Neural and Behavioral Evidence for Infants' Sensitivity to the Trustworthiness of Faces

Sarah Jessen[1,2] and Tobias Grossmann[1,3]

Abstract

■ Face evaluation is a key aspect of face processing in humans, serving important functions in regulating social interactions. Adults and preschool children readily evaluate faces with respect to a person's trustworthiness and dominance. However, it is unclear whether face evaluation is mainly a product of extensive learning or a foundational building block of face perception already during infancy. We examined infants' sensitivity to facial signs of trustworthiness (Experiment 1) and dominance (Experiment 2) by measuring ERPs and looking behavior in response to faces that varied with respect to the two facial attributes. Results revealed that 7-month-old infants are sensitive to facial signs of trustworthiness but not dominance. This sensitivity was reflected in infants' behavioral preference and in the modulation of brain responses previously linked to emotion detection from faces. These findings provide first evidence that processing faces with respect to trustworthiness has its origins in infancy and shed light on the behavioral and neural correlates of this early emerging sensitivity. ■

INTRODUCTION

Human faces provide a wealth of socially relevant information regarding a person's gender, age, and race (Jack & Schyns, 2015; Calder & Young, 2005). Adults also readily evaluate a person's character with respect to its trustworthiness, dominance, and competence on the basis of facial appearance (Todorov, Olivola, Dotsch, & Mende-Siedlecki, 2015; Oosterhof & Todorov, 2008; Todorov, Said, Engell, & Oosterhof, 2008). Face evaluation along these dimensions affects decision-making and cooperative behavior and thereby serves important functions in regulating human social interactions (Todorov et al., 2015). It has been argued that especially face evaluation regarding someone's trustworthiness is of adaptive significance as it helps to decide who might be friend or who might be foe and thus guides whom to approach and whom to avoid (Todorov, 2008; Fiske, Cuddy, & Glick, 2007). For example, in economic games participants are less willing to trust an individual with an untrustworthy-looking face (Tingley, 2014; Rezlescu, Duchaine, Olivola, & Chater, 2012; Chang, Doll, van 't Wout, Frank, & Sanfey, 2010; Schlicht, Shimojo, Camerer, Battaglia, & Nakayama, 2010; Stirrat & Perrett, 2010; van 't Wout & Sanfey, 2008) but are more likely to give money to a person with a trustworthy-looking face (Rezlescu et al., 2012). While having an impact on decision-making, facial evaluation of another person's trustworthiness as such is thought to reflect automatic processes as it occurs rapidly, unintentionally, and requires very little exposure time to the face (Stewart et al., 2012; Todorov, Pakrashi, & Oosterhof, 2008; Willis & Todorov, 2006). At the mechanistic level, trustworthiness evaluations are considered to rely on an overextension of our ability to respond to facial expressions. In particular, it has been shown that trustworthy faces structurally resemble happy facial expressions, whereas untrustworthy faces are more likely to be perceived as angry (Engell, Todorov, & Haxby, 2010; Said, Sebe, & Todorov, 2009). The notion that trustworthiness evaluations rely on processes implicated in emotion perception has also been confirmed in neuroimaging studies using ERPs (Marzi, Righi, Ottonello, Cincotta, & Viggiano, 2014; Dzhelyova, Perrett, & Jentzsch, 2012) and fMRI (Engell et al., 2010).

Given the pervasiveness, readiness, and importance of face evaluation in guiding human social behavior, it appears vital to investigate its developmental origins. Recently, it has been shown that children evaluate a face's character in a similar manner as adults do, providing first insights into the ontogeny of face evaluation (Caulfield, Ewing, Bank, & Rhodes, 2015; Cogsdill, Todorov, Spelke, & Banaji, 2014). Children of ages 3 and above tend to classify trustworthy-looking faces as nice, dominant-looking faces as strong, and competent-looking faces as smart (Cogsdill et al., 2014), and from 5 years on, children can also explicitly judge faces as more or less trustworthy (Caulfield et al., 2015). Importantly, Cogsdill and colleagues (2014) showed that children's face-to-trait inferences might reflect more general valence-based decisions; children (and adults) were shown to apply the mean versus nice evaluation not only to faces varying in trustworthiness but also in dominance and competence. However, this

---

[1] Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany, [2] University of Lübeck, [3] University of Virginia

Journal of Cognitive Neuroscience 28:11, pp. 1728–1736
doi:10.1162/jocn_a_00999

work leaves unclear whether face evaluation is mainly a product of extensive learning during early development because children had 3 or more years of experience with faces or is a more foundational building block of face perception also found in preverbal infants. Moreover, it also leaves unclear what the neural mechanisms of face evaluation in development are.

From a developmental perspective, human infants have been shown to evaluate social agents on the basis of their behavior toward other individuals and also on the basis of certain physical characteristics. Specifically, 6-month-old infants prefer an agent that has helped another agent in achieving a goal but avoid an agent that has hindered another agent in achieving a goal (Hamlin, Wynn, & Bloom, 2007). This is also reflected at the brain level, where a particular ERP component, namely, the P400, was larger in response to prosocial (helping) agents when compared with antisocial (hindering) agents (Gredebäck et al., 2015). In contrast, 10- to 13-month-old infants, but not 8-month-old infants, have been shown to use size cues to reason about dominance when observing social encounters (Thomsen, Frankenhuis, Ingold-Smith, & Carey, 2011). Although this suggests that infants younger than 10 months might not be able to extract dominance cues from an agent's appearance, a recent study suggests that infants as young as 6 months are sensitive to dominance relations when characterized via group size (Pun, Birch, & Baron, 2016). It has been argued that group size as a cue is more salient and potentially more informative than other physical cues such as size, therefore allowing for a dominance discrimination at a younger age. Together, prior work suggests that infants are sensitive to cues that index trustworthiness from early in infancy, whereas sensitivity to dominance may depend on the specific cues used to convey dominance. Importantly, it is unknown whether infants are sensitive to facial signs of trustworthiness and dominance.

We therefore examined infants' sensitivity to facial signs of trustworthiness (Experiment 1) and dominance (Experiment 2). First, based on prior work (Thomsen et al., 2011; Hamlin et al., 2007), we predicted that 7-month-old infants are sensitive to trustworthiness but not dominance. Trust evaluations are of primary importance for survival and thus seen from early in life as they allow us to assess who is friend and who is foe, whereas dominance assessments appear more complex and require highly salient cues. More specifically, we hypothesized that, similar to prior work (Hamlin et al., 2007), infants will show a preference for trustworthy compared with untrustworthy faces. Second, we predicted that, similar to what has been shown for adults (Marzi et al., 2014; Dzhelyova et al., 2012), differences in facial trustworthiness will be reflected in brain processes implicated in emotional face processing in infants (ERP components: P400 and Nc). Third, we predicted that, as for behavioral cues to helpful behavior (Hamlin et al., 2007), differences in trustworthiness of the face will be reflected in a modulation of the P400.

TRUSTWORTHINESS

Methods

Participants

Twenty-nine 7-month-old infants participated in this study ([mean ± *SD*] age = 213 ± 9 days, range = 119–229, 15 girls). Sample size was determined a priori and based on prior comparable research (Jessen & Grossmann, 2015; Peltola, Leppänen, Mäki, & Hietanen, 2009). For the EEG analysis, two infants were excluded from the final sample because of failure to contribute at least 10 artifact-free trials per condition. For the analysis of the preferential looking paradigm, three infants were excluded from the final sample because they did not complete all three trials (one of them was among the two infants also excluded from the final EEG sample). For the preferential touching analysis, only infants who touched one of the pictures in at least one of the three trials were included ($n = 17$).

All infants were born full term (38–42 weeks gestational age) with a birth weight of at least 2500 g. The parents gave written informed consent, and the study was approved by the ethics committee at the University of Leipzig and conducted according to the declaration of Helsinki.

Stimuli

Face stimuli were selected from an existing database of computer-generated faces (Oosterhof & Todorov, 2008). These faces had been generated using FaceGen Modeller 3.2 (Singular Inversions, 2007, Toronto, Canada), and varied in trustworthiness according to models developed by Oosterhof and Todorov (2008). We selected three male white identities (005, 010, and 016), of which we each used a neutral version, a version classified as untrustworthy (−3 *SD* from the average neutral face), and a version classified as trustworthy (+3 *SD* from the average neutral face), leading to a total of nine different faces (see Figure 1, top row). Note that, although faces in which trustworthiness or untrustworthiness is extremely exaggerated (beyond ±3 *SD*) have been shown to be perceived as happy or angry by adults (see Oosterhof & Todorov, 2008), the facial stimuli used in the current study were within this critical ±3 *SD* range and are thus still perceived as emotionally neutral by adult raters (see Oosterhof & Todorov, 2008).

To ensure that the faces were indeed perceived as intended and previously reported in the literature (Oosterhof & Todorov, 2008), we asked a group of 24 adult participants (mean age = 24 ± 3 years, 12 women) to judge the faces on a 7-point-Likert scale with respect to trustworthiness (1 = *not trustworthy at all*, 7 = *very trustworthy*). As expected, untrustworthy faces were perceived as least trustworthy ([mean ± *SD*] 3.18 ± 0.75), trustworthy faces as most trustworthy (5.19 ± 0.96), and neutral faces received intermediate scores (4.25 ± 0.79; all differences

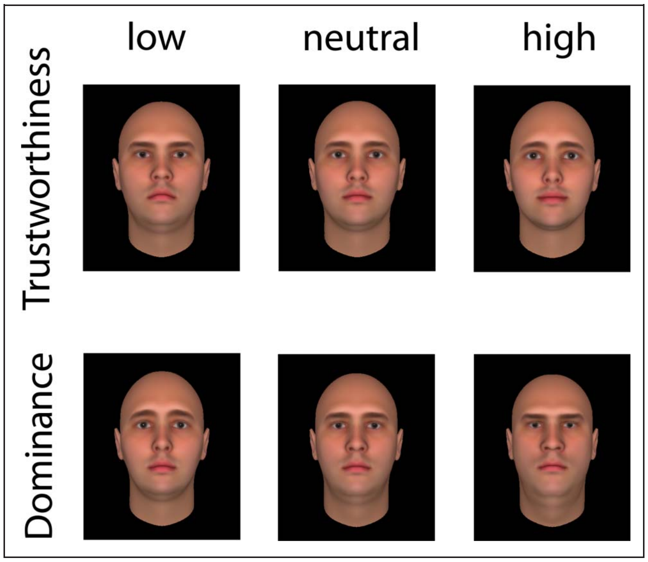

Figure 1. Example of stimulus material. Face stimuli varied in either trustworthiness (Experiment 1) or dominance (Experiment 2). For each experiment, three types of faces were presented: (1) faces that had previously been classified as low on a given trait (untrustworthy or subdominant), (2) faces that had previously been classified as medium (or neutral) with respect to a given trait, and (3) faces that had previously been judged to be high on a given trait (dominant or trustworthy).

were highly significant, $p < .001$, using a repeated-measures ANOVA and *t* tests for post hoc analysis).

For the preferential looking paradigm, a trustworthy (+3 *SD*), neutral, and untrustworthy face (−3 *SD*) from a fourth identity (017) was chosen from the same database. Pictures were printed to a size of 13 × 18 cm and glued to thick cardboard. On the back, stripes of Velcro were attached to fix the pictures at an equal distance on a wooden board during the experiment.

Design

The EEG experiment consisted of three conditions: trustworthy, neutral, and untrustworthy. For each condition, 90 faces were presented, 30 from each identity, leading to a total of 270 stimuli. The order of stimulus presentation was pseudorandomized, ensuring that the same condition was not repeated more than once. Furthermore, trials were split up into 10 miniblocks consisting of 27 trials each (nine per condition and three per identity). Miniblocks were presented consecutively without interruption. Each participant received an individual randomization.

Each trial started with the presentation of a white fixation star presented in the center of the screen on a black background for 300 msec. This was followed by the actual stimulus images for 800 msec. After picture offset, an ISI followed, during which a black screen was shown for a randomly varying duration between 800 and 1200 msec.

For the preferential looking paradigm, the three faces (trustworthy, neutral, untrustworthy) were presented pairwise, leading to a total of three pairs (trustworthy vs. untrustworthy, trustworthy vs. neutral, neutral vs. untrustworthy). The presentation order of the three pairs was counterbalanced across participants, and the side on which each face was presented (left or right) was also counterbalanced. Each pair was presented for 30 sec.

Procedure

After arriving in the lab, infant and parents were given time to familiarize with the new environment, and parents were informed about the experiment and then signed a consent form. The infant was sitting on their parent's lap while the EEG recording was prepared. For recording, an elastic cap (EasyCap, Eaton, OH) in which 27 Ag-Ag-Cl-electrodes were mounted according to the 10–20 system was used. An additional electrode was attached below the infant's right eye for computing the EOG. The EEG was recorded with a sampling rate of 500 Hz using a REFA-8 amplifier (Twente Medical Systems, Oldenzaal, The Netherlands).

The experiment took place in a soundproof, electrically shielded chamber, in which the infant was seated on their parent's lap. Stimuli were presented on a CRT monitor with screen resolution of 1024 × 786 and a refresh rate of 60 Hz at a distance of approximately 90 cm from the infant. The parent was instructed not to interact with the child during the experiment.

The infant's looking behavior during the EEG experiment was monitored using a small camera mounted on top of the monitor. When the infant became inattentive, video clips with colorful moving abstract shapes accompanied by ring tones were played to redirect the infant's attention to the screen. The experiment continued until the maximum number of trials was presented or the infant became too fussy.

After the EEG cap and gel were removed, the preferential looking paradigm followed. The parent was asked to sit down with the infant on their lap on a blanket on the floor, while the experimenter sat opposite the infant. If the infant did not want to sit on the lap, they were also allowed to sit or kneel on the blanket. The pictures were attached to a wooden board 25 cm apart (measuring from the inner corners of the picture). For five infants, the pictures were attached 38 cm apart. The wooden board with the pictures attached was presented to the infant at a distance where the infant could comfortably touch both pictures (see Figure 4). Before the initiation of a trial, the pictures were covered with a black cloth. A trial started with the removal of the cloth and lasted for 30 sec. During the trial, the parent was instructed to close their eyes or look sideways to prevent any influence of the parent on the infant's reactions. The experimenter monitored the infant's attention during the trial. If the infant looked away from the board, she tapped on the center of the board to redirect the attention of the infant to the experiment. The trials were video recorded to allow for offline coding of the infant's behavior (looking and touching).

EEG Analysis

We analyzed the data using Matlab (The MathWorks, Inc., Natick, MA), the Matlab toolbox FieldTrip (Oostenveld, Fries, Maris, & Schoffelen, 2011), and SPSS (IBM SPSS Statistics for Windows, Version 22.0; IBM Corp., Armonk, NY). Data were re-referenced offline to the mean of TP9 and TP10 and bandpass-filtered between 0.2 and 20 Hz. Trials were segmented into 1-sec epochs lasting from 200 msec before stimulus onset to 800 msec after stimulus onset. In five participants, one electrode was noisy and therefore interpolated using spherical spline interpolation (Perrin, Pernier, Bertrand, & Echallier, 1989). To detect trials contaminated by artifacts, the standard deviation was computed in a sliding window of 200 msec. If the standard deviation exceeds 80 mV at any electrode, the entire trial was discarded. Additionally, the trials were inspected visually for any remaining artifacts. Furthermore, the video recording of the infants during the experiments was analyzed, and all trials in which the infant did not attend to the screen were excluded from further analysis. (In eight infants, this was not possible becuase of a technical error during video recording. However, if anything, this should decrease the signal-to-noise ratio, and hence, we decided to include those eight infants in the analysis.) Infants contributed on average 35 ± 17 (mean ± *SD*) trials per condition in the EEG analysis (trustworthy: 35 ± 18, neutral: 35 ± 17, untrustworthy: 35 ± 18).

We analyzed the N290, P400, and Nc ERP components. The N290 and P400 were analyzed at occipital electrodes (O1 and O2), and the mean amplitude was computed in a time window from 200 to 300 msec for the N290 and 360 to 500 msec for the P400. The Nc amplitude was examined at frontal electrodes (F3, FZ, F4, FC5, FC6) between 400 and 600 msec. For the N290 and P400, a repeated-measures ANOVA with the factors Trustworthiness (untrustworthy, neutral, and trustworthy) and Hemisphere (left, right) was computed. For the Nc, a repeated-measures ANOVA was computed with the factor Trustworthiness only. Student's *t* tests were computed to further analyze interaction effects, and effect sizes are reported as partial eta-square (η^2) for ANOVAs and Cohen's *d* for *t* tests.

Behavioral Analysis

All videos were coded by a rater who was blind to the design of the study. To check coder agreement, the videos from six participants were recoded by a second coder and interrater reliability was assessed using Pearson's correlation coefficient ($r = .86$). Duration of looking and touching to either picture was scored over a duration of 30 sec.

To compare looking durations between the pictures, the total looking duration across all three trials was computed for all three pictures. As each picture was shown twice (once in combination with each of the other two pictures; for instance, trustworthiness vs. neutral and trustworthiness vs. untrustworthiness), the summed looking duration was computed from two values (e.g., $\text{duration}_{\text{trustworthiness}} = \text{duration}_{\text{trustworthiness\_trial1}} + \text{duration}_{\text{trustworthiness\_trial2}}$). This summed looking duration was divided by the total looking duration to all pictures (e.g., $\text{duration}_{\text{trustworthiness}}/[\text{duration}_{\text{trustworthiness}} + \text{duration}_{\text{neutral}} + \text{duration}_{\text{untrustworthiness}}]$). This procedure was employed to compute percentage looking duration for all three conditions. Percentage touching duration was computed in an identical manner.

On the basis of the adult ratings presented above, we expected to see a linear increase in looking/touching duration from untrustworthy to trustworthy faces in infants and therefore entered the computed values (percentage looking and touching) into an *F* test to test for linear trends.

To further validate the obtained results, we conducted binomial tests contrasting the number of infants preferring trustworthy over untrustworthy, trustworthy over neutral, and neutral over untrustworthy faces.

Results

N290

We did not observe any significant effects in the time window from 200 to 300 msec at occipital electrodes (all $p\text{s} \geq .22$).

P400

We found a significant interaction between Trustworthiness and Hemisphere at occipital electrodes between 360 and 500 msec, $F(1.85, 48.19) = 3.26, p = .05, \eta^2 = 0.11$ (see Figure 2). Although there was a significant effect of Trustworthiness at the occipital electrode over the right hemisphere ($F(1.97, 51.21) = 3.37, p = .043, \eta^2 = 0.11$; [mean ± *SD*] trustworthy: 5.52 ± 12.58 μV, neutral: 11.42 ± 16.25 μV, untrustworthy: 9.04 ± 14.47 μV), there was no effect at the occipital electrode over the left hemisphere ($p = .85$). Post hoc tests revealed a larger amplitude in response to neutral compared with trustworthy faces, $t(26) = -2.44, p = .022, d = -0.47$. None of the others contrasts were significant (all $p\text{s} > .1$).

Nc

We observed a significant main effect of Trustworthiness between 400 and 600 msec at frontal electrodes, $F(1.94, 50.40) = 3.66, p = .034, \eta^2 = 0.12$ (see Figure 3), revealing a significantly larger Nc amplitude for neutral compared with untrustworthy, $t(26) = 2.27, p = .032, d = 0.44$, and

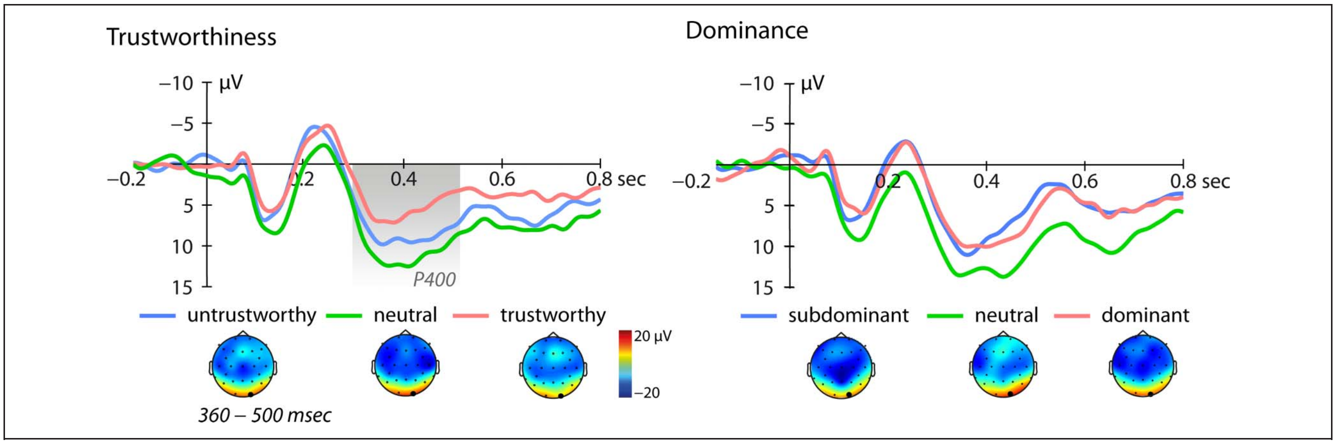

Figure 2. ERP responses at right occipital electrode. The left graph shows the ERP responses in the trustworthiness experiment. The right graph shows ERP responses in the dominance experiment. Neutral faces elicited a larger P400 than trustworthy faces. The bottom row shows the topographical distribution of the ERP responses for the three conditions between 360 and 500 msec.

neutral compared with trustworthy faces, $t(26) = 2.26, p = .033$, $d = 0.43$ (trustworthy: −7.96 ± 11.18 μV; neutral: −14.38 ± 15.03 μV; untrustworthy: −8.46 ± 11.57 μV).

Behavioral Results

Our analysis revealed a linear relation between Trustworthiness and infants' looking preference, $F(1, 25) = 5.96$, $p = .022$, $\eta^2 = 0.19$ (see Figure 4). Specifically, as shown in Figure 4, infants looked longest at the trustworthy faces and shortest at untrustworthy faces and spent intermediate amounts of time looking at neutral faces. Binomial tests performed for the behavioral comparison between the face pairs confirmed this result (18 of 26 infants showed a preference for trustworthy faces over untrustworthy faces: $p = .023$; 18 of 26 infants showed a preference for neutral over untrustworthy faces: $p = .023$; 17 of 26 infants showed a preference for trustworthy over neutral faces: $p = .046$).

We did not observe any significant effect of Trustworthiness on touching duration ($p = .33$).

DOMINANCE

Methods

Participants

Thirty-four 7-month-old infants participated in the study (age = 215 ± 9 days, range = 200–229, 16 girls). Sample size was matched to Experiment 1 and comparable to prior studies (Jessen & Grossmann, 2015; Peltola et al., 2009). None of the infants had participated in Experiment 1. For the EEG-analysis, two infants were excluded from the final sample because of failure to contribute at least 10 artifact-free trials per condition. For the analysis of the preferential looking paradigm, seven infants were excluded from the final sample because they did not complete all three trials (one of them was among the two infants also excluded from the final EEG sample).

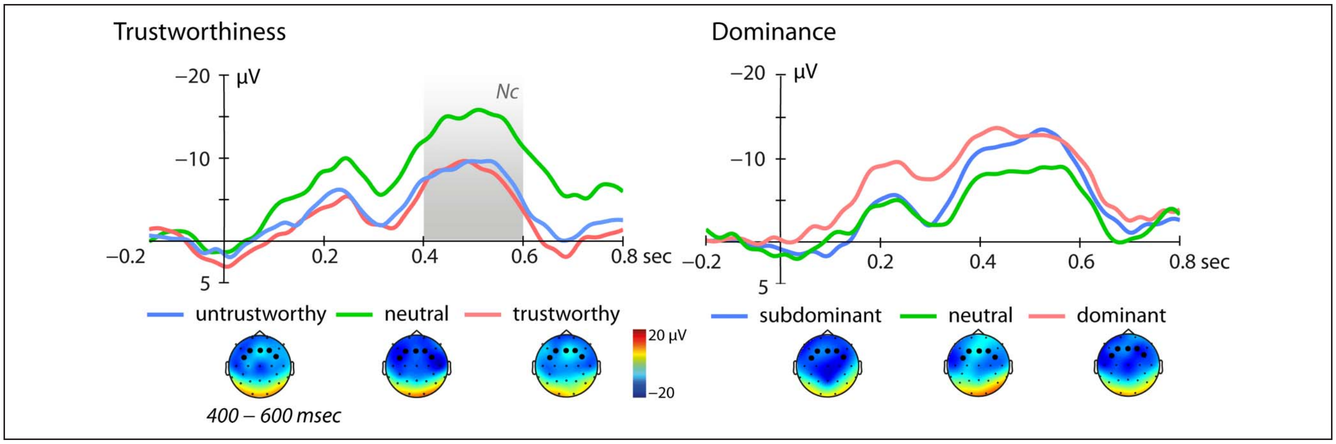

Figure 3. ERP responses at frontal electrodes. The left graph shows the ERP responses in the trustworthiness experiment. The right graph shows ERP responses in the dominance experiment. Neutral faces elicited a larger Nc compared with both trustworthy and untrustworthy faces. The bottom row shows the topographical distribution of the ERP responses for the three conditions between 400 and 600 msec.

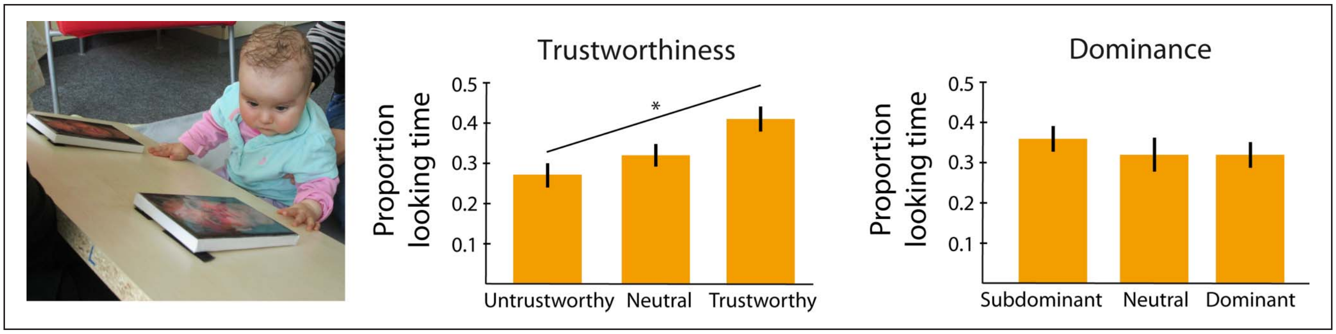

Figure 4. Behavioral results. The photograph on the left shows the setup of the behavioral test. Percentage looking times (mean percentage ± *SE*) across all three trials are shown for the trustworthiness experiment (center) and dominance experiment (right). For trustworthiness, infants showed a linear increase in looking time from untrustworthy to trustworthy faces (center). No effects were seen for faces varying in dominance (right).

As in Experiment 1, all infants were born full term (38–42 weeks gestational age) with a birth weight of at least 2500 g. The parents gave written informed consent, and the study was approved by the ethics committee at the University of Leipzig and conducted according to the declaration of Helsinki.

Stimuli

Face stimuli were selected from the same database as for Experiment 1 (Oosterhof & Todorov, 2008). For Experiment 2, however, we selected faces varying in dominance (Oosterhof & Todorov, 2008). We selected the same three male White identities, and for each identity a neutral version, a version classified as subdominant (−3 *SD* from the average neutral face), and a version classified as dominant (+3 *SD* from the average neutral face).

The same group of 24 adults as for Experiment 1 was asked to judge the faces on a 7-point-Likert scale with respect to dominance (1 = *not dominant at all*, 7 = *very dominant*). As expected, dominant faces received the highest values (4.83 ± 0.85), followed by neutral faces (3.85 ± 0.95), and subdominant faces, which received the lowest values (2.30 ± 0.63; all differences highly significant [ps < .001] using a repeated-measures ANOVA and t tests for post hoc analysis).

Similar to Experiment 1, the dominant, neutral, and subdominant version of a fourth identity (017) was chosen for the preferential looking paradigm.

Design

The design was identical to Experiment 1, except that dominant, neutral, and subdominant faces were shown instead of faces varying in trustworthiness.

EEG Analysis

Preprocessing of the EEG data was identical to Experiment 1. Infants contributed on average 33 ± 15 (mean ± *SD*) trials per condition in the EEG analysis (dominant: 33 ± 15; neutral: 33 ± 16; subdominant: 34 ± 16).

Again, we analyzed the N290 and P400 at occipital electrodes (O1, O2) between 200 and 300 msec and 360 and 500 msec, respectively, and the Nc at frontal electrodes (F3, FZ, F4, FC5, FC6) between 400 and 600 msec. All analyses remained identical to Experiment 1.

Behavioral Analysis

The procedure for the behavioral analysis was identical to Experiment 1. As in Experiment 1, videos from seven participants were coded by a second coder. Interrater reliability was assessed using Pearson's correlation coefficient yielding a correlation of $r = .92$.

Results

N290 and P400

There were no significant effects neither on the N290 nor on the P400 amplitude [N290: Dominance × Hemisphere: $F(1.81, 56.08) = 0.28$, $p = .74$, $\eta^2 = 0.009$; P400: Dominance × Hemisphere: $F(1.56, 48.28) = 0.41$, $p = .62$, $\eta^2 = 0.013$].

Nc

There was no significant effect on the Nc amplitude, $F(1.96, 60.81) = 1.74$, $p = .19$, $\eta^2 = 0.053$.

Behavioral Results

There were no significant behavioral effects, $F(1, 26) = 1.46$, $p = .24$, $\eta^2 = 0.053$.

Discussion

The current study examined the developmental origins of face evaluation by measuring infants' behavioral and neural responses to facial signs of trustworthiness (Experiment 1) and dominance (Experiment 2). Our results revealed that, by the age of 7 months, infants distinguish between faces on the basis of their trustworthiness but not their dominance. At the behavioral level, infants prefer to look at trustworthy faces, while dispreferring to look at untrustworthy faces. At the neural level, discriminating between facial trustworthiness was reflected in brain responses (P400 and Nc) previously linked to emotional face processing and also seen in response to behavioral cues of trustworthiness (P400; Gredebäck et al., 2015; Leppanen, Moulson, Vogel-Farley, & Nelson, 2007). These findings suggest that sensitive responding to facial cues of trustworthiness is a foundational building block of face processing from early in human development. Critically, the current data further suggest that this ability likely represents (a) an overextension of infants' sensitivity to emotional facial expressions and (b) points to the emergence of a flexible system that assesses an agents' trustworthiness from behavioral and facial cues.

Confirming our prediction, we found that 7-month-old infants were sensitive to facial signs of trustworthiness but not dominance. This is in line with the argument that trust evaluations are primary compared with other kinds of evaluative processes, because assessing who is friend or foe is thought to be of prime importance for survival (Fiske et al., 2007). Therefore, the current infant data add important developmental evidence for this notion. Furthermore, the present finding is in agreement with a set of behavioral studies reporting that infants around this age are sensitive to an agent's helpfulness from behavioral cues (Hamlin et al., 2007), but only later become sensitive to an agent's dominance (Mascaro & Csibra, 2012; Thomsen et al., 2011), unless the agent is interpreted as part of a group (Pun et al., 2016). Similar to prior studies (Hamlin & Wynn, 2011; Hamlin et al., 2007), infants in the current study preferred the trustworthy individual and dispreferred the untrustworthy individual, suggesting similar approach and withdrawal tendencies.

Interestingly, our data provide some hints that infants may show a similar behavioral tendency for subdominant faces as they show for trustworthy faces. Although there were no significant differences in the behavioral data of the dominance experiment, inspecting looking time means (see Figure 4) indicates that infants looked longer at subdominant than neutral or dominant faces. One possible explanation for this pattern might be that trustworthy and subdominant faces share certain physical characteristics (Oosterhof & Todorov, 2008), which are preferred by infants. Relatedly, future studies, using eye tracking for instance, are needed to investigate which aspect of the face infants use when sensitively responding to different character traits.

In contrast to prior work, our preference for trustworthy individuals was only reflected in infants' looking but not in their touching behavior. This might be related to the fact that previous work used small and graspable nonhuman characters, whereas we used relatively large facial stimuli and infants might be less likely to touch a face to express a preference but rather show prolonged looking behavior.

Our ERP results showed that discriminating between facial trustworthiness was reflected in a modulation of the P400 and Nc. Both ERP components have been implicated in emotion processing from faces (Peltola et al., 2009; Leppanen et al., 2007), supporting the notion that trustworthiness detection relies on an overextension of the ability to sensitively respond to facial expressions (Engell et al., 2010; Said, Sebe, et al., 2009). With regard to the current findings, it is important to mention that infants' sensitivity to emotional facial expressions is in place by 7 months of age. Specifically, 7-month-old infants, but not 5-month-old infants, have been shown to discriminate between different emotional facial expressions and show an attentional bias for fearful facial expressions (Jessen & Grossmann, 2016; Peltola, Hietanen, Forssman, & Leppänen, 2013; Peltola et al., 2009). These studies show that this fear bias can be observed in infants' looking time and in their ERP responses. More specifically, 7-month-old infants look longer at fearful compared with happy faces (Peltola et al., 2009) and are slower to disengage attention from fearful faces compared with happy faces or nonemotional novel facial expressions (Peltola et al., 2013; Peltola, Leppanen, Palokangas, & Hietanen, 2008). At the neural level, the Nc has been shown to distinguish between positive (happy) and negative (fearful) facial expressions with a larger amplitude to fear, indexing a greater allocation of attention (Jessen & Grossmann, 2015; Peltola et al., 2009).

The present finding is noteworthy because these ERP effects were obtained in response to emotionally neutral faces that only varied with respect to their trustworthiness. Considering that the Nc is an index of attention allocation and a greater Nc amplitude reflects increased attention to a facial stimulus, the Nc effects in the current study suggest that neutral faces evoke the greatest attentional response in infants, whereas trustworthy and untrustworthy faces result in smaller Nc amplitudes indexing attenuated allocation of attention. This specific pattern for the Nc, while providing evidence for the discrimination of trustworthy and untrustworthy faces from neutral faces, is difficult to interpret. One possibility is that this is explained by findings showing that adults show nonlinear brain responses (amygdala) that distinguish between neutral and both highly trustworthy and highly untrustworthy faces, whereas no difference was observed between highly trustworthy and highly untrustworthy faces (Stewart et al., 2012; Said, Dotsch, & Todorov, 2010; Said, Baron, & Todorov, 2009) This might also relate to the fact that neutral faces are considered to be the more prototypical faces and a deviation in trustworthiness in either direction elicits similar brain responses (Said et al., 2010).

Besides a modulation of the Nc, we also observed an effect for the P400 at right occipital electrodes, with neutral faces eliciting the largest P400 amplitude, followed by untrustworthy faces and then trustworthy faces. To find that this effect is lateralized to the right hemisphere is in line with prior work in adults, showing differential responses to trustworthiness from faces only in the right hemisphere (Manssuer, Roberts, & Tipper, 2015; Dzhelyova et al., 2012). Prior work with 7-month-old infants (Jessen & Grossmann, 2015; Leppanen et al., 2007) shows that negative facial expressions elicit a larger P400 when compared with positive facial expressions. In the current study, untrustworthy faces also elicit a larger P400 than trustworthy faces, suggesting that this effect may reflect valence-related processes triggered by faces varying in trustworthiness. Furthermore, the infant P400 has been shown to reflect the detection of an agent's behaviorally acquired status as a prosocial or antisocial individual (Gredebäck et al., 2015). This suggests that P400 modulations reflect infants' sensitivity to facial and behavioral signs of trustworthiness. However, the exact modulation of the P400 in the current study, while generally consistent with the emotion perception ERP work mentioned above (Jessen & Grossmann, 2015; Leppanen et al., 2007), is inconsistent with prior work on behavioral cues to trustworthiness, showing a greater P400 to prosocial agents than antisocial agents.

This might be explained by the fact that infants tend to show greater effects for stimuli that are novel and unusual (i.e., untrustworthy faces) when robustly learned (or acquired) as would be the case for the trustworthiness of faces, whereas infants tend to show greater effects for stimuli that are more familiar and common (i.e., helpful agents) when the information was newly learned (or acquired; Sirois & Mareschal, 2004). More generally, with respect to this proposal, it needs to be acknowledged that the direction of the amplitude modulation for the Nc and P400 is not easy to interpret, because it has yielded conflicting results in previous studies. In particular, although Gredebäck and colleagues (2015) report a larger P400 in response to prosocial over antisocial agents, previous studies on emotion perception typically report a larger P400 for negative compared with positive facial expressions (Jessen & Grossmann, 2015; Leppanen et al., 2007). Furthermore, in contrast to Gredebäck et al.'s (2015) study, we included a neutral condition, which further complicated a direct comparison between studies. Similarly, the direction of the Nc amplitude modulation is difficult to interpret. More specifically, the amplitude of the Nc is not only influenced by the emotional expression of a face (Jessen & Grossmann, 2015; Peltola et al., 2009) but also by its familiarity (de Haan & Nelson, 1999; but see Snyder, Webb, & Nelson, 2002) and by how similar in appearance two faces are (de Haan & Nelson, 1997). Therefore, future research is needed that explicitly tests the sensitivity of these ERP components in infants to newly learned and acquired signs of trustworthiness of a person.

Clearly, more work is needed to clarify this issue. Taken together with prior work, the current findings suggest that the ability to discriminate between the trustworthiness of faces represents an overextension of infants' sensitivity to emotional facial expressions and provide evidence for the emergence of a flexible system that assesses an agents' trustworthiness from behavioral and facial cues.

To conclude, the current study sheds new light on the nature and development of face evaluation by providing first evidence that evaluating faces with respect to their trustworthiness has its origins in infancy. Together with prior work, the current findings lend support to the view that, from early in development, humans form intuitive impressions about others' trustworthiness, an ability that likely reflects humans' preparedness to become apt cooperators (Tomasello, 2014).

Reprint requests should be sent to Sarah Jessen, Department of Neurology, University of Lübeck, Ratzeburger Allee 160, 23562 Lübeck, Germany, or via e-mail: sarah.jessen@neuro.uni-luebeck.de.

REFERENCES

Calder, A. J., & Young, A. W. (2005). Understanding the recognition of facial identity and facial expression. *Nature Reviews: Neuroscience, 6,* 641–651.

Caulfield, F., Ewing, L., Bank, S., & Rhodes, G. (2015). Judging trustworthiness from faces: Emotion cues modulate trustworthiness judgments in young children. *British Journal of Psychology*. doi:10.1111/bjop.12156.

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,* 87–105.

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

de Haan, M., & Nelson, C. A. (1997). Recognition of the mother's face by six-month-old infants: A neurobehavioral study. *Child Development, 68,* 187–210.

de Haan, M., & Nelson, C. A. (1999). Brain activity differentiates face and object processing in 6-month-old infants. *Developmental Psychology, 35,* 1113–1121.

Dzhelyova, M., Perrett, D. I., & Jentzsch, I. (2012). Temporal dynamics of trustworthiness perception. *Brain Research, 1435,* 81–90.

Engell, A. D., Todorov, A., & Haxby, J. V. (2010). Common neural mechanisms for the evaluation of facial trustworthiness and emotional expressions as revealed by behavioral adaptation. *Perception, 39,* 931–941.

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

Gredebäck, G., Kaduk, K., Bakker, M., Gottwald, J., Ekberg, T., Elsner, C., et al. (2015). The neuropsychology of infants' pro-social preferences. *Developmental Cognitive Neuroscience, 12,* 106–113.

Hamlin, J. K., & Wynn, K. (2011). Young infants prefer prosocial to antisocial others. *Cognitive Development, 26,* 30–39.

Hamlin, J. K., Wynn, K., & Bloom, P. (2007). Social evaluation by preverbal infants. *Nature, 450,* 557–559.

Jack, R. E., & Schyns, P. G. (2015). The human face as a dynamic tool for social communication. *Current Biology, 25,* R621–R634.

Jessen, S., & Grossmann, T. (2015). Neural signatures of conscious and unconscious emotional face processing in human infants. *Cortex, 64,* 260–270.

Jessen, S., & Grossmann, T. (2016). The developmental emergence of unconscious fear processing from eyes during infancy. *Journal of Experimental Child Psychology, 142,* 334–343.

Leppanen, J. M., 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,* 232–245.

Manssuer, L. R., Roberts, M. V., & Tipper, S. P. (2015). The late positive potential indexes a role for emotion during learning of trust from eye-gaze cues. *Social Neuroscience, 10,* 635–650.

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,* 63–72.

Mascaro, O., & Csibra, G. (2012). Representation of stable social dominance relations by human infants. *Proceedings of the National Academy of Sciences, U.S.A., 109,* 6862–6867.

Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J.-M. (2011). FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational Intelligence and Neuroscience, 2011,* 156869.

Oosterhof, N. N., & Todorov, A. (2008). The functional basis of face evaluation. *Proceedings of the National Academy of Sciences, U.S.A., 105,* 11087–11092.

Peltola, M. J., Hietanen, J. K., Forssman, L., & Leppänen, J. M. (2013). The emergence and stability of the attentional bias to fearful faces in infancy. *Infancy, 18,* 905–926.

Peltola, M. J., Leppänen, J. M., Mäki, S., & Hietanen, J. K. (2009). Emergence of enhanced attention to fearful faces between 5 and 7 months of age. *Social Cognitive and Affective Neuroscience, 4,* 134–142.

Peltola, M. 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,* 60–68.

Perrin, F., Pernier, J., Bertrand, O., & Echallier, J. F. (1989). Spherical splines for scalp potential and current density mapping. *Electroencephalography and Clinical Neurophysiology, 72,* 184–187.

Pun, A., Birch, S. A., & Baron, A. S. (2016). Infants use relative numerical group size to infer social dominance. *Proceedings of the National Academy of Sciences, U.S.A., 113,* 2376–2381.

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

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,* 519–528.

Said, C. P., Dotsch, R., & Todorov, A. (2010). The amygdala and FFA track both social and non-social face dimensions. *Neuropsychologia, 48,* 3596–3605.

Said, C. P., Sebe, N., & Todorov, A. (2009). Structural resemblance to emotional expressions predicts evaluation of emotionally neutral faces. *Emotion, 9,* 260–264.

Schlicht, E. J., Shimojo, S., Camerer, C. F., Battaglia, P., & Nakayama, K. (2010). Human wagering behavior depends on opponents' faces. *PloS One, 5,* e11663.

Sirois, S., & Mareschal, D. (2004). An interacting systems model of infant habituation. *Journal of Cognitive Neuroscience, 16,* 1352–1362.

Snyder, K., Webb, S. J., & Nelson, C. A. (2002). Theoretical and methodological implications of variability in infant brain response during a recognition memory paradigm. *Infant Behavior & Development, 25,* 466–494.

Stewart, L. H., Ajina, S., Getov, S., Bahrami, B., Todorov, A., & Rees, G. (2012). Unconscious evaluation of faces on social dimensions. *Journal of Experimental Psychology: General, 141,* 715–727.

Stirrat, M., & Perrett, D. I. (2010). Valid facial cues to cooperation and trust: Male facial width and trustworthiness. *Psychological Science, 21,* 349–354.

Thomsen, L., Frankenhuis, W. E., Ingold-Smith, M., & Carey, S. (2011). Big and mighty: Preverbal infants mentally represent social dominance. *Science, 331,* 477–480.

Tingley, D. (2014). Face-off: Facial features and strategic choices. *Political Psychology, 35,* 35–55.

Todorov, A. (2008). Evaluating faces on trustworthiness: An extension of systems for recognition of emotions signaling approach/avoidance behaviors. *Annals of the New York Academy of Sciences, 1124,* 208–224.

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,* 519–545.

Todorov, A., Pakrashi, M., & Oosterhof, N. N. (2008). Evaluating faces on trustworthiness after minimal time exposure. *Social Cognition, 27,* 813–833.

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

Tomasello, M. (2014). The ultra-social animal. *European Journal of Social Psychology, 44,* 187–194.

van 't Wout, M., & Sanfey, A. G. (2008). Friend or foe: The effect of implicit trustworthiness judgments in social decision-making. *Cognition, 108,* 796–803.

Willis, J., & Todorov, A. (2006). First impressions: Making up your mind after a 100-ms exposure to a face. *Psychological Science, 17,* 592–598.