study 3 and 4 explored whether this sensitivity emerges earlier in development by focusing on preverbal infants. More specifically, Study 3 and 4 investigated whether sensitivity to the physical facial information that drives trustworthiness judgements in adults is already present in 6-month-old infants. In these studies, both infants' overt attentional responses and their neural sensitivity to trustworthiness cues were measured using, respectively, a preferential looking paradigm and neurophysiological methods to record EEG. As outlined in Paragraph 1.4, so far only four studies have investigated infants' ability to discriminate between different level of face trustworthiness (Jessen & Grossmann, 2016; Jessen & Grossmann, 2017; Jessen & Grossmann, 2019; Sakuta et al., 2018). In spite of earlier demonstration that infants are better at discriminating female faces, as compared to male ones, because the former are more familiar to them than the latter (see Quinn, Yahr, Kuhn, Slater, & Pascalils, 2002; Ramsey-Rennels & Langlois, 2006; Ramsey-Rennels & Davis, 2008), all of the the available studies on infants' sensitivity to trustworthiness used artificial male face identities as stimulus material, thus limiting the generalizability of the obtained evidence to real-life situation. To overcome this limitation, Study 3 and 4 used more ecological female faces as stimulus material to investigate neural sensitivity and attentional responses to facial cues of trustworthiness.

More specifically, Study 3 used the same Event Related Potential (ERP) paradigm and preferential looking task adopted by Jessen & Grossmann (2016) to explore infants' neural sensitivity to trustworthiness cues of natural, female faces. Because the obtained results were inconsistent with those previously obtained by the authors, Study 4 used a new, infants' friendly, Electroencephalographic (EEG) visual discrimination paradigm known as the Fast Periodic Visual Stimulation paradigm (FPVS, see Norcia, Appelbaum, Ales, Cottereau, & Rossion, 2015) to further investigate neural discrimination of high-trustworthy and low-trustworthy faces. The FPVS paradigm consists in a fast stimulus presentation at a constant frequency able to elicit change in voltage amplitudes in the electrical activity of the brain at that same frequency of stimulation. This paradigm was chosen as it proved to be particularly effective in studying visual discrimination of faces with infants, as the response becomes observable within a short amount of time. Two face identities were presented at a rate of 6 Hz, and every 1.2 Hz the trustworthiness level of the presented faces changed (a trustworthy face every 4 untrustworthy faces, or vice versa, in a counterbalanced order). We hypothesized that the sensitivity of the paradigm would have allowed us to detect infants' ability to discriminate between different levels of expressed trustworthiness.

Chapter 3 presents the validation of a set of stimuli that will be used in a series of cross-cultural studies to explore whether the evidence gathered from Study 2 generalize across cultures and face ethnicities. The only available studies on this topic were conducted with adults, and showed that Caucasian and Asian participants use similar information to judge face trustworthiness (F. Xu et al., 2012), and that perception of trustworthiness is generalizable across face ethnicity (Birkás et al., 2014). To test whether children's ability to discriminate physical cues to trustworthiness varies across cultures and across face race, we created a 5-steps continuum of female Asian faces and a 5-steps continuum of female Caucasian faces (taken from the 7-steps continuum used in Study 2). We plan to use these continua as stimulus material to test a group of Italian preschool-aged children and a group of Japanese children form the same age group for their ability to perceptually discriminate among the faces and to make trustworthiness judgements of the same faces (Study 5).

# Chapter 1

## Study 1

**Individual differences in perceptual sensitivity to facial cues to trustworthiness: the role of social attitudes**[1]

As highlighted in the Introduction of this thesis, faces are an effective channel for social communication. Adults easily decode and respond to different social cues conveyed by a person's face, and use them to produce evaluations about that person's intentions, mental state or character. We've already noted how adults are able to make judgements based on faces after a very short exposure time (e.g., Bar, Neta, & Linz, 2006; Todorov, Pakrashi, & Oosterhof, 2009; Willis & Todorov, 2006), and how they can disentangle between fine-grained differences in facial cues, attributing different social judgements according to these little variations (Jones, Kramer, & Ward, 2012). Much research has been devoted to characterizing the cognitive and neural processes underlying adults' ability to evaluate facial trustworthiness, i.e. whether a stranger individual can be safely approached or avoided (Adolphs, 2002; Willis & Todorov, 2006), but only few studies tackled the issue of whether and how individuals differ in their sensitivity to the facial cues that drive those inferences and/or in their proneness to perform trait inferences from faces.

This is rather surprising, given that individual differences in social personality characteristics are known to shape face perception in general. For example, Cheung and colleagues (Cheung, Rutherford, Mayes, & McPartland, 2010) reported a differential

[1] The data presented here were discussed in the paper: 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, 45,* 224-236.

sensitivity to stimulus inversion at the level of the N170 event-related potential component for extraverts compared to introverts, which was marked by an enhanced N170 for inverted compared to upright faces in the former group, but not in the latter. Because the N170 inversion effect is a marker of perceptual expertise, its association with extraversion was interpreted by the Authors as suggesting that personality characteristics may affect social motivation, which would in turn affect the amount of perceptual experience that individuals acquire with faces through social interactions. Therefore, both these aspects would be diminished in typically developing individuals who score high in introversion, as well as in individuals who show clinical levels of social impairments (e.g., Dawson, Webb, & McPartland, 2005; Gepner, de Gelder, & Schonen, 1996; Grelotti, Gauthier, & Schultz, 2002; Klin et al., 1999). Not by chance, a study by Kirihara et al. (2012) reported an abnormal N170 in schizophrenics with low extraversion scores in response to emotional faces.

Recent evidence suggests that personality traits affect not only the processing of faces in general, but also sensitivity to social cues conveyed by faces. Individuals who score high on anxiety traits are reported to retain a more fine-grained representation of untrustworthy faces in visual working memory in comparison to individuals who score low on this dimension (Meconi, Luria, & Sessa, 2014). Moreover, they tend to overestimate social threat of ambiguous facial expressions, such as neutral faces or faces expressing low intensity of anger (Gutiérrez-García & Calvo, 2017), as a consequence of over-sensitivity towards sources of threat.. Self-protection motives (defined as those mental states that tune perceptual and cognitive processes to threat-related environmental stimuli) are also known to positively impact performance accuracy in identifying facial signals of trustworthiness (Young, Slepian, & Sacco, 2015). Finally, evidence of a relation between an individual's social attitude and their disposition to derive trustworthiness judgments form faces is also found in individuals with Autism Spectrum Disorders (ASD), who show abnormal face-based judgements of trustworthiness in association with impaired social relation and communication abilities (Adolphs, Sears, & Piven, 2001; Ewing, Caulfield, Read, & Rhodes, 2015; Forgeot d'Arc et al., 2016). Similarly, individuals with Williams Syndrome show hypersocial behavior in association with an atypical positive bias in trustworthiness judgments of unknown people (Bellugi, Adolphs, Cassady, & Chiles, 1999).

Although some studies explored the association between individual variability in personality dimensions and sensitivity to facial cues to trustworthiness, no studies at all have investigated whether such variability also impacts how facial cues to trustworthiness are perceptually represented in long-term memory, that is how faces displaying these cues are stored in memory according to their differences and similarities. In accordance with Valentine's theory (Valentine, 1991) of how facial experience is stored in memory, when navigating into the social environment our visual system builds a representational model that maps the perceived properties of the faces we encounter into a multidimensional face-space that accounts for stimulus variation. This mapping continues throughout the life span, and is constrained by the amount and quality of experience one person gains with faces (e.g., Gao, Maurer, & Nishimura, 2010; Humphreys & Johnson, 2007; Rhodes & Jeffery, 2006): the more experience we acquire with specific face types, the more fine-grained their representation will be in our perceptual space (see review by Valentine, Lewis, & Hills, 2016).

Although there is evidence that interindividual variability in face-space characteristics is related to individual differences in face recognition skills (e.g., Dennett, McKone, Edwards, & Susilo, 2012), to the best of our knowledge no studies have explored whether individual differences in personality dimensions that are central to an individual's social drive are associated to corresponding variations in the organization of their face-space. To fill this gap in the literature, Study 1 investigated whether individual differences in extraversion-introversion levels are associated with differences in perceptual sensitivity and mental representation of facial cues to trustworthiness. We focused on the extraversion–introversion dimension, as it represents the personality attribute that best explains an individual's social motivation. Indeed, people who score low on extraversion are described as socially inhibited, tend to avoid social circumstances and consider time they spend alone as more gratifying than time they spend with others. In contrast, people who score high on extraversion assiduously pursue social interactions and find time spent with others as more gratifying than time spent alone. Moreover, as already discussed, available evidence suggests that extraversion–introversion scores measured through self-report questionnaires modulate neural correlates of face processing (Cheung et al., 2010; Fink, 2005).

Our stimulus material was retrieved from Oosterhof and Todorov's (2008) Database, that includes different computer-generated face identities slightly varying on 7 levels of perceived trustworthiness. A group of typically developing adults was tested on a Perceptual Similarity Task, where they were asked to compare two probe faces to a simultaneously presented target face and select the one they judged to be more similar to the target. Perceptual sensitivity to facial cues to trustworthiness was measured as subjects' accuracy and response times in individuating the correct probe face, that is the face which is closer to the target along the trustworthiness continuum. We also collected pairwise dissimilarity scores with the aim to perform a multidimensional scaling analysis (MDS), which provided a measure of how the perceptual representation of facial cues to trustworthiness is organized in adults' long term memory. This method is used to visualize stimuli as points in a two-dimensional space, whose distance describes how similar/dissimilar they were perceived, clustered and organized (Robert, 2007; Shepard, 1980). We choose this method as it has been used in previous studies exploring the structure of face representation in adults and children (e.g., Nishimura, Maurer, & Gao, 2009), allowing researchers to describe how participants represented perceived similarities and differences across faces (Edelman, 1998).

Based on the hypothesis that individual differences in social motivation relate to individuals' sensitivity to facial cues to trustworthiness, we conjectured that individuals who scored high in extraversion would better and/or faster address similarities between the probe and the target face. Moreover, we hypothesized that, if individual differences in social motivation also affect the way faces varying in trustworthiness levels are represented in long term memory, individuals scoring high on extraversion would show a more fine-grained representation compared to those of score low on extraversion. A second task, called Multi-arrangement Dissimilarity Task (Kriegeskorte & Mur, 2012), served us to test for the stability and consistency of the similarity judgements obtained from the perceptual similarity task. Participants were required to actively organize sets and subsets of faces from the trustworthiness continuum in a two-dimensional space and in subsequent trials, and the physical distance between the faces was used to derive pairwise dissimilarity scores. The item set context in which dissimilarities are judged thus varies across trials, yielding a deeper reflection of the participants' mental representation.

Explicit subjective judgements of perceived trustworthiness (trustworthy vs. untrustworthy) and emotional expression (happy vs. angry) were also acquired for each level of trustworthiness intensity from each subjects to explore the presence of an association with extraversion levels, and whether variations in perceived trustworthiness correspond to variations in the intensity of perceived emotion (Caulfield, Ewing, Burton, Avard, & Rhodes, 2014; Todorov et al., 2008).

## Materials and methods

### Participants

The sample size was based on previous studies investigating the relation between individual differences in face processing skills and personality traits (Cheung et al., 2010; Meconi et al., 2014). In addition, a power analysis for a multiple regression model with two predictors (target trustworthiness and extraversion score) revealed that about 34 participants would be needed to have an 80% chance to observe a significant effect with an alpha level of .05 and a medium effect size. Therefore, our sample included 34 young adult participants (33 females; mean age = 24.93 years; range = 22-32). All had normal or corrected-to-normal visual skills and did not report any history of psychiatric or neurological disorders. Three additional participants were tested but had to be excluded from the final sample since they were classified outliers in response times by using both the interquartile method (upper q+3*iqr) and the standard deviation (+/-2 std from the mean). All participants signed an informed written consent before being tested. The protocol was carried out in accordance with the ethical standards of the Declaration of Helsinki (BMJ 1991; 302: 1194) and approved by the Ethics Committee of the University of Milano-Bicocca.

Stimuli

To select the stimulus set, an independent sample of 25 young adults (19 females; mean age = 24.96 years; range = 19-35) provided 9-step ratings of perceived trustworthiness for 11 computer-generated emotionally neutral male identities retrieved from the Todorov database (Oosterhof & Todorov, 2008). These face stimuli were created by using FaceGen Modeller 3.2 (Singular Inversions, www.facegen.com) based on data-driven, computational models of trustworthiness judgments. For each face identity, the authors generated seven variations along the trustworthiness dimension composing a continuum ranging from -3SD (maximal untrustworthiness) to +3SD (maximal trustworthiness), with the neutral version located at 0SD. Subjects were asked to rate each -3SD and +3SD steps of the 11 selected identities on a digital questionnaire via SurveyMonkey (SurveyMonkey Inc.) on a scale ranging from 1 (“I wouldn’t trust this person at all”) to 9 (“I would definitely trust this person”). We then selected the face identity that yielded the highest (*M*= 6.41, *SD*= 1.41) and lowest (*M*= 2.91, *SD*= 1.42) trustworthy ratings (i.e., Identity fi_002 of the Todorov database) (Figure 1.1.1).

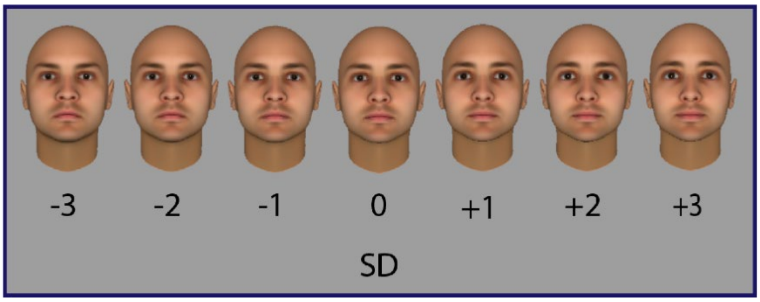


*Figure 1.1. The trustworthiness continuum composed of seven variations of the computer-generated face identity (fi_002; Oosterhof & Todorov, 2008) used in the tasks. The figure is taken from Baccolo & Macchi Cassia (2019).*

Apparatus and Procedure

Participants were tested in a quiet and dedicated room. They seated 60 cm in front of a 17.3-inch touch screen monitor with a resolution of 1080p, onto which the stimuli were presented in color; all participants completed the perceptual similarity task followed by the multi-arrangement dissimilarity task. A MATLAB (Mathworks Inc.) script interfaced with Mousetracker (Freeman & Ambady, 2010) was used to present the stimuli and collect participants' responses. Before the testing session, participants completed the Italian version of the Big Five Questionnaire (BFQ; Caprara, Barbaranelli, Borgogni, & Perugini, 1993) delivered on Qualtrics (Qualtrics, Provo, Utah, USA, https://www.qualtrics.com). At the end of the testing session they filled in two questionnaires, presented in a counterbalanced order, aimed to acquire explicit subjective judgements of perceived trustworthiness and emotional expression for each of the seven face stimuli composing the trustworthiness continuum.

*BFQ*

The Italian version of the BFQ (Caprara et al., 1993) is a self-report questionnaire composed of 44 items conceived to measure the Big Five dimensions of personality: extraversion, agreeableness, conscientiousness, neuroticism, and openness to experience. The BFQ have high internal consistency, temporal stability, convergent and discriminant validity. Participants responded on a 5-point Likert-type scale. We scored only the 24 items contributing to the Extraversion scale, and checked for the internal consistency of the scale, which was high ($\alpha$ = 0.916). Extraversion scores were then converted into z-scores.

*Trustworthiness and Happiness ratings*

Participants provided 9-step ratings (1 = "I wouldn't trust this person at all"/"This person does not look happy at all"; 9 = "I would definitely trust this person"/"This person looks very happy") of perceived trustworthiness (trustworthy vs. untrustworthy) and emotional expression (happy vs. angry) by filling in two separate Qualtrics-delivered questionnaires, presented in a counterbalanced order.

*Perceptual Similarity Task*

Participants were told that one target face would appear on the screen and that they would have to recognize which of the two probe faces appearing right after the initial presentation was more similar to the target. Participants controlled the start of each trial by pressing a START button appearing centrally at the bottom of the screen, which was replaced by the target face upon participants' pressure. After 1000 ms, two probes appeared on the right and left side of the upper portion of the screen, which remained on the screen until a response was made. Participants were instructed to keep the cursor where the START button appeared, and decide which of the two probes was more similar to the target by drifting the cursor towards the chosen probe. Five-hundred ms after a response was made, the START button reappeared on the screen. If a response was not made by 3000 ms after the probes appeared, the participant was invited to respond faster in following trials by mean of a pop-up message. The two probes and the target were all different variations on the trustworthiness continuum (Figure 1.2). Each of the seven variations was presented as target 15 times, for a total of 105 trials. The left/right position of each probe was randomized across participants. On 9 catch-trials the two probes were equally distant from the target along the trustworthiness continuum, while on all the remaining, experimental, trials one probe was closer to target than the other. Catch trials were used to obtain a full-factorial design were each variation of the continuum was compared the same number of times with all the others; they were included in the analyses on the dissimilarity scores but excluded from the analyses on response accuracy and response times. Responses on experimental trials were scored as correct when participants selected the probe that was closer to the target along the continuum. To ensure participants fully understood the task instructions, they performed a few practice trials with an example face identity.

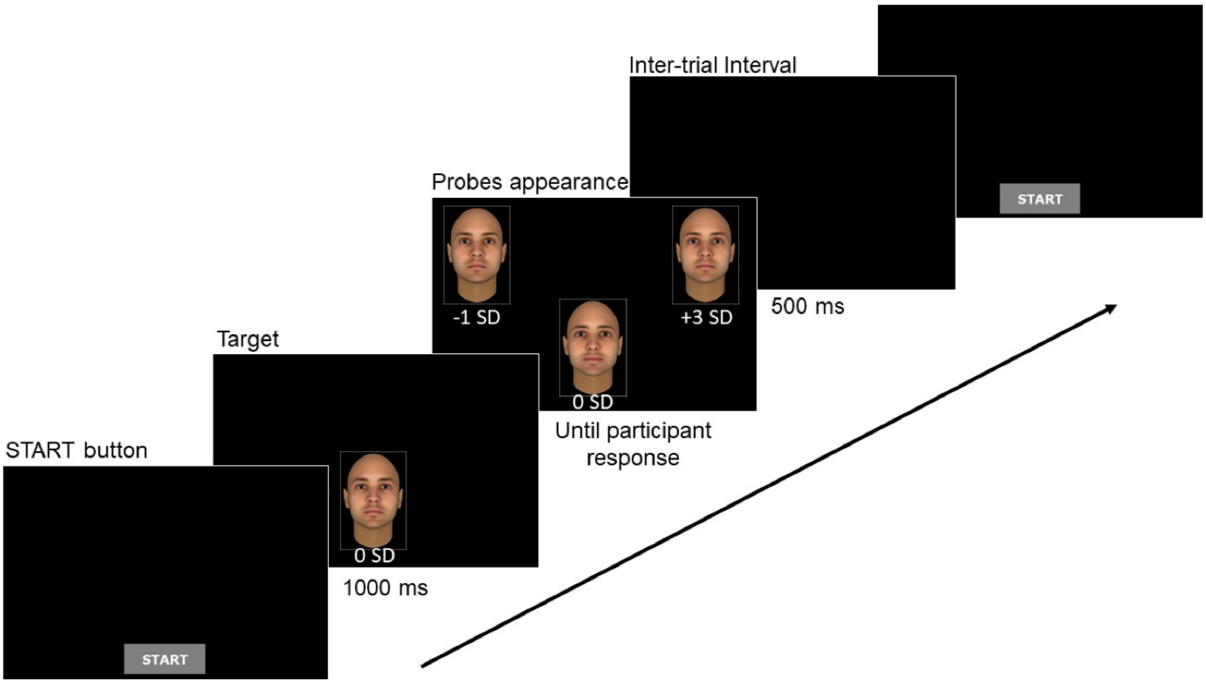


*Figure 1.2. Example trial from the perceptual similarity task. The face-probe presented on the left is closer to the target than the to the face-probe presented on the right. SD values for each stimulus are showed for clarity. The figure is taken from Baccolo & Macchi Cassia (2019).*

*Multi-arrangement dissimilarity task*

The multi-arrangement dissimilarity task was adapted from Kriegeskorte & Mur (2012). Participants were instructed to arrange the whole set and multiple subsets of the seven face stimuli composing the trustworthiness continuum presented as icons on a computer screen according to their similarity, by means of mouse drag-and-drop operations. This method works through an optimization process (Lift-the-Weakest Algorithm for Adaptive Design of Item Subsets), which keeps track of information on pairwise dissimilarities specified for each pair of items in each trial, and creates item subsets presented in subsequent trials to corroborate those dissimilarities values characterized by the weakest evidence. At the beginning of each trial, the faces were initially shown in a random order at regular intervals around a circular arena. Participants were asked to dislocate each face within the arena by dragging and dropping it with the mouse, and they were told that the position of each face in the arena indicated its similarity relationship with every other face in the arena. At the beginning of the first trial, the icons of all the seven faces appeared at the border of the arena; during the following trials, the Lift-the-Weakest Algorithm for Adaptive Design of Item Subsets selected different subsets of the seven faces based on weakest evidence of dissimilarity (see Kriegeskorte & Mur, 2012, for a detailed description of the algorithm), which appeared at the borders of the arena. The participants' task remained unchanged throughout the task, as they had to arrange the faces within the arena to reflect their perceptual similarity (Figure 1.3). Data acquisition ended after 5 minutes from the start, and not before participants completed at least the first trial. On average, participants completed 4.5 trials (range= 1-11).

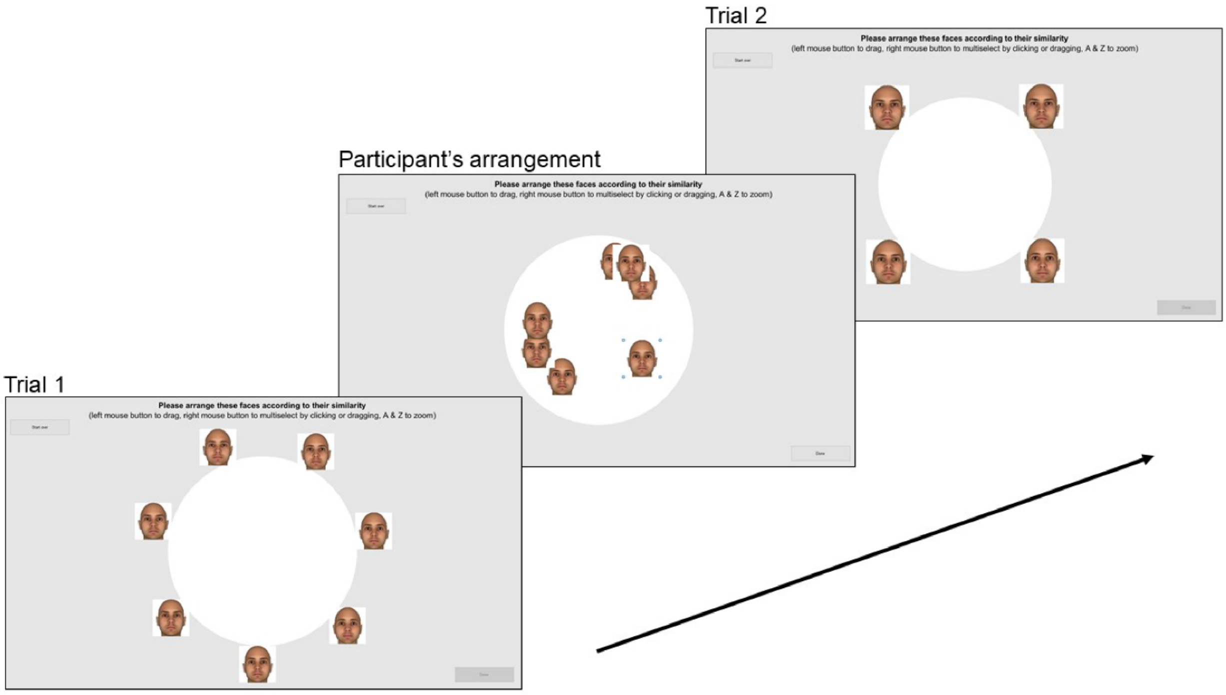


*Figure 1.3. Three subsequent example trials of the Multi-arrangement Dissimilarity Task, respectively: Trial 1, starting point (a), participant's arrangement (b), Trials 2, starting point with subsets of faces selected by the Lift-the-Weakest Algorithm for Adaptive Design of Item Subsets according to the weakest evidence of dissimilarity (c) (Kriegeskorte & Mur, 2012). The figure is taken from Baccolo & Macchi Cassia (2019).*

## Results

### *Perceptual Similarity Task*

*Response accuracy.* To test whether participants' accuracy in selecting which probe was more similar to the target face was modulated by the position of the target on the trustworthiness continuum and the participant's extraversion score, we conducted a repeated-measures Analysis of Covariance (ANCOVA) with target trustworthiness (-3SD, -2SD, -1SD, 0SD, +1SD, +2SD, +3SD) as the within-subjects factor and extraversion score entered as covariate. A significant main effect of target, $F(6, 246) = 5.345$, $p < .001$, $p\eta^2 = 0.115$, *power* = 0.995 was found: post-hoc analysis (Bonferroni corrected) revealed that participants were more accurate on -3 SD target trials 0 SD, $p = .002$, and +2 SD target trials, $p = .020$, and on +3 SD

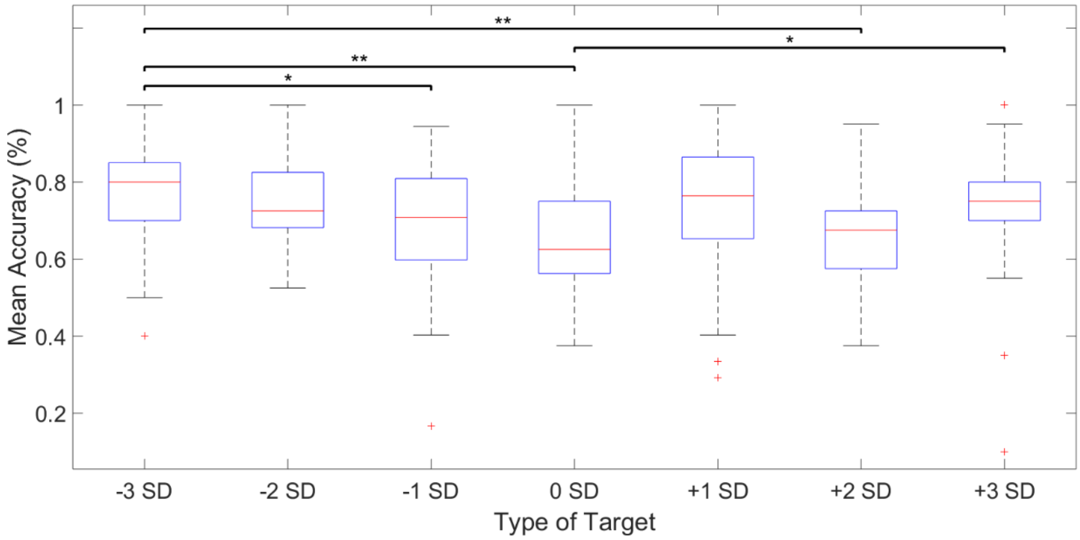


*Figure 1.4. Boxplot of mean accuracy for each target face of the trustworthiness continuum. Red crosses represent outliers. *p < .05, **p < .01, ***p < .001. The figure is taken from Baccolo & Macchi Cassia (2019).*

compared to 0 SD target trials, $p$ = .019 (Figure 1.4). A test of within-subjects contrasts revealed a significant quadratic trend, $F(1,41) = 17.370$, $p < .001$, $p\eta^2 = 0.982$.

*Response times.* A repeated-measures ANCOVA with target trustworthiness (-3SD, -2SD, -1SD, 0SD, +1SD, +2SD, +3SD) as the within-subject factor and extraversion score as covariate conducted on mean correct response times (RTs) revealed a significant main effect of extraversion score, $F(1,41) = 5.039$, $p = .030$, $p\eta^2 = 0.109$, *power* = 0.592, which we followed up by means of correlational analysis. This analysis revealed a negative association between participants' RTs and their extraversion score, $r = -0.33$, $p = .031$, 95% CI [-0.574, -0.034], as the higher their level of extraversion, the faster they were in selecting the probe that was more similar to the target faces, irrespective of the intensity of the target's trustworthiness (Figure 1.5).

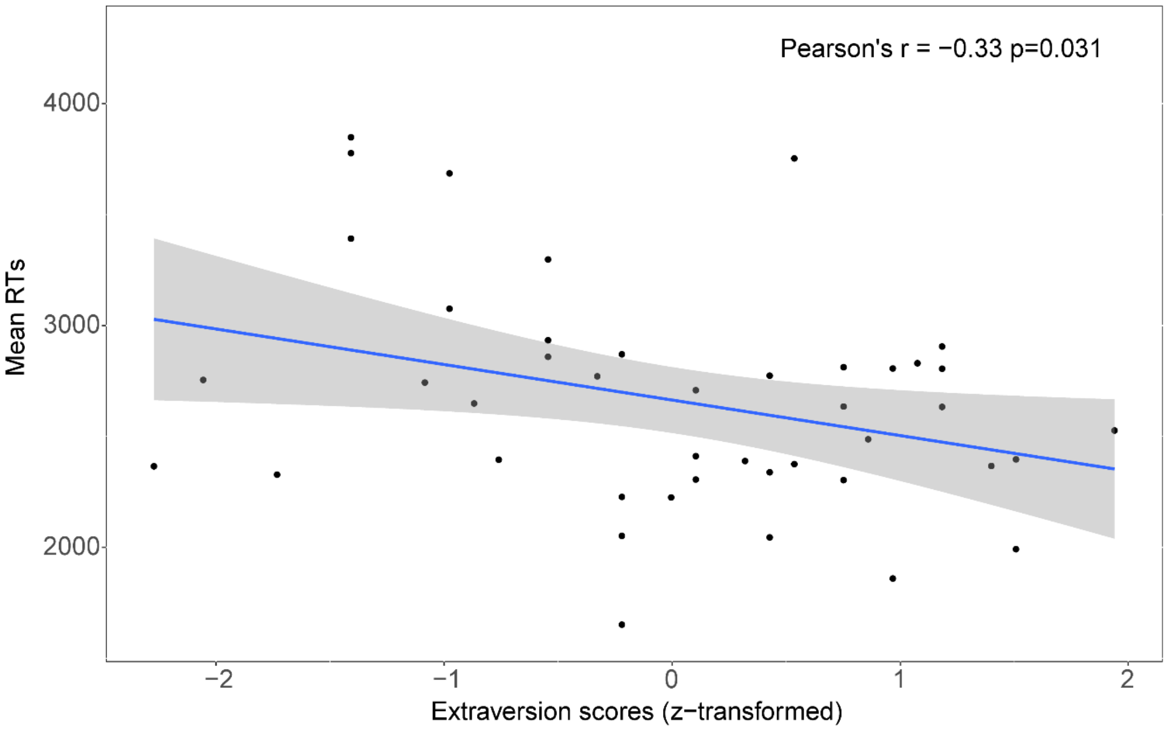


*Figure 1.5. Mean RTs in the perceptual similarity task plotted as a function of participants' Extraversion scores, r = -0.33, p = .031. The figure is taken from Baccolo & Macchi Cassia (2019).*

*Dissimilarity scores.* Pairwise dissimilarity scores were obtained for each participant from correct and incorrect responses on all trials, by assigning a distance score of 0 (minimum dissimilarity) to the face pair composed by the target face and the selected probe, and a distance score of 1 (maximum dissimilarity) to the face pair composed by the target face and the non-selected probe. For each face pair, the summed score was scaled to 0-1 by dividing it by the number of pairwise judgements acquired (N = 10). Dissimilarity scores for each subject were used to derive a 7x7 Representational Dissimilarity Matrix (RDM) representing how dissimilar each face pair was perceived to be. Each column and row represents the dissimilarity judgement for one step of the trustworthiness continuum with respect to every other steps of the, with the diagonal representing the perceived dissimilarity within the same step of the trustworthiness continuum, which is therefore a diagonal of zeros (Kriegeskorte, Mur, & Bandettini, 2008). The resulting RDMs were averaged across subjects to build a single RDM representing the average perceived dissimilarity between face pairs differing in trustworthiness intensity (Figure 1.6).

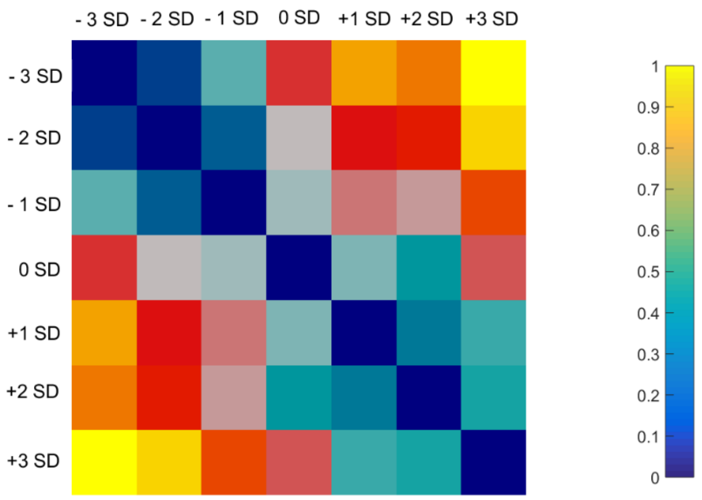


*Figure 1.6. Representational Dissimilarity Matrix (RDM) derived from participants' responses in the perceptual similarity task and depicting the average perceived dissimilarity between pairwise combinations of faces from the trustworthiness continuum (yellow (light grey) = maximum dissimilarity, blue (dark grey) = minimum dissimilarity). The figure is taken from Baccolo & Macchi Cassia (2019).*

With the aim to represent participants' perceived similarity between the seven steps of the trustworthiness continuum, pairwise dissimilarity scores were averaged across participants and used for a multidimensional scaling (MDS) analysis with a MATLAB script adapted from Kriegeskorte and Mur (2012). Kruskal stress formula (Kruskal & Wish, 1978) was used to compute the goodness-of-fit of the MDS solution, which turned out to be 0.087. With the aim to explore how participants clustered together the faces in the MDS solution, a cluster analysis with the dendrogram MATLAB function (Sireci & Geisinger, 1992) was performed (Figure 1.7).

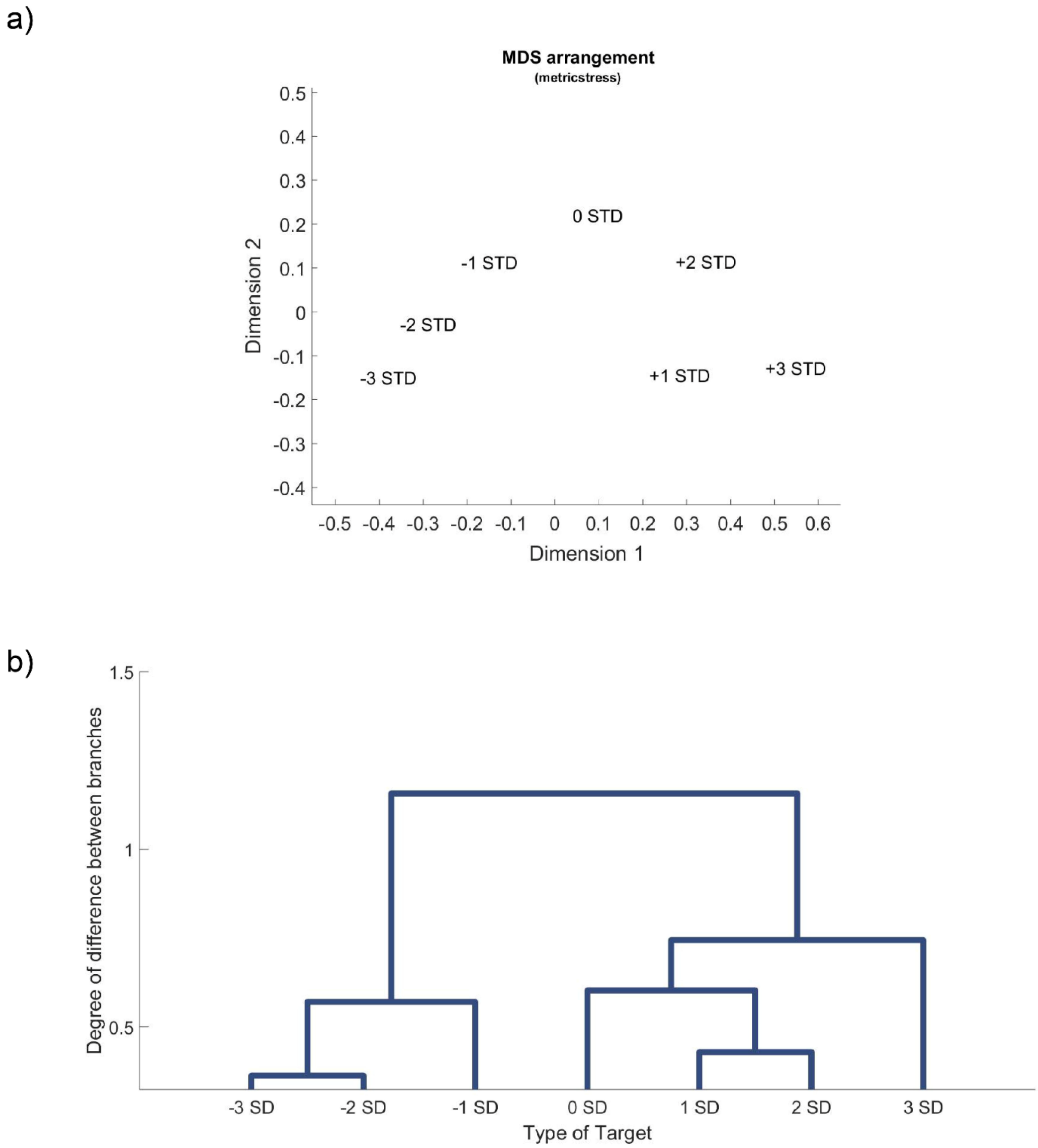


*Figure 1.7. The MDS solution derived from average dissimilarity scores acquired during the perceptual similarity task (a), and the corresponding hierarchical plot (b). The figure is taken from Baccolo & Macchi Cassia (2019).*

With the aim to verify to which extend participants agreed in attributing dissimilarity scores for each step of the trustworthiness continuum, a repeated-measure ANOVA on cosine distances between participants' pairwise dissimilarity scores for each trustworthiness step was computed. Cosine distance is one minus the cosine of the angle between two vectors of an inner product space: Two vectors with identical orientation have a cosine distance of 0, two orthogonal vectors have a cosine distance of 1. Therefore, cosine distance ranges between 0 and 1. Since, in a RDM, each row contains the dissimilarity scores obtained for one trustworthiness step with respect to every other trustworthiness steps, cosine distance was computed between each row of the RDM of every single subject and the corresponding RDM's row of the of every other subject. A repeated-measures ANOVA on cosine distances with trustworthiness intensity as the within-subjects factor reached statistical significance, $F(6,5412) = 125.735$, $p < .001$, $p\eta^2 = 0.122$, *power* = 1, adding evidence that participants consistently agreed in attributing dissimilarity scores for very (+/- 3SD) and moderately (+/- 2SD) trustworthy or untrustworthy faces than for all the others, *p*s < .05 (Figure 1.8).

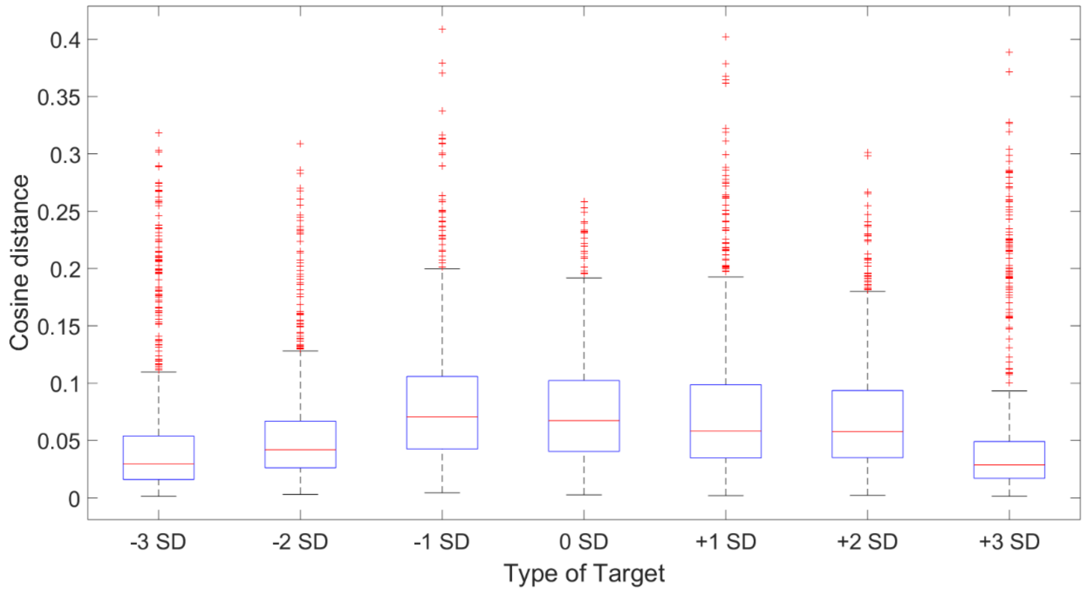


*Figure 1.8. Boxplot of cosine distances between subjects' pairwise dissimilarity scores plotted as a function of trustworthiness intensity. Red crosses represent outliers. The figure is taken from Baccolo & Macchi Cassia (2019).*

To explore whether variability in the way subjects represented facial cues to trustworthiness was related to their extraversion scores, for each participant pairs, cosine distance between the entire RDMs (namely, dissimilarity scores between all face pairs) and the euclidean distance between their extraversion scores was computed. Our conjecture was that if the expectation that subjects represent facial cues to trustworthiness as a function of their levels of extraversion, the greater the difference among extraversion scores, the greater the difference among RDMs. Against this expectation, a correlation between euclidean distances of subjects' extraversion scores and the cosine distances of the same subjects' RDMs computed on a total of 903 pairs, namely all possible pairwise combinations of the 43 participants, proved to be negative and significant, but very weak, $r = -0.08$, $p < .01$, 95% CI [-0.151, -0.021]. Indeed, as can be observed from Figure 1.9, subjects agreed in the attribution of dissimilarity judgements independently of their extraversion scores: 95% of datapoints remained below a cosine distance of 0.177, with 0.067 and 0.045 as mean and standard deviation.

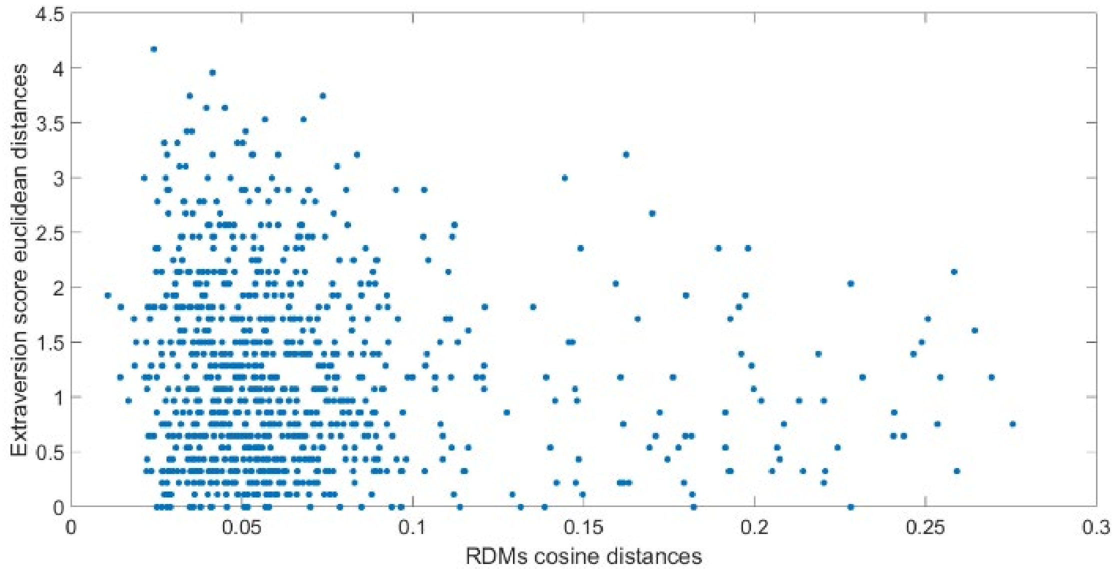


*Figure 1.9. Cosine distances between the RDMs of subjects pairs as a function of the euclidean distances between subjects' extraversion scores. The figure is taken from Baccolo & Macchi Cassia (2019).*

*Multi-arrangement Dissimilarity Task*

An average RDM was obtained also from the pairwise dissimilarity scores derived from participants' performance in the Multi-arrangement Dissimilarity task. Correlational analysis revealed a significant positive association between this matrix and the one obtained from the dissimilarity scores derived from the Perceptual Similarity task, $r = 0.83$, $p < .001$, 95% CI [0.628, 0.930]. This indicates that the dissimilarity scores derived from the two tasks are comparable, suggesting that participants' perception of dissimilarity among the seven faces varying in their level of expressed trustworthiness was consistent and reliable.

*Trustworthiness and Happiness ratings*

To determine how participants' explicit judgments of perceived trustworthiness varied as a function of the faces position along the trustworthiness continuum, we performed a repeated-measures ANCOVA with trustworthiness intensity (-3SD, -2SD, -1SD, 0SD, +1SD, +2SD, +3SD) as the within-subjects factor, and extraversion score entered as covariate. A main effect of trustworthiness intensity was found, $F(6,264) = 71.157$, $p < .001$, $p\eta^2 = 0.634$, *power* 1.0. Within-subjects contrasts showed a significant linear trend, $F(1,41) = 166.056$, $p < .001$, $p\eta^2 = 0.802$, *power* 1.0, indicating that participants' judged the trustworthiness intensity of the seven faces form the continuum as varying as a function of the stimulus position along the continuum itself.

With the aim to address whether subjects' assigned explicit judgements of perceived emotion in relation to the position of faces in the trustworthiness continuum, we performed a repeated-measures ANCOVA with trustworthiness intensity as the within-subjects factor and extraversion score entered as covariate. The ANCOVA revealed a main effect of trustworthiness intensity, $F(2,264) = 199.571$, $p < .001$, $p\eta^2 = 0.819$, *power* 1.0. Also in this case, within-subjects contrasts showed a significant linear trend, $F(1,41) = 522.098$, $p < .001$, $p\eta^2 = 0.922$, *power* 1.0, indicating that participants' judgments on emotion valence and intensity of the seven faces varied as a function of the position of the faces along the trustworthiness continuum.

To explore the relationship between explicit judgments of perceived trustworthiness and explicit judgments of perceived happiness of the seven faces, we ran a correlational analysis between the two judgments scores, which proved significant, $r = 0.99$, $p < .001$, 95% CI [0.918, 0.998].

Perceived trustworthiness was positively related to perceived happiness: participants' judgements of happiness increased by 1.17683 points for each increase in judgments of trustworthiness. Nevertheless, happiness in the faces was judged as overall less intense than trustworthiness, as revealed by a paired-sample *t*-test on participants' judgments along the two dimensions, which proved significant, $t(6)= 3.963$, $p = .007$.

**Discussion**

In Study 1, we were interested in exploring whether and how variability in adults' social attitudes is associated to differences in sensitivity to facial cues to trustworthiness, and in the way such cues are represented in long term memory. As discussed in the Introduction, we focused on the extraversion dimension as it represents the personality attribute that best explains an individual's social motivation. Indeed, people who score high on extraversion are described as socially active, tend to engage in social circumstances and consider time they spend with others as more gratifying than time they spend alone, which would in turn affect the amount of perceptual experience that such individuals acquire with faces through social interactions. We therefore hypothesized that individuals with higher extraversion scores would be more efficient in assessing perceptual similarity between faces that slightly varied in the level of trustworthiness they expressed. Indeed, analyses of participants' response times in the perceptual similarity task confirmed that individuals with higher extraversion scores were faster at successfully disentangle between fine-grained differences in trustworthiness intensity expressed by the probe faces. This finding is in line with evidence form previous studies showing that differences on the introversion-extraversion dimension affect EEG responses to emotional faces (Fink, 2005).

The acquisition of perceptual expertise in processing faces across development might in fact be bound to social motivation, which might impact the amount of time an individual spends with other people (Cheung et al., 2010), and consequently the amount of experience one gains with faces. Not by chance, visual experience in form of adaptation to different levels of face trustworthiness are already known to improve perceptual sensitivity to this face dimension (Keefe et al., 2013). Our results showing a negative correlation between extraversion scores and response times to successfully detect fine-grained variations in the intensity of expressed facial trustworthiness might therefore be explained as an effect of greater perceptual expertise in extraverts than in introverts. Indeed, an individual's social (and cultural) experience plays an important role in shaping the spontaneous tendency to infer other people's traits from facial attributes (Over & Cook, 2018). Developmental research suggests that early in development infants are extremely sensitive to social cues related to other people's appearance or actions, and modulate their own behavior based on such cues (Mascaro & Csibra, 2014; Van de Vondervoort & Hamlin, 2017). Even though research on sensitivity to facial cues to trustworthiness has focused mainly on adults, there is evidence that preverbal infants can discriminate between different levels of expressed trustworthiness (Jessen & Grossmann, 2016; Jessen & Grossmann, 2017). By the age of 3 years, children can use facial cues to generate explicit judgments about how "nice" or "mean" a person looks (Cogsdill et al., 2014). This suggests that, although mature personality traits may emerge later in childhood, sensitivity to those facial cues that are later used to generate trustworthiness judgements emerges early in development, to be continuously refined throughout the lifespan as a result of each individual's own experience within their social environment. Indeed, individual differences in two temperamental traits that have been linked to adults' Extraversion – approach to novelty and fearfulness – are known to modulate the amount and quality of experience infants gain with social stimuli, consequently affecting neural sensitivity to such stimuli (e.g., de Haan, Belsky, Reid, Volein, & Johnson, 2004; Rajhans, Missana, Krol, & Grossmann, 2015; Taylor-Colls & Pasco Fearon, 2015). Following from this evidence, our finding of faster responding to facial cues to trustworthiness in extraverts relative to introvert adults may represent a developmental outcome of distinctive ontogenetic experience with the social environment.

Importantly, differences on the extraversion dimension did not affect response accuracy in the perceptual similarity task. Instead, variance in subjects' accuracy depended on the intensity of the trustworthiness cues expressed by the target face, following a quadratic trend: Higher accuracy values were found when the target face belonged to one of the continuum extremes (- 3 SD, + 3 SD), independently of the valence, compared to the neutral (0 SD) face. Therefore, participants' ability to perceptually discriminate faces seems to be affected by the intensity of the facial cues that are relevant to trustworthiness judgments in such a way that those faces including more intense social cues enjoy a processing advantage over those including less intense cues. This is in line the widely reported finding of an attentional and processing advantage of angry (e.g., LoBue, 2009) and fearful/threatening (e.g., Holmes, Green, & Vuilleumier, 2005) faces over neutral ones. Nevertheless, our results were not limited to faces with a negative valence, as also the extremely trustworthy face enjoyed the same processing advantage. This finding of valence-independent sensitivity to trustworthiness cues is more in line with those data showing a processing advantage for emotional faces over neutral ones, irrespective of their valence (Roesch, Sander, Mumenthaler, Kerzel, & Scherer, 2010). Accordingly, imaging studies revealed that the amygdala shows stronger activation in response to both positively valenced and negatively valenced faces that lie at extremes of the trustworthiness continuum than to faces near the center of the continuum, following a quadratic response pattern similar to that observed for accuracy rates in the current study (Said, Baron,

& Todorov, 2009; Said, Dotsch, & Todorov, 2010). This response pattern was linked to the fact that neutral faces are perceived as more typical than faces that lie at the extremes of the trustworthiness continuum: the amygdala response would therefore be explained by deviations in typicality rather than by deviations in valence (Said et al., 2010). A similar pattern of valence-independent sensitivity to trustworthiness cues was found in preverbal infants, who show 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).

A second aim of Study 1 was to investigate whether variability in social motivation affects the way faces that slightly varied in the level of expressed trustworthiness are represented in a multidimensional space. We obtained no evidence that participants' extraversion scores modulated how cues to trustworthiness are represented in long term memory. Rather, as evident from the MDS and cluster analysis, all participants represented faces as a function of the intensity of facial cues to trustworthiness, grouping them in a cluster of untrustworthy faces (-3, -2, -1 SD) and a cluster of trustworthy faces (+1, +2, +3 SD), to which the neutral face (0 SD) is attached. Therefore, the neutral face was perceived to be more similar to the trustworthy cluster, and this resonates well with findings showing that that neutral faces are often associated to emotional states (e.g., Adams, Nelson, Soto, Hess, & Kleck, 2012; Lee, Kang, Park, Kim, & An, 2008), suggesting that absolute neutrality is not a feature of face perception.

Congruent with the idea that neutral faces embody ambiguous cues, our data suggested that subjects agreed less in assessing similarities of the neutral, -1 SD and the +1 SD faces. Indeed, cosine distance values, our measure of participants' disagreement, followed a parabolic trend that reached the highest values in the central hub of the continuum, where, congruently, accuracy values were the lowest. Rather, subjects' agreement was higher when attributing dissimilarity judgements for faces that lied at trustworthiness extremes (namely +3 and -3 SD), consistent with the observation that more intense facial cues to trustworthiness are more easily processed than less intense ones, independently of valence.

It is important to note that individual differences in extraversion scores shaped sensitivity to facial cues to trustworthiness but not explicit judgements on this face dimension. Rather, explicit judgements of perceived trustworthiness increased from the -3 SD face to the +3 SD face following the intensity changes in the continuum as it was conceived by Oosterhof and Todorov (Oosterhof & Todorov, 2008). Interestingly, trustworthiness judgements did not differ for each single trustworthiness intensity: subjects did not assign significantly different judgements for -3SD and -2SD, -2SD and -1SD, +1SD and +2SD, +2SD and +3SD. This could be due to the fact that even the most fine-grained changes in physical cues are enough to provoke strong judgements on a person's trustworthiness, independently from the intensity of the physical cue displayed.

Explicit judgements of perceived trustworthiness were strongly correlated with judgements of perceived happiness in our data. This is congruent with previous demonstrations of a robust association between perceived trustworthiness and the attribution of emotional states (Adams, Ambady, Neil Macrae, & Kleck, 2006; Oosterhof & Todorov, 2008; Todorov et al., 2008). This result is also in line with the emotion overgeneralization hypothesis (Said, Sebe, & Todorov, 2009; Zebrowitz, Fellous, Mignault, & Andreoletti, 2003), according to which spontaneous trustworthiness judgements are overgeneralized responses to facial cues that resemble emotional expressions. Nevertheless, although strongly correlated, happiness judgments were significantly lower than trustworthiness judgments, suggesting that, albeit being an important cue to trustworthiness, variance in emotional valence cannot fully explain the variance of trustworthiness evaluation. For example, other characteristics, like face typicality and/or attractiveness (Said et al., 2010), could play a role in building our impression of others' approachability (e.g., Jones et al., 2012; Sofer, Dotsch, Wigboldus, & Todorov, 2015).

It is important to recognize that Study 1 has few important limitations. First of all, due to the large number of trials needed to achieve dissimilarity judgements for all possible pairwise combinations of faces, we were forced to use one single face identity, which could have partly impaired the generalizability of our results. Secondly, computer-generated faces might not satisfactorily reflect individuals' expertise with faces (Crookes et al., 2015), as we obviously do not encounter artificial faces in our daily social interactions. The use of more ecological face stimuli is advisable for future studies, as they might evoke different patterns of performance. Finally, the fact that we did not observe any effect of individual differences on response accuracy levels in the Perceptual Similarity task could be related to the fact that the task itself was not challenging enough. Participants were instructed to be as fast as possible in providing their response through initial instructions and through a reminder that appeared on the screen each time 3000 ms had elapsed after the probes appeared and no response was made. However, a shorted stimulus presentation duration would have made the task more difficult and maybe more sensitive to individual differences in performance. We opted to avoid a response deadline because one aim of the task was to collect dissimilarity judgements from all possible pairwise combinations between trustworthiness levels from the continuum. To do so, a response on each trial was needed, while a response deadline could have caused null responses to occur.

In summary, Study 1 provided new evidence that individual differences in personality traits that are relevant to social motivation, namely extraversion, are able to modulate perceptual discrimination of facial cues to trustworthiness. Individuals who scored higher on extraversion were faster in attributing similarity judgements to faces displaying fine-grained differences in the intensity of expressed trustworthiness. Still, response accuracy or the way trustworthiness facial cues are represented in long term memory were not shaped by individual differences in extraversion scores. Taken together, these results add further evidence to the ontogenetic hypothesis that social motives related to one's personality can have an impact on the amount of social experience that drives the development of perceptual sensitivity to facial cues to trustworthiness (see Young et al., 2015).

# Chapter 2

**The ontogenetic roots of sensitivity to facial cues to trustworthiness**

In Chapter 1, we illustrated how humans display fine-grained sensitivity to non-verbal cues from faces, such as facial cues to trustworthiness, and discussed how individual differences related to social attitudes can shape this sensitivity. We proposed a developmental framework, within which social motives related to personality traits would impact an individual's amount of social experience and thus the level of developed sensitivity to facial cues to trustworthiness. Within this view, it would be reasonable to assume that different amounts of exposure to faces accumulated at different times in development might modulate the degree of sensitivity to such cues, just like visual experience in form of adaptation to different levels of face trustworthiness improves perceptual sensitivity to this face dimension (Keefe et al., 2013).

Sensitivity to social information delivered by faces of the individuals who are approaching us originates in the very early stages of development. Infants can disentangle between different emotional facial expressions (e.g., happiness and fear) by the age of 7 months (e.g., Grossmann, 2010; Nelson, 1987; Quadrelli, Conte, Macchi Cassia, & Turati, 2019), and can regulate their behaviour accordingly (Cohn & Tronick, 1983; Hirshberg, 1990). If a large part of the literature investigated how sensitivity to facial information associated to emotional expressions emerges early in the development to refine in the first 18 months of life, only a few studies investigated the developmental trajectory of the sensitivity to facial cues to trustworthiness and the ability to generate approach/avoidance behaviours based on such cues. Little evidence from studies conducted with preverbal infants indicate that, at least to some extent, sensitivity to facial cues that are relevant to social inferences is present in the first year of life. For example, 6-month-old infants show attentional biases towards threat-related facial configurations like fear (Peltola et al., 2008), and 4- to 6-month-olds prefer to approach individuals whose face displays an happy expression, and avoid individuals exhibiting negative facial expressions (Serrano et al., 1995). Moreover, infants prefer prosocial individuals to antisocial ones, who may represent a possible source of threat (Hamlin & Wynn, 2011; Van de Vondervoort & Hamlin, 2017), and they would rather approach an individual with which their mother interacted positively (Fein, 1975). This is similar to what 6- to 11-year-old children do when they prefer to trust those individuals who help others rather than those who don't (Fu et al., 2015). Interestingly, Gredebäck and colleagues (Gredebäck et al., 2015) showed that infants' preference for prosocial others (Hamlin & Wynn, 2011) is also reflected in infants' electrocortical activity, as they found that the temporo-parietal P400 event-related potential (ERP) component differentiated between agents that were helping others and agents who were hindering others, indicating that the infants' brain react differently to the two events.

Evidence in favour of the hypothesis that sensitivity to facial cues that are used by adults to generate trustworthiness judgements emerge as early as 7 months of age is found in three studies by Jessen & Grossmann (2016, 2017, 2019). A study investigating the electrophysiological response to faces varying in levels of expressed trustworthiness showed that infants could discriminate between neutral faces and high or low trustworthy faces (Jessen & Grossmann, 2016), even when face-stimuli were subliminally presented (Jessen & Grossmann, 2017). By using a preferential looking paradigm, the authors showed that 7-months-old infants also preferred to orient their attention towards trustworthy faces than to neutral or untrustworthy faces (Jessen & Grossmann, 2016). Accordingly, infants oriented more efficiently toward objects that were paired with trustworthy faces than untrustworthy faces in a gaze-cueing paradigm (Jessen & Grossmann, 2019). Similar results were obtained by a study conducted with 6- to 8-month-old Japanese infants, who showed a preference for trustworthy Caucasian faces over untrustworthy Caucasian faces, yet this pattern was observed only for highly dominant faces (Sakuta et al., 2018). Taken together, these studies seem to suggest that we are tuned to perceptual cues of other people's approachability since the very early stages of development.

A Study by Cogsdill and colleagues (Cogsdill, Todorov, Spelke, & Banaji, 2014) revealed that when they are 3 year-olds, children use these facial cues to trustworthiness to generate explicit judgements about how 'mean' or 'nice' a person looks, and by 6 years of age these judgements reach the same consistency as the one of adults'. Eight, 10-, and 12-year-olds display within-age consistency when judging trustworthiness (rated on a 3-point scale) of a series of faces, and older children, but not younger ones, even display similarities and consistencies with adults' judgements (Ma, Xu, & Luo, 2016). Children aged 5 and 10 who are playing an economic trust game more favourably place their trust in individuals whose look is trustworthy than to those who look untrustworthy (Ewing et al., 2015).

Overall, evidence from children suggests that sensitivity to facial cues to trustworthiness and the ability to use such cues to make explicit judgements of trustworthiness might not require extended social experience, but would have emerged as a result of evolutionary pressure as an adaptive response to threat (Cogsdill, Todorov, Spelke, & Banaji, 2014; LoBue, 2009).

Based on this evidence, Studies 2, 3, and 4 have investigated the emergence of sensitivity to facial cues to trustworthiness in preschoolers and school-aged children (Study 2), as well as in infancy (Study 3 and 4). More specifically, Study 2 investigated developmental differences in children's and adults' fine-grained sensitivity to facial cues to trustworthiness, focusing on the perceptual representation of such cues, and taking into account individual differences in emotion comprehension. Study 3 and 4 extended the investigation of the developmental origins of perceptual sensitivity to facial cues to trustworthiness to the first year of life. Study 3 replicated and extended the evidence obtained by Jessen and Grossman (2016) with 7-month-old infants by using more ecological face stimuli (images of real faces as opposed of computer-generated faces), and by investigating the role of temperamental traits in shaping neural sensitivity to facial trustworthiness cues. Study 4 extended the results of Study 3 by using a newly developed EEG (electroencephalographic) paradigm known as Fast Periodic Visual Stimulation (Rossion, 2014), which will be extensively described in the designated method section.

## Study 2[2]

**Adults' and children's sensitivity to facial cues to trustworthiness: developmental differences and the role of emotional development**

Study 2 aims at tackling a few questions that have not been covered by previous studies on the development of human sensitivity to facial trustworthiness.

First, we aimed at exploring how perceptual sensitivity to slight differences in facial cues to trustworthiness develops across childhood and into adulthood. Previous studies explored infants' (Jessen & Grossmann, 2016, 2017) and children's (e.g., Cogsdill & Banaji, 2015; Cogsdill et al., 2014;  Ma et al., 2016) sensitivity to facial cues to trustworthiness by using computer-generated faces obtained from data-driven modelling (but see Cogsdill & Banaji, 2015) lying at the opposites of the trustworthiness continuum, and thus expressing opposite and extreme levels of trustworthiness. Due to the high high level of distinctiveness of the trustworthiness opposites, participants' might have been facilitated in accomplishing the task and their performance might have been inflated. Moreover, asking children to distinguish between very trustworthy and very untrustworthy faces may limit our chances to measure their fine-grained sensitivity to facial cues to trustworthiness in real-life situations, as we regularly

[2] The data presented here were discussed in a paper that has been recently accepted for publication: Baccolo, E., Macchi Cassia, V. (2019). Age-related differences in sensitivity to facial trustworthiness: Perceptual representation and the role of emotional development. *Child Development*.

process minor facial cues and decode even the slightest difference in our interlocutors' facial expressions. Finally, albeit strictly manipulated and controlled, computer-generated faces may not fully reflect participants' expertise at face processing, including perceptual discrimination (e.g., see Balas & Pacella, 2017; Crookes et al., 2015). For all these reasons, in Study 2 we opted for using as stimulus material a set of seven real face images that were created by parametrically manipulating one face identity, such that the level of perceived trustworthiness slightly varied along a seven-steps continuum from low trustworthiness to high trustworthiness. This allowed us to trace developmental differences in children's perceptual sensitivity to fine-grained variations in facial cues to trustworthiness (see Method) embedded in an exemplar of a face category (i.e., female face) that is highly familiar to young children (see Quinn, Yahr, Kuhn, Slater, & Pascalils, 2002; Ramsey-Rennels & Langlois, 2006).

It is important to stress that the aim of the study was to investigate children's sensitivity to those internal facial features whose configuration vehiculate trustworthiness judgements in adults, rather than which type of face identity (and the related morphology) is perceived as more or less trustworthy. In fact, all our studies are framed in the line of research that does not consider trustworthiness as a stable trait that is associated to specific individual facial morphologies, but rather a dynamic trait emerging from a specific configuration of facial features (e.g. up/downturned eyebrows and upward/downturned lips) that, irrespective of facial identity, consistently trigger social perception of trustworthiness (e.g., see Jack & Schyns, 2015). The way our stimulus set was created reflects these intents as, by using a data-driven approach, we didn't have to specify a-priori which Action Units (AUs) should vary to manipulate perceived trustworthiness, but rather use trustworthiness references obtained from averaging very trustworthy/untrustworthy faces to extrapolate the facial cues that are crucial in in influencing a person's trustworthiness judgement independently of face identity and gradually morph a neutral face towards these references. A similar approach was used by

Todorov and colleagues (Oosterhof & Todorov, 2008) when creating their own set of stimuli. The authors used this method because, as stated by Oosterhof and colleagues (Todorov, Olivola, Dotsch, & Mende-Siedlecki, 2015), "*it is far from clear how one should define what counts as a facial feature (mouth versus corner of mouth versus lips versus image pixel, etc.). This difficulty is further compounded by the fact that some features may not have labels and, moreover, that both perceivers and experimenters may be unaware of the effects of these features on social perception. Todorov and his colleagues have advocated an alternative, data-driven approach to model social perception of faces in an unbiased fashion [...] capable of capturing the variance in facial structure that leads to specific social attributions*" (p. 5). In line with this approach, we selected from the Chicago Face Database (Ma, Correll, & Wittenbrink, 2015) a set of faces that were previously rated as the most trustworthy and the most untrustworthy, and we averaged them to obtain a trustworthy and an untrustworthy prototype. An average neutral face was morphed towards the two references in order to obtain the a seven-steps trustworthiness continuum. The idea behind using averaged, prototype, faces as references was to extract those features that are important for trustworthiness perception across individual face identities.

Perceptual sensitivity to facial cues to trustworthiness was inferred from 5-year-olds', 7-year-olds' and adults' performance in an *Oddmanout task* (as in Nishimura, Maurer, & Gao, 2009), where participants were presented with a set of three faces taken from the trustworthiness continuum and were asked to select the face that they perceived to be the most different from the rest. Pairwise dissimilarity scores were computed as a measure of children's and adults' perception of similarity/dissimilarity among all the seven faces composing the trustworthiness continuum, and allowed us to estimate their ability to discriminate the facial information that slightly varied along the continuum.

Dissimilarity scores also allowed us to tackle a second question not covered in previous studies, i.e. if children and adults differ in the way they represent physical cues to trustworthiness in long term memory. According to the well-known face-space model proposed by Valentine (1991; also see review by Valentine, Lewis, & Hills, 2016), throughout development our visual system extrapolates and maps the characteristics of the faces we encounter in our everyday social interactions into a multidimensional perceptual space. Within this space, faces are represented according to their perceived similarity along different dimensions, such that variance in their position reflects differences among critical cues that are important for discriminating among individual faces. The structure of the face-space is shaped by the amount of experience one gains with faces, and refines with 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), but major reorganizations take place 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 also undergoes important refinements across this same age range (Rodger et al., 2015). Moreover, by the age of 7 years children do not differ from adults in terms of the number of dimensions along which they represent faces, even though when making similarity judgments children rely more heavily on one single dimension, such as hair type or eyes colours, while adults use all dimensions equally (Nishimura et al., 2009; Pedelty, Levine, & Shevell, 1985).

In Study 2, age-related changes in long-term memory representation of facial cues to trustworthiness were investigated by using dissimilarity scores derived from the *Oddmanout* *task*. For each of the three age groups (younger children, older children, and adult), these scores were used to build a Representational Dissimilarity Matrix (RDM), which shows how participants represented similarity across faces, and to perform a cluster analysis which showed how participants grouped faces together according to their perceived similarity (Sireci & Geisinger, 1992). In light of earlier demonstrations that children aged 3 years are able to make explicit approach/avoidance judgements 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.

One last issue Study 2 aimed to explore related to the impact of emotional development on children's ability to make trustworthiness judgements. Recent evidence has shown that individual differences in personality and social behavior affect how facial cues to trustworthiness are detected and used to generate social trait inferences (see Study 1, but also Meconi, Luria, & Sessa, 2014; Young, Slepian, & Sacco, 2015). Diminished abilities to disentangle socially relevant cues from faces are associated to impairments in social cognition and mentalizing abilities. For example, adults with ASD who have atypical emotion recognition skills show abnormal face-based judgements of trustworthiness (Adolphs, Sears, & Piven, 2001; Forgeot d'Arc et al., 2016).

Despite this evidence supports the alleged relation between emotion recognition skills and social judgements from facial cues, to the best of our knowledge, so far no studies have investigated whether individual variations in emotion comprehension abilities are associated to corresponding variations in children's perceptual sensitivity to facial cues to trustworthiness as well as in their attitude to use such cues to generate trustworthiness judgements. In the first years of life, social cognition is subject to impressive developmental modifications, and this is especially true for the so called *Theory of Mind*, which is the ability to understand other people's mental states (emotions, thoughts, desires, motives) and to respond to them appropriately (Astington & Dack, 2008). If at 3 and 4 years of age children can use other individuals' facial expressions to infer their emotional states, it is only by the age of 5 years of age that they first develop more critical components of emotion understanding (e.g., the situational causes of the outward expression of emotion). Other fundamental components of emotion understanding, like the relationship between one's beliefs and his/her emotional states, appear during school years, and reach adult-like levels only in early adolescence (Pons, Harris, & de Rosnay, 2004).

In light of this evidence, Study 2 explored the relationship between children' s development of emotion development and their sensitivity to facial cues to trustworthiness by focusing on the 5- to 7-year age range. More precisely, we explored whether 5- and 7-year olds' emotion comprehension skills affected their perceptual representation of facial cues to trustworthiness and/or their ability to make explicit trustworthiness judgements. We did so by correlating children's scores in the Test of Emotion Comprehension (TEC; Pons & Harris, 2000) to their performance in the *Oddmanout task* and in a second task – i.e., the *Pairwise 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 (as in 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. On the contrary, in the *Pairwise Preference task* children were asked to indicate which face they would better trust within a pair of faces randomly selected from our trustworthiness continuum. Participant's response was used to compute a trustworthiness preference score for each single face.

In summary, Study 2 have three main goals: (1) to explore whether perceptual sensitivity to face information subtending social perception of trustworthiness changes across childhood and into adulthood, (2) to investigate the presence of age-related differences in the structure of the mental representation of facial cues to trustworthines, and (3) to explore whether children's emotion understanding abilities modulate their social perception of trustworthiness from faces.

Five- and 7-year-old children were selected as target age groups in order for our data to be comparable with those obtained by previous studies exploring children's ability to make trustworthiness judgements, which targeted this same age range (Caulfield et al., 2016; Ewing et al., 2015). Furthermore, the 5- to 8-years age range is also critical for the development of face representation, including the representation of facial expressions of emotions (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 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

A Power Analysis for a univariate ANOVA with three groups (5-year-olds, 7-year-olds and adults) was conducted to estimate sample size, which showed that about 64 participants should lead to an 80% chance to obtain a significant effect with an alpha level of .05 and a large effect size. The final sample was composed 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). Recruitment for children took place in preschools and schools within a major city area. All children came from middle-class Caucasian families (except one Hispanic) and lived in a racially homogeneous neighbourhood. Recruitment for adults took place either at the university or by word of mouth in the community, and all subjects came from middle-class families. An additional 17 children (10 5-year-olds) were excluded from the final sample due to being distracted during the test. All procedures were carried out in accordance with the Ethics Standards of the Declaration of Helsinki (BMJ 1991; 302: 1194) and were accepted by the Ethics Committee of the University of Milan-Bicocca. All children's parents gave informed written consent and children gave their verbal assent before starting the study. Adults signed an informed consent.

Stimuli

Stimuli represented a continuum of trustworthiness ranging from 1 (very untrustworthy) to 7 (very trustworthy), interleaved by a neutral face, and characterized as variations of one female facial identity (see Figure 2.1). 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. The averaged neutral face was morphed by 3 steps (30%, 60% and 100%) towards the very untrustworthy reference and by 3 steps (30%, 60% and 100%) towards the very trustworthy reference, thus obtaining a 7-step trustworthiness continuum which included the neutral face. Figure 2.1 shows the output of an image difference analysis describing the physical variations among the seven faces included in the continuum. MATLAB (Mathworks Inc.) function *imshowpair* and the *diff* method were used to create a difference image between the most untrustworthy face (namely, face 1) and all other faces from the continuum. 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 corners of the mouth (i.e., downturned in untrustworthy faces and upturned in trustworthy faces), and the nostrils (i.e, nose wrinkling).

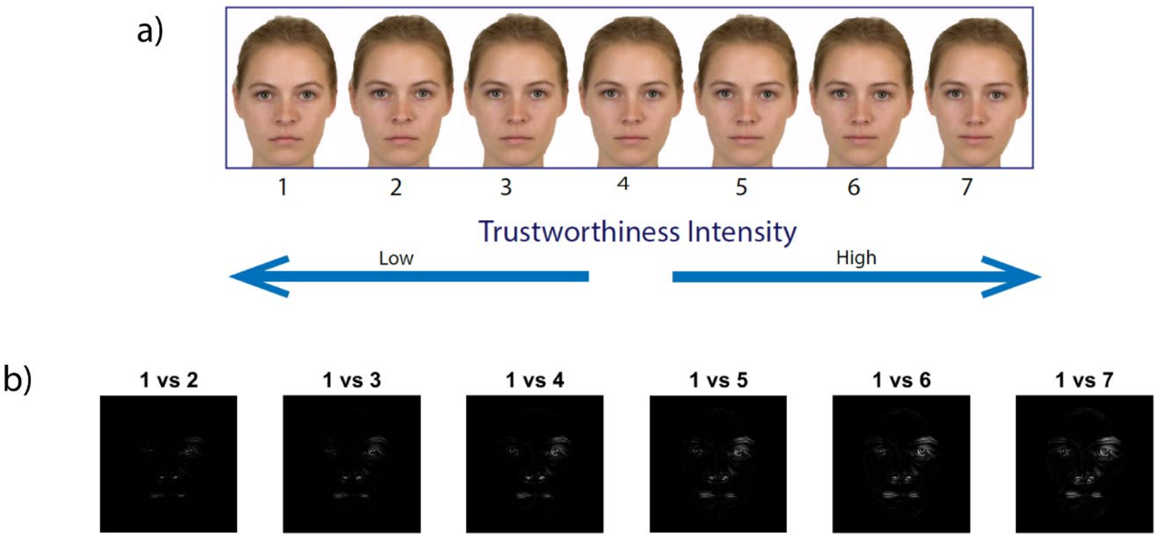


*Figure 2.1. The seven female faces composing the trustworthiness continuum, each composed of seven variations of one single female face identity, where Face 4 is the average of three neutral faces, Face 3, 2 and 1 are the 30%, 60% and 100% morphs of the same face towards the averaged untrustworthy extreme, and Face 5, 6 and 7 are the 30% and 60% and 100% morphs of the same face towards the averaged trustworthy extreme (a). Results of the difference analyses between the most untrustworthy face (i.e., face 1) and all other faces from the continuum, performed to describe the aspects of the face that change the most along the continuum (b).*

*Stimuli validation.* In order to make sure that the faces actually expressed a continuum of trustworthiness from low to high, an independent sample of 42 adults (34 females; mean age = 23.36 years; range = 18 - 35) was asked to evaluate 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"). A repeated-measures ANOVA with trustworthiness intensity as the within-subject factor was performed on the mean scores resulting from adults' ratings, and attained statistical significance, $F(6,246) = 20.295$, $p < .001$, $p\eta^2 = 0.331$. This was followed up with a test of within-subjects contrasts that revealed a significant linear trend, $F(1,41) = 58.760$, $p < .001$, $p\eta^2 = 0.589$. This indicates that participants' trustworthiness judgements varied monotonically as a function of the faces' position along the trustworthiness continuum. Moreover, we intended to test whether the experimental stimuli could elicit explicit judgements on other face dimensions. Indeed, according to the emotion overgeneralisation hypothesis (Said et al., 2009; Zebrowitz et al., 2003), trustworthiness judgements arise from an overgeneralisation of spontaneous responses to emotional expressions. Likewise, trustworthiness judgements might be influenced by the evaluation of other 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, a second group of adults (N = 46, 25 females; mean age = 23.98 years; range = 19 - 35) was asked to evaluate each step of the trustworthiness continuum on perceived emotion, typicality and attractiveness. For each step of the continuum, subjects were asked if they considered it to be emotional, typical and attractive. These three questions were asked in a random order. If subjects indicated that a face was expressing an emotion, they had to specify which kind of emotion (happy, angry, sad, scared or other) and give a score related to the intensity of the selected emotion on a scale that ranged 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 considered the face to be typical or attractive, they had to specify 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”). In case of negative answers (e.g., if participants evaluated the face to be not emotional, not typical or not attractive), scores were treated as zero values. With the aim to explore differences between the average intensity trustworthiness judgements elicited by each step of the continuum and the average intensity of emotional, typicality and attractiveness judgements, six independent *t*-tests (i.e., one for each emotion - happiness, angriness, sadness and fear - plus typicality and attractiveness judgements) were performed. The results showed that faces were generally judged to be more trustworthy than emotional (all *ps* $< .001$). On the other hand, intensity of judgements 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 results suggest that all steps of the continuum are perceived as poorly emotional. On the other hand, overall intensity did not differ between trustworthiness, typicality and attractiveness judgements. Moreover, an increase in intensity judgements of trustworthiness elicited by each of the seven steps of the continuum was correlated to an increase in perceived attractiveness, $r = .965$, $p < .001$, 95% confidence interval (CI) [0.775, 0.995], and a decrease in perceived angriness, $r = -.875$, $p = .01$, 95% confidence interval (CI) [-0.981, -0.356]. This means that, even though faces were perceived as poorly emotional, the less trustworthy faces of the continuum were scarcely perceived as expressing negatively valenced emotions, congruently to what was previously reported in the literature (Said, Sebe, et al., 2009; Zebrowitz et al., 2003). Furthermore, and once again congruently with the literature identifying attractiveness as an heuristic feature for trustworthiness judgements (Ma et al., 2016), the highly trustworthy faces were judged to be more attractive.

Apparatus and Procedure

Subjects performed all tasks individually and in a quiet, designated room (in the case of 5-year-old and the 7-year-old children the room was one of the school's classrooms). The *Oddmanout task* was always performed before the *Pairwise Preference task*. While performing both tasks, subjects seated 60 cm from a 17.3-inch touch-screen monitor with a resolution of 1080p. ASF (Schwarzbach, 2011) and MATLAB Psychtoolbox for Windows (Brainard, 1997) were used to control stimulus presentation and response collection. For all children, trial presentation was controlled manually by the experimenter, who started the trial as soon as the child was looking at the monitor. Furthermore, for all children the *Test of Emotion Comprehension* (TEC) (Pons & Harris, 2000) was administered between the two tasks.

*Oddmanout task*

The *Oddmanout task* was designed to measure the perceived dissimilarity between faces of the continuum, which varied in the level of the trustworthiness they expressed. 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. The instructions for adults were to select the face they judged to be more different from the others making use of the touch-screen interface of the computer. Also, children's response was collected though the touch-screen interface, after they had been told a story of a young princess who was locked in a castle tower by a witch. The experimenter should set free the princess with the help of the child, as he was able to steal the tower keys and get to the princess' room, but the witch made two avatars of the real princess in order to mislead him. Based on this story, the experimenter asked the child to help him find which one was the real princess, by looking attentively at the three faces and choose the one who appeared as more different from the others. In order to make sure that children understood all instructions, they were asked to repeat to the experimenter what their task was. Moreover, before starting with the task, children performed 5 practice trials. Since the aim of the task was to acquire pairwise dissimilarity scores for all possible triplet combinations of the seven face stimuli, participants viewed 35 trials, which corresponds to the binomial coefficient that is obtained when selecting 3 faces out of a total of 7 faces, without considering triplets repetition or order. Therefore, each trustworthiness step was shown 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 trustworthiness steps on the screen were randomized across trials. A central fixation cross determined the beginning of each trial (Figure 2.2). This stayed on the screen for 1000 ms in case of adults, or until the experimenter determined the beginning of the trial for children. The stimuli stayed on the screen until participants' response.

*Pairwise Preference task*

The aim of the *Pairwise Preference task* was to achieve explicit trustworthiness judgements for each of the seven faces of the continuum. At each trial, participants were asked to identify the face they trusted more between two simultaneously presented faces that were randomly selected from the continuum, by using the touch-screen interface of the computer. To facilitate the understanding of the task, children were told a second story: after succeeding in saving the princess, the child and the experimenter got lost in a supermarket. Here they encountered two identical girls: one being the real princess, one being the princess' avatar created by the witch, that claimed to know where the exit door was. The child was asked to help the experimenter to identify the real princess, i.e. the "good one", as she wanted to help them find the exit, while the other, the "mean one", wanted to disguise them. Before the beginning of the task, children were asked to explain what they were meant to do, to make sure the task was fully understood.

Since the aim of the task was to acquire participants' preference for all possible pairwise combinations of the seven steps of the continuum, participants viewed 21 trials, which corresponds to the binomial coefficient that is obtained when selecting 2 faces out of a total of 7 faces, without considering pairs repetition or order. Each step of the continuum was compared to all other steps for a total of 6 times; the position of the faces on the screen was randomized across trials. A central fixation cross determined the beginning of each trial (Figure 2.2). This stayed on the screen for 1000 ms in case of adults, or until the experimenter determined the beginning of the trial for children.

*Test of Emotion Comprehension*

Children were also administered with a *Test of Emotion Comprehension* (TEC) (Albanese & Molina, 2008) in between the *Oddmanout task* and the *Pairwise Preference task*. This test evaluates nine dimensions of emotion understanding, namely the recognition of emotions based on facial expressions, the understanding of external emotional 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 relation between morality and emotional experiences (Pons & Harris, 2000). The TEC is a reliable instrument from a psychometric point of view. It is characterized by good test–retest reliability ($r = .84$) after 3-months (Pons, Harris, & Doudin, 2002), and has good internal consistency, as reported Cronbach's alpha values range between .61 and .97 (Albanese & Molina, 2008; Farina & Belacchi, 2014; Kårstad, Kvello, Wichstrøm, & Berg-Nielsen, 2014). The test can be delivered to children aged 3 to 11 years, and comes in the form of a booklet of illustrations that tells a series of stories. Each story represents a different circumstance with four possible outcomes that are represented by different face expressions (happy, sad, angry, afraid). Children need to associate an emotion to the circumstance by selecting the matching facial expression, either by pointing to the corresponding drawing or by using words. The test duration ranges between 20 to 30 minutes. For each child, the final score is obtained by summing the partial scores obtained in the nine dimensions of emotion understanding and can range between 0 and 9; these scores were then transformed into *z scores*.

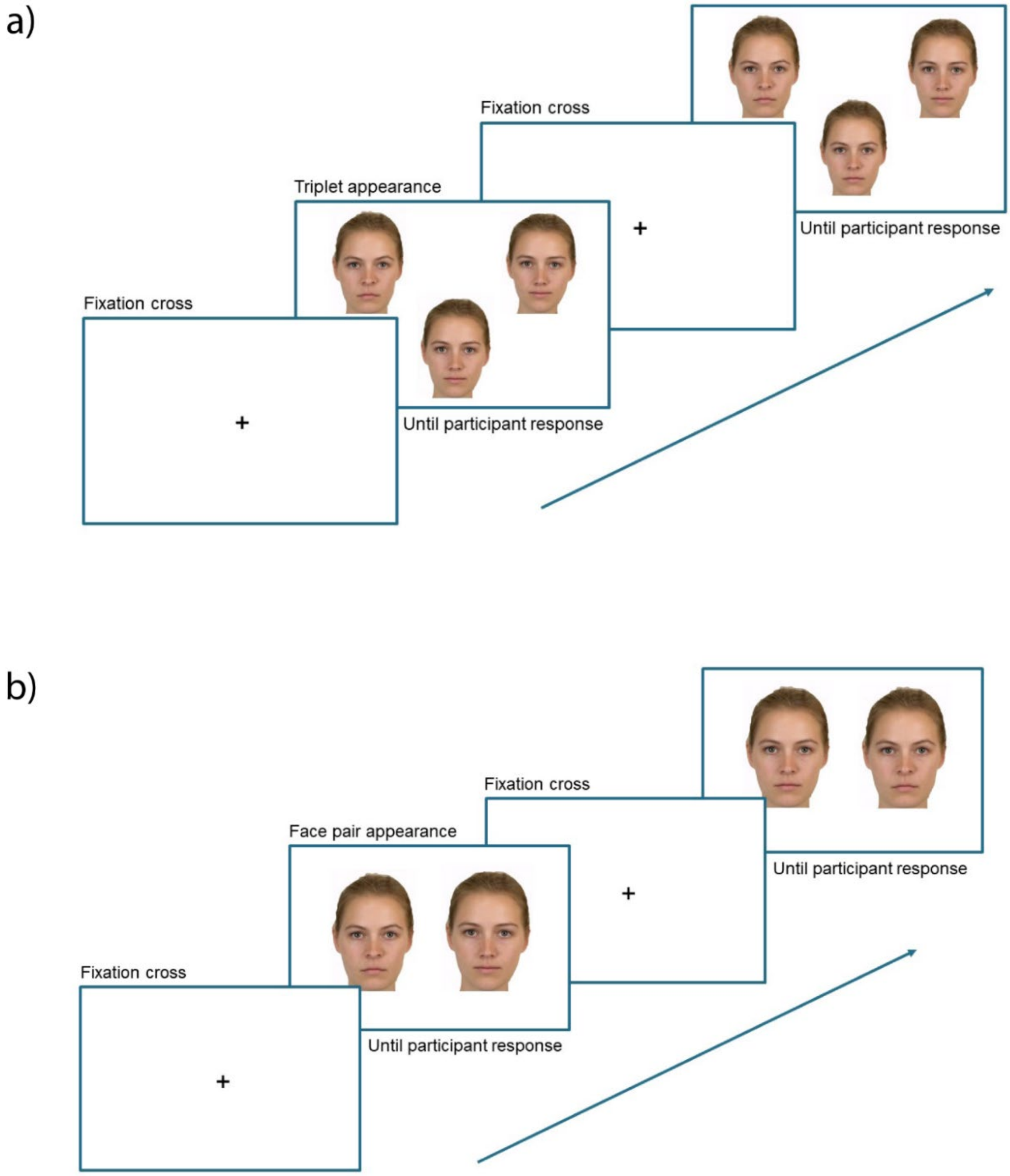


*Figure 2.2. Example trials from the Oddmanout task (a). As soon as the fixation cross disappeared, three faces were randomly selected from the trustworthiness continuum and presented simultaneously. Participants selected the one that looked more different. Example of trials from the*

*Pairwise Preference task (b). As soon as the fixation cross disappeared, two faces were randomly selected from the trustworthiness continuum and participants selected the face they trusted more.*

**Results**

Oddmanout task: Perceived dissimilarity between face pairs

In order to confirm that both the younger and the older children correctly understood the task, one-sample *t*-tests for each age group (5-year-olds, 7-year-olds, adults) was carried out on percent accuracy of those trials (N = 2) where one of the three presented faces was best discernible from the other two (like trials showing faces 7-6-1 or 1-2-7). The analysis revealed that performance in these trials was higher than chance (50%) for all age groups: 5-year-olds (M = 76%), $t(28) = 5.477$, $p < .001$; 7-year-olds (92%), $t(30) = 12.490$, $p < .001$; 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 dissimilarity scores, one for each face pair within the triplet. As the participant indicated that one face was the most different within the triad, we attributed a distance score of 0 (i.e., minimum level of dissimilarity) to the face pair composed of the non-selected faces, and a score of 1 (i.e., maximum dissimilarity) to the face pairs composed of the selected face and the non-selected ones. For each participant, dissimilarity scores for each face pair were summed and then divided by the number of trials where that face pair was showed (i.e., N = 5 for all face pairs) to obtain a scale ranging from 0 to 1. These final values were used to obtain a 7x7 Representational Dissimilarity Matrix (RDM), which 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 symmetrical along the diagonal.

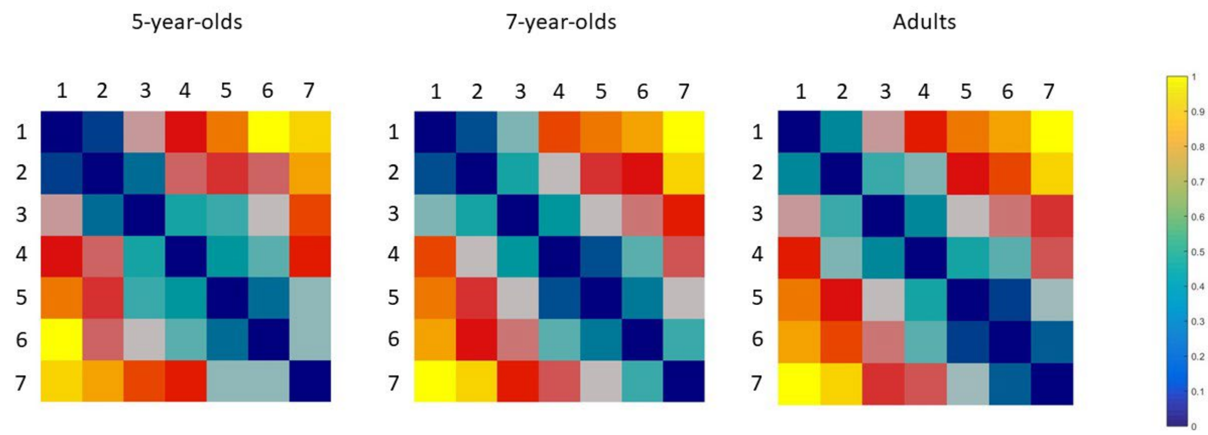


*Figure 2.3. Representational dissimilarity matrices of the averaged pairwise dissimilarity scores acquired in the Oddmanout task, one for each age group. Yellow denotes maximum dissimilarity, while blue denotes minimum dissimilarity. The matrices are symmetrical across the diagonal.*

Within each age group, individual RDMs were averaged across subjects to obtain an RDM for each age group (see Figure 2.3). 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. The three RDMs showed to strongly 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.* With the aim to further test age-related differences in the way participants represented in memory perceptual difference in facial cues to trustworthiness, agglomerative hierarchical cluster analyses were carried out by taking as input pairwise dissimilarity scores separately for each age group (see Everitt, 2011). Even though correlational analyses showed that general configuration of RDMs was similar across age groups, the cluster analysis can be used to better describe the way subjects aggregated the experimental stimuli on the basis of their perceptual similarities. The *average linkage* method in MATLAB (Mathworks Inc.) groups items by creating a multilevel hierarchy to form a hierarchical tree based on average distance between items. Figure 2.4 depicts the three dendrograms (one for the 5-year-olds, one for the 7-year-olds, and one for the adults, respectively) derived from the cluster analyses carried out on the pairwise dissimilarity scores. The X-axis it shows the logical order of the dissimilarity judgements; the Y-axis it shows the degree of perceived difference between trustworthiness intensities, which is the distance that the function *linkage* computes between couple of items. If clusters linkage is lower than 70% of the maximum linkage, clusters are shown in scales of grey.

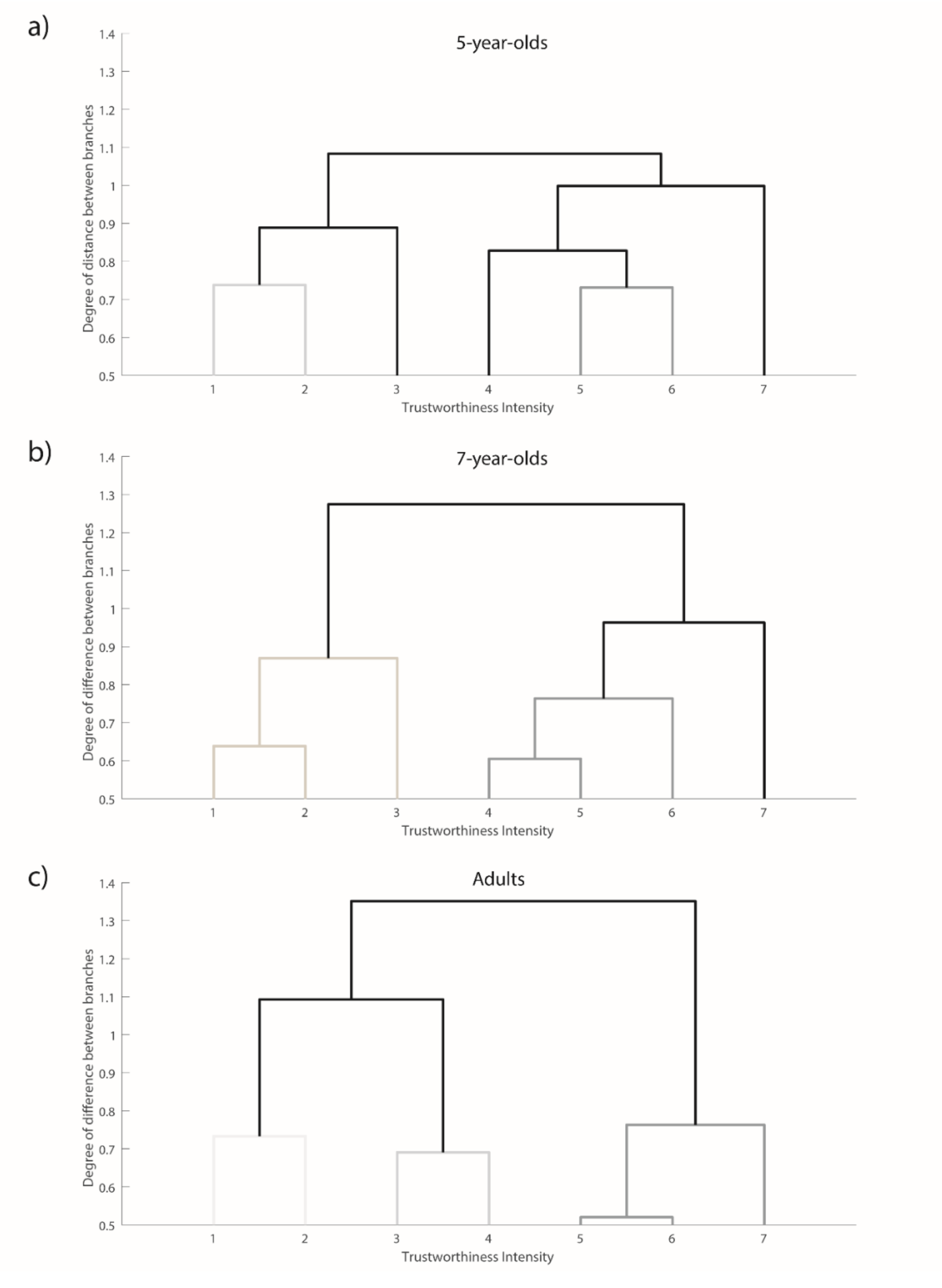


*Figure 2.4. Results of the cluster analyses on the pairwise dissimilarity scores obtained from the Oddmanout task represented in the form of hierarchical plots, separately for the 5-year-olds (a), the 7-year-olds (b), and the adults (c). Different clusters are shown in different grayscales.*

*Intra-group consistency of Pairwise Dissimilarity Scores*. With the aim to test age-related differences in the way participants represented perceptual differences in facial cues to trustworthiness, intra-group consistency in the way subjects' attributed perceived dissimilarity scores was computed within each of the three age groups. Specifically, cosine distances were computed on all possible pairwise combinations of vectorized RDMs (which is the upper part of the matrices, along the diagonal) of single participants and separately for 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. To test whether consistency in attributing dissimilarity scores differed according to age, we performed a univariate ANOVA on cosine distances with age as between-subjects factor, which revealed a significant effect, $F(2,1429) = 136.006$, $p < 0.001$, $p\eta^2 = 0.160$. Indeed, pairwise dissimilarity scores proved to be more consistent in adults ($M_{\text{cosine distance}} = 0.065$, $SD = 0.020$) than in 5-year-olds ($M_{\text{cosine distance}} = 0.093$, $SD = 0.027$), $p < .001$, and 7-year-olds ($M_{\text{cosine distance}} = 0.081$, $SD = 0.032$), $p < .001$. On the other hand, the 7-year-olds proved to be more consistent than the 5-year-olds, $p < .001$. The empirical cumulative distribution of cosine distances separately for the three age groups is shown in Figure 2.5.

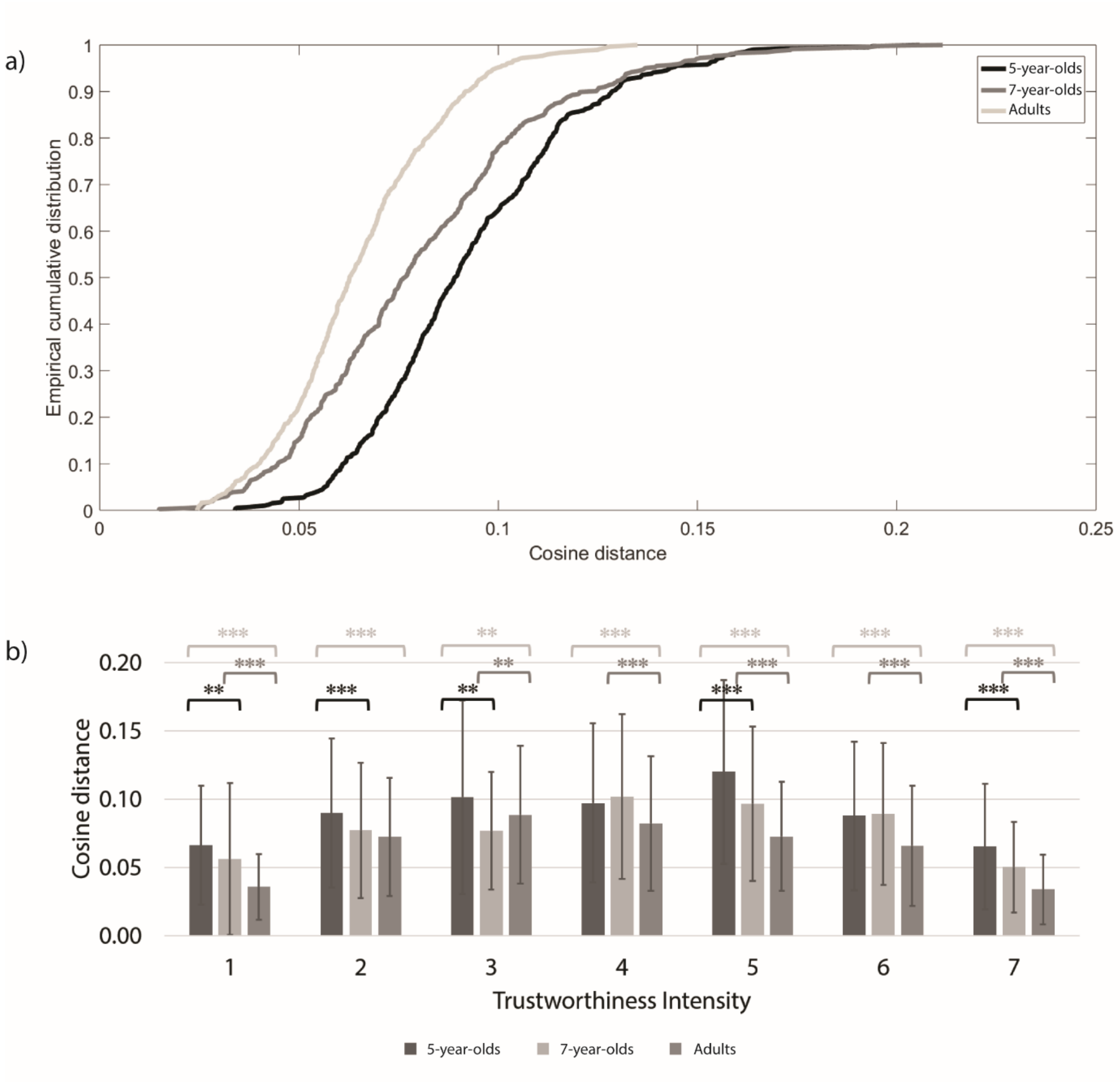


*Figure 2.5. Empirical cumulative distributions for the three ages' intra-group cosine distances of dissimilarity scores (a). Intra-group cosine distances for the three ages' dissimilarity scores and for each of the seven trustworthiness intensities (b). **p < .01 ***p < .001*

In order to investigate whether dissimilarity scores within each age group were equally consistent across participants for all trustworthiness intensitities, cosine distances on dissimilarity scores were computed separately for each trustworthiness intensity. In fact, each row (or column) of the RDM contains the pairwise dissimilarity scores between one trustworthiness step and each of the other trustworthiness steps. Separately for each age group, cosine distances were computed between each RDM row of each single subject and the corresponding RDM rows of of all other subjects. A reapeated-measures ANOVA with trustworthiness intensity as the whithin-subjects factor and age group as the between-subjects factor was computed on the resulting values. Both main effects proved to be significant (trustworthiness intensity: $F(6,8574) = 251.291$, $p < .001$, $p\eta^2 = 0.150$, power = 1; age group: $F(2,1429) = 110.788$, $p < .001$, $p\eta^2 = 0.134$, power = 1), and also the interaction between the two factors, $F(12,2850)= 17.596$, $p < .001$, $p\eta^2 = 0.024$, power = 1. A test of within-subjects contrasts showed the presence of a significant quadratic trend for trustworthiness intensity for all age groups, *ps* < .001 (see Figure 2.5).

<u>Pairwise Preference task</u>: explicit trustworthiness judgements

*Response accuracy.* Percent accuracy was computed from single subjects's response on each trial of the *Pairwise Preference task* as the percentage of trials in which participants selected the face with higher levels of trustworthiness in the continuum. One-sample *t*-tests, one for each age group, were computed to test whether subjects regularly selected the face with higher levels of trustworthiness. All tests revealed to be significant, thus showing that, for all age groups, accuracy levels were significantly above chance (5-year-olds: $t(28) = 7.177$, $p < .001$; 7-year-olds: $t(30) = 11.049$, $p < .001$; adults: $t(33) = 12.575$, $p < .001$). On the other hand, a significant effect was found also in a univariate ANOVA with age group as the between-subjects factor, $F(2,91) = 4.704$, $p = .011$, $p\eta^2 = 0.094$, power = 0.776. This would suggest that age groups differed in the accuracy whit which they choose the more trustworthy face within each face pair. Post-hoc comparisons (Bonferroni corrected) showed that the 5-year-olds (*M* = 72.25%, *SD* = 16.69) were less accurate than adults (*M* = 84.17%, *SD* = 15.85), *p* = .014, and revealed they also were marginally less accurate than the 7-year-olds (*M* = 82.18%, *SD* = 16.22), *p*= .06 (see Figure 2.6).

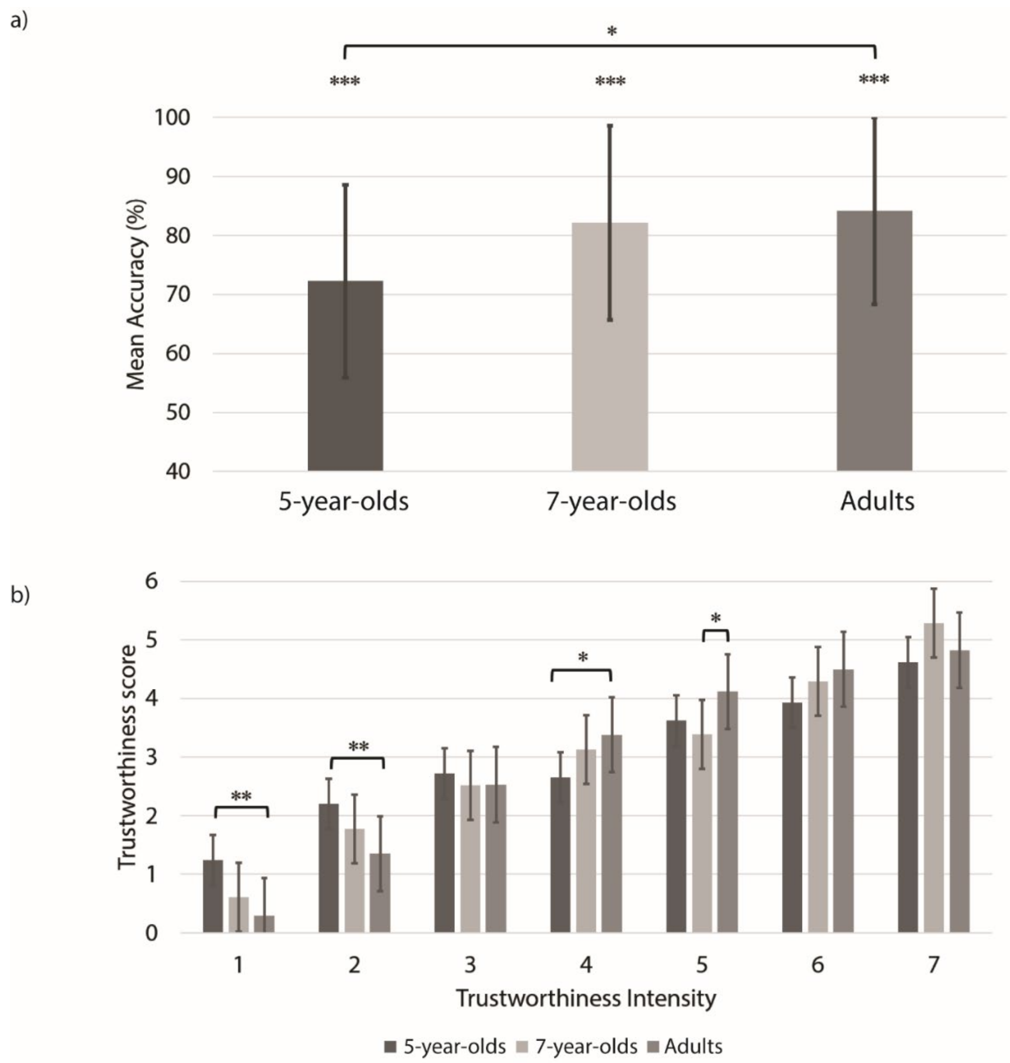


*Figure 2.6. Average percent accuracy for each of the three age groups derived from the Pairwise Preference task (a). Average trustworthiness scores derived from the Pairwise Preference task for each of the seven trustworthiness steps (b). *p < .05 ***p < .001*

*Trustworthiness Scores.* With the aim to measure how consistently each step of the trustworthiness continuum was chosen as the more trustworthy in a face pair, for each trial of the *Pairwise Preference task* we computed a trustworthiness score for every trustworthiness step. As soon as the participant indicated a face to be the more trustworthy than the other in the pair, we attributed a score of 1 to the chosen face, and a score of 0 to the non-chosen face. Trustworthiness scores of each single step of the continuum were summed across trials and separately for each participant. As each step of the continuum was shown for a total of 6 times, the trustworthiness score of each face ranged from 0 (in case it was never chosen to be the most trustworthy) to 6 (in case it was always chosen to be the most trustworthy). With the aim to test whether 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 (namely whether the higher the displayed trustworthiness cues, the higher the score), a repeated-measures ANOVA with trustworthiness intensity as the within-subjects factor and age group as the between-subjects factor was performed. A main effect of trustworthiness intensity, $F(6,546) = 119.327$, $p < .001$, $p\eta^2 = 0.576$, power = 1, and an interaction between trustworthiness intensity and age, $F(12,174) = 2.821$, $p < 0.01$, $p\eta^2 = 0.06$, power = 0.98 (see Figure 2.6) were found. Post-hoc analysis (Bonferroni corrected) showed that 5-year-olds judged face 1 and 2 to be more trustworthy with respect to adults' judgements on the same items ($p = .004$ and $p = .006$), while adults judged face 5 to be more trustworthy with respect to what specified by 7-year-olds ($p = .017$). Nevertheless, all age groups showed a significant linear trend in the attribution of trustworthiness scores (all *p*s < .001) as revealed by a within-subjects contrast analysis.

Test of Emotion Comprehension (TEC)

In order to check for the existence of the expected developmental differences in children's performance at TEC (as in Pons et al., 2004), an independent-samples *t*-test was performed on the *z*-transformed TEC scores of 5-year-olds and 7-year-olds. This proved to be significant, $t(58)= 5.45$, p < 0.001, suggesting that 7-year-olds ($M = 8.03$, $SD = 0.91$) performed better at the test than 5-year-olds ($M = 6.17$, $SD = 1.65$). In light of this age-related overall difference in children's emotion comprehension abilities, correlational analyses between TEC scores and *Oddmanout task* as well as *Pairwise Preference task* performance scores were used to investigate possible individual differences in the way perceptual cues to trustworthiness are represented in memory and are used to attribute explicit trustworthiness judgements 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, intragroup cosine distance between the vectorized RDMs (namely dissimilarity scores between all face pairs) and the euclidean distance between TEC scores was performed separately for the two age groups, 5-year-olds and 7-year-olds. The euclidean distance between two points is the length of the linear segment that connects them,, and is employed as an index of similarity between couple of elements. With the aim to obtain a measure of similarity between all TEC scores separately for each age group, euclidean distance was computed between all possible pairwise combinations of TEC scores. 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). In the case of the 5-year-olds group, the analysis revealed a significant, but very weak, positive correlation, $r = 0.053$, $p < .001$, 95% confidence interval (CI) [-0.045, 0.149]. In the case of the 7-year-olds group, the analysis revealed no significant correlation, $r = - 0.018$, $p = .685$, 95% confidence interval (CI) [-0.106, 0.0698].

*TEC and Pairwise Preference accuracy.* With the aim to test if individual differences in emotion recognition abilities influenced the capability to generate explicit trustworthiness judgements separately for the 5-year-olds and the 7-year-olds, subjects' accuracy in the *Pairwise Preference task* was correlated to z-transformed scores at TEC. The analysis revealed a positive correlation in the case of 5-year-olds, $r = 0.44$, $p = .017$, 95% confidence interval

(CI) [0.087, 0.694], but not for the 7-year-olds, $r = 0.02$, $p = .921$, 95% confidence interval (CI) [-0.338, 0.371].

**Discussion**

Study 2 aimed at exploring the development of perceptual sensitivity to those facial cues that drive trustworthiness judgements in adults, and the way these are represented in long term memory, taking into account the role of individual differences in emotion comprehension. With the aim to measure age-related differences in perceptual sensitivity to facial cues to trustworthiness, 5-year-olds, 7-year-olds and adults were tested in an *Oddmanout task*, which was conceived to provide measures of perceived similarity/dissimilarity among faces that only slightly varied in the level of expressed trustworthiness. Analysis of participants' performance showed that, at the age of 5 years, children already represent faces as a function of the intensity of the trustworthiness they express. Indeed, mean RDMs representing perceived dissimilarity between face pairs proved to highly correlate across age groups, suggesting that there were no qualitative differences in participants' sensitivity to variations in facial cues to trustworthiness. Nevertheless, our findings also showed that this sensitivity became more fine-grained with development, as intra-group consistency in the attribution of dissimilarity judgements increased with age. More specifically, intra-group consistency was higher for adults' judgements than for 5-year-olds' and 7-year-olds' judgements, and intra-group consistency was higher for the judgments made by the 7-year-olds' than for those made by the 5-year-olds. The *Oddmanout task* was also allowed us to explore whether there were age-related differences in the consistency among participants' dissimilarity scores for each trustworthiness intensity. Adults' scores were more consistent than those of the 5-year-olds for all trustworthiness intensities, and for almost all trustworthiness levels adults were also more consistent than the

7-year-olds. These results are congruent with earlier evidence that within-age consistency in the trustworthiness judgements increases with age (Ma et al., 2016).

Cluster analysis performed on dissimilarity scores acquired from the *Oddmanout task* confirmed the finding of age-related improvement in participants' proficiency at attributing dissimilarity judgements for faces that only slightly vary in the level of expressed trustworthiness. Even though both children and adults represented faces as a function of the level of trustworthiness they expressed, the hierarchical structure of the representation of cues to trustworthiness becomes more fine-grained with increasing age. Five-year-old children's dissimilarity judgments formed one cluster composed of the two most untrustworthy faces (faces 1 and 2) and another cluster consisting of two trustworthy faces (faces 5 and 6). Going up in the hierarchy, these two clusters expanded to comprise the three untrustworthy faces (faces 1, 2, 3) on one side, and the three trustworthy faces (faces 5, 6, 7) on the other, with the neutral face included in the latter cluster. The structure of children's representation becomes more diversified for the 7-year-olds, who, already at the bottom of the hierarchy, show the three untrustworthy faces (faces 1, 2, 3) forming one cluster, and the two moderately trustworthy faces (faces 5 and 6) form another cluster to which, once again, the neutral face (face 4) is attached, and to which the very trustworthy face (face 7) is hierarchically related. The hierarchical organization of the clusters becomes maximally differentiated in adulthood, when three different clusters are visible, which comprise 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). Going up along the hierarchy, a sovra-cluster appears, which includes both the more untrustworthy faces (faces 1 and 2) and the more neutral faces (faces 3 and 4). These results are in line with Valentine's (Valentine et al., 2016) theory, as they confirm the notion that the representation of facial characteristics that are relevant for distinguishing among individual faces and/or face types becomes more fine-grained and differentiated across development.

In addition to age-related differences in the structure of the mental representation of facial cues to trustworthiness, results of Study 2 also showed consistency in performance across age groups. Indeed, the consistency with which participants attributed dissimilarity judgements showed similar variations across trustworthiness levels for all age groups. Analyses on intra-group cosine distances of dissimilarity scores showed a significant quadratic trend for both 5-year-olds, 7-year-olds and adults, 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 result is similar to that obtained in Study 1, where adults discriminated more easily the faces located at the opposite extremes of the trustworthiness continuum, irrespective of their valence, compared to those located in the central area of the continuum. In line with these findings, 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 (Sarah Jessen & Grossmann, 2016).

As already discussed in relation to the results of Study 1, the finding that faces displaying more intense physical cues to trustworthiness enjoy a processing advantage over those displaying less intense cues, irrespective of the valence of such cues, might be related to the adaptive advantage of being tuned to socially connoted faces than to socially neutral ones, just like processing of emotional faces, which convey adaptively relevant information, is usually faster than processing neutral faces, even when only little face information conveying emotional intensity is available (Roesch et al., 2010). Another possible explanation for the stronger agreement participants showed in attributing dissimilarity judgments to faces at the extreme opposites of the trustworthiness continuum than to faces at the centre of the continuum might be that the former were perceived as less prototypical than the latter. Indeed, previous studies showed that amygdala response in adult participants is better predicted by deviations in typicality than by deviation in valence (Said et al., 2010). Nevertheless, this explanation does not apply to our data, as our stimuli validation procedure showed 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.

Explicit trustworthiness judgements derived from the *Pairwise Preference task* showed an age-related pattern similar to the one observed for perceptual sensitivity to physical cues to trustworthiness recorded during the *Oddmanout task.* Both the 5-year-olds and the 7-year-olds performed above chance when selecting the more trustworthy face in a pair, and that for all age groups the pairwise preference scores varied linearly as a function of the position of the face along the trustworthiness continuum. At the same time, performance accuracy was higher for the adults compared to the 5-year-old children, but not the 7-year olds, suggesting that, by the age of 7 years, children reach adult-like level of performance. This result is in line with previous evidence that children aged 7 years are as sensitive as adults in attributing explicit trustworthiness judgements (Cogsdill et al., 2014; Cogsdill & Banaji, 2015). Interestingly, the younger children were not only less accurate and consistent in their attribution of trustworthiness judgements, but they also tended to overestimate the trustworthiness intensity of the more untrustworthy faces (1 and 2), by attributing to these faces higher trustworthiness scores compared to 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 confirm that sensitivity to facial cues to trustworthiness is not yet fully developed at this young age.

In fact, we found that, at 5 years but not at 7 years, children differed in their ability to generate trustworthiness judgements from faces as a function of their emotion comprehension skills, as measured through the *Test of Emotion Comprehension* (TEC). Five-year-olds who scored higher at TEC performed better when asked to detect the more trustworthy face in a pair in the *Pairwise Preference task*. In contrast, we found no evidence that children's emotion understanding skills modulated how they represented facial cues to trustworthiness, given the null correlation between TEC scores and children's intragroup cosine distances between the vectorized RDMs built on pairwise dissimilarity scores derived from the *Oddmanout task*.

The fact that performance in the *Pairwise Preference task* was modulated by emotion understanding skills selectively for the 5-year-olds seems to imply that development of emotional intelligence is intertwined with children's perception of social traits from faces. Indeed, 5-year-olds' performance in the *Pairwise Preference task* was only slightly worse than that of the 7-year-olds (5-year-olds: $SD$ = 16.69; 7-year-olds: $SD$ = 16.22), and similarly variable (5-year-olds: $SD$ = 16.69; 7-year-olds: $SD$ = 16.22), but, in line with Pons et al. (2004) they scored significantly lower than the older children at TEC, and showed larger variance in their response to the questionnaire (5-year-olds: $SD$ = 1.65; 7-year-olds: $SD$ = 0.91). This shows that emotion comprehension skills are unevenly distributed in younger children, as the age of 5 is when critical components of emotion comprehension (e.g., understanding of the outward expression of emotion and their causes) starts to emerge, while other fundamental components (e.g., the ability to understand the mentalistic origins of emotions), appear only later in time (Pons et al., 2004). The positive correlation between performance at TEC and the *Pairwise Preference task* observed at this young age therefore suggests that, in the earlier critical stages of the development of emotion understanding, 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.

It is important to note that Study 2 has a number of limitations that could be addressed by future studies. First of all, future studies shall test the generalizability of our results by using a new set of averaged stimuli created starting from a different pool of face identities. Indeed, even if the use of averaged identities reduced the effect of idiosyncratic facial characteristics on perceptual sensitivity to physical cues of trustworthiness and explicit judgements of trustworthiness, the use of one single average face identity might have hindered the generalizability of our results.

Second, as individual differences in emotional development affected trust perception, future studies may explore whether variability in temperamental traits and self-regulation abilities impact sensitivity to facial cues to trustworthiness in preverbal infants, who have been shown to discriminate faces based on those cues (Jessen & Grossmann, 2016; Jessen & Grossmann, 2017; Sakuta et al., 2018). This question is the focus of Study 3.

Third, future studies could test how the findings from the present study generalize across cultures and face ethnicities. Previous literature showed that perception of trustworthiness generalizes across face ethnicities (Birkás et al., 2014), and that Caucasian and Asian adult participants use similar facial cues to generate trustworthiness judgements (Xu et al., 2012). It would be interesting to explore whether the same applies to children, in light of the fact that, starting from preschool years, children rely on racial information when making social judgements (e.g., Bigler & Liben, 2007; Killen & Stangor, 2001). Furthermore, it should be taken into account how cultural differences affect perception of facial trustworthiness, as cultural practices might influence display rules. This question is the focus of Study 5 (Chapter 3).

In summary, Study 2 shows that, already at 5 years of age, children represent perceived differences between faces that slightly vary in the level of expressed trustworthiness along a continuum of intensity. However, this representation refines with age, becoming more fine-grained. Similarly, even though the 5-year-old children were overall less accurate than the adults in making explicit trustworthiness judgements and overestimated the trustworthiness intensity of the more untrustworthy faces, the trustworthiness judgements of both the 5-year-ols and the adults were linearly distributed along the trustworthiness continuum in the same way. Finally, at the age of 5 years 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 indicating that the ability to discriminate facial cues associated to trustworthiness and to use these cues to generate trustworthiness judgements is present 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 understanding abilities.

## Study 3

**Neural and behavioural responses to variations in facial cues to trustworthiness in infants.**

Study 3 extended the investigation of the developmental origins of perceptual sensitivity to facial cues to trustworthiness to the first year of life by measuring 7-month-old infants' neural and behavioral responses to faces displaying different levels of trustworthiness intensity. The study replicated and extended the evidence obtained by Jessen and Grossman (2016) by using more ecological face stimuli (images of real faces as opposed of computer-generated faces), and by investigating variations in infants' neural sensitivity to facial cues to trustworthiness associated to individual differences in temperamental traits.

As already discussed, all previous studies employed as stimuli artificial computer-generated faces that, though allowing for a rigorous manipulation of the features of interest, do not reflect infants' daily life expertise (see Crookes et al., 2015). For this reason, in Study 3 we used as experimental stimuli real female faces that were created with the same data-driven morphing procedure described in the method section of Study 2. More specifically, we selected three trustworthiness steps from the continuum: the most untrustworthy (Face 1), the neutral (Face 4) and the most trustworthy (Face 7).

In fact, the first aim of Study 3 was to explore whether, by using more ecological stimuli, infants' electrophysiological responses would differentiate not only between the two faces at the extreme opposites of the trustworthiness continuum and the neutral one (as in Jessen & Grossmann, 2016, 2017), but also between the very trustworthy and the very untrustworthy face. Accordingly, in Study 3a young adults and 6-months-old infants were presented with Low Trustworthy, Neutral and High Trustworthy faces while their stimulus-locked Event-Related Potential (ERPs) activity was recorded through a 128-channel high-density EEG system. We included a sample of adult participants with the aim to further validate our stimuli by replicating previous evidence of a larger electrophysiological response to low trustworthy faces at the level of the N170 and LPP (Dzhelyova, Perrett, & Jentzsch, 2012; Lischke, Junge, Hamm, & Weymar, 2018; Marzi, Righi, Ottonello, Cincotta, & Viggiano, 2014). This granted us with a reference for infants' electrophysiological responses to the new stimuli, and a ground to discuss possible age-related qualitative differences in such responses.

A second aim of Study 3 was to verify whether neural sensitivity to facial cues to trustworthiness was reflected in infant's behaviour by investigating their attentional preference. More specifically, we aimed at testing whether we would have replicated the same pattern of preference observed by Jessen and Grossmann (Jessen & Grossmann, 2016).

Therefore, Study 2b tested 6-month-old infants in a standard preferential looking task in which different pairwise combinations of the three trustworthiness steps used in study 2a (Low Trustworthy, Neutral and High Trustworthy faces) were presented side by side on the computer screen.

The third aim of Study 3 was to explore for the first time whether variability in infants' temperamental traits modulates sensitivity to facial cues to trustworthiness.

In Study 1 and 2 we obtained evidence that, in both adults and children, aspects of the beholder's personality associated to social motivation and social cognition affect trustworthiness perception. In infancy, there is mounting evidence that individual differences in temperamental traits and self-regulation abilities impact sensitivity to environmental stimuli. For example, the likelihood to experience negative feelings, as measured by the Negative Affect scale of the Infant Behavioral Questionnaire scale (IBQ-R, Gartstein & Rothbart, 2003), shows positive correlation with the amplitude of the face-specific P400 component in response to angry expressions (Quadrelli et al., 2019). Similarly, higher sensitivity to signals to threat as an effect of fearful temperament was associated to larger Nc to fearful with respect to happy faces at 7 months (de Haan et al., 2004). Finally, higher negative emotionality scores were associated to larger Nc responses to happy faces in 3 to 13-month-old infants (Martinos, Matheson, & de Haan, 2012).

In light of this evidence, in Study 3a we expected to observe a relation between sensitivity to facial cues to trustworthiness and variability in IBQ-R's Negative Affect. Moreover, the subscale of Surgency (the likelihood to experience and display high levels of activity) was taken into account as it is considered to be a precursor to adults' Extraversion (BFQ, Caprara, Barbaranelli, Borgogni, & Perugini, 1993), a character trait that was shown to modulate adults' sensitivity to facial cues to trustworthiness in Study 1. For a matter of comparability to infants' data, adults were asked to fill in the Big Five Questionnaire to test whether variability in Neuroticism (corresponding to infants' Negative Affect) and Extraversion scores have an impact on adults' neural sensitivity to facial cues to trustworthiness.

## Study 3a

## Materials and methods

Participants

The adult sample included 20 undergraduate or graduate university students (7 males, $M_{age}$ = 25.25 years; SD = 3.06; range: 20 – 32 years), who were recruited from the community by word of mouth on a voluntary basis or received course credits for their participation. We tested eight additional adults, who were excluded from the final sample due to EEG artifacts. All participants did not report any history of neurological or psychological disorders and had normal or corrected- to normal vision. Before the testing session, they signed an informed consent and completed the Italian version of the Big Five Questionnaire (BFQ; Caprara, Barbaranelli, Borgogni, & Perugini, 1993), a self-report questionnaire designed to measure the Big Five dimensions of personality: Extraversion, Agreeableness, Conscientiousness, Neuroticism, and Openness to Experience.

The infant sample was composed of 21 6-month-old healthy infants (10 males, $M_{age}$ = 200 days; *SD* = 7.43; range: 184 – 211 days). All of them were born full-term with normal birth weight (> 2, 400 g). Eighteen additional infants were tested and excluded from the final sample due to fussiness (n = 1) and excessive EEG artifacts resulting in less than 10 artifact-free trials per condition (n = 17). Infants were recruited through written invitations that were sent to parents based on birth records of neighbouring cities. Parents gave their written informed consent, and the mother or the primary caregiver completed the Infant Behavior Questionnaire-Revised in its very short form (IBQ-R VSF; Putnam, Helbig, Gartstein, Rothbart, & Leerkes, 2014). The questionnaire included queries aimed to assess the frequency of specific temperament-related behaviors observed within the last week; we focused on two subscales: Negative Affect (NA, tendency to experience negative feelings and difficulty being soothed) and Surgency (SU, tendency to show high levels of activity and positive emotions and to act impulsively), which are thought to be the equivalent of the two personality dimensions of Neuroticism and Extraversion.

The protocol of the study was carried out in accordance to the ethical standards of the Declaration of Helsinki (BMJ 1991; 302:1194), and was approved by the Ethics Committee of the University of Milano-Bicocca (Protocol number: 236).

Stimuli

Stimuli consisted of the cropped versions of 3 of the 7 variations of the female face identity composing the trustworthiness continuum used in Study 2. Specifically, for the sake of comparability to Jessen and Grossmann's study (2016), we selected the most untrustworthy face (Face 1), the neutral face (Face 4) and the most trustworthy face (Face 7). These faces were cropped into an oval shape (see Figure 2.7).

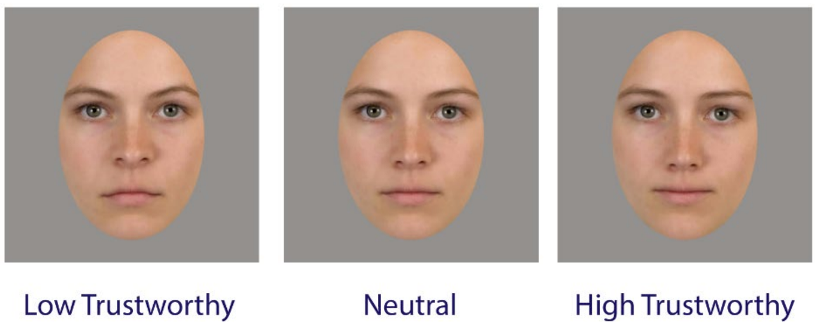


*Figure 2.7. The three trustworthiness variations of the female face identity used as stimuli in Study 3a.*

*Stimuli validation.* As one of the goals of the study was to test whether the results obtained by Jessen and Grossmann (2016) would generalize to more ecological, realistic face images, we measured the extent to which our face stimuli differed from the artificial computer-generated stimuli created by Oosterhof and Todorov (Oosterhof & Todorov, 2008) used in this earlier study. To this end, a group of young adults ($N$ = 37, 4 males, $M_{age}$ = 26.84 years; $SD$ = 6.01;

range: 21 – 46 years) was asked to rate the perceived trustworthiness intensity of our three face stimuli and the three face stimuli used in the Jessen and Grossmann's (2016) study using a 9-point scale where 1 indicated that they would not at all approach the depicted person, and 9 indicated that they would very much approach the depicted person. The three faces from each stimulus sets were presented in a random order in separate blocks, with the order of the blocks counterbalanced across participants. The same adult participants were also asked to rate the perceived similarity of all possible pairwise combinations between the three faces taken from the two different stimulus sets on a 9-point scale, where 1 indicated the two faces look identical and 9 indicated that the two faces look completely different. In this case also, the three faces from each stimulus sets were presented in a random order in separate blocks, with the order of the blocks counterbalanced across participants.

We compared participants' ratings of perceived trustworthiness intensity for the two stimulus sets by performing a repeated-measures ANOVA with stimulus set (current set versus Todorov set) and trustworthiness level (low, neutral, high) as within-subjects factors. The ANOVA revealed a significant Stimulus Set x Trustworthiness level interaction, $F(2,72) = 8.275$, $p = .001$, $p\eta^2 = .187$, power = .934: although participants' ratings differed significantly across the three trustworthiness intensities for both stimulus sets (all *ps* < .002) according to the position of the faces along the continuum, the Low trustworthy face of the current set ($M = 4.270$, $SD = .253$) was perceived as significantly more trustworthy than the Low trustworthy face from the Todorov set ($M = 2.865$, $SD = .213$, $p < .001$).

We compared participants' ratings of perceived similarity for the two stimulus sets by performing a repeated-measures ANOVA with stimulus set (current set versus Todorov set) and pairwise comparison (Low vs High, Neutral vs High, Neutral vs Low) as within-subjects factors. The ANOVA revealed a significant effect of stimulus set, $F(1,36) = 20.427$, $p < .001$, $p\eta^2 = .450$, power = 1.00, as participants rated the face pairs from the current stimulus set ($M =$

4.117, $SD = .288$) as more similar than those from the Todorov set ($M = 5.252$, $SD = .301$), and a main effect of pairwise comparison, $F(2,72) = 28.596$, $p < .001$, $p\eta^2 = .443$, power = 1.00, as, irrespective of the stimulus set, a test of within-subjects contrasts revealed a significant linear trend, $F(1,36) = 50.598$, $p < .001$, $p\eta^2 = 0.584$, with the Low vs High trustworthy face pair rated as more discernible ($M = 5.608$, $SD = .341$) than the Neutral vs Low trustworthy face pair ($M = 4.811$, $SD = .354$), which, in turn, was rated as more discernible than the Neutral vs High trustworthy face pair ($M = 3.635$, $SD = .236$).

These data confirm that, for both stimulus sets, the level of perceived trustworthiness increased linearly from the Low trustworthy to the High trustworthy face. Nevertheless, the Low trustworthy face of the current set was perceived to be more trustworthy with respect to the Low trustworthy face from the Todorov set. Even though the number of studies that investigate how gender affect perception of trustworthiness are very limited, this effect could be due to the fact that female faces are evaluated more positively than male faces (see for example Sutherland, Young, Mootz, & Oldmeadow, 2015), even though the reason why this effect was limited to the Low trustworthy face should be further investigated in future studies.

A second aspect revealed by the stimulus validation is that faces of the current set were perceived as more similar to one another than the faces of Todorov set. This means that our stimuli, even though perceived to express different levels of trustworthiness, are characterized by more fine-grained differences than the faces of Todorov set. Fort this reason, we might be able to measure a finer sensitivity to variations on this face dimension with respect to what was measured by Jessen and Grossmann (2016).

A final aspect emerging from the validation is that the Neutral face was perceived as more similar to the High rather than the Low trustworthy face, irrespective of stimulus set. This result is in line with what emerged from the cluster analysis of Study 1 (which used the Todorov set), where the Neutral face was perceived to be more similar to the trustworthy cluster.

Nevertheless, in Study 2 (which used the current set) the Neutral face was part of the untrustworthy cluster. These data confirm the hypothesis discussed in Study 1, according to which absolute neutrality is not a feature of face perception (e.g., Adams, Nelson, Soto, Hess, & Kleck, 2012; Lee, Kang, Park, Kim, & An, 2008). Rather, neutral faces might be perceived as ambiguous, thus driving both slightly positive or negative judgements.

Procedure

Data acquisition took place in an audiometric and electrically shielded cabin. Stimuli were presented on a 24-inch monitor, while participants seated at approximately 60 cm away from the screen. Infants seated on their parent's legs. Stimulus presentation was controlled through E-Prime software v2.0 (Psychology Software Tools Inc., Pittsburgh, PA). Adults were asked to sit as still as possible during the entire procedure. The same instruction was given to parents, who were also asked to avoid interfering with the experimental session by communicating with the child through verbal or non-verbal means. For infants, the whole experimental session was recorded through an infrared video camera connected to the data acquisition computer and hidden over the stimulus presentation monitor. The live image of the infant's face and body allowed the experimenter to pause or terminate the session as soon as the infant became fussy. Each infant was presented with all three trustworthiness intensities, which were presented one at a time in a pseudo-randomized . The experimental session terminated when infants attended to a maximum of 270 trials, or got tired. Within each trial the stimulus was presented for 1000 ms, and the following interstimulus interval varied randomly between 900 and 1100 ms. The experimenter presented a looming fixation point between trials to reorient infants' attention to the monitor as soon as they got distracted.

<u>EEG recording and processing</u>

Electroencephalographic (EEG) data were recorded through a 128-electrode HydroGel Geodesic Sensor Net (Electrical Geodesic Inc., Eugene, OR), with the vertex electrode (Cz) used as reference. Signals were amplified using an EGI NetAmps 300 amplifier with a sampling rate of 500 Hz and an online band-pass filter of 0.1–100 Hz. Impedances were checked online before the session started and they were considered as adequate if lower than 50 KΩ. EEG data were processed online through NetStation v4.6.4 (Eugene, OR). Continuous signals were band-pass filtered at 0.30 Hz, and segmentation into epochs was centered on stimulus onset with a 100 ms baseline and 1,000 ms of stimulus presentation. Re-reference was conducted to the algebraic mean of all the channels. As soon as signal surpassed ± 200 μV, channels were automatically dismissed in a window of 80 ms on the segmented data. If any artifacts survived this procedure, they were hand-edited. Further trials were excluded if more than 15% of the channels (N ≥ 18) turned out to be bad (as in Halit, de Haan, & Johnson, 2003). For the remaining trials, individually reject channels were replaced using spherical spline interpolation. A repeated-measures ANOVA with trustworthiness level (Neutral, Low, High) as the between-subjects factor revealed that, not for the adults, nor for the infants, the mean number of trials contributing to the average ERP differed between the High Trustworthy face condition (adults: 44.4, *SD* = 11.87; infants: 13.90, *SD* = 4.17) the Low Trustworthy face condition (adults: 43.45, *SD* = 12.49; infants: 13.62, *SD* = 3.47) and the Neutral face condition (adults: 42.75, *SD* = 13.73; infants: 12.86, *SD* = 3.41) (*ps* > .39).

*Adults.* For the adults, we analysed the face sensitive N170 (120–180 ms) and the attentional LPP component (520–720 ms), which were both reported to be modulated by changes in face trustworthiness (Lischke, Junge, Hamm, & Weymar, 2018; Marzi et al., 2014). In accordance with a study by Yang and colleagues (Yang, Qi, Ding, & Song, 2011), we also analysed the C1

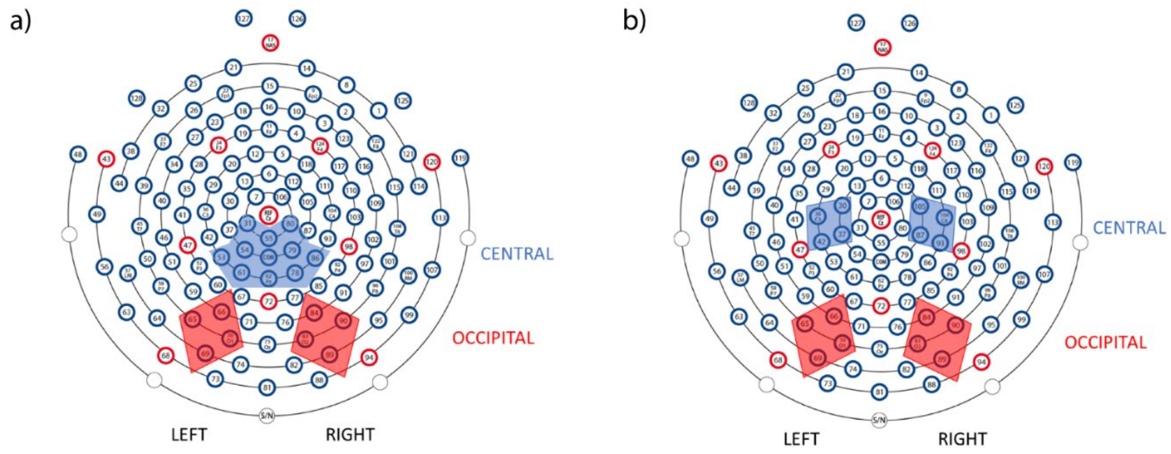


*Figure 2.8. The selected electrode locations used with adults (a) and infants (b)*

(70–100 ms) component. One cluster of electrodes was selected for each hemisphere where the N170 was more clearly visible (left: 65, 66, 69, 70; right: 83, 84, 89, 90). C1 and LPP were observed over fronto-central electrode sites (31, 53, 54, 55, 61, 62, 78, 79, 80, 86) (see Figure 2.8). The time windows for the four components were chosen based on previous ERP reports (e.g., Lischke et al., 2018; Yang et al., 2011) and visual examination of the components' peak for each participant.

*Infants.* Like in Jessen and Grossmann (2016), we analysed two face-sensitive components, the N290 (200-260 ms), and the P400 (330-430 ms), and the attentive Nc (300-600 ms). One cluster of electrodes was selected for each hemisphere where the N290 and the P400 were more clearly visible (left: 65, 66, 69, 70; right: 83, 84, 89, 90) (see Figure 2.8). A prominent P1 (120-170 ms) was also visible in the occipital regions, and was therefore included in the analyses.. A well-defined Nc component was observed over fronto-central electrode sites, and two clusters of electrodes, one for each hemisphere, were selected for this component (left: 36, 30, 37, 42; right: 87, 93, 104, 105). The time windows for the four components were chosen based on previous infant ERP reports (e.g., Leppänen, Moulson, Vogel-Farley, & Nelson, 2007) and visual examination of the components' peak for each participant.

Following Jessen and Grossmann (2016), for each of the considered components, peak latency (ms) and mean amplitude (µV) values were extracted and entered in the statistical analyses.

**Data analysis**

IBM 25.0 (IBM Corporation, Armonk, NY, USA) and RStudio.1.0.136 (RStudio Team (2015). RStudio: Integrated Development for R. RStudio, Inc., Boston, MA URL http://www.rstudio.com/) were used for data analyses.

Adults

Peak latency and mean amplitude of the N170 were analysed through a 3 x 2 repeated measures Analysis of covariance (ANCOVA) with trustworthiness level (Neutral, Low, High) and Hemisphere (left, right) as within-subjects factors and Neuroticism and Extraversion scores derived form the BFQ entered as covariates, whereas for the C1 and the LPP the same repeated-measures ANCOVA was conducted without including hemisphere as a factor.

Infants

Peak latency and mean amplitude of the P1, N290, P400 and Nc were analysed through a 3 x 2 repeated measures Analysis of covariance (ANCOVA) with trustworthiness level (Neutral, Low, High) and Hemisphere (left, right) as within-subjects factors Negative Affect and Surgency scores derived from the IBQ-R entered as covariates.

Following previous studies investigating the role of individual differences in infants temperament in modulating facial emotion processing (Quadrelli, Conte, Macchi Cassia, & Turati, 2019; Taylor-Colls & Pasco Fearon, 2015), each temperament effect was followed up by correlational analyses between the scores of each temperamental trait and ERP difference scores computed by subtracting peak latencies values or mean amplitudes values for the Neutral face from those recorded for High Trustworthy (i.e., High Trustworthy-Neutral), the Neutral from those recorded for the Low Trustworthy face (i.e., Low Trustworthy-Neutral) and the

Low Trustworthy from those recorded for the High Trustworthy (i.e., High-Low Trustworthy). Only significant correlations were reported.

**Results**

Adults

*N170 Amplitude*. The analysis revealed a main effect of trustworthiness level, $F(2,34) = 8.052$, $p = .001$, $p\eta^2 = .321$, power = .940. A test of within-subjects contrasts showed a significant linear trend, $F(1, 17) = 26.987$, $p < .001$, with the Low Trustworthy face evoking larger response ($M = -.304$, $SD = .499$) compared to the High Trustworthy ($M = -.055$, $SD = .401$) and the Neutral ($M = .369$, $SD = .463$) faces (see Figure 2.9).

*N170 Latency*. The analysis showed no significant effects.

*C1 Amplitude.* The analysis showed a main effect of trustworthiness level, $F(2, 34) = 4.820$, $p = .019$, $p\eta^2 = .221$, power = .717, and a test of within-subjects contrasts revealed a significant linear trend, $F(1, 17) = 5.506$, $p = .031$, with the Neutral face ($M = -.277$, $SD = .201$) evoking larger response compared to the High Trustworthy ($M = -.216$, $SD = .176$) and the Low Trustworthy ($M = .186$, $SD = .249$) faces (see Figure 2.9).

*C1 Latency.* The analysis showed no significant effects.

*LPP Amplitude.* The analysis showed a main effect of trustworthiness level, $F(2, 34) = 3.472$, $p = .042$, $p\eta^2 = .170$, power = .610. A test of within-subjects contrasts revealed a significant linear trend, $F(1,17) = 8.018$, $p = .012$, due to the Low Trustworthy face eliciting larger amplitudes ($M = 1.973$, $SD = .258$) compared to the High Trustworthy ($M = 1.602$, $SD = .247$) and Neutral ($M = 1.458$, $SD = .201$) faces (see Figure 2.9).

*LPP Latency.* The analysis showed no significant effects.

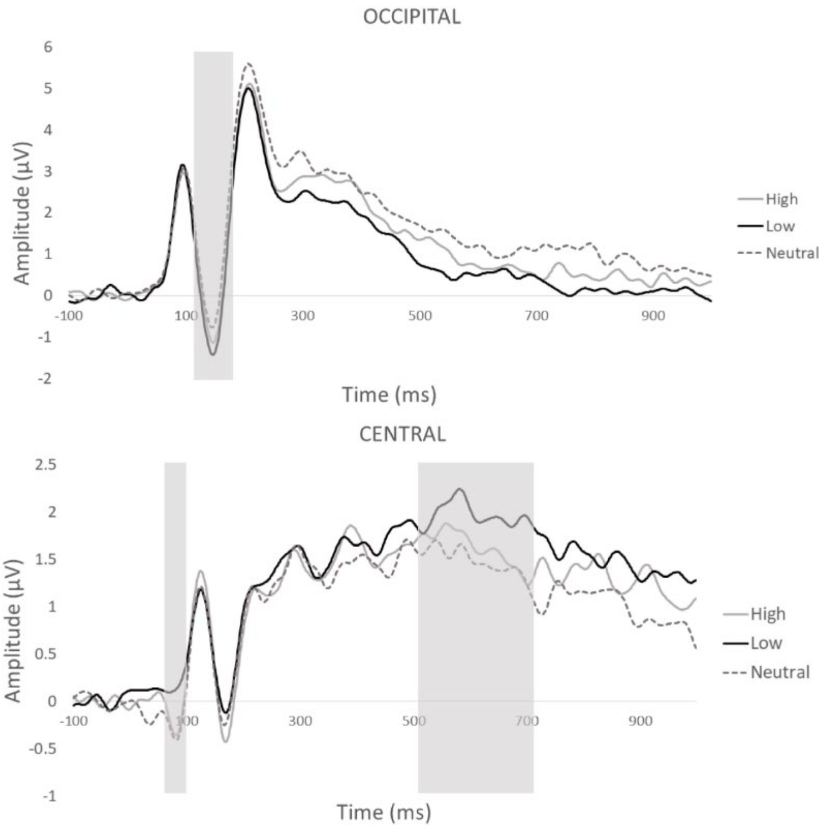


*Figure 2.9. Waveform plots depicting grand-average ERPs for the N170, LPP and C1 components in response to the Low Trustworthy (solid black line), Neutral (dashed grey line), and High Trustworthy (solid grey line).*

<u>Infants</u>

*P1 Amplitude.* The analysis showed no significant effect.

*P1 Latency.* The analysis showed a significant interaction between trustworthiness level and hemisphere, $F(2,36) = 4.021$, $p = .027$, $p\eta^2 = .183$, power = .681. However, this result proved spurious, as no significant comparisons were observed between levels of trustworthiness at the electrodes of interest.

*N290 Amplitude.* The analysis showed an interaction between hemisphere and Negative Affect, $F(1,18) = 20.304$, $p < .001$, $p\eta^2 = .530$, power = .989, which was investigated through correlational analyses performed separately for each hemisphere and Negative Affect scores. Correlational analysis showed a significant negative association between N290 amplitude at the left hemisphere and Negative Affect scores, $r$ = -0.33, $p$ = .031, 95% confidence interval (CI) [-0.835, -0.274] (Figure 2.10).

*N290 Latency.* The analysis showed no significant effect.

*P400 Amplitude.* The analysis showed a significant Trustworthiness level x Hemisphere interaction, $F(2,36) = 3.360$, $p = .046$, $p\eta^2 = .157$, power = .598. Post-Hoc tests revealed that the Low Trustworthy face was characterized by higher amplitudes in the left, $M$ = 21.449, $SD$ = 2.355 with respect to the right, $M$ = 16.550, $SD$ = 2.309, hemisphere, $p$ = .011 (LSD) (Figure 2.10).

*P400 Latency.* The analysis showed a significant Trustworthiness Level x Surgency interaction, F(2,36) = 3.962, $p$ = .028, $p\eta^2 = .180$, power = .674., which was investigated through correlations of differential scores between trustworthiness levels and Surgency scores. Surgency scores positively correlated with differential scores obtained from subtracting the Low Trustworthy to High Trustworthy (i.e., High-Low) faces latency values, $r$ = 0.49, $p$ = .024, 95% confidence interval (CI) [0.073, 0.760]. The ANCOVA also revealed a significant interaction between trustworthiness level and hemisphere, $F(2,36) = 3.386$, $p = .045$, $p\eta^2 = .158$, power = .601. Only in the left hemisphere, High Trustworthy face latencies were faster ($M$ = 367.167, $SD$ = 7.148) than Neutral face latency values ($M$ = 378.762, $SD$ = 6.322), $p$ = .038 (LSD) (Figure 2.10).

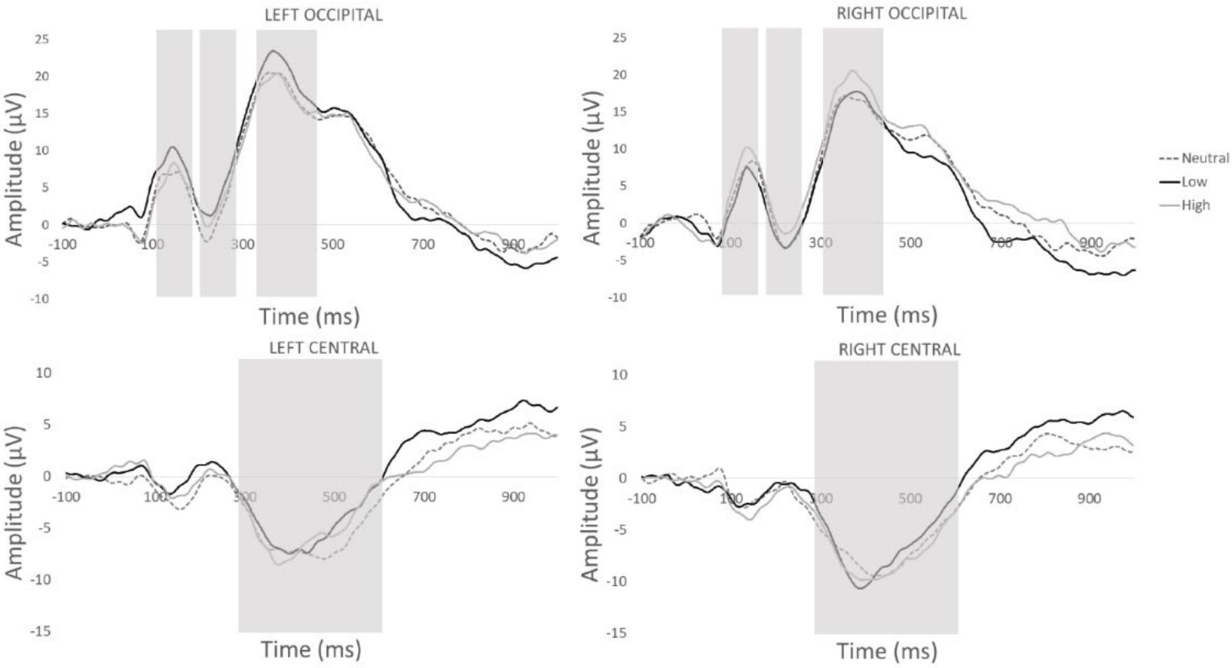


*Figure 2.10. Waveform plots depicting grand-average ERPs for the P1, the N290, the P400 and the Nc components in response to the Low Trustworthy (solid black line), Neutral (dashed gray line), and High Trustworthy (solid gray line).*

Nc Amplitude. The analysis showed no significant effects.

Nc Latency. The analysis showed a significant interaction between trustworthiness level and Negative Affect, $F(2, 36) = 4.369$, p = .024, $p\eta^2 = .195$, power = .684, which was investigated through correlations of differential scores between trustworthiness levels and Negative affect scores (see Figure 2.11). Negative Affect scores negatively correlated with differential scores obtained from subtracting Neutral face to the High Trustworthy face (i.e., High-Neutral) latency values, $r$ = - 0.54, $p$ = .0117, 95% confidence interval (CI) [-0.787, -0.139]. Furthermore, we observed a significant interaction between trustworthiness level and hemisphere, $F(2,36) = 4.245$, $p = .022$, $p\eta^2 = .191$, power = .706. Only in left hemisphere, latencies related to the High Trustworthy face proved to be faster ($M$ = 404.929, $SD$ = 9.730) than latencies related to the Neutral face ($M$ = 445.119, $SD$ = 11.518), $p$ = .004 (LSD), and the Low Trustworthy face ($M$ = 439.286, $SD$ = 12.063), $p$ = .027. Furthermore, the High

Trustworthy face showed faster latencies in the left hemisphere, $M = 404.929$, $SD = 9.730$, with respect to the right hemisphere, $M = 425.738$, $SD = 9.420$, $p = 0.23$ (LSD) (Figure 2.10).

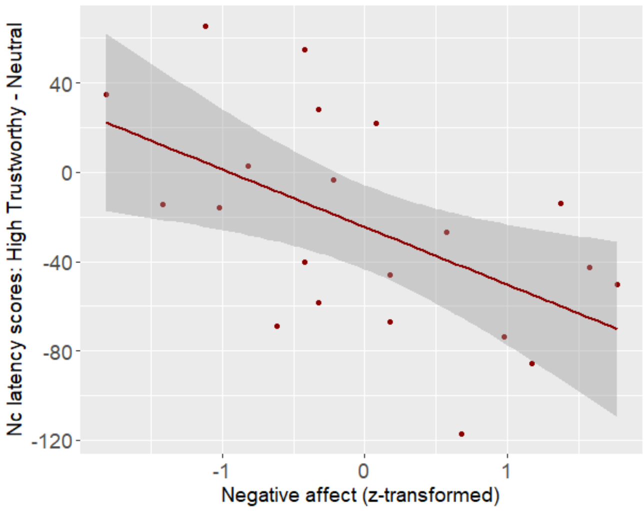


*Figure 2.11. Negative correlation between Negative Affect scores and differential scores of the High Trustworthy minus Neutral face Nc latency values.*

**Study 3b**

**Materials and methods**

<u>Participants</u>

Twenty-seven 6-month-olds composed the final sample (11 males, $M_{age} = 217.89$ days; $SD = 9.45$; range: 200 – 231 days). All infants were all born full-term and had a normal weight at birth (> 2, 250 g). We tested one additional infant but discarded the data due to fussiness.

Participants' recruitment followed the same method used in Study 3a; parents gave their written informed consent before the testing session started.

The protocol of the study was carried out in accordance to the ethical standards of the Declaration of Helsinki (BMJ 1991; 302:1194) and was approved by the Ethics Committee of the University of Milano-Bicocca (Protocol number: 236).

Stimuli

Stimuli consisted of the uncropped version of the three trustworthiness intensities used in Study 3a: the High Trustworthy face (HT), the Low Trustworthy face (LT), and the Neutral face (N) (see Figure 2.12). We presented the stimuli in their uncropped version as we wanted to maximize the ecological validity of the face images in order to facilitate the appearance of infants' visual preference response.

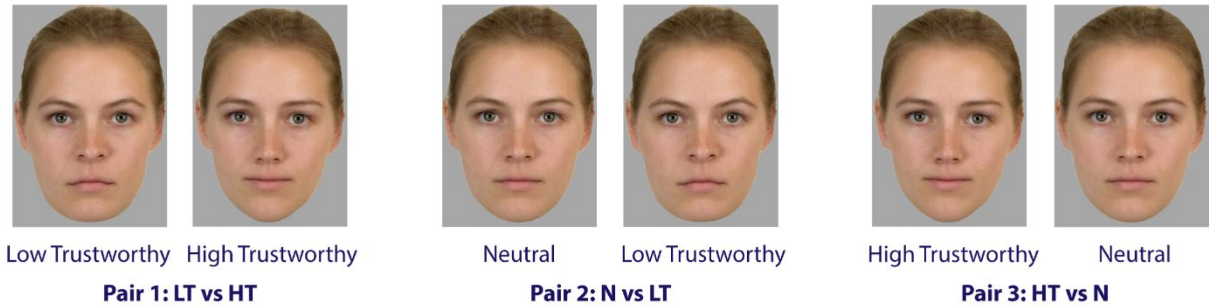


*Figure 2.12. The three pairs of stimuli presented to the infants in the Visual Preference task of Study 3b.*

Apparatus and Procedure

Infants were tested using a standard preferential-looking paradigm (Fantz, 1958). All infants viewed three 10-s bilateral presentations of three stimulus pairs, for a total of three trials. The three pairs were obtained by all the possible pairwise combinations of the three trustworthiness intensities, and consisted of the HT and LT stimuli (Pair 1), the LT and the N stimuli (Pair 2), and the HT and the N stimuli (Pair 3) (see Figure 2.12). The three pairwise combinations were shown in a random order, with the constrain that, because each face image appeared twice on the screen, its side (left, right) position on the screen was counterbalanced within and across participants. For example, if the HT stimulus was shown on the left side of the screen when paired with the N stimulus, it appeared on the right side when paired with the LT stimulus, or viceversa. Each trial started with an animation that attracted the infant's attention to the centre of the screen; as soon as the infant looked at it the experimenter started the trial. The order of pair presentation, as well as the initial left/right position of the stimuli, was counterbalanced among participants.

Stimuli were presented on a 24-inch monitor was used for stimulus presentation, while participants seated at approximately 60 cm away from the screen. Infants always seated on their parent’s laps. E-Prime software v2.0 (Psychology Software Tools Inc., Pittsburgh, PA) was employed to present experimental stimuli. Parents were asked to avoid interfering with the experimental session by communicating with the child with both verbal or non-verbal means. The whole experimental session was recorded by an infrared video camera connected to the data acquisition computer and hidden over the monitor, which allowed the experimenter to look at live images of the infant’s eyes and body, and to pause or terminate the session as soon as the infant became fussy. Videotapes of eye movements were recorded and subsequently analyzed frame by frame to the nearest 40 ms by a coder who was blind to the specific position of the stimuli on each trial. The coder recorded, separately for each stimulus and each trial, the total fixation time, i.e., the sum of all fixations. As a measure of inter-observer reliability, the duration of single fixations was recorded by a second coder for all of the infants in the sample. The level of agreement was high, $r = .988$, $p < .001$.

**Results**

In order to test whether infants showed a preference for either of the two stimuli within each pair, we computed three different scores using three different procedures reported in the literature: a pairwise preference score (after Macchi Cassia, Kuefner, Westerlund, & Nelson, 2006), a total preference score (after Jessen & Grossmann, 2016), and a delta score (after Montoya, Westerlund, Troller-Renfree, Righi, & Nelson, 2017).

<u>Pairwise preference score.</u> Following the procedure adopted in Macchi Cassia et al. (2006), each infant’s looking time at the HT stimulus for Pairs 1 and 3 and at the N stimulus for Pair 2 was divided by the total time spent looking at either stimuli within each pair and then converted

into a percentage score. Hence, only scores significantly above 50% indicated a preference for the considered stimuli. Three preliminary one-way analyses of variance (ANOVAs), one for each of the considered stimuli (HT in Pair 1, HT in Pair 3 and N in Pair 2), performed on the preference scores manifested by the three groups of infants who saw Pair 1, 2 or 3 as the first stimulus pair within the testing session, revealed that order of pair presentation did not affect infants' visual preferences, all *ps* $> .557$. To determine whether the obtained pairwise preference scores differed from chance (50%) for each of the three stimulus pairs, three separate one-sample *t*-tests were applied, one for each pair. All tests failed to reach significance (all *p*s $> .613$), suggesting that infants' looking times were equally distributed across the stimuli within each pair (see Figure 2.13).

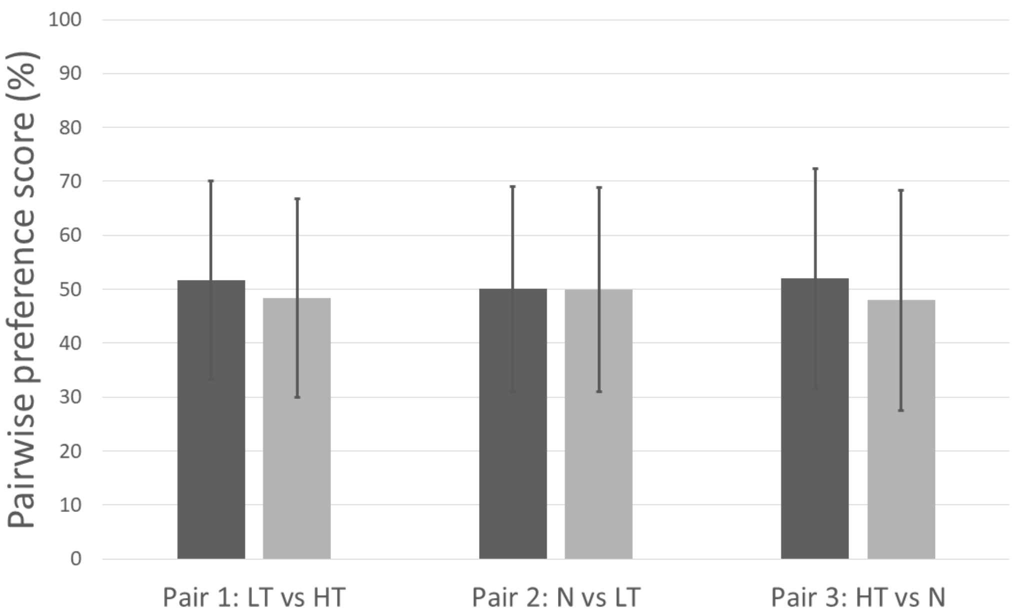


*Figure 2.13. Percent of total looking time spent fixating on each stimulus within each pair.*

Total preference score. Since each stimulus was shown twice across trials, a total looking time for each stimulus was obtained by summing the time spent looking at that stimulus in the two trials in which it was presented (Pairs 1 and 3 for the HT stimulus, Pairs 1 and 2 for the LT stimulus, and Pairs 2 and 3 for the N stimulus). Following Jessen & Grossmann (2016), the obtained total looking times on the HT stimulus, the LT stimulus and the N stimulus were divided by the overall total time spent looking at all the stimuli across the three trials and then converted into a percentage score. A repeated-measure ANOVA with trustworthiness level (HT, N, LT) as the between-subjects factor proved non-significant, $F(2,52) = .041$, $p = .960$, again suggesting that none of the three stimuli preferentially attracted infants' attention across trial.

Delta score. Following the procedure adopted in Montoya et al. (2017), for each infant and for each pair we computed the difference between the time spent looking at one stimulus and the time spent looking at the other, and divided the obtained delta by the total fixation time on both stimuli of the pair. Specifically, the delta was computed by subtracting looking time on the LT stimulus from the time spent looking at the HT stimulus for Pair 1, by subtracting looking time on the LT stimulus from looking time on the N stimulus for Pair 2, and by subtracting looking time on the N stimulus from looking time on the HT stimulus for Pair 3. A repeated-measure ANOVA with trustworthiness level (HT, N, LT) as the within-subjects factor proved non-significant, $F(2,52) = .072$, $p = .890$, indicating once again that none of the three stimuli elicited a preferential visual response, irrespective of the stimulus with which it was paired.

**Discussion**

Study 3 aimed to explore perceptual sensitivity to those facial cues that drive adults' trustworthiness judgements in infancy, at an age, i.e., 6 months, when infants can discriminate between other facial cues from faces, such as different kinds of facial expressions signalling emotional states, like happiness and fear (Grossmann, 2010; Nelson, 1987). Six-months-old infants' neural (Study 3a) and behavioural (Study 3b) responses were recorded while infants were presented with realistic images of female faces previously judged by expressing different levels of trustworthiness by adult observers.

We also included in the study a sample of young adults with the aim of checking that our stimuli did in fact evoke the same ERP modulations observed in previous studies on the face-specific N170 and the attentional LPP component. Indeed, in accord with previous evidence (Dzhelyova et al., 2012), we observed a linear increase in N170 activity, in the direction of higher response for the more untrustworthy face: N170 amplitude was lower for the Neutral face, larger for the High Trustworthy face and maximal for the Low Trustworthy face. This finding adds to earlier demonstrations that facial cues to trustworthiness are extracted already at the structural encoding stage of face processing, possibly as a result of the adaptive advantage that sensitivity to such cues provides to the individual. The same pattern of results was evident at the level of the attentional LPP, whose amplitude was also modulated by the intensity and the valence of the facial cues to trustworthiness in the current study, in a way which is not dissimilar to from has been previously found in the literature (Lischke, Junge, Hamm, & Weymar, 2018; Marzi et al., 2014). The amplitude of the fronto-central LPP increased linearly across the Neutral face, the High Trustworthy and the Low Trustworthy face, for which the amplitude was greatest. Other studies observed a similar intense allocation of attention for low trustworthy faces (Lischke et al., 2018; Marzi et al., 2014; Yang et al., 2011), and interpreted such response as an adaptive reaction to environmental threat. An opposite pattern of results was found for the early occurring C1, which was also reported to be sensitive to thread-related stimuli by previous studies (Pourtois, Grandjean, Sander, & Vuilleumier, 2004; Stolarova, Keil, & Moratti, 2006). In our study, the C1 was larger in response to the High Trustworthy and the Neutral faces in comparison to the Low Trustworthy face, a result that is similar to the one observed by Yang and colleagues (Yang et al., 2011), and that further proves that structural properties of facial cues to trustworthiness are processed very fast.

More crucially, we observed a modulation of neural activity in response to our stimuli also in 7-months-old infants. Indeed, we observed an earlier-peaking P400 in response of the High Trustworthy face with respect to the Neutral face in the left hemisphere. Shorter P400 latencies in response to face stimuli were previously reported to mirror early face processing (de Haan, Johnson, & Halit, 2003). Shorter latencies in favour of the High Trustworthy face therefore might indicate that high trustworthy faces enjoy a processing advantage over those faces that do not include cues related to trustworthiness perception. The finding that the effect was confined to the left hemisphere is partially congruent with the "Approach-Withdrawal hypothesis" proposed by Davidson (1983), which claims that environmental stimuli that favour approach behaviours are processed in the left hemisphere, and those that favour avoidance behaviours are processed in the right hemisphere. It is important to note that, in adults, similar lateralized responses typically emerge at the level of attentional components who are thought to originate from the prefrontal cortex (PFC), and thus should be reflected in infants in the Nc. Indeed, our results showed that the Nc peaked faster in response to the High Trustworthy face with respect to the Neutral face only in the left hemisphere, and that the High Trustworthy face evoked overall faster latencies in the left hemisphere compared to the right hemisphere. Therefore it is possible that in infants, like in adults, the High Trustworthy face is perceived as an appealing stimulus, which would recruit the left PFC that is reported to be part of a system that facilitates approach behaviour to engaging stimuli. According to this view, we should have also observed faster Nc latencies in favour of the Low Trustworthy face in the right hemisphere. The fact that this was not the case might be explained by the fact that, under typical rearing conditions, infants are more frequently exposed to positive rather than negative expressions. As previously argued by Farroni and colleagues (Farroni, Menon, Rigato, & Johnson, 2007), selective responses to positively valenced faces in infants might thus emerge as a result of infant's social experience. Although it does not explain the lateralization effect, another possible explanation for the response observed for the High Trustworthy face in our sample has to do with infants' early sensitivity to prosocial stimuli. There is evidence that five-month-old-infants discriminate between prosocial and non-prosocial others and prefer the former, avoiding the latter (Hamlin & Wynn, 2011). Moreover, it has been shown that 6-month-old infants are sensitive to events where agents are hindering or helping others (Gredebäck et al., 2015). In particular, the P400 component showed significantly higher amplitudes in favour of the helping agent with respect to the hindering agent. Following from this evidence, positively valenced faces like those associated to perception of High Trustworthiness may be perceived from the earlier stages of development as signalling a friendly approach, with sensitivity to negatively-valenced faces emerging later in development, as an effect of experience or development of cortical development (see Johnson, Senju, & Tomalski, 2015; Leppänen, Cataldo, Bosquet Enlow, & Nelson, 2018). Nevertheless, since some studies reported that a bias in favour of negative, threat-related stimuli (both face stimuli and non-face stimuli, like objects or voices) is in place already in infants (for a review, see Vaish, Grossman, & Woodward, 2008), the fact that we did not observe a differentiated response in favour of the Low Trustworthy face should be put at stake in future studies, and further discussed.

Our results showed both similarities and differences with respect to those previously reported by Jessen and Grossman (2016). Congruent findings lie in a general modulation of the P400 and the Nc, but not the N290, in response to variations in the level of expressed trustworthiness. However, the pattern of the effects that we obtained differed from that obtained by the authors. As already discussed, the P400 and the Nc in our study revealed a left lateralized latency advantage in favour of the High Trustworthy face with respect to the Neutral face, which is evidence of faster allocation of attention to the former stimulus as compared to the latter. In contrast, Jessen and Grossman reported a right-lateralized amplitude advantage at the level of the P400 in favour of the Neutral faces with respect to the High Trustworthy faces. The authors noted that their finding is incongruent with the previously reported evidence of larger P400 amplitudes for prosocial agents (Gredebäck et al., 2015), which, instead, is more in line with our finding of shorter P400 latencies for the High Trustworthy face.

A second difference between our results and those reported by Jessen and Grossmann (2016) relates to the Nc, which in our data peaked faster to the High Trustworthy face than to the Neutral face, but not the Low Trustworthy face. The authors instead reported larger Nc amplitudes for the Neutral face compared to both the High Trustworthy and the Low Trustworthy faces. We believe this discrepancy could arise from the differences between the stimuli used in two studies, i.e., artificial, computer-generated faces in Jessen and Grossman's study vs. realistic face images in the current study. Indeed, computer-generated faces might be processed differently with respect to more ecological ones (see Crookes et al., 2015). Second, we tried to quantify the perceptual differences between the two sets of experimental stimuli, and found that the Low Trustworthy faces used by Jessen and Grossman and taken from the Todorov database were perceived to be less trustworthy than our Low Trustworthy face. This might explain why we failed to replicate the authors' finding of neural discrimination between the Neutral and both the Low Trustworthy and the Hight Trustworthy face at the level of the Nc. Moreover, Jessen and Grossman's stimuli were perceived to be generally more distinguishable from each other with respect to our stimuli, which were perceived as more similar and thus might be more difficult to discriminate based on their subtle differences. Interestingly, differences between the stimuli of the current set and Todorov set did not affect adults' data, as we perfectly replicated results from previous studies investigating adults' neural sensitivity to face trustworthiness using Todorov stimuli (e.g., Marzi et al., 2014). This could be explained by the fact that adults gained more expertise with faces, which might allow them to easily generalize across different face-types expressing similar levels of trustworthiness. On the contrary, infants' reduced experience with faces might makes it more difficult for them to generalize across stimuli set, thus affecting the neural processes underlying discrimination of such stimuli. The way different kind of stimuli can affect neural responses to face trustworthiness, especially in infancy, should be put under scrutiny in future studies.

The presence of finer differences between our stimuli compared to those used by Jessen and Grossman, and their consequent perceived similarity, might also be the reason why we failed to replicate the previously reported behavioural preference for the High Trustworthy face over the Neutral and the Low Trustworthy face in Study 3b. Indeed, in our study infants' failed to show a preference for any of the three presented stimuli

An important aspect of the results obtained by Study 3a is the impact of temperamental traits on infants' neural sensitivity to facial cues to trustworthiness. Although the N290 was not modulated by variations in trustworthiness intensity, its overall amplitude over the left hemisphere was negatively associated to infants' score at the Negative Affect scale (i.e., to lower chances of experiencing negative feelings and less difficulty in being soothed). This is consistent with previously reported evidence that the N290 component is sensitive to variability in infants' communicative behaviours (Key, Stone, & Wiliams, 2009). In fact, larger N290 amplitudes and shorter N290 latencies in response to social stimuli are associated to infants' higher levels of expressive communication.

A second interesting result is that higher scores at the Surgency scale were associated to a faster processing of the Low Trustworthy with respect to the High Trustworthy faces at the level of the P400 component, as revealed by shorter P400 latencies. This results indicate that infants showing higher levels of activity and positive emotions are faster at allocating attention to the Low Trustworthy face with respect to the High. This result in in line with data showing that infants who manifest an attentional bias for thread-related facial expressions such as fear by 7 months of age will later show greater social involvement, with higher chances to empathize with others and understand other people's needs at 48 months of age (Peltola, Yrttiaho, & Leppänen, 2018). Moreover, this result show that infants who show low levels of activity and positive emotions are slower at allocating attention to the Low but faster at allocating attention to the High Trustworthy face, which is in line with the modulatory effect that Negative affect (tendency to experience negative feelings and difficulty being soothed) showed at the level of the Nc. In fact, infants scoring high in Negative Affect were faster in allocating attention to the High Trustworthy face with respect to the Neutral face, as revealed by Nc shorter latencies. Negative Affect is described as the tendency to experience negative emotions, such as nervousness or stress. High degrees of stress in early life are known to impact attentional mechanisms, being associated, for example, to avoidance of threat-related or negative environmental stimuli, such as facial expressions of fear (KL Humphreys, Kircanski, Colich, & Gotlib, 2016) or angriness (Nelson, Westerlund, McDermott, Zeanah, & Fox, 2013). In our study, we did not observe avoidance-like responses to a threat-related stimulus like the Low Trustworthy face in association to high Negative Affect scores. Nevertheless, infants experiencing higher degrees of distress might be faster in allocating attention to the High Trustworthy face as a copying strategy to reduce stress itself (In-Albon, Kossowsky, & Schneider, 2010).

In Summary, Study 3 explored infants' sensitivity to facial cues that are used by adults and older children to generate trustworthiness judgements using realistic images of a female face identity, in order to tap the perceptual expertise that infants naturally acquire with faces in their social environment. The stimulus set was previously validated as described in the Method section of Study 2 and were used to tackle adults' and children sensitivity to slight variations in facial cues to trustworthiness. They were also used to test a group of adults in order to check if they were capable to elicit neural responses similar to those elicited by the highly validated stimulus sets used in previous studies (e.g., Marzi et al., 2014; Jessen & Grossmann, 2016). Our infant data confirm the few existing evidence (e.g., Jessen & Grossmann, 2016; Sakuta, Kanazawa, & Yamaguchi, 2018) of early sensitivity to facial cues to trustworthiness, and extend such evidence by showing that individual differences in temperamental traits modulate neural activity underling the processing of facial trustworthiness.

# Study 4

**Neural responses to changes in facial cues to trustworthiness: a FPVS study**

Study 4 explored infants' neural sensitivity to the perceptual cues that differentiate the High Trustworthy face from the Low Trustworthy face by using an EEG Oddball paradigm and a Fast Periodic Visual Stimulation procedure. Indeed, Jessen and Grossman (2016) reported evidence of neural discrimination between a neutral face and both the trustworthy and the untrustworthy faces, but not between these two faces. As already discussed, this evidence is in line with previous demonstration of a quadratic, valence-independent pattern of amygdala activation, with maximal activation in response to both positively-valenced and negatively-valenced faces lying at extremes of the trustworthiness continuum, and weaker activation in response to faces that are close to the center of the continuum (Said, Baron, & Todorov, 2009; Said, Dotsch, & Todorov, 2010). Indeed, this pattern is also in line with the results of Study 1 and Study 2, in which participants showed better discrimination performance for more intense facial cues to trustworthiness irrespective of their valence. As already noted, this might be related to the adaptive advantage of being tuned to socially relevant cues than to socially neutral cues, and is in accord with the finding that the processing of emotional faces, which convey adaptively relevant information, is usually faster than the processing of neutral expressions (e.g., Schuboe, Gendolla, Meinecke, & Abele, 2006; Hirai et al., 2017).

Unlike previous results from Jessen and Grossman (2016), our results from Study 3a, in which an ERP procedure was used, showed that High Trustworthy and Low Trustworthy faces were discriminated at the level of the P400 latency with the amplitude of the difference being modulated by individual differences in temperamental traits related to the infant's level of activity and positive emotions (i.e., Surgency). Moreover, the P400 peaked faster in response to the High Trustworthy face compared to both the Neutral face and the Low Trustworthy face in the left hemisphere.

In light of the discrepancies between the current (Study 3a) and the past (Jessen and Grossmann, 2016) ERP evidence, Study 4 used a different EEG visual discrimination paradigm, i.e., the Fast Periodic Visual Stimulation paradigm, to explore neural discrimination between High trustworthy and Low trustworthy faces in 6-month-old infants, and the role of temperamental traits in modulating neural sensitivity to trustworthiness cues. The Fast Periodic Visual Stimulation (FPVS) paradigm (Rossion, 2014) has been successfully used as a sensitive measure of visual discrimination of faces and face-like stimuli (e.g., de Heering & Rossion, 2015; Liu-Shuang, Norcia, & Rossion, 2014; Rossion, Torfs, Jacques, & Liu-Shuang, 2015; Peykarjou, Hoehl, Pauen, & Rossion, 2017).

In the FPVS paradigm the stimuli of interest are repeatedly displayed at a fixed periodic frequency rate, which leads to a periodic change in voltage amplitude of the electroencephalographic signal at the same frequency of stimulation. Discrimination between the face categories of interest is measured through an Oddball paradigm, in which one face category, the Standard stimulus, is presented repeatedly and a different stimulus, the Oddball, is presented once every four repetitions of the Standard stimulus. Therefore, a discrimination response can be recorded at the frequency at which the change in face category occurs (i.e., the Oddbal frequency). For example, if the Standard faces are presented at a baseline stimulation rate of 6 Hz (6 faces per second), and the Oddball is shown every four stimuli, a discrimination response could be observed at a frequency of 1.2 Hz (6 Hz/5). A periodic response in the form of a high Signal to Noise Ratio (SNR) at the Oddball frequency is considered as a measure of visual discrimination. This method of recording EEG responses is preferred to others because of its objectivity, as the experimenter can stimulate at a pre-fixed frequency, and because it provides a response that is quantifiable in a very short amount of time, making it particularly suitable to be used with infants, who are typically scarcely collaborative.

In Study 4, young adults and 6-months-old infants were tested in a FPVS task, in which real images of a High Trustworthy and a Low Trustworthy variation of two female face identities were presented at a rate of 6 Hz following an Oddball paradigm, with the trustworthiness level (High versus Low) changing every 1.2 Hz. As in Study 3a, a group of adults was also tested using the same paradigm in order to validate the stimuli and the procedure, and provide qualitative investigation of age-related differences in neural sensitivity to facial cues to trustworthiness.

In light of the results of Study 3a, where individual differences in temperamental traits was found to modulate sensitivity to facial cues to trustworthiness, in the current study as well we analyzed infants' and adults' electrophysiological responses as a function of inter-individual variability in infants' temperamental traits related to Negative Affect and Surgency, and adults' personality traits related to Neuroticism and Extraversion.

**Materials and methods**

Participants

Twenty-two young adults composed the final adult sample (8 males, $M_{age}$ = 24.95 years; *SD* = 3.17; range: 20 – 32 years). All of them were either undergraduate or graduate university students compensated with course credits or who participated on a voluntary basis by word of mouth within the community. We also tested three additional adults, which were excluded as a consequence of excessive EEG artefacts (*N* = 2) or lack of enough survived conditions (*N* = 1). All subjects did not report any history of neurological or psychological disorders and declared to have normal or corrected-to normal vision. Participants gave their informed consent and completed the Italian version of the Big Five Questionnaire (BFQ; Caprara, Barbaranelli, Borgogni, & Perugini, 1993). As in Study 3, with the aim to obtain measures of individual differences in personality related dimensions that could be compared to individual differences in infant temperamental traits, only the items contributing to the Neuroticism (tendency to experience negative or anxious states) and Extraversion (tendency to joyfully engage in human interactions) scales were considered.

Thirthy 6-month-old healthy infants composed the final infant sample (17 males, $M_{age}$ = 196.7 days; *SD* = 7.72; range: 182 – 211 days). All infants were born full-term and had a normal weight at birth (> 2, 413 g). We also tested 22 additional infants, which were excluded due to fussiness (*N* = 8), excessive artefacts (*N* = 9) or because an entire condition had to be excluded because, for all presented trials, infants did not attend to the screen for at least 50% of the entire trial duration or did not show a Baseline response (see the Procedure, *N* = 5).

Infants were recruited through written invitations sent to parents based on birth records of neighbouring cities. Parents gave their written informed consent, and the mother or the primary caregiver completed the Infant Behavior Questionnaire-Revised in its very short form (IBQ-R VSF; Putnam, Helbig, Gartstein, Rothbart, & Leerkes, 2014). The questionnaire included queries aimed to assess the frequency of specific temperament-related behaviors observed within the last week; we focused on two subscales: Negative Affect (NA, tendency to experience negative feelings and difficulty being soothed) and Surgency (SU, tendency to show high levels of activity and positive emotions and to act impulsively), which are thought to be the equivalent of the two personality dimensions of Neuroticism and Extraversion.

The protocol of the study was carried out in accordance to the ethical standards of the Declaration of Helsinki (BMJ 1991; 302:1194), and was approved by the Ethics Committee of the University of Milano-Bicocca (Protocol number: 236/2016).

Stimuli

Stimuli were four coloured images of two cropped female face identities displaying two levels of trustworthiness: Low Trustworthiness (LT) and High Trustworthiness (HT) (see Figure 2.14). The stimuli were created with WebMorph (DeBruine, 2017) according to the same procedure used to generate the stimuli used in Study 2 and 3, in which we morphed an averaged neutral face towards an averaged untrustworthy and an averaged trustworthy face. We opted for not using the same facial identity used in these previous studies because repetition of the same facial identity would have likely caused visual adaptation, thus reducing participants' EEG response (Rossion, 2014). We selected three dark-haired and three blond-haired neutral faces from the Karolinska Directed Emotional Faces (KDEF) Database (Lundqvist, Flykt, & Öhman, 1998), and averaged each triplet to obtain two neutral facial identities, one dark-haired (Identity 1) and one blonde (Identity 2).

We also selected and averaged the 3 most untrustworthy faces and the 3 most trustworthy faces available in the Chicago Face Database (Ma, Correll, & Wittenbrink, 2015) in order to use them as references for transforming the two neutral face identities towards the untrustworthy/trustworthy extremes. The averaged neutral faces were morphed by a 50% step towards the very untrustworthy reference and by a 50% step towards the very trustworthy reference to obtain a Low Trustworthy (LT) version and a High Trustworthy (HT) version of each face identity.

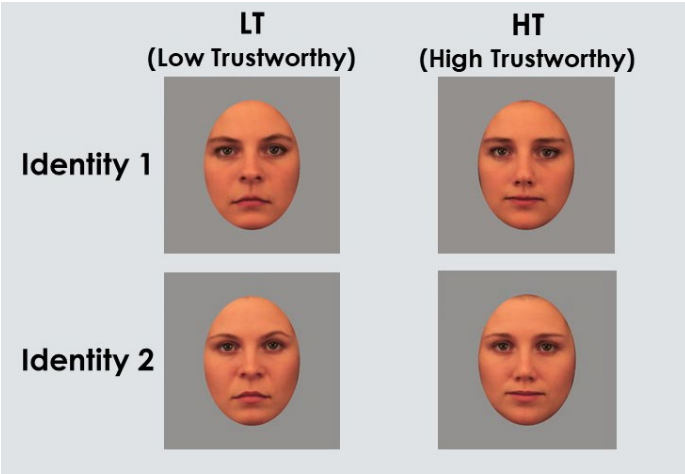


*Figure 2.14. The High Trustworthy and the Low Trustworthy versions of the two face identities used as stimuli in Study 4.*

*Stimuli validation.* Stimuli were validated by asking an independent sample of 40 adults (33 females; mean age = 23 years; range = 19 - 30) to rate the perceived trustworthiness intensity, perceived emotion and attractiveness of the four face stimuli on a scale ranging from 1 (e.g., "I wouldn't trust this person at all", "This face is hardly typical/attractive") to 9 (e.g., "I would definitely trust this person", "This face is quite typical/attractive"). Participants rated the four stimuli on the trustworthiness dimension first, and were subsequently asked to rate them on perceived emotion, attractiveness and typicality with the three questions presented in random order. If participants indicated that the face was expressing an emotion, they were prompted to select the expressed emotion (happy, angry, sad, scared or other) and score the intensity of the selected emotion on 1-to-9 scale. In case of negative answers (e.g., if participants evaluated the face to be not emotional, not typical or not attractive), scores were treated as zero values.

A repeated-measures ANOVA with Trustworthiness Intensity (LT, HT) and Identity (1, 2) as within-subjects factors performed on the mean scores resulting from adults' ratings revealed a main effect of trustworthiness, $F(1,39) = 30.347$, $p < .001$, $p\eta^2 = 0.438$: independently of face identity, the Low Trustworthy face was perceived as less trustworthy (*M*= 3.49, *SD* = .261) than the High Trustworthy face (*M* = 4.98, *SD* = .237). A repeated-measures ANOVA with Perceived Dimension (trustworthiness, happiness, angriness, sadness, fear, attractiveness, typicality) and Trustworthiness Intensity (LT, HT) as within-subjects factors revealed a main effect of perceived dimension, $F(6,234) = 85.272$, $p < .001$, $p\eta^2 = 0.686$. Faces were rated as overall more trustworthy than emotional or attractive (all *ps* < .001), whereas the intensity of perceived trustworthiness did not differ from those of perceived typicality ($p = 1.000$), irrespective of face identity. The ANOVA also revealed a Perceived Dimension x Trustworthiness Intensity interaction, $F(6,234) = 10.869$, $p < .001$, $p\eta^2 = 0.218$: the HT face and the LT face did not differ in terms of perceived typicality and attractiveness, but, even though the stimuli were perceived as overall poorly emotional, the HT face was perceived as more happy (*M* = .588, *SD* = .173), $p = .002$, less angry (*M* =.00, *SD* = .00), $p < .001$, and more sad (*M* =1.100, *SD* = .229), $p = .007$, than the LT face (respectively, *M* =.00, *SD* = .00; *M* =2.025, *SD* = .363; *M* =.275, *SD* = .133). This is in line with what we observed in the validation procedure of the stimuli used in Study 2 and 3a, and is in accord with previous demonstration that trustworthiness judgements should be interpreted as overgeneralized responses to facial cues resembling emotional expressions (Said, Sebe, et al., 2009; Zebrowitz et al., 2003).

<u>Procedure</u>

Data acquisition took place for all participants in an audiometric and electrically shielded cabin. A 24-inch monitor was used for stimulus presentation, while participants seated at approximately 60 cm away from the screen. Infants always seated on their parent's legs. E-Prime software v2.0 (Psychology Software Tools Inc., Pittsburgh, PA) was employed to present experimental stimuli. Adults were asked to sit as still as possible during the entire procedure. The same instruction was given to parents, who were also asked to avoid interfering with the experimental session by communicating with the child with both verbal or non-verbal means. In the case of infants, the whole experimental session was recorded by an infrared video camera connected to the data acquisition computer and hidden over the monitor. The live image of the infant's face and body allowed the experimenter to pause or terminate the session as soon as the infant became fussy.

Through a MATLAB (Mathworks Inc.) custom script (adapted from Barry-Anwar, Hadley, Conte, Keil, & Scott, 2018), stimuli were presented periodically at a Baseline frequency of 6 Hz, and following an Oddball task, where every fifth face (Oddball frequency of 1.2 Hz) the trustworthiness level changed. During the stimulation, faces were presented by a sinusoidal contrast modulation (from 0% to 100%) and varied in size and screen position (of 1 angle degree from the screen centre) to avoid neural responses related to low level stimuli characteristics (see Figure 2.15).

Subjects viewed two different conditions: one where the Standard stimulus was LT and the Oddball was HT (High-Oddball condition), and one where the Standard stimulus was HT and the Oddball was LT (Low-Oddball condition). The order at which conditions were presented was counterbalanced across participants. Each condition was composed of a total of 4 trials, each lasting 15 000 ms, where a cycle of 4 Standard stimuli and one Oddball was repeated 18 times. For each trial, each facial identity was shown 9 times, and the same facial identity could not be shown more than 3 times in a row. For infants, the experimenter used attention getters between trials to reorient infants' attention, and started each trial as soon as the infant was looking at the screen. The entire presentation lasted between two to three minutes.

Adults were instructed to press the mouse button as soon as they noticed a color change of the fixation cross (e.g., Xu, Liu-Shuang, Rossion, & Tanaka, 2017). This task aimed at maintaining participants' attention throughout the duration of the experimental session. A light grey fixation cross was presented in the centre of the screen and briefly (500 ms) changed colour to light green in a random fashion, from a minimum of 2 and a maximum of 4 times per trial. Participants performed well in this task, with a mean accuracy of 77.15% (*SD* = 8.06).

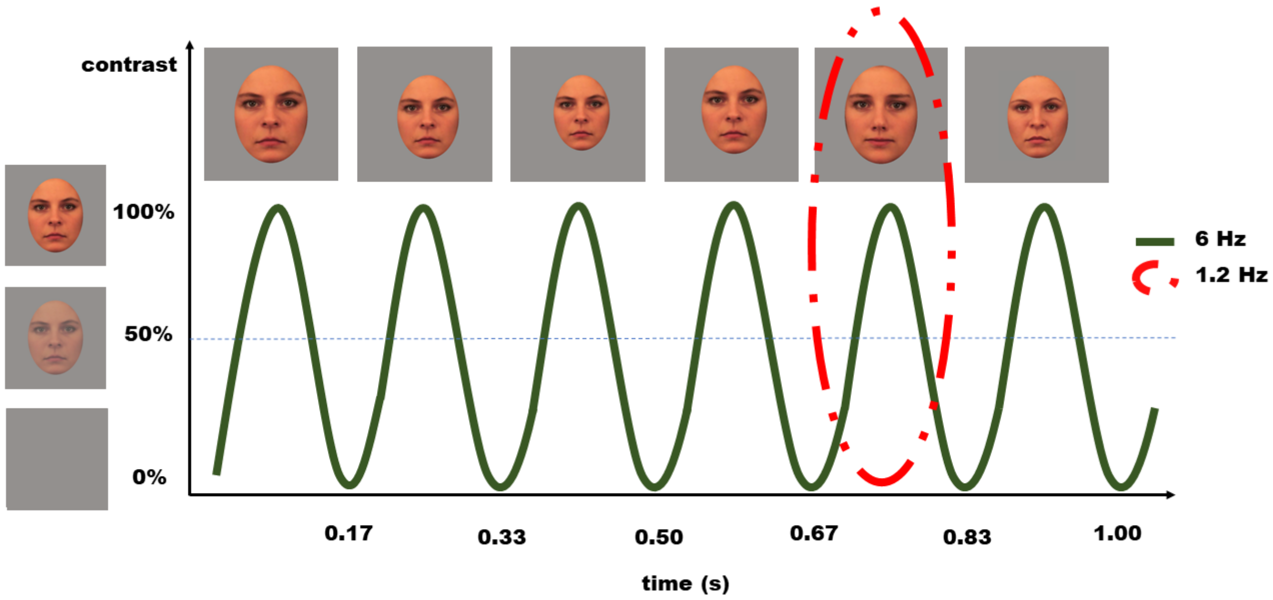


*Figure 2.15. Example of Fast periodic visual stimulation (FPVS) during the Oddball task (High-Oddball condition). Six faces were shown within one second (6 Hz Baseline frequency), and the trustworthiness intensity (HT or LT) changed every 5 faces (1.2 Hz Oddball frequency, see the red circle). Stimuli were presented by a sinusoidal contrast modulation (from 0% to 100%) and varied in size and screen position (of 1 angle degree from the screen centre).*

<u>EEG recording and processing</u>

Electroencephalographic (EEG) data were recorded through a 128-electrode HydroGel Geodesic Sensor Net (Electrical Geodesic Inc., Eugene, OR), with the vertex electrode (Cz) used as reference. Signals were amplified using an EGI NetAmps 300 amplifier with a sampling rate of 500 Hz and an online band-pass filter of 0.1–100 Hz. . Impedances were checked online before the session started and they were considered as adequate if lower than 50 KΩ.

EEG data were processed through a MATLAB (Mathworks Inc.) custom script adapted from Barry-Anwar and colleagues (Barry-Anwar et al., 2018). We bandpass filtered the data using a high-pass cut-off of 0.02 Hz with a $3^{rd}$ order Butterworth filter and a 18dB/octet slope and a low-pass cut-off of 30 Hz with a $6^{th}$ order Butterworth filter and a slope of 36 dB/octer. After filtering, we segmented the data starting 100 ms before the commencement of the trial until the end of each 15 000 ms trial. We removed from the analyses the peripheral band of electrodes (17, 43, 48, 49, 56, 63, 68, 73, 81, 88, 94, 99, 107, 113, 119, 120, 125, 126, 127, and 128) as they are typically characterized by noise. We detected bad channels trial by trial, using three metrics: the median absolute voltage value, the standard deviation of voltage values and the maximum difference in voltage values at each channel. As a first step, we removed those channels with values equal or larger than 2 standard deviations from the median absolute voltage. As a second step, we removed all remaining channels with values equal or larger than 2.5 standard deviations from the updated median absolute voltage. Bad channels values were substituted with the average voltage from up to 6 neighbouring channels. If any of the neighbouring channels were also identified as bad channels, fewer than 6 channels were used. At least one channel of one condition was interpolated for all participants, except for one adult. An average of 5.8 % of channels for adults and an average of 5.3 % of channels for infants were interpolated. Channels were re-referenced to the average reference.

Those trials where the infant did not look at the screen for at least 50% of the entire 15-sec trial duration were removed. Across all infants, 8.75% of all trials (21 out of 240 trials) had to be excluded following this criterion. After that, we inspected trials for artefacts and noise, and we excluded them if noisy for more than 50% of the trial duration. Across adults, 9.09% of all trials (16 out of 176) were removed with this criterion, while across infants 6.67% of all trials (16 out of 240) were removed. In case less than 50% of the entire trial duration was characterized by artefacts or noise, those noisy segments were replaced with zero values. This way we were able to remove noisy segments without having to discard the entire trial and without affecting the length of the trials themselves. Nevertheless, the presence of zero values prevented the possibility to average trials at a group level. Consequently, we first computed SNRs and Baseline Corrected Amplitudes (BCAs, see e.g., Peykarjou, Hoehl, Pauen, & Rossion, 2017) on single trials and then averaged SNRs and BCAs values across trials.

Frequency domain analysis

In order to obtain segments that contained an integer number of cycles (4 Standard faces and 1 Oddball face), we removed the 100 ms baseline before the start of each trial. Each single trial underwent a windowed Fast Fourier Transform (FFT) at a frequency resolution of 1/15 s. SNR and BCA were computed on each trial's frequency domain values. SNR values were computed as the ratio of the amplitude at the frequency of interest to the average of the 20 neighbouring frequency bins (10 on each side, excluding the immediately adjacent bin on each side). BCA values were computed as the amplitude at the frequency of interest minus the average of the 20 neighbouring frequency bins (10 on each side, excluding the immediately adjacent bin on each side). If participants did not show an SNR of at least 1.5 for the Baseline frequency in the medial occipital electrodes (namely, 70, 75 and 83) the corresponding trial was excluded from analyses. Consequently, subjects were included in the final analyses if they contributed with at least one or more 15 s trial to each condition (High-Oddball condition, Low-Oddball condition). Adults' average of good trials was 6.64 ($SD$ = 1.62; range 3 – 8 trials) out of 8 trials. Infants' average of good trials was 6.57 ($SD$ = 1.33; range 3 – 8 trials) out of 8 trials.

For the Baseline response, for both adults and infants, harmonics in the occipital clusters proved to be significant up to the 5$^{th}$ harmonic (i.e., 30 Hz). Therefore, analyses on SNRs and BCAs were performed on the sum of these harmonics (Retter & Rossion, 2016).

For the Oddball frequency, adults' responses proved to be significant up to the second harmonic, therefore we analysed the aggregated response at 1.2 and 2.4 Hz. Infants' Oddball response did not prove to be significant nor at 1.2 Hz or at 2.4 Hz. Rather, significant responses were observed from 9.6 Hz to 14.4 Hz. Therefore, we analysed the aggregated response of these harmonics, excluding the 12 Hz Baseline frequency.

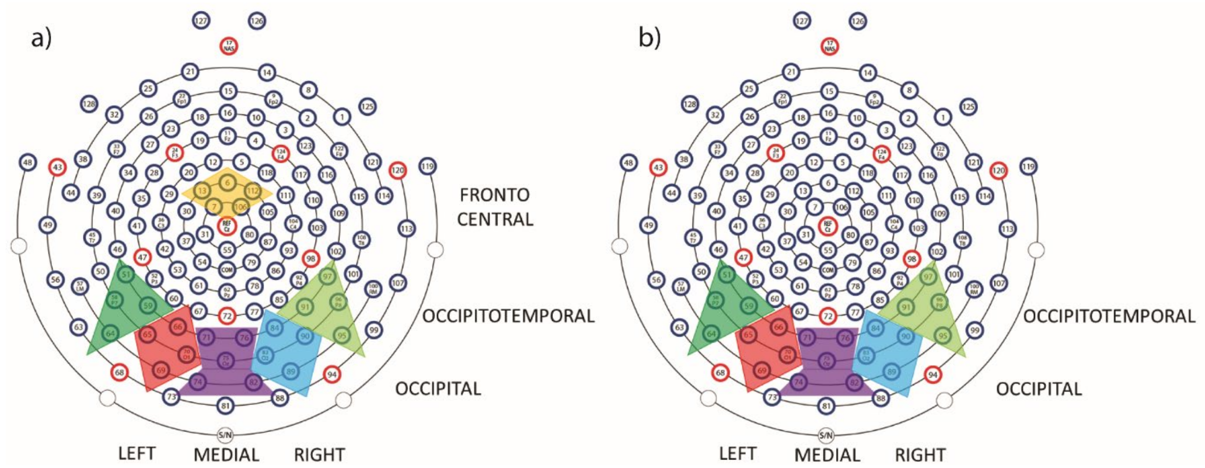


*Figure 2.16. Regions of Interest (ROIs) for the adults (a) and the 6-month-old infants (b).*

<u>Regions of Interest (ROIs)</u>

Based on previous FPVS studies, we considered five regions of interests across posterior regions: left, medial and right occipital regions, and left and right occipitotemporal regions (e.g., de Heering & Rossion, 2015; Liu-Shuang, Norcia, & Rossion, 2014, see Figure 2.16a). For the adult sample, a frontocentral cluster was added based on the observation of the topographical distribution of the response (see Figure 2.16b). SNR and BCA values within each ROI were averaged across channels and used as input for the analyses separately for the two conditions (High-Oddball condition, Low-Oddball condition).

**Results**

Adults

*Baseline Frequency.* The Baseline response at occipital regions was analysed through one-sample *t*-tests (one-tailed) against 1 (noise level) on SNRs of all occipital ROIs (see Barry-Anwar et al., 2018). The response proved to be significant for both conditions across all occipital sites, all *ps* < .001 (left Occipital: $M_{High\text{-}Oddball}$= 2.432, $SD_{High\text{-}Oddball}$= .970, $M_{Low\text{-}Oddball}$= 2.414, $SD_{Low\text{-}Oddball}$= 1.089; medial occipital: $M_{High\text{-}Oddball}$= 2.749, $SD_{High\text{-}Oddball}$= 1.009, $M_{Low\text{-}Oddball}$= 2.758, $SD_{Low\text{-}Oddball}$= 1.100; right Occipital: $M_{High\text{-}Oddball}$= 2.234, $SD_{High\text{-}Oddball}$= .776, $M_{Low\text{-}Oddball}$= 2.265, $SD_{Low\text{-}Oddball}$= 0.823), see Figure 2.17.

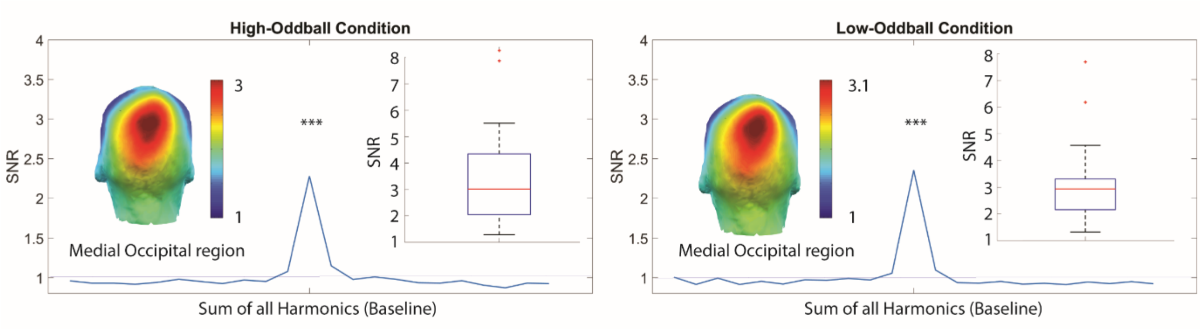


*Figure 2.17. Adults' SNR at the Baseline frequency (sum of all harmonics of interest) for the High-Oddball condition and the Low-Oddball condition. Topographic plots represent SNR over posterior regions. The frequency graph represents the mean SNR in the medial occipital region, which peaked at the Baseline frequency (sum of harmonics of interest). The boxplot represents the distribution of SNR values across participants in the same brain region.*

To explore whether the EEG signal at the Baseline frequency differed according to the condition or the selected ROIs, we conducted a repeated-measures ANOVA on BCA values with Condition (High-Oddball condition, Low-Oddball condition) and ROI (left, medial and right occipital) as within-subjects factors. The analysis yielded a significant main effect of ROI, $F(2,38) = 9.212$, $p = .001$, $p\eta^2 = .327$, as BCA were significantly larger in the medial occipital region ($M = 1.495$, $SD = .197$) than in both the left occipital ($M = 1.115$, $SD = .162$), $p = .006$,

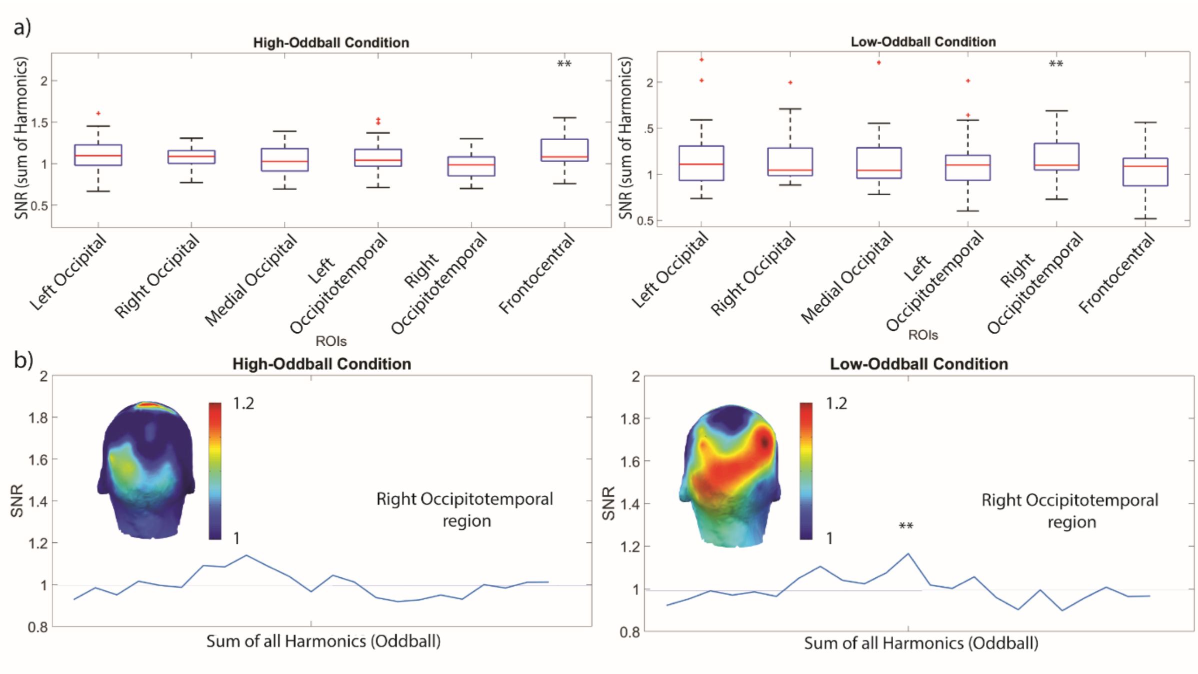


*Figure 2.18. Adults' SNR at the Oddball frequency (sum of all harmonics of interest) for the High-Oddball condition and the Low-Oddball condition. Boxplots (a) represent the distribution of SNR values across participants in all ROIs of interest. Topographic plots (b) represent SNR over posterior regions. The frequency graph represents the mean SNR in the right occipitotemporal region.*

and the right occipital regions ($M = 1.18$, $SD = .159$), $p = .001$, irrespective of trustworthiness intensity.

*Oddball Frequency.* The Oddball response was analysed through one-sample *t*-tests (one-tailed) against 1 (noise level) on SNRs of all ROIs, and the statistical threshold was established through Bonferroni correction for multiple comparisons ($p = .008$). SNR was significantly higher than noise in the frontocentral cluster of the High-Oddball condition, $M = 1.133$, $SD = .181$, $t(21) = 3.434$, $p = .001$, and in the right occipitotemporal cluster of the Low-Oddball condition, $M = 1.174$, $SD = .242$, $t(21) = 3.365$, $p = .0015$. To explore whether the signal at the Oddball frequency differed according to condition, ROI, or scores on the Neuroticism and Extraversion scales of the IBQ-R, we conducted a repeated-measures ANCOVA on BCA values with condition (High-Oddball condition, Low-Oddball condition) and ROI (left, medial and

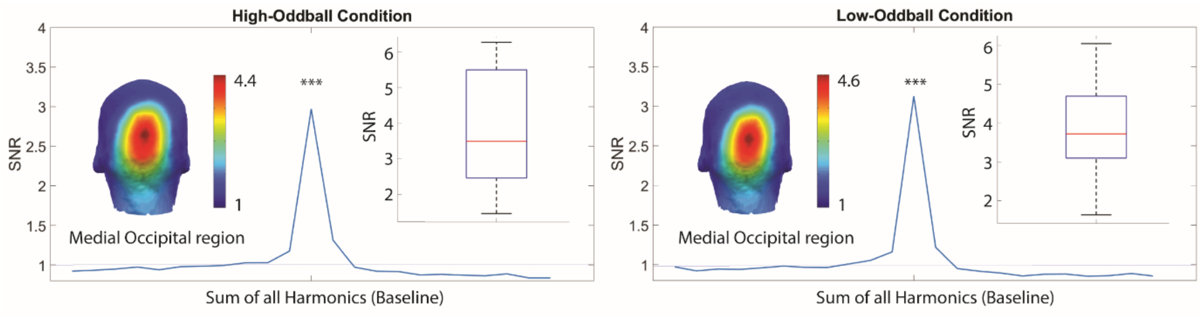


*Figure 2.19. Infants' SNR at the Baseline frequency (sum of all harmonics of interest) for the High-Oddball condition and the Low-Oddball condition. Topographic plots represent SNR over posterior regions. The frequency graph represents the mean SNR in the medial occipital region, which peaked at the Baseline frequency (sum of harmonics of interest). The boxplot represents the distribution of SNR values across participants in the same brain region.*

right occipital, left and right occipitotemporal and frontocentral) as within-subjects factors and Negative Affect and Surgency scales entered as covariates. The analysis revealed a significant Condition x ROI interaction, $F(5,95) = 5.051$, $p = .002$, $p\eta^2 = .210$, as BCA values for the two conditions differed only in the right occipitotemporal cluster, BCA being higher for the Low-Oddball condition proved to be significantly higher, $M = .163$, $SD = .05$, than for the High-Oddball condition, $M = -.041$, $SD = .037$, $p = .003$, see Figure 2.18.

<u>Infants</u>

*Baseline Frequency.* The one-sample *t*-tests (one-tailed) against 1 (noise level) on SNRs proved to be significant for both conditions across all occipital ROIs, all $ps < .001$ (left Occipital: $M_{High\text{-}Oddball} = 2.184$, $SD_{High\text{-}Oddball} = .939$, $M_{Low\text{-}Oddball} = 2.098$, $SD_{Low\text{-}Oddball} = .799$; medial occipital: $M_{High\text{-}Oddball} = 3.556$, $SD_{High\text{-}Oddball} = .991$, $M_{Low\text{-}Oddball} = 3.502$, $SD_{Low\text{-}Oddball} = 1.187$; right Occipital: $M_{High\text{-}Oddball} = 2.159$, $SD_{High\text{-}Oddball} = .724$, $M_{Low\text{-}Oddball} = 2.282$, $SD_{Low\text{-}Oddball} = 0.784$), see Figure 2.19.

To explore whether the signal at the Baseline frequency differed according to condition or ROIs, a repeated-measures ANOVA on BCA values with condition (High-Oddball condition, Low-Oddball condition) and ROI (left, medial and right occipital) as within-subjects factors was conducted. The ANOVA yielded a significant main effect of ROI, $F(2,58) = 26.090$, $p < .001$, $p\eta^2 = .474$, which was due to BCA being significantly higher in the medial occipital region ($M = 6.504$, $SD = .577$) than in both the left occipital ($M = 3.032$, $SD = .452$), $p < .001$, and the right occipital regions ($M = 3.054$, $SD = .369$), $p < .001$, irrespective of condition.

*Oddball Frequency.* The Oddball response was analysed through one-sample *t*-tests (one-tailed) against 1 (noise level) on SNRs across all ROIs, and the statistical threshold was established through Bonferroni correction for multiple comparisons ($p = .01$). SNR for the Low-Oddball condition was significantly higher than noise in the right occipital region ($M = 1.059$, $SD = .129$), $t(29) = 2.530$, $p = .0085$, whereas SNR for the LT-Oddball condition was marginally higher than noise in the medial occipital cluster ($M = 1.052$, $SD = .118$), $t(29) = 2.399$, $p = .0115$. To explore whether the EEG signal at the Oddball frequency differed according to condition, ROI, or infants' scores at the Negative Affect and Surgency scales of the IBQ-R, a repeated-measures ANCOVA on BCA values was performed with condition (High-Oddball condition, Low-Oddball condition) and ROI (left, medial and right occipital, left and right occipitotemporal as within-subjects factors, and Negative Affect and Surgency entered as covariates. This analysis lead to a significant effect of condition, $F(1,27) = 5.051$, $p = .018$, $p\eta^2 = .190$, due to BCA values for the Low-Oddball condition being significantly higher ($M = .044$, $SD = .027$), than those for the High-Oddball condition ($M = -.033$, $SD = .020$), see Figure 2.20.

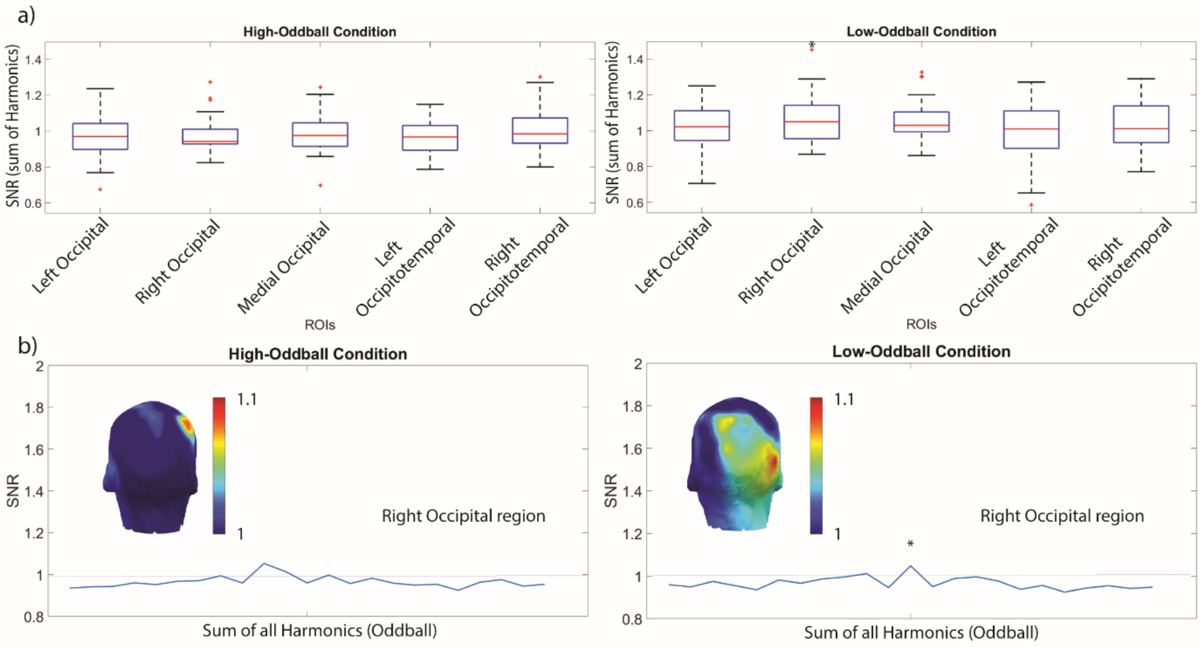


*Figure 2.20. Infants' SNR at the Oddball frequency (sum of all harmonics of interest) for the High-Oddball condition and the Low-Oddball condition. Boxplots (a) represent the distribution of SNR values across participants in all ROIs of interest. Topographic plots (b) represent SNR over posterior regions. The frequency graph represents the mean SNR in the right occipital region.*

## Discussion

Study 4 aimed at further exploring 6-month-old infants' sensitivity to facial cues to trustworthiness by focusing on their ability to discriminate between a very trustworthy face and a very untrustworthy face, taking into account the impact of individual variability related to temperamental traits. In order to do so, a newly developed visual discrimination paradigm that is particularly suitable for infants was used in combination with an oddball paradigm, in which truthful images of two female face identities displaying low or high levels of trustworthiness were present at a fast and constant periodic rate, and a different stimulus, the Oddball, was presented once every four standard stimuli.

Data from a group of adults confirmed that the stimuli and the procedure were suitable to produce a periodic change in voltage amplitude of the electroencephalographic signal at the same frequency of stimulation over the occipital regions, which was irrespective of the condition (i.e. whether the Oddball was a High or a Low Trustworthy face). More importantly, the fact that SNR of the Oddball frequency was significantly higher than noise in both the conditions confirmed adults' ability to discriminate between the two levels of trustworthiness, although such discrimination occurred in different brain areas depending on the condition: in a frontocentral cluster when the Oddball was a High Trustworthy face, and in a right occipitotemporal cluster when the Oddball was a Low Trustworthy face.

The frontocentral area has been previously found to be involved in the processing of emotionally relevant stimuli (e.g., Eimer & Holmes, 2007; Hilimire, Mienaltowski, Blanchard-Fields, & Corballis, 2013), suggesting that this brain area might play a crucial role in processing socially-valenced information embedded in human faces. The fact that the High Trustworthy face was more easily discriminated when embedded in a stream of Low Trustworthy faces at the frontocentral cluster is somewhat consistent with the results of Study 3a concerning the adults' C1. Indeed, the amplitude of the C1 was larger in response to the High Trustworthy and the Neutral faces compared to the Low Trustworthy face, just like in Yang and colleagues (Yang et al., 2011). Taken together, these results suggest that the early perceptual processing of high trustworthy faces might take place in frontocentral areas of the adults' brain.

On the contrary, the Low Trustworthy faces were discriminated from the stream of High Trustworthy faces in the right occipitotemporal cortex, as supported also from the observation that BCA values in the right occipitotemporal region were significantly higher in the Low-Oddball condition than in the High-Oddball condition. The right occipitotemporal cortex is known to be involved in different aspects of face processing, from identity discrimination to emotion recognition (e.g, Gauthier et al., 2000; Harry, Williams, Davis, & Kim, 2013; Kanwisher, McDermott, & Chun, 1997), as revealed also by other FPVS studies (Jacques et al., 2016). Our finding of enhanced processing of the Low Trustworthy face over the occipitotemporal cortex is in line with the evidence gathered in Study 3a of a larger occipitotemporal N170 in response to Low Trustworthy faces compared to High Trustworthy faces. In addition, the findings that the discriminative response was confined to the right occipitotemporal cortex is congruent with the valence hypothesis (Silberman & Weingartner, 1986), according to which the right hemisphere is involved in the processing of more negatively-valenced stimuli, while the left hemisphere processes more positively-valenced stimuli. Within this view, the right occipitotemporal cortex might play an important role in discriminating negatively-valenced facial cues, such as untrustworthiness, especially when untrustworthy faces are presented in a stream of trustworthy ones. Nevertheless, the valence hypothesis remains an highly debated one (Demaree, Everhart, Youngstrom, & Harrison, 2005), and our data do not fully support it, as discrimination in the High-Oddball was not left-lateralized. Future studies should put these results at stake by further investigating the role of different brain areas in processing different aspects of face trustworthiness in the adult population.

Data from the infants' sample confirmed results of Study 3a by showing that 6-month-olds are already sensitive to those facial cues that drive trustworthiness judgements in adults. Like for the adults, data on the Baseline frequency showed that EEG activity in the occipital regions oscillated at the frequency of presentation of faces that varied in the level of expressed trustworthiness, irrespective of the condition. As revealed from the analyses on the Oddball frequency, discrimination of the two face categories was found in the right occipital region for the condition where the Low Trustworthy face was Oddball, similarly to the adults' sample. Nevertheless, no evidence of discrimination was observed for the condition where the Oddball face was High Trustworthy. This differential processing for the two conditions was confirmed by the analysis on BCA values, that were overall higher for the Low-Oddball condition than for the High-Oddball condition.

This pattern of results could be explained making reference to the evidence provided by visual search studies. Indeed, various studies on both adults and children report a search asymmetry in favour of thread-related stimuli such as angry faces, when presented within happy or neutral face distractors (Hansen & Hansen, 1988; LoBue, 2009; Mather & Knight, 2006; Öhman, Lundqvist, & Esteves, 2001). This visual search asymmetry is interpreted as resulting from adaptive mechanisms that would prompt attentional response to threat-related environmental stimuli. In the case of infants, whose experience with faces is still relatively poor, the attentional advantage for threat-related stimuli might make the Low Trustworthy face more easily distinguishable when embedded in a fast-presentation stream of High Trustworthy faces than the reverse.

It should be noted, however, that the results obtained from the infant sample are incongruent with the ERP evidence obtained in Study 3a, showing faster allocation of attention to the High Trustworthy face in the left occipital and prefrontal areas. Moreover, in Study 3a we observed a modulatory effect of Surgency scores on the P400 latency sensitivity to differences between the High and the Low Trustworthy faces, while in Study 4 individual differences in temperamental traits did not affect infants' EEG responses.

A possible explanation for the observed discrepancies between the results of Study 3a and Study 4 might lie in the different nature of the two paradigms. First, the ERP paradigm measured the effort (amplitudes) and the speed (latencies) with which faces varying in the level of expressed trustworthiness were processed. On the other hand, the FPVS paradigm measured the ability of the infants' brain to discriminate between different levels of trustworthiness. While with ERPs the focus is on brain's response to each singular stimulus presentation, in the FPVS paradigm the focus is on brain's responses to the differences between standard and oddball stimuli, which can only be observed if the observer detects the communalities between standards and the difference of oddballs from the standards, and generalizes across oddballs.

In light of these differences, it is reasonable to assume that the FPVS paradigm would provide more sensitive measures of neural discrimination compared to ERPs.

In addition to this, it should also be considered that the stimulus presentation procedures adopted in the ERP and the FPVS paradigms are very much different. ERP stimuli were presented for 1000 ms each, while FPVS stimuli were presented for only 166 ms each (6 stimuli per second). Therefore, results from Study 3a might highlight that the left occipital and frontal areas process the High Trustworthy face faster than the Neutral and the Low Trustworthy faces when these stimuli are presented one at a time and at a slower pace (1000 ms each). On the other hand, Study 4 reveals that in the context of a fast presentation, the right occipital area discriminates between the two face categories when the Low Trustworthy face is the one to be distinguished in a stream of High Trustworthy ones, possibly as an effect of the search asymmetry previously discussed.

The fact that the ERP paradigm measured the effort and the speed with which faces varying in the level of expressed trustworthiness were processed, while the FPVS paradigm measured the ability to discriminate between different levels of trustworthiness might also explain why Surgency scores modulated infants' EEG responses in Study 3a, but not in Study 4. In fact, in Study 3a variability in Surgency scores affected the time of processing of the High Trustworthy and the Low Trustworthy faces, with higher Surgency scores corresponding to faster processing of the Low Trustworthy face with respect to the High Trustworthy face. On the other hand, in Study 4 variability in Surgency scores did not affect the ability to perceptually discriminate between the two face types. These results are in line with what we observed in Study 1, where variability in adults' Extraversion scores modulated the speed at which different levels of facial cues to trustworthiness are discriminated by in a behavioural discrimination task, without affecting discrimination accuracy.

Even though the neural and cognitive mechanisms that might explain the differential results among the two paradigms should be further investigated, results from Studies 3a and 4 are consistent in showing that 6-month-old infants are sensitive to those facial cues that, later in development, are used to generate trustworthiness judgements. In light of the evidence provided by Study 4 that 6-month-olds discriminate faces at the extremes of the trustworthiness continuum, a goal for future investigations could be to use the FPVS paradigm to probe infants' finer discrimination of trustworthiness cues by testing their ability to generalize across high and low trustworthy faces presented among neutral standard faces.

# Chapter 3

## Study 5

### Cross-cultural differences in perceptual sensitivity to facial cues to trustworthiness

In the Introduction and in Chapter 1 of this thesis, we reported and discussed evidence about adults' fine-grained ability to detect and respond to subtle differences in facial cues to trustworthiness. In Chapter 2, we reported evidence showing that sensitivity to these cues emerges in early stages of development (i.e., the first year of life), thus suggesting that it might require none, or a limited amount of experience. In light of this evidence, a viable way to explore the impact of experience in shaping sensitivity to facial cues to trustworthiness is to examine whether and how such sensitivity generalize across cultures.

Previous studies on the role of culture in shaping sensitivity to face information revealed that faces belonging to ethnicities that are underrepresented in the individual's social environment are recognized with greater difficulty with respect to those belonging to the individual's own ethnicity (i.e., the own-race bias; e.g, Bothwell, Brigham, & Malpass, 1989; Walker & Tanaka, 2003). This is usually explained as an effect of the perceptual expertise acquired as a result of repeated experience with a specific face category (Chiroro & Valentine, 1995).

Other evidence of the important role played by cultural experience comes from cross-cultural studies on emotion recognition abilities. These studies found both universal constants and differences in emotion recognition abilities across different cultures. If the hypothesis of the universality of the six basic human emotions (happiness, surprise, fear, anger, sadness, surprise) states that these emotions are expressed and recognized through the same facial cues across cultures (Ekman, Sorenson, & Friesen, 1969), a number of studies have now proven that significant differences exist in evaluation of certain emotion categories (Biehl et al., 1997), in the evaluation of displayed intensity of emotions (Ekman et al., 1987; Yrizarry, Matsumoto, & Wilson-Cohn, 1998) and in the cues (e.g., eyes vs mouth) used to process emotional information (Yuki, Maddux, & Masuda, 2007).

In the Introduction of this thesis, we have already discussed how sensitivity to facial cues that drives trustworthiness judgements could result from the interaction between biological mechanisms and experience-based mechanisms. Indeed, such sensitivity could emerge in response of evolutionary pressures to be able to distinguish between friendly or hostile intentions of others, to be then refined and shaped in the course of development by the social context and personality traits. In light of these considerations, we would expect that individuals from different cultural background use similar facial cues to draw information about one’s trustworthiness, even though perceptual expertise in discriminating among these cues might vary for different face categories (e.g., own-race versus other-race faces).

Only few studies have investigated cross-cultural differences in trustworthiness judgments from facial cues (Birkás, Dzhelyova, Lábadi, Bereczkei, & Perrett, 2014; Xu et al., 2012), and the results suggest that such judgements are not modulated by the participants’ cultural background. Xu and colleagues (Xu et al., 2012) showed that Caucasian and Asian adults use similar facial information to judge trustworthiness from Caucasian faces, as they both relied on facial attractiveness, inner brow ridge or skin shade. In addition, Birkás and colleagues (Birkás et al., 2014) reported that Caucasian and Asian participants generalize their perception of trustworthiness to faces of different ethnicities (African, Caucasian, South Asian and East Asian). Importantly, the only study conducted on developmental population by Sakuta and colleagues (Sakuta et al., 2018) showed that 6- to 8-month-old Japanese infants presented with Caucasian faces preferred the trustworthy face over the untrustworthy face, suggesting that they could disentangle facial cues to trustworthiness from other-race faces. Nevertheless, the study by Sakuta et al., (2018) failed to provide a cross-cultural comparison, as it did not include a group of Caucasian infants tested with Asian faces, thus leaving open the question of whether perceptual expertise has a role in modulating sensitivity to variations in facial cues to trustworthiness already in infancy.

Notwithstanding the relevance of these data, none of the existing studies has investigated whether there are cross-cultural differences in sensitivity to fine-grained differences in facial cues to trustworthiness of truthful facial identities, as they used computer-generated faces displaying extreme trustworthiness intensities. Moreover, none of the previous studies investigated cross-cultural differences in developmental population by using a fully factorial cross-cultural design, in which face ethnicity and the ethnicity of the beholder are simultaneously manipulated.

In Study 5, the same procedure adopted in Study 2 and 4 was used to create and validate a 5-step trustworthiness continuum of one female Asian face varying in the level of expressed trustworthiness. As for the Caucasian trustworthiness continuum used in Study 2, the validation procedure was performed under own-race conditions, that is by asking a group of Japanese adults to rate each face on the trustworthiness dimension as well as on the dimensions of perceived emotion, typicality and attractiveness. In addition, to provide a cross-cultural comparison of adults' tendency to make trustworthiness judgments from own-race and other-race faces, Italian and Japanese adults were asked to rate the trustworthiness intensity of the 5 faces composing the Asian trustworthiness continuum and a subset of 5 faces taken from the Caucasian trustworthiness continuum used in Study 2.

Once validated on the adult population, the Caucasian and the Asian 5-step trustworthiness continuums will be used to test cross-cultural differences in sensitivity to facial cues to trustworthiness in preschoolers and school-aged children, with the aim of testing whether the evidence gathered in Study 2 generalizes across cultures and face ethnicity.

**Materials and methods**

Participants

The sample included 53 Italian young adults (40 females; mean age = 22.45 years; range = 19-32) and 57 Japanese young adults (42 females; mean age = 22.47 years; range = 20-34). All participants signed an informed written consent before testing. The protocol was carried out in accordance with the ethical standards of the Declaration of Helsinki (BMJ 1991; 302: 1194) and approved by the Ethics Committee of the University of Milano-Bicocca.

Stimuli

Stimuli were realistic coloured images of 5 variations of one Asian female face identity and 5 variations of one Caucasian female face identity that were selected from those used in Study 2. The 5 variations of the Asian female identity were created by using the same procedure adopted in Study 2 and 4 to create the Caucasian stimuli, with two exceptions. First, the neutral Asian face was generated from the average of four (as opposed to only three in Study 2 and 4) Asian identities taken from the Chicago Face Database (Ma, Correll, & Wittenbrink, 2015) that were rated as neutral on the trustworthiness dimension. Second, the averaged neutral face was morphed towards the trustworthy and untrustworthy averaged references obtained from the three most trustworthy and the three most untrustworthy Asian identities available in the Database by 2 steps, i.e., 30% and 60%, as opposed to the 3 steps used in Study 2, because the face resulting from the 100% morphing looked very distorted. As for the Caucasian continuum, we selected a subset of 5 faces from those used in Study 2 by excluding the two trustworthiness extremes (+ 100% and – 100%) of the original 7-step continuum (see Figure 3.1).

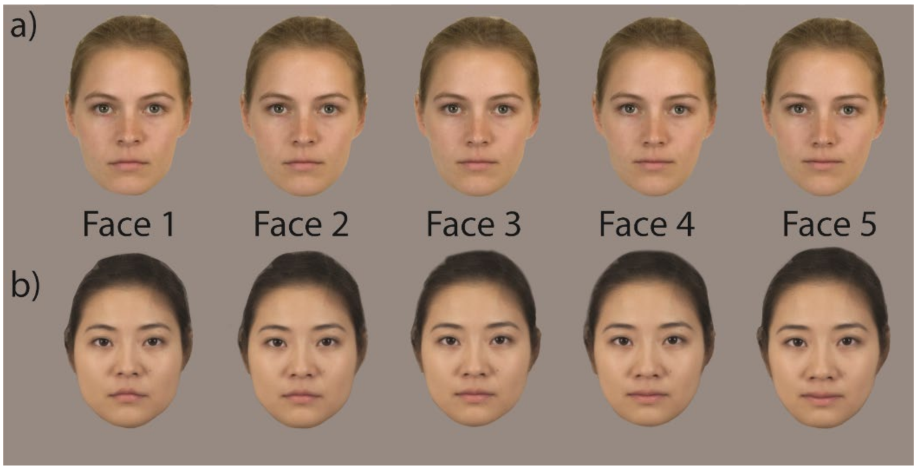


*Figure 3.1. The Caucasian (a) and the Asian (b) trustworthiness continuum, each composed of five variations of one single female face identity where Face 3 is the average of neutral faces, Face 2 and 1 are the 30% and 60% morphs of the same face towards the averaged untrustworthy extreme, and Face 4 and 5 are the 30% and 60% morphs of the same face towards the averaged trustworthy extreme.*

Apparatus and Procedure

Participants were asked to provide their ratings through an online Questionnaire delivered on Qualtrics (Qualtrics, Provo, Utah, USA, https://www.qualtrics.com).

Participants in both groups were instructed to rate perceived trustworthiness for each Caucasian and Asian face on a scale ranging from 1 ("I wouldn't trust this person at all") to 9 ("I would definitely trust this person"). Face ethnicity was blocked across trials, and the order in which trial blocks were presented was counterbalanced across participants.

After providing their trustworthiness ratings, Japanese participants were also asked to rate each face on the dimensions of on perceived emotion, typicality and attractiveness. For each face, participants were asked if they considered the face as emotional, typical and attractive, with the three questions presented in random order. As in Study 2, if participants responded that the face was expressing an emotion, they were asked to indicate which emotion (happy, angry, sad, scared or other) and to rate the intensity of the perceived 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 perceived the face to be typical or attractive, they were asked to rate how much typical or attractive it was on a 1-to-9 scale. In case of negative answers (e.g., if participants perceived the face as emotionally neutral, not typical or not attractive), scores were treated as zero values.

**Results**

*Validation of the Asian trustworthiness continuum in the Japanese population.* In order to validate the Asian trustworthiness continuum created through the morphing procedure we entered the mean scores resulting from the trustworthiness ratings provided for each face of the continuum by the Japanese participants into a repeated-measures ANOVA with trustworthiness intensity as the within-subject factor. The ANOVA attained statistical significance, $F(4,224) = 11.483$, $p < .001$, $p\eta^2 = 0.170$, and the effect was followed up through a test of within-subjects contrasts, which proved significant, $F(1,56) = 23.428$, $p < .001$, $p\eta^2 = 0.295$. The presence of a significant linear trend in participants' ratings indicated that the higher the position along the trustworthiness continuum, the higher the trustworthiness scores attributed by the participants.

To explore whether the participants' average rating of the five Asian faces along the trustworthiness dimension differed from the average ratings of the faces on the emotional, the typicality or the attractiveness dimension, six independent *t*-tests (i.e., one for each emotion - happiness, angriness, sadness and fear - plus typicality and attractiveness judgements) were performed (Bonferroni-corrected threshold: $p < .007$). The comparison attained significance for the emotional and the attractiveness dimensions, showing that, on average, the faces were judged as more trustworthy than emotional (all *ps* $< .001$) or attractive ($p < .003$), suggesting that trustworthiness judgments were more promptly available than judgments about the other targeted dimensions. On the other hand, average ratings of perceived trustworthiness did not differ from the average ratings of perceived typicality, $t(8) = 2.22$, $p = .022$. To further explore these effects we correlated participants' ratings of each single face of the continuum on the trustworthiness dimension with the ratings provided on the other dimensions (Bonferroni-corrected threshold: $p < .007$). Trustworthiness ratings did not correlate with other dimensions (all *ps* > .040), except for an almost significant correlation with attractiveness ratings, $r = .951$, $p = .013$, 95% confidence interval (CI) [.425, .997].

*Cross-cultural comparison.* In order to explore whether ratings of trustworthiness intensity varied as a function of the participants' cultural background and the ethnicity of the face to be rated, the scores resulting from the trustworthiness ratings provided by the Japanese and the Italian participants were entered into a repeated-measures ANOVA with trustworthiness level (1, 2, 3, 4, 5) and face ethnicity (Caucasian, Asian) as within-subjects factors, and culture (Italian, Japanese) as the between-subjects factor. The analysis revealed a main effect of trustworthiness level, $F(4, 432) = 30.190$, $p < .001$, $p\eta^2 = .218$, which was followed up with a test of within-subjects contrasts that revealed a significant linear trend in participants' ratings, $F(1,108) = 75.059$, $p < .001$, $p\eta^2 = 0.410$, indicating that the higher the position along the trustworthiness continuum, the higher the trustworthiness scores attributed by the participants. No main effect or interactions involving face ethnicity and/or culture were observed.

**Discussion**

Study 5 aimed at validating a set of stimuli to be used to test cross-cultural differences in sensitivity to facial cues to trustworthiness in children.

The validation procedure adopted to test whether the faces composing the Asian trustworthiness continuum evoked trustworthiness judgments according to their position along the continuum in Japanese participants confirmed that this was the case. Indeed, since Japanese adults are supposedly maximally familiar with own-race Asian faces, we predicted that, if the stimuli we created truly reflected a fine-grained continuum of trustworthiness, the scores resulting from trustworthiness ratings would have followed a linear trend, with lower scores attributed to Face 1 (Low Trustworthy) and higher scores attributed to Face 5 (High Trustworthy). The analysis on the Trustworthiness scores provided by the Japanese participants confirmed our prediction. Moreover, the Japanese participants perceived the Asian faces as more intensively expressing trustworthiness than emotions or attractiveness, suggesting that trustworthiness judgments were more promptly available in our participants than judgments about the other targeted dimensions. Although ratings of trustworthiness did not differ from typicality ratings, correlational analyses performed on the ratings provided on each single face of the continuum failed to show significant associations between participants' perception of trustworthiness intensity and their perception of how much typical the face appeared to be, whereas the association was almost significant for attractiveness ratings. Overall, we believe these findings are in line with previous demonstrations that, in Caucasian adults, typicality and attractiveness are important cues to face trustworthiness perception (Hu, Abbasi, Zhang, & Chen, 2018; Oosterhof & Todorov, 2008; Schmidt, Leventsten, & Ambadar, 2012).

These results are in line with those obtained from the validation procedure performed for the Caucasian trustworthiness continuum under own-race testing conditions in Study 2, with the only exception being the correlation between trustworthiness ratings and angriness ratings, which was significant for Caucasian faces but not for Asian faces. Indeed, the angriness ratings provided by Caucasian participants for each of the Caucasian faces in Study 2 varied according to trustworthiness intensity, as the less trustworthy faces of the continuum were scarcely perceived as expressing negatively valenced emotions. This might suggest that Asians do not use emotional cues to angriness to judge face trustworthiness. On the other hand, all faces from the Asian continuum, as well as all faces from the Caucasian continuum, were perceived as being equally typical.

The comparison between Italian and Japanese's trustworthiness judgments from own-race and other-race faces allowed to test whether there are cross-cultural differences in sensitivity to fine-grained differences in facial cues to trustworthiness of truthful facial identities, namely to test whether sensitivity to facial cues to trustworthiness and to make trustworthiness judgements based on such cues is a cross-cultural phenomenon and/or is influenced from perceptual expertise with own-race faces.

Caucasian (Italian) and Asian (Japanese) adults tended to attribute trustworthiness judgements in a similar fashion and independently of face ethnicity. The analysis revealed that trustworthiness intensity was perceived to linearly increase along the continuum independently of face ethnicity (Caucasian, Asian) or cultural group (Italian, Japanese). It is well-known that other-race faces are recognized with less accuracy and with slower reaction times, as the ability to recognize a face is subject to an other-race effect (e.g., Ge et al., 2009). The fact that trustworthiness perception based on facial cues was not subject to this effect suggests that, contrary to identity recognition, such ability might be cross-cultural, as it is not influenced by the experience an individual gains with a certain face category. Nevertheless, future studies should test whether differences across cultures are observed when increasing the difficulty of the task, such as when measuring the accuracy of discrimination or reaction times in discrimination tasks where subjects need to disentangle between fine-grained differences of different levels of trustworthiness.

The present stimuli will be used to assess cross-cultural differences of sensitivity to facial cues to trustworthiness from a developmental perspective. The study will involve a group of preschoolers, a group of school-aged children and a group of adults of Italian and Japanese origins. Subjects will be asked to perform a simultaneous matching-to-sample task and an identity recognition task separately for the two face ethnicities. These tasks will be used to assess sensitivity to fine-grained differences of trustworthiness levels, taking into account the experience related to culture and age accumulated by each individual. The comparison between accuracy of discrimination or reaction times for each level of trustworthiness of the two face ethnicities across cultural groups, and/or age groups, would add further evidence as to whether facial cues to trustworthiness transcend cultural differences, maybe as an effect of the adaptive relevance of this facial information.

# General Discussion

The present thesis aimed at investigating the ontogeny of the sensitivity to fine-grained differences in facial cues to trustworthiness, i.e., those cues that are used to infer whether a person is approaching us safely or hostilely. As specified in the Introduction of this thesis, the term *trustworthiness* was not meant as a stable feature of a single face identity or morphology, but rather as the combination of those perceptual cues of the face (Action Units - AUs) that, independently of identity, are crucial in influencing a person's trustworthiness (approach/avoidance) judgement. Therefore, stimuli used in all the presented studies (both Todorov's stimuli and our stimuli) were always created using a data-driven approach, which allows to extrapolate those facial cues that are crucial in driving a trustworthiness judgement, without having to specify a-priori which facial AUs should vary. These cues were used to morph an averaged neutral face identity to obtain a trustworthiness continuum where only slight differences differentiated one face from another.

This thesis aimed at addressing three main questions: (1) Are there individual differences related to social personality that influence fine-grained sensitivity to trustworthiness facial cues in the adult population? (Study 1). (2) What is the developmental trajectory of this sensitivity? (Study 2, 3, 4). (3) Are there cross-cultural differences in the attribution of trustworthiness judgements? (Study 5).

Evidence coming from Study 1 suggests that individual differences in fine-grained perceptual sensitivity and mental representation of facial features related to trustworthiness judgements are shaped by individual differences in social motivation. In this study, we focused on the extraversion dimension as individuals with high extraversion scores are known to spend more time with other people, which might in turn specialize their experience with faces during development (Li et al., 2010). Subjects with higher extraversion scores were faster at successfully disentangling between fine-grained differences in the level of trustworthiness expressed by faces. Nevertheless, response accuracy or the way trustworthiness facial cues are represented in the long-term memory were not shaped by individual differences in the extraversion dimension.

Study 2 further explored the hypothesis that differential ontogenetic experience in the social domain can affect perceptual sensitivity to face trustworthiness by focusing on the developmental trajectory of such sensitivity. More specifically, the study focused on how perceptual sensitivity and mental representation of fine-grained differences in facial information subtending social perception of trustworthiness develops across childhood, as a function of differences in emotional development. Results showed that, by the age of 5 years, children already represent faces according to the intensity of displayed trustworthiness, but this representation becomes increasingly fine-grained across development, reaching adult-like levels of performance by age 7. Moreover, even though children aged 5 proved to perform less accurately than adults in attributing trustworthiness judgements and tended to overestimate trustworthiness levels of untrustworthy faces, their explicit trustworthiness judgements increased linearly along the continuum, just like adults' judgments. Finally, at the age of 5, accuracy in generating trustworthiness judgements was associated to greater emotion comprehension skills. All data converged in indicating that sensitivity to facial cues to trustworthiness and the ability to employ these cues to generate trustworthiness judgements is present in preschool years, and reaches adult-like levels at the age of 7, developing together with emotion understanding abilities.

In light of the evidence provided by Study 2 that preschoolers are already sensitive to those facial cues that drives trustworthiness judgements in adults, Study 3 and 4 explored whether this sensitivity is already apparent in the first year of life. Study 3 used the same ERP and preferential looking paradigms employed by Jessen and Grossmann (2016) to explore infants' neural and behavioral sensitivity to trustworthiness cues in natural, female faces. A subset of the stimuli composing the trustworthiness continuum used in Study 2 (a Low, a Neutral and a High Trustworthy face) was used to test a groups of adults in order to check whether they elicited the same neural responses evoked by the highly validated set of stimuli used in previous studies (Marzi et al., 2014; Jessen & Grossmann, 2016). Data from the adults' sample showed that faces evoke different responses based on the intensity of the cues to trustworthiness they include already in the early structural encoding stage of processing, and confirmed previous demonstration of adults' enhanced attentional response to Low Trustworthy faces (Lischke, Junge, Hamm, & Weymar, 2018; Marzi, Righi, Ottonello, Cincotta, & Viggiano, 2014; Lischke et al., 2018; Yang, Qi, Ding, & Song, 2011).

We observed a modulation of neural activity in response to our stimuli also in 7-months-old infants. Indeed, both P400 and Nc peaked faster in response of the High Trustworthy face with respect to the Neutral face in the left hemisphere. Moreover, infant data showed that individual differences in temperamental traits impact neural sensitivity to variations in physical cues to trustworthiness. Indeed, infants scoring high in Surgency allocated faster attention to the Low Trustworthy face with respect to the High Trustworthy face, and those scoring high in Negative Affect were faster in allocating attention to the High Trustworthy face with respect to the Neutral face. Behavioral data derived from the preferential looking paradigm were not in line with ERP data, as infants did not show an attentional preference for any of the trustworthiness levels. Neural and behavioral incongruencies with Jessen and Grossman's study were discussed as an effect of differences between the stimuli used in the two studies.

In light of the differences between results of Study 3 and those previously obtained by Jessen and Grossman, Study 4 focused on the trustworthiness extremes (Low and High trustworthy faces) to further test whether infants can discriminate between these two trustworthiness levels by using a new, infants' friendly EEG visual discrimination paradigm, namely the Fast Periodic Visual Stimulation paradigm (FPVS, see Norcia et al., 2015). The FPVS paradigm consists in a fast presentation where the stimuli of interest are repeatedly displayed at a fixed periodic frequency rate, which leads to a periodic change in voltage amplitude of the electroencephalographic signal at the same frequency of stimulation. A group of adults was also tested using the same paradigm in order to validate the stimuli and the procedure. Adults' data showed that the activity in the occipital regions oscillated at the frequency of presentation of faces that vary in the level of expressed trustworthiness, independently of the condition (i.e. whether the Oddball is a High or a Low Trustworthy face). More importantly, the fact that SNR of the Oddball frequency was significantly higher than noise in both conditions confirms adults' ability to discriminate between the two levels of trustworthiness.

The infants' data confirmed Study 3 evidence showing that 7-month-olds are already sensitive to those facial cues that drive trustworthiness judgements in adults. Like the adults' sample, for infants as well EEG activity in the occipital regions oscillated at the frequency of presentation of faces that varied in the level of expressed trustworthiness. Analyses on the Oddball frequency revealed that infants discriminated the Low Trustworthy faces from the High trustworthy faces in the right occipital region for the condition where the Low Trustworthy face was Oddball. Nevertheless, no evidence of discrimination was observed for the condition where the Oddball face was High Trustworthy. Differences in the results from the two conditions were interpreted in light of the evidence coming from visual search studies (e.g., Hansen & Hansen, 1988; LoBue, 2009; Mather & Knight, 2006; Öhman, Lundqvist, & Esteves, 2001). Indeed, these studies, which involved both adults and children, report a search asymmetry in favour of thread-related stimuli such as angry faces, when presented within happy or neutral face distractors. This asymmetry is interpreted as an effect of adaptive mechanisms that would stem to promptly respond to threat-related environmental stimuli.

It is important to note that the pattern of results emerged from Study 3 and 4 were different. Indeed, the ERP study revealed overall faster allocation of attention in favor of the High Trustworthy face, localized in the left occipital and prefrontal areas, which was interpreted as an adaptive reaction related to fast processing of pro-social stimuli (i.e., stimuli that guarantee safety). On the other hand, Study 4 provided evidence of visual discrimination only in the condition where the Low Trustworthy face was the Oddball, which can be thought as a processing advantage in favor of thread-related stimuli. Moreover, Study 3 showed a modulatory effect of Surgency scores on differential P400 latencies for the High and Low Trustworthy faces, while Surgency scores did not prove to modulate discrimination of the same face categories in Study 4.

As discussed in Chapter 2, a possible explanation for the observed discrepancies between the results of Study 3 and Study 4 might lie in the different nature of the two paradigms (ERP and FPVS Oddball paradigm). In fact, the ERP paradigm measured the effort and the speed with which faces varying in the level of expressed trustworthiness were processed, while the FPVS paradigm measured the ability to discriminate between different levels of trustworthiness. Following from these considerations, differences in Surgency scores might affect the time of processing, but not the ability to process differences among the two face categories, just like variability in the Extraversion dimension in adults affected response times, but not discrimination accuracy in a perceptual sensitivity task addressing facial trustworthiness (Study 1).

Notwithstanding the discussed differences between the results of Study 3 and 4, data from both studies show that 6-month-old infants are sensitive to those facial cues that, later in development, are used to generate trustworthiness judgements.

Finally, Chapter 3 presented a validation of stimuli that will be used to explore the presence of developmental cross-cultural differences in the perception of face trustworthiness. A 5-steps continuum of female Asian faces was created to be used with a sub-sample of our Caucasian 7-faces as experimental stimuli to acquire trustworthiness judgements of a group of Italian and a group of Japanese adults. First, the analysis on the trustworthiness scores of the Japanese sample confirmed the prediction that stimuli reflected a fine-grained continuum of trustworthiness, as the intensity of the scores followed a linear trend, with lower scores attributed to Face 1 (Low Trustworthy) and higher scores attributed to Face 5 (High Trustworthy).

Second, the cross-cultural analysis showed that Caucasian (Italian) and Asian (Japanese) adults tended to attribute trustworthiness judgements in a similar fashion and independently of face ethnicity (Caucasian, Asian). This result would suggest that trustworthiness perception based on facial cues is cross-cultural, as it is not influenced by the experience an individual gains with a certain face category. Nevertheless, future studies should test whether differences across cultures are observed when increasing the difficulty of the task, such as when measuring the accuracy of discrimination or reaction times in discrimination tasks where subjects need to disentangle between fine-grained differences between different levels of trustworthiness.

The studies presented in this thesis have a number of limits. First, all studies used as stimulus material one or maximum two female face identities, and this might have hindered the generalizability of the results. There are three main reasons behind the choice of using such stimuli. First, the use of one face identity allowed us to measure fine-grained sensitivity to facial cues of trustworthiness, as subjects had to evaluate faces that only slightly differed from one another. The use of an averaged identity was thought to reduce the influence of idiosyncratic facial characteristics on both perceptual sensitivity to trustworthiness cues and explicit trustworthiness judgements. Second, using more than one face identity in Study 1 and 2 would have led to an excessive number of trials to guarantee all pairwise comparisons among intensities. Third, as reported for stimuli description of Study 5, the morphing towards trustworthiness references is not always successful, as faces tend to easily distort. Therefore, creating two trustworthiness continuums of the same facial ethnicity proved to be impossible. Moreover, we always used a female face as our goal was to measure fine-grained sensitivity to a face type with which young children are highly experienced with (i.e., female faces, see Quinn, Yahr, Kuhn, Slater, & Pascalils, 2002; Ramsey-Rennels & Langlois, 2006). Future studies shall test whether our results generalize to other face identities, including truthful male faces created using the same procedure. Moreover, by enhancing stimulus variability the tasks would become more interesting especially for infants participants, and the stimulus set more ecologically valid.

A second limit of this thesis concerns the data-driven approach used to create the experimental stimuli. This method has the huge advantage of not having to specify a-priori which facial characteristics should vary to express different levels of trustworthiness. Nevertheless, in all studies we could not systematically identify the role of different facial cues in driving trustworthiness judgements. Future studies should put more effort in detecting the unique configural set of facial characteristics that drives these judgements, independently of face identity (as first attempted by Gill et al., 2014).

In conclusion, results from studies 1-5 provide evidence that sensitivity to facial cues to trustworthiness manifests in the very first year of life, to be then refined by experience over the course of development. Results suggest that differential ontogenetic experience determined by individual differences in social attitudes, emotion recognition abilities and temperamental traits may influence fine-grained sensitivity to trustworthiness facial cues. Moreover, the attribution of trustworthiness judgements from facial cues of different ethnicities might not be determined by the cultural context where an individual grows and develops. Future studies should examine this latter aspect in depth.

# References


Adams, R. B., Albohn, D., & Kveraga, K. (2017). Social Vision: Applying a Social-Functional Approach to Face and Expression Perception. *Current Directions In Psychological Science*. https://doi.org/10.1177/0963721417706392

Adams, Reginald B, Ambady, N., Neil Macrae, C., & Kleck, R. (2006). Emotional expressions forecast approach-avoidance behavior. *Motivation and Emotion*, *30*(2), 177–186. https://doi.org/10.1007/s11031-006-9020-2

Adams, Reginald B, Nelson, A. J., Soto, J. A., Hess, U., & Kleck, R. E. (2012). Emotion in the neutral face: a mechanism for impression formation? *Cognition & Emotion*, *26*(3), 431–441. https://doi.org/10.1080/02699931.2012.666502

Adolphs, R, Sears, L., & Piven, J. (2001). Abnormal processing of social information from faces in autism. *Journal Of Cognitive Neuroscience*, *13*(2), 232–240. Retrieved from http://dx.doi.org/10.1162/089892901564289

Adolphs, Ralph. (2002). Recognizing Emotion From Facial Expressions: Psychological and Neurological Mechanisms. *Behavioral and Cognitive Neuroscience Reviews*, *1*(1), 21–62. https://doi.org/10.1177/1534582302001001003

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

Ballew, C. C. I. I., & Todorov, A. (2007). Predicting Political Elections from Rapid and Unreflective Face Judgments. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(46), 17948. https://doi.org/10.1073/pnas.0705435104

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

Barry-Anwar, R., Hadley, H., Conte, S., Keil, A., & Scott, L. S. (2018). The developmental time course and topographic distribution of individual-level monkey face discrimination in the infant brain. *Neuropsychologia*, *108*, 25–31. https://doi.org/10.1016/j.neuropsychologia.2017.11.019

Bellugi, U., Adolphs, R., Cassady, C., & Chiles, M. (1999). Towards the neural basis for hypersociability in a genetic syndrome. *Neuroreport*, *10*(8), 1653–1657. https://doi.org/http://dx.doi.org/10.1097/00001756-199906030-00006

Biehl, M., Matsumoto, D., Ekman, P., Hearn, V., Heider, K., Kudoh, T., & Ton, V. (1997). Matsumoto and Ekman's Japanese and Caucasian Facial Expressions of Emotion (JACFEE): Reliability Data and Cross-National Differences. *Journal of Nonverbal Behavior*, *21*, 3–21. https://doi.org/10.1023/A:1024902500935

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/10.0.4.87/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*, 56–61. https://doi.org/10.1016/j.paid.2014.05.012

Bond, J., Charles, F., Berry, D. S., & Omar, A. (1994). The Kernel of Truth in Judgments of Deceptiveness. *Basic & Applied Social Psychology*, *15*(4), 523–534. Retrieved from 10.0.4.183/s15324834basp1504_8

Bond, J., Charles, F., & DePaulo, B. M. (2006). Accuracy of Deception Judgments: Appendix B. *Personality & Social Psychology Review (Lawrence Erlbaum Associates)*, *10*(3), 1. https://doi.org/10.1207/s15327957pspr1003_2B

Bothwell, R. K., Brigham, J. C., & Malpass, R. S. (1989). Cross-Racial Identification. *Personality & Social Psychology Bulletin*, *15*, 19–25. https://doi.org/10.1177/0146167289151002

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

Caprara, G. V., Barbaranelli, C., Borgogni, L., & Perugini, M. (1993). The "big five questionnaire": A new questionnaire to assess the five factor model. *Personality and Individual Differences*, *15*(3), 281–288. https://doi.org/https://doi.org/10.1016/0191-8869(93)90218-R

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

Caulfield, F., Ewing, L., Burton, N., Avard, E., & Rhodes, G. (2014). Facial trustworthiness judgments in children with ASD are modulated by happy and angry emotional cues. *Plos One*, *9*(5), e97644–e97644. https://doi.org/10.1371/journal.pone.0097644

Chang, L. J., Doll, B. B., van't Wout, M., Frank, M. J., & Sanfey, A. G. (2010). Seeing is believing: Trustworthiness as a dynamic belief. *Cognitive Psychology*, *61*(2), 87–105. https://doi.org/10.1016/j.cogpsych.2010.03.001

Cheung, C. H. M., Rutherford, H. J. V, Mayes, L. C., & McPartland, J. C. (2010). Neural responses to faces reflect social personality traits. *Social Neuroscience*, *5*(4), 351. https://doi.org/doi:10.1080/17470911003597377

Chiroro, P., & Valentine, T. (1995). An Investigation of the Contact Hypothesis of the Own-race Bias in Face Recognition. *The Quarterly Journal of Experimental Psychology Section A*, *48*, 879–894. https://doi.org/10.1080/14640749508401421

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

Davidson, R. (1983). Hemispheric specialization for cognition and affect. In A. Gale & J. Edwards (Eds.), *Physiological Correlates of Human Behaviour* (pp. 203–216). London: Academic Press.

Dawson, G., Webb, S. J., & McPartland, J. (2005). Understanding the nature of face processing impairment in autism: insights from behavioral and electrophysiological studies. *Developmental Neuropsychology*, *27*(3), 403–424. https://doi.org/https://doi.org/10.1207/s15326942dn2703_6

de Gelder, B., Vroomen, J., Pourtois, G., & Weiskrantz, L. (1999). Non-conscious recognition of affect in the absence of striate cortex. *Neuroreport*. https://doi.org/10.1097/00001756-199912160-00007

de Haan, M., Belsky, J., Reid, V., Volein, A., & Johnson, M. H. (2004). Maternal personality and infants' neural and visual responsivity to facial expressions of emotion. *Journal Of Child Psychology And Psychiatry, And Allied Disciplines*, *45*(7), 1209–1218. https://doi.org/https://doi.org/10.1111/j.1469-7610.2004.00320.x

de Haan, M., Johnson, M. H., & Halit, H. (2003). Development of face-sensitive event-related potentials during infancy: a review. *International Journal of Psychophysiology*, *51*(1), 45–58. https://doi.org/10.1016/S0167-8760(03)00152-1

de Heering, A., & Rossion, B. (2015). Rapid categorization of natural face images in the infant right hemisphere. *Elife*, *4*, e06564–e06564. https://doi.org/10.7554/eLife.06564

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

Delgado, M., Frank, R., & Phelps, E. (2005). Perceptions of moral character modulate the neural systems of reward during the trust game. *Nature Neuroscience*. https://doi.org/10.1038/nn1575

Demaree, H., Everhart, D., Youngstrom, E., & Harrison, D. (2005). Brain lateralization of emotional processing: historical roots and a future incorporating "dominance." *Behavioral and Cognitive Neuroscience Reviews*, *4*, 3–20. https://doi.org/10.1177/1534582305276837

Dennett, H. W., McKone, E., Edwards, M., & Susilo, T. (2012). Face Aftereffects Predict Individual Differences in Face Recognition Ability. *Psychological Science*, *23*(11), 1279. https://doi.org/10.1177/0956797612446350

Dimberg, U., & Thunberg, M. (2007). Speech anxiety and rapid emotional reactions to angry and happy facial expressions. *Scandinavian Journal of Psychology*, *48*(4), 321–328. https://doi.org/10.1111/j.1467-9450.2007.00586.x

Dotsch, R., & Todorov, A. (2012). Reverse Correlating Social Face Perception. *Social Psychological and Personality Science*. https://doi.org/10.1177/1948550611430272

Dzhelyova, M., Perrett, D. I., & Jentzsch, I. (2012). Temporal dynamics of trustworthiness perception. *Brain Research*, *1435*, 81–90. https://doi.org/10.1016/j.brainres.2011.11.043

Edelman, S. (1998). Representation is representation of similarities. *The Behavioral And Brain Sciences*, *21*(4), 449–467.

Ekman, P, WV, F., O'Sullivan, M., Chan, A., Diacoyanni-Tarlatzis, I., Heider, K., … PE, R.-B. (1987). Universals and cultural differences in the judgments of facial expressions of emotion. *Journal Of Personality And Social Psychology*. https://doi.org/10.1037/0022-3514.53.4.712

Ekman, Paul, & Oster, H. (1979). Facial Expressions of Emotion. *Annual Review of Psychology*, *30*(1), 527. https://doi.org/10.0.4.122/annurev.ps.30.020179.002523

Ekman, Paul, Sorenson, E. R., & Friesen, W. V. (1969). Pan-Cultural Elements in Facial Displays of Emotion. *Science*, *164*(3875), 86–88. https://doi.org/10.1126/science.164.3875.86

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

Fantz, R. L. (1958). Pattern vision in young infants. *The Psychological Record*, *8*, 43–47. https://doi.org/10.1126/science.140.3564.296

Farina, E., & Belacchi, C. (2014). The Relationship between Emotional Competence and Hostile/Prosocial Behavior in Albanian Preschoolers: An Exploratory Study. *School Psychology International*, *35*(5), 475–484. https://doi.org/10.1177/0143034313511011

Farroni, T., Menon, E., Rigato, S., & Johnson, M. (2007). The perception of facial expressions in newborns. *European Journal of Developmental Psychology*, *4*(1), 2–13. https://doi.org/10.1080/17405620601046832

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

Fenske, M. J., Raymond, J. E., Kessler, K., Westoby, N., & Tipper, S. P. (2005). Attentional Inhibition Has Social-Emotional Consequences for Unfamiliar Faces. *Psychological Science*, *16*(10), 753. https://doi.org/10.1111/j.1467-9280.2005.01609.x

Ferrari, C., Lega, C., Vernice, M., Tamietto, M., Mende-Siedlecki, P., Vecchi, T., … Cattaneo, Z. (2016). The Dorsomedial Prefrontal Cortex Plays a Causal Role in Integrating Social Impressions from Faces and Verbal Descriptions. *Cerebral Cortex (New York, N.Y.: 1991)*. https://doi.org/10.1093/cercor/bhu186

Fink, A. (2005). Event-related desynchronization in the EEG during emotional and cognitive information processing: differential effects of extraversion. *Biological Psychology*, *70*(3), 152–160. https://doi.org/http://dx.doi.org/10.1016/j.biopsycho.2005.01.013

Fiske, S. T., Cuddy, A. J. C., & Glick, P. (2007). Universal dimensions of social cognition:

warmth and competence. *Trends In Cognitive Sciences*, *11*(2), 77–83. https://doi.org/https://doi.org/10.1016/j.tics.2006.11.005

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

Fox, E., Lester, V., Russo, R., Bowles, R. J., Pichler, A., & Dutton, K. (2000). Facial Expressions of Emotion: Are Angry Faces Detected More Efficiently? *Cognition & Emotion*, *14*(1), 61–92. https://doi.org/10.1080/026999300378996

Freeman, J. B., & Ambady, N. (2010). MouseTracker: software for studying real-time mental processing using a computer mouse-tracking method. *Behavior Research Methods*, *42*(1), 226–241. https://doi.org/10.3758/BRM.42.1.226

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://10.0.3.248/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

Gartstein, M. A., & Rothbart, M. K. (2003). Studying infant temperament via the Revised Infant Behavior Questionnaire. *Infant Behavior and Development*, *26*(1), 64–86. https://doi.org/10.1016/S0163-6383(02)00169-8

Ge, L., Zhang, H., Wang, Z., Quinn, P. C., Pascalis, O., Kelly, D., … Lee, K. (2009). Two Faces of the Other-Race Effect: Recognition and Categorisation of Caucasian and Chinese Faces. *Perception*, *38*, 1199–1210. https://doi.org/10.1068/p6136

Gepner, B., de Gelder, B., & Schonen, S. (1996). Face processing in autistics: Evidence for a generalised deficit? *Child Neuropsychology*, *2*(2), 123–139. https://doi.org/10.1080/09297049608401357

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

Gredebäck, G., Kaduk, K., Bakker, M., Gottwald, J., Ekberg, T., Elsner, C., … Kenward, B. (2015). The neuropsychology of infants' pro-social preferences. *Developmental Cognitive Neuroscience*, *12*, 106–113. https://doi.org/10.1016/j.dcn.2015.01.006

Grelotti, D. J., Gauthier, I., & Schultz, R. T. (2002). Social interest and the development of cortical face specialization: What autism teaches us about face processing. *Developmental Psychobiology*, *40*(3), 213–225. https://doi.org/https://doi.org/10.1002/dev.10028

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

Gutiérrez-García, A., & Calvo, M. G. (2017). Social anxiety and threat-related interpretation of dynamic facial expressions: Sensitivity and response bias. *Personality and Individual Differences*, *107*, 10–16. https://doi.org/10.1016/j.paid.2016.11.025

Halit, H., de Haan, M., & Johnson, M. H. (2003). Cortical specialisation for face processing: face-sensitive event-related potential components in 3- and 12-month-old infants. *Neuroimage*, *19*(3), 1180–1193. https://doi.org/10.1016/S1053-8119(03)00076-4

Hamlin, J. K., & Wynn, K. (2011). Young infants prefer prosocial to antisocial others. *Cognitive Development*, *26*, 30–39. https://doi.org/10.0.3.248/j.cogdev.2010.09.001

Hansen, C. H., & Hansen, R. D. (1988). Finding the Face in the Crowd: An Anger

Superiority Effect. *Journal of Personality and Social Psychology*, *54*, 917–924. https://doi.org/10.1037/0022-3514.54.6.917

Harris, P. L. (2007). Trust. *Developmental Science*, *10*(1), 135–138. https://doi.org/10.1111/j.1467-7687.2007.00575.x

Harry, B., Williams, M. A., Davis, C., & Kim, J. (2013). Emotional expressions evoke a differential response in the fusiform face area. *Frontiers in Human Neuroscience*, *7*. https://doi.org/10.3389/fnhum.2013.00692

Haxby, J. V, Gobbini, M. I., Rhodes book editor, G., Calder book editor, A., Johnson book editor, M., & Haxby book editor, J. V. (2011). *Distributed Neural Systems for Face Perception*. *Oxford Handbook of Face Perception:* https://doi.org/10.1093/oxfordhb/9780199559053.013.0006

Hehman, E., Flake, J., & Freeman, J. (2015). Static and Dynamic Facial Cues Differentially Affect the Consistency of Social Evaluations. *Personality & Social Psychology Bulletin*. https://doi.org/10.1177/0146167215591495

Hehman, E., Flake, J. K., Xie, S. Y., Stolier, R. M., & Freeman, J. B. (2019). Toward a comprehensive model of face impressions: What we know, what we do not, and paths forward. *Social and Personality Psychology Compass*. https://doi.org/10.1111/spc3.12431

Hershler, O., & Hochstein, S. (2005). At first sight: a high-level pop out effect for faces. *Vision Research*. https://doi.org/10.1016/j.visres.2004.12.021

Heyman, G. D. (2008). Children's Critical Thinking When Learning From Others. *Current Directions In Psychological Science*, *17*(5), 344–347. https://doi.org/10.1111/j.1467-8721.2008.00603.x

Hilimire, M. R., Mienaltowski, A., Blanchard-Fields, F., & Corballis, P. M. (2013). Age-related differences in event-related potentials for early visual processing of emotional faces. *Social Cognitive And Affective Neuroscience*. https://doi.org/10.1093/scan/nst071

Hirai, M., Muramatsu, Y., Mizuno, S., Kurahashi, N., Kurahashi, H., & Nakamura, M. (2017). Preserved search asymmetry in the detection of fearful faces among neutral faces in individuals with Williams syndrome revealed by measurement of both manual responses and eye tracking. *Journal of Neurodevelopmental Disorders*, *9*, 8. https://doi.org/10.1186/s11689-017-9190-0

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

Holmes, A., Green, S., & Vuilleumier, P. (2005). The involvement of distinct visual channels in rapid attention towards fearful facial expressions. *Cognition & Emotion*, *19*(6), 899–922. https://doi.org/http://10.0.4.56/02699930441000454

Hoshiyama, M., Kakigi, R., Watanabe, S., Miki, K., & Takeshima, Y. (2003). Brain responses for the subconscious recognition of faces. *Neuroscience Research*, *46*(4), 435–442. https://doi.org/10.1016/S0168-0102(03)00121-4

Hu, Y., Abbasi, N. U. H., Zhang, Y., & Chen, H. (2018). The Effect of Target Sex, Sexual 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, Kate, & Johnson, M. H. (2007). The development of "face-space" in infancy. *Visual Cognition*, *15*(5), 578–598. https://doi.org/10.1080/13506280600943518

Humphreys, KL, Kircanski, K., Colich, N., & Gotlib, I. (2016). Attentional avoidance of fearful facial expressions following early life stress is associated with impaired social functioning. *Journal Of Child Psychology And Psychiatry, And Allied Disciplines*. https://doi.org/10.1111/jcpp.12607

In-Albon, T., Kossowsky, J., & Schneider, S. (2010). Vigilance and Avoidance of Threat in the Eye Movements of Children with Separation Anxiety Disorder. *Journal of Abnormal Child Psychology*, *38*(2), 225–235. https://doi.org/10.1007/s10802-009-9359-4

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

Jacques, J., Corentin, J., Joan, L.-S., Hélène, B., Sophie, C.-C., Louis, M., & Bruno, R. (2016). A face-selective ventral occipito-temporal map of the human brain with intracerebral potentials. *Proceedings of the National Academy of Sciences of the United States of America*, *113*, E4088. https://doi.org/10.1073/pnas.1522033113

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. (2019). Neural evidence for the impact of facial trustworthiness on object processing in a gaze-cueing task in 7-month-old infants. *Social Neuroscience*, *9*, 1–9. https://doi.org/10.1080/17470919.2019.1651764

Jessen, Sarah, & 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, Sarah, & Grossmann, T. (2017). Neural evidence for the subliminal processing of facial trustworthiness in infancy. *Neuropsychologia*, *126*, 46–53. https://doi.org/10.1016/j.neuropsychologia.2017.04.025

Johnson, M.H., Dziurawiec, S., Ellis, H., & Morton, J. (1991). Newborns' preferential tracking of face-like stimuli and its subsequent decline. *Cognition*, *40*(1–2), 1–19. https://doi.org/10.1016/0010-0277(91)90045-6

Johnson, Mark H., Senju, A., & Tomalski, P. (2015). The two-process theory of face processing: Modifications based on two decades of data from infants and adults. *Neuroscience and Biobehavioral Reviews*, *50*, 169–179. https://doi.org/10.1016/j.neubiorev.2014.10.009

Jones, A. L., Kramer, R. S. S., & Ward, R. (2012). Signals of personality and health: the contributions of facial shape, skin texture, and viewing angle. *Journal Of Experimental Psychology. Human Perception And Performance*, *38*(6), 1353–1361. https://doi.org/10.1037/a0027078

Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: A module in human extrastriate cortex specialized for face perception. *Journal of Neuroscience*, *17*, 4302–4311. https://doi.org/10.1523/JNEUROSCI.17-11-04302.1997

Karmiloff-Smith, A., Casey, B., Massand, E., Tomalski, P., & Thomas, M. (2014). Environmental and Genetic Influences on Neurocognitive Development: The Importance of Multiple Methodologies and Time-Dependent Intervention. *Clinical Psychological Science: A Journal Of The Association For Psychological Science*. https://doi.org/doi.org/10.1177/2167702614521188

Kårstad, S. B., Kvello, Ø., Wichstrøm, L., & Berg-Nielsen, T. S. (2014). What do parents know about their children’s comprehension of emotions? Accuracy of parental estimates in a community sample of pre-schoolers. *Child: Care, Health & Development*, *40*(3), 346–353. https://doi.org/10.1111/cch.12071

Keefe, B., Dzhelyova, M., Perrett, D., & Barraclough, N. (2013). Adaptation improves face trustworthiness discrimination. *Frontiers In Psychology*. https://doi.org/10.3389/fpsyg.2013.00358

Key, A. P. F., Stone, W., & Wiliams, S. M. (2009). What Do Infants See in Faces? ERP Evidence of Different Roles of Eyes and Mouth for Face Perception in 9-Month-old

Infants. *Infant and Child Development*, *18*(2), 149–162. https://doi.org/10.1002/icd.600

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

Kim, M.-K., Yoon, H.-J., Shin, Y.-B., Lee, S.-K., & Kim, J.-J. (2016). Neural basis of distorted self-face recognition in social anxiety disorder. *NeuroImage: Clinical*, *12*, 956–964. Retrieved from http://10.0.3.248/j.nicl.2016.04.010

King-Casas, B., Tomlin, D., Anen, C., Camerer, C. F., Quartz, S. R., & Montague, P. R. (2005). Getting to Know You: Reputation and Trust in a Two-Person Economic Exchange. *Science*, *308*(5718), 78–83. https://doi.org/10.1038/nn1575

Kirihara, K., Kasai, K., Tada, M., Nagai, T., Kawakubo, Y., Yamasaki, S., … Araki, T. (2012). Neurophysiological impairment in emotional face processing is associated with low extraversion in schizophrenia. *Progress In Neuro-Psychopharmacology & Biological Psychiatry*, *37*(2), 270–275. https://doi.org/10.1016/j.pnpbp.2012.02.012

Klin, A., Sparrow, S. S., de Bildt, A., Cicchetti, D. V, Cohen, D. J., & Volkmar, F. R. (1999). A normed study of face recognition in autism and related disorders. *Journal Of Autism And Developmental Disorders*, *29*(6), 499–508. https://doi.org/https://doi.org/10.1023/A:1022299920240

Kramer, R. S. S., & Ward, R. (2010). Internal facial features are signals of personality and health. *Quarterly Journal of Experimental Psychology*, *63*(11), 2273–2287. https://doi.org/10.1080/17470211003770912

Kriegeskorte, N., & Mur, M. (2012). Inverse MDS: Inferring Dissimilarity Structure from Multiple Item Arrangements. *Frontiers In Psychology*, *3*, 245. https://doi.org/10.3389/fpsyg.2012.00245

Kriegeskorte, N., Mur, M., & Bandettini, P. (2008). Representational similarity analysis -

connecting the branches of systems neuroscience. *Frontiers In Systems Neuroscience*, *2*, 4. https://doi.org/10.3389/neuro.06.004.2008

Kruskal, J. B., & Wish, M. (1978). *Multidimensional Scaling*. Beverly Hills, Calif: SAGE Publications, Inc. https://doi.org/http://dx.doi.org/10.4135/9781412985130

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

Leppänen, J., Cataldo, J., Bosquet Enlow, M., & Nelson, C. A. (2018). Early development of attention to threat-related facial expressions. *Plos One*. https://doi.org/10.1371/journal.pone.0197424

Leppänen, J., Moulson, M. C., Vogel-Farley, V. K., & Nelson, C. A. (2007). An ERP Study of Emotional Face Processing in the Adult and Infant Brain. *Child Development*, *78*(1), 232–245. https://doi.org/10.1111/j.1467-8624.2007.00994.x

Leppänen, J., & Nelson, C. A. (2009). Tuning the developing brain to social signals of emotions. *Nature Reviews. Neuroscience*. https://doi.org/10.1038/nrn2554

Li, J., Tian, M., Fang, H., Xu, M., Li, H., & Liu, J. (2010). Extraversion predicts individual differences in face recognition. *Communicative & Integrative Biology*. https://doi.org/10.4161/cib.3.4.12093

Li, Q., Heyman, G. D., Mei, J., & Lee, K. (2017). Judging a Book by Its Cover: Children's Facial Trustworthiness as Judged by Strangers Predicts Their Real-World Trustworthiness and Peer Relationships. *Child Development*, *90*, 562–575. https://doi.org/10.1111/cdev.12907

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

Liu-Shuang, J., Norcia, A. M., & Rossion, B. (2014). An objective index of individual face discrimination in the right occipito-temporal cortex by means of fast periodic oddball stimulation. *Neuropsychologia*, *52*, 57–72. https://doi.org/10.1016/j.neuropsychologia.2013.10.022

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

Lundqvist, D., Flykt, A., & Öhman, A. (1998). The Karolinska Directed Emotional Faces – KDEF, CD ROM from Department of Clinical Neuroscience, Psychology section, Karolinska Institutet. ISBN 91-630-7164-9.

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

Macchi Cassia, V., Kuefner, D., Westerlund, A., & Nelson, C. A. (2006). A behavioural and ERP investigation of 3-month-olds' face preferences. *Neuropsychologia*, *44*(11), 2113–2125. https://doi.org/10.1016/j.neuropsychologia.2005.11.014

Marsh, A. A., Ambady, N., & Kleck, R. E. (2005). The effects of fear and anger facial expressions on approach- and avoidance-related behaviors. *Emotion (Washington, D.C.)*, *5*(1), 119–124. https://doi.org/10.1037/1528-3542.5.1.119

Martinos, M., Matheson, A., & de Haan, M. (2012). Links between Infant Temperament and Neurophysiological Measures of Attention to Happy and Fearful Faces. *Journal of Child Psychology and Psychiatry*, *53*(11), 1118–1127. https://doi.org/10.1111/j.1469-

7610.2012.02599.x

Marzi, T., Righi, S., Ottonello, S., Cincotta, M., & Viggiano, M. P. (2014). Trust at first sight: evidence from ERPs. *Social Cognitive And Affective Neuroscience*, *9*(1), 63–72. https://doi.org/10.1093/scan/nss102

Mascaro, O., & Csibra, G. (2014). Human infants' learning of social structures: the case of dominance hierarchy. *Psychological Science*, *25*(1), 250–255. https://doi.org/10.1177/0956797613500509

Mather, M., & Knight, M. R. (2006). Angry faces get noticed quickly: Threat detection is not impaired among older adults. *Journals of Gerontology - Series B Psychological Sciences and Social Sciences*, *61*, 54. https://doi.org/10.1093/geronb/61.1.P54

McCullough, M. E., & Reed, L. I. (2016). What the face communicates: clearing the conceptual ground. *Current Opinion in Psychology*, *7*, 110–114. https://doi.org/10.1016/j.copsyc.2015.08.023

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

Meltzoff, A. N., & Moore, M. K. (1977). Imitation of Facial and Manual Gestures by Human Neonates. *Science*, *198*(4312), 75. https://doi.org/10.1126/science.897687

Montoya, L., Westerlund, A., Troller-Renfree, S., Righi, G., & Nelson, C. A. (2017). The effect of heterogeneous race exposure during infancy. *Cognitive Development*, *42*, 74–83. https://doi.org/10.1016/j.cogdev.2017.02.014

Morton, J., & Johnson, M. H. (1991). CONSPEC and CONLERN: a two-process theory of infant face recognition. *Psychological Review*, *98*(2), 164–181. https://doi.org/10.1037/0033-295X.98.2.164

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., Westerlund, A., McDermott, J., Zeanah, C., & Fox, N. (2013). Emotion recognition following early psychosocial deprivation. *Development And Psychopathology*. https://doi.org/10.1017/S0954579412001216

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://10.0.3.248/j.jecp.2009.02.005

Norcia, A., Appelbaum, L., Ales, J., Cottereau, B., & Rossion, B. (2015). The steady-state visual evoked potential in vision research: A review. *Journal Of Vision*. https://doi.org/10.1167/15.6.4

Öhman, A., Lundqvist, D., & Esteves, F. (2001). The face in the crowd revisited: A threat advantage with schematic stimuli. *Journal of Personality and Social Psychology*, *80*, 381–396. https://doi.org/10.1037/0022-3514.80.3.381

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

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

Peltola, M J, Yrttiaho, S., & Leppänen, J. M. (2018). Infants' attention bias to faces as an early marker of social development. *Developmental Science*, *21*, e12687.

https://doi.org/10.1111/desc.12687

Peltola, Mikko J, Leppanen, J. M., Palokangas, T., & Hietanen, J. K. (2008). Fearful Faces Modulate Looking Duration and Attention Disengagement in 7-Month-Old Infants. *Developmental Science*, *11*(1), 60–68. https://doi.org/10.1111/j.1467-7687.2007.00659.x

Peykarjou, S., Hoehl, S., Pauen, S., & Rossion, B. (2017). Rapid Categorization of Human and Ape Faces in 9-Month-Old Infants Revealed by Fast Periodic Visual Stimulation. *Scientific Reports*. https://doi.org/10.1038/s41598-017-12760-2

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*.

Pons, F., Harris, P. L., & Doudin, P.-A. (2002). Teaching Emotion Understanding. *European Journal of Psychology of Education*, *17*(3), 293–304. https://doi.org/10.1007/BF03173538

Porter, S., ten Brinke, L., & Gustaw, C. (2010). Dangerous decisions: the impact of first impressions of trustworthiness on the evaluation of legal evidence and defendant culpability. *Psychology, Crime & Law*, *16*(6), 477–491. https://doi.org/10.1080/10683160902926141

Pourtois, G., Grandjean, D., Sander, D., & Vuilleumier, P. (2004). Electrophysiological correlates of rapid spatial orienting towards fearful faces. *Cerebral Cortex*, *14*(6), 619–633. https://doi.org/10.1093/cercor/bhh023

Putnam, S. P., Helbig, A. L., Gartstein, M. A., Rothbart, M. K., & Leerkes, E. (2014). Development and Assessment of Short and Very Short Forms of the Infant Behavior

Questionnaire–Revised. *Journal of Personality Assessment*, *96*(4), 445–458. https://doi.org/10.1080/00223891.2013.841171

Quadrelli, E., Conte, S., Macchi Cassia, V., & Turati, C. (2019). Emotion in motion: Facial dynamics affect infants' neural processing of emotions. *Developmental Psychobiology*, *61*, 843–858. https://doi.org/10.1002/dev.21860

Quinn, P. C., Yahr, J., Kuhn, A., Slater, A. M., & Pascalils, 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

Rajhans, P., Missana, M., Krol, K. M., & Grossmann, T. (2015). The association of temperament and maternal empathy with individual differences in infants' neural responses to emotional body expressions. *Development And Psychopathology*, *27*(4), 1205–1216. https://doi.org/10.1017/S0954579415000772

Ramsey-Rennels, J. L., & Davis, R. E. (2008). Facial experience during the first year. *Infant Behavior & Development*, *31*(4), 665–678. https://doi.org/10.0.3.248/j.infbeh.2008.04.009

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

Retter, T. L., & Rossion, B. (2016). Uncovering the neural magnitude and spatio-temporal dynamics of natural image categorization in a fast visual stream. *Neuropsychologia*, *91*, 9–28. https://doi.org/10.1016/j.neuropsychologia.2016.07.028

Rezlescu, C., Duchaine, B., Olivola, C., & Chater, N. (2012). Unfakeable facial configurations affect strategic choices in trust games with or without information about past behavior. *Plos One*. https://doi.org/10.1371/journal.pone.0034293

Rhodes, G., & Jeffery, L. (2006). Adaptive norm-based coding of facial identity. *Vision*

*Research*, *46*(18), 2977–2987. https://doi.org/10.1016/j.visres.2006.03.002

Rhodes, G., Jeffery, L., Watson, T. L., Clifford, C. W. G., & Nakayama, K. (2003). Fitting the Mind to the World: Face Adaptation and Attractiveness Aftereffects. *Psychological Science*, *14*(6), 558. https://doi.org/10.1046/j.0956-7976.2003.psci_1465.x

Robert, H. (2007). Warren S. Torgerson, Theory and methods of scaling. *Behavioral Science*, *4*(3), 245–247. https://doi.org/10.1002/bs.3830040308

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

Rossion, B. (2014). Understanding individual face discrimination by means of fast periodic visual stimulation. *Experimental Brain Research*. https://doi.org/10.1007/s00221-014-3934-9

Rossion, Bruno, Torfs, K., Jacques, C., & Liu-Shuang, J. (2015). Fast periodic presentation of natural images reveals a robust face-selective electrophysiological response in the human brain. *Journal of Vision*, *15*. https://doi.org/10.1167/15.1.18

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 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

Sakuta, Y., Kanazawa, S., & Yamaguchi, M. (2018). Infants prefer a trustworthy person: An early sign of social cognition in infants. *Plos One*. https://doi.org/10.1371/journal.pone.0203541

Schaller, M., & Skowronski, J. J. (2012). *Evolutionary Bases of First Impressions*. New York: Guilford Press.

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

Schofield, C. A., Coles, M. E., & Gibb, B. E. (2007). Social anxiety and interpretation biases for facial displays of emotion: Emotion detection and ratings of social cost. *Behaviour Research and Therapy*, *45*(12), 2950–2963. https://doi.org/10.1016/j.brat.2007.08.006

Schuboe, A., Gendolla, G. H. E., Meinecke, C., & Abele, A. E. (2006). Detecting emotional faces and features in a visual search paradigm: Are faces special? *Emotion*, *6*, 246–256. https://doi.org/10.1037/1528-3542.6.2.246

Schupp, H., Ohman, A., Junghöfer, M., Weike, A., Stockburger, J., & Hamm, A. (2004). The facilitated processing of threatening faces: an ERP analysis. *Emotion (Washington, D.C.)*, *4*(2), 189–200. https://doi.org/10.1037/1528-3542.4.2.189

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

Serrano, J. M., Iglesias, J., & Loeches, A. (1995). Brief report: Infants' responses to adult static facial expressions. *Infant Behavior and Development*, *18*(4), 477–482. https://doi.org/10.1016/0163-6383(95)90036-5

Shepard, R. N. (1980). Multidimensional Scaling, Tree-Fitting, and Clustering. *Science*, *210*(4468), 390. https://doi.org/https://doi.org/10.1126/science.210.4468.390

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

Silberman, E. K., & Weingartner, H. (1986). Hemispheric lateralization of functions related to emotion. *Brain and Cognition*, *5*, 322–353. https://doi.org/10.1016/0278-2626(86)90035-7

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

Stolarova, M., Keil, A., & Moratti, S. (2006). Modulation of the C1 visual event-related component by conditioned stimuli: Evidence for sensory plasticity in early affective perception. *Cerebral Cortex*, *16*(6), 876–887. https://doi.org/10.1093/cercor/bhj031

Sutherland, C. A. M., Young, A. W., Mootz, C. A., & Oldmeadow, J. A. (2015). Face gender and stereotypicality influence facial trait evaluation: Counter-stereotypical female faces are negatively evaluated. *British Journal of Psychology*, *106*(2), 186–208. https://doi.org/10.1111/bjop.12085

Taylor-Colls, S., & Pasco Fearon, R. M. (2015). The Effects of Parental Behavior on Infants' Neural Processing of Emotion Expressions. *Child Development*, *86*(3), 877–888. https://doi.org/http://10.0.4.87/cdev.12348

Todorov, A., Dotsch, R., Porter, J. M., Oosterhof, N. N., & Falvello, V. B. (2013). Validation of data-driven computational models of social perception of faces. *Emotion*, *13*(4), 724–738. https://doi.org/10.1037/a0032335

Todorov, A., Mende-Siedlecki, P., & Dotsch, R. (2013). Social judgments from faces. *Current Opinion in Neurobiology*, *23*(Social and emotional neuroscience), 373–380. https://doi.org/10.1016/j.conb.2012.12.010

Todorov, A., Olivola, C. Y., Dotsch, R., & Mende-Siedlecki, P. (2015). Social Attributions from Faces: Determinants, Consequences, Accuracy, and Functional Significance. *Annual Review of Psychology*, *66*(1), 519–564. https://doi.org/10.1146/annurev-psych-113011-143831

Todorov, A., Pakrashi, M., & Oosterhof, N. N. (2009). Evaluating faces on trustworthiness after minimal time exposure. *Social Cognition*, *27*(6), 813–833. https://doi.org/https://doi.org/10.1521/soco.2009.27.6.813

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

Vaish, A., Grossman, T., & Woodward, A. (2008). Not All Emotions Are Created Equal: The Negativity Bias in Social-Emotional Development. *Psychological Bulletin*, *134*(3), 383–403. https://doi.org/10.1037/0033-2909.134.3.383

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

van Honk, J., Tuiten, A., van den Hout, M., Koppeschaar, H., Thijssen, J., E, de H., & Verbaten, R. (1998). Baseline salivary cortisol levels and preconscious selective attention for threat. A pilot study. *Psychoneuroendocrinology*, *23*(7), 741–747. https://doi.org/10.1016/S0306-4530(98)00047-X

Walker, P. M., & Tanaka, J. W. (2003). An Encoding Advantage for Own-Race versus Other-Race Faces. *Perception*, *32*, 1117–1125. https://doi.org/10.1068/p5098

Williams, L. M., Liddell, B. J., Rathjen, J., Brown, K. J., Gray, J., Phillips, M., … Gordon, E. (2004). Mapping the time course of nonconscious and conscious perception of fear: an integration of central and peripheral measures. *Human Brain Mapping*, *21*(2), 64–74. https://doi.org/10.1002/hbm.10154

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

Wilson, J. P., & Rule, N. O. (2017). Advances in Understanding the Detectability of Trustworthiness From the Face: Toward a Taxonomy of a Multifaceted Construct. *Current Directions in Psychological Science*, *26*(4), 396–400. https://doi.org/10.1177/0963721416686211

Wilson, J., & Rule, N. (2015). Facial Trustworthiness Predicts Extreme Criminal-Sentencing Outcomes. *Psychological Science*, *26*(8), 1325–1331.

https://doi.org/10.1177/0956797615590992

Wojciszke, B., Bazinska, R., & Jaworski, M. (1998). On the dominance of moral categories in impression formation. *Personality & Social Psychology Bulletin*, *24*(12), 1251–1263. https://doi.org/10.1177/01461672982412001

Xu, B., Liu-Shuang, J., Rossion, B., & Tanaka, J. (2017). Individual Differences in Face Identity Processing with Fast Periodic Visual Stimulation. *Journal Of Cognitive Neuroscience*, *29*, 1368–1377. https://doi.org/10.1162/jocn_a_01126

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

Yang, D., Qi, S., Ding, C., & Song, Y. (2011). An ERP study on the time course of facial trustworthiness appraisal. *Neuroscience Letters*, *496*(3), 147–151. https://doi.org/10.1016/j.neulet.2011.03.066

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

Yrizarry, N., Matsumoto, D., & Wilson-Cohn, C. (1998). American-Japanese Differences in Multiscalar Intensity Ratings of Universal Facial Expressions of Emotion. *Motivation and Emotion*, *22*, 315–327. https://doi.org/10.1023/A:1021304407227

Yuki, M., Maddux, W. W., & Masuda, T. (2007). Are the windows to the soul the same in the East and West? Cultural differences in using the eyes and mouth as cues to recognize emotions in Japan and the United States. *Journal of Experimental Social Psychology*, *43*, 303–311. https://doi.org/10.1016/j.jesp.2006.02.004

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*, *7*(3), 194–215. https://doi.org/10.1207/S15327957PSPR0703_01