Implicit responses to face trustworthiness measured with fast periodic visual stimulation

Sara C. Verosky Department of Psychology, Oberlin College, Oberlin, OH, USA

Katja A. Zoner Department of Psychology, Oberlin College, Oberlin, OH, USA

Corinne W. Marble* Department of Psychology, Oberlin College, Oberlin, OH, USA

Margaret M. Sammon* Department of Psychology, Oberlin College, Oberlin, OH, USA

Charlotte O. Babarinsa Department of Psychology, Oberlin College, Oberlin, OH, USA

People rapidly and spontaneously form trustworthiness impressions based on facial appearance. Studies using functional magnetic resonance imaging find that activity in the amygdala and other brain regions tracks with face trustworthiness, even when participants are not explicitly asked to judge face trustworthiness. The current study investigated whether it would be possible to detect implicit responses using another method: fast periodic visual stimulation (FPVS). While scalp electroencephalogram (EEG) was recorded, participants viewed sequences of faces in which a single base face was presented at a rate of 6 Hz and oddball faces with different identities were presented every fifth face (6 Hz/5 = 1.2 Hz). Within a given sequence, the oddball faces were all either less trustworthy-looking or trustworthy-looking. The base face either matched the oddball faces on trustworthiness or did not match, so that the experiment had a 2 (trustworthiness of oddball) × 2 (match between base/oddball faces) design. Although participants' task was unrelated to the faces, the trustworthiness of the oddball faces had a strong influence on the response at 1.2 Hz and its harmonics. There was a stronger response for sequences with less trustworthy- versus trustworthy-looking oddball faces over bilateral occipitotemporal sites, medial occipital sites, and beyond. In contrast, the match in trustworthiness between the base face and the oddball faces had only a minimal effect. The effect of oddball type was observed after a short recording time, suggesting that FPVS offers an efficient means of capturing implicit neural responses to face trustworthiness.

Introduction

People form trait impressions of others based solely on facial appearance (Todorov, Olivola, Dotsch, & Mende-Siedlecki, 2015). Although people are able to judge faces along many different trait dimensions, these judgments are highly correlated with each other (Oosterhof & Todorov, 2008). Judgments of face trustworthiness have been found to approximate a valence dimension that underlies much of the variance in these judgments (Oosterhof & Todorov, 2008; Sutherland et al., 2013). People are able to make trustworthiness judgments after extremely brief exposures to faces, with judgments made in 100 ms or less correlating highly with judgments made under unlimited time (Rule, Ambady, & Adams, 2009; Todorov, Pakrashi, & Oosterhof, 2009; Willis & Todorov, 2006). Moreover, a behavioral study using the "who said what" memory paradigm found that people spontaneously form impressions of others based on face trustworthiness, even in contexts where trustworthiness was not made salient (Klapper, Dotsch, van Rooij, & Wigboldus, 2016).

Trait judgments based on facial appearance are driven by physical features or sets of physical features

Citation: Verosky, S. C., Zoner, K. A., Marble, C. W., Sammon, M. M., & Babarinsa, C. O. (2020). Implicit responses to face trustworthiness measured with fast periodic visual stimulation. *Journal of Vision*, *20*(7):29, 1–12, https://doi.org/10.1167/jov.20.7.29.

https://doi.org/10.1167/jov.20.7.29 Received January 12, 2020; published July 29, 2020 ISSN 1534-7362 Copyright 2020 The Authors

(Todorov et al., 2015; Zebrowitz & Montepare, 2008). For example, the resemblance of neutral faces to emotional facial expressions has been found to influence the types of impressions people form (Adams, Nelson, Soto, Hess, & Kleck, 2012; Oosterhof & Todorov, 2008, 2009; Said, Sebe, & Todorov, 2009; Zebrowitz, Kikuchi, & Fellous, 2010). Computer modeling and behavioral studies suggest that judgments of face trustworthiness specifically share a perceptual basis with angry versus happy expressions (Engell, Todorov, & Haxby, 2010; Oosterhof & Todorov, 2008, 2009).

The amygdala plays an important role in trustworthiness judgments from faces. Early evidence for the role of the amygdala came from a study of amygdala lesion patients, where patients with bilateral amygdala lesions evaluated faces as more trustworthy than control participants (Adolphs, Tranel, & Damasio, 1998). Subsequent studies using functional magnetic resonance imaging (fMRI) demonstrated that activity in the amygdala tracks with face trustworthiness, regardless of whether participants are explicitly asked to judge face trustworthiness (Engell, Haxby, & Todorov, 2007; Winston, Strange, O'Doherty, & Dolan, 2002; for a meta-analysis, see Mende-Siedlecki, Said, & Todorov, 2013). Activity in the occipital and temporal regions also tracks with face trustworthiness, but correlational evidence suggests that activity in these regions is modulated by the amygdala (Todorov & Engell, 2008).

While fMRI studies give insight into brain regions involved in the social evaluation of faces, studies using event-related potentials (ERPs) have examined the time course of these judgments. These studies find effects of face trustworthiness as early as 100 ms after stimulus onset (Marzi, Righi, Ottonello, Cincotta, & Viggiano, 2014; Yang, Qi, Ding, & Song, 2011), with continued effects over time (Dzhelyova, Perrett, & Jentzsch, 2012; Rudoy & Paller, 2009). Although most of the ERP studies examining responses to face trustworthiness involve participants making explicit trustworthiness judgments, a study using a visual oddball paradigm found an effect of trustworthiness during an unrelated task (Kovacs-Balint, Stefanics, Trunk, & Hernadi, 2014). In this study, rare untrustworthy-looking oddball faces elicited the visual mismatch negativity component, but rare trustworthy-looking oddball faces did not.

The goal of the current study was to investigate whether it would be possible to detect implicit responses to face trustworthiness using another method: fast period visual stimulation (FPVS) in conjunction with electroencephalography (EEG). With FPVS, when the brain is stimulated repeatedly at a particular frequency, this results in a response at that frequency that can be recorded using EEG (Regan, 1966). Although the data are recorded in the time domain, they are transformed into the frequency domain for analysis, allowing the signal at the frequency of interest to be precisely quantified. This method offers two important advantages: it has a high signal-to-noise ratio and it typically does not rely on participants performing a particular behavioral task (Rossion, 2014).

Although FPVS has primarily been used to investigate processing of lower-level visual stimuli, recently it has begun to be used to investigate face processing (Rossion, 2014). An initial study examining the processing of facial identity found adaptation to a stream of images of an identical face versus a stream of images of different faces (Rossion & Boremanse, 2011). This identity adaptation was strongest over the right occipitotemporal electrode sites. Previous work demonstrates that the right occipitotemporal cortex plays an important role in face processing (Bentin, Allison, Puce, Perez, & McCarthy, 1996; Kanwisher, McDermott, & Chun, 1997; McCarthy, Puce, Gore, & Allison, 1997). Identity adaptation was later found to be maximal when the faces were presented at a rate of close to 6 Hz (Alonso-Prieto, Van Belle, Liu-Shuang, Norcia, & Rossion, 2013).

In the subsequently developed fast-periodic visual oddball paradigm, the same versus different identity conditions were combined into a single stream of faces (Liu-Shuang, Norcia, & Rossion, 2014). In this stream of faces, a base face is presented at a fixed rate. Every fifth face, this base face is replaced by a face of a different identity. The response at the oddball frequency and its harmonics represents differentiation between different identities and it is present over bilateral occipitotemporal electrode sites, with the strongest responses typically observed over the right occipitotemporal sites (Dzhelyova & Rossion, 2014; Liu-Shuang et al., 2014; Liu-Shuang, Torfs, & Rossion, 2016; Xu, Liu-Shuang, Rossion, & Tanaka, 2017). Meanwhile, the response at the stimulation frequency and its harmonics reflects the general visual response, and it is strongest over medial occipital sites (Dzhelyova & Rossion, 2014; Liu-Shuang et al., 2014, 2016; Xu et al., 2017).

Recently, it was shown that FPVS can be used to detect implicit responses to the physical attractiveness of faces (Luo, Rossion, & Dzhelyova, 2019). In this study, participants viewed sequences of faces where the attractiveness of the faces either alternated between less attractive and attractive or did not alternate. The authors found a stronger response at the frequency of alternation for sequences where attractiveness alternated, indicating implicit discrimination of attractiveness. Because attractiveness and trustworthiness judgments tend to be correlated (Oosterhof & Todorov, 2008), Luo and colleagues' study provides support for the hypothesis that it will be possible to detect implicit responses to face trustworthiness. However, a meta-analysis of fMRI studies (Mende-Siedlecki et al., 2013) found different regions were active for judgments of attractiveness versus trustworthiness, likely due to differences in face typicality, and it serves as a reminder that responses to these two trait dimensions could also differ.

The current study used the fast-periodic visual oddball paradigm to investigate implicit responses to face trustworthiness. While EEG was recorded, participants viewed oddball sequences of faces where a base face was presented repeatedly at a rate of 6 Hz, and oddball faces with different identities were presented every fifth face (6 Hz/5 = 1.2 Hz). Within each sequence, trustworthiness was manipulated in two ways. First, all of the oddball faces in the sequence were either less trustworthy-looking or trustworthy-looking. Second, the base face was either less trustworthy-looking or trustworthy-looking, so that it either matched or did not match the trustworthiness of the oddball faces. This meant the experiment had a 2 (trustworthiness of oddball) × 2 (match between base/oddball faces) design. During the experiment, participants were asked to monitor the color of a fixation cross, a task that did not involve attending to face trustworthiness.

We investigated whether the trustworthiness of the faces would influence the size of the face individuation response at 1.2 Hz and its harmonics. We hypothesized that there would be two main effects of trustworthiness. Based on fMRI studies demonstrating increasing activity in the amygdala and occipital/temporal regions with decreasing face trustworthiness (Engell et al., 2007; Todorov & Engell, 2008; Winston et al., 2002), we hypothesized that the face individuation response would be larger for sequences with less trustworthy-looking versus trustworthy-looking oddball faces. Based on the logic that changes in face trustworthiness would emphasize changes in facial identity, we also hypothesized that the face individuation response would be larger for sequences where the base face and oddball faces differed in trustworthiness versus had the same trustworthiness. To foreshadow our results, while the trustworthiness of the oddball faces had a large effect on the face individuation response, the match between the base face and oddball faces did not. We examined the face individuation response over bilateral occipitotemporal sites and medial occipital sites. We also examined the general visual response at 6 Hz and its harmonics in these regions.

Methods

Participants

Twenty-five undergraduates and members of the Oberlin community participated in the experiment. Participants gave consent in accordance with a protocol approved by the Oberlin College Institutional Review Board and they were compensated for their time with partial course credit or payment. All participants reported normal or corrected-to-normal vision. One participant was excluded due to low response accuracy (> 3 standard deviations [SD] below the mean). The final sample consisted of 24 participants, 13 female, 11 male; $M_{\text{age}} = 19.2$; $SD_{\text{age}} = 1.1$. Behavioral data from one participant were lost due to technical difficulties.

Stimuli

The experiment relied on participants being able to differentiate between facial identities; therefore, we used photographs of faces rather than computer-generated faces, because computer-generated faces can be more difficult to tell apart. Twenty faces with neutral expressions were selected from the larger Karolinska Directed Emotional Faces set (Lundqvist, Flykt, & Ohman, 1998). The faces were selected based trustworthiness ratings collected by Oosterhof and Todorov (2008), available in the form of z-scores on the Todorov lab website (http://tlab.princeton.edu/). The five male and five female faces with the lowest trustworthiness ratings were selected as the less trustworthy-looking faces, and the five male and five female faces with the highest trustworthiness ratings were selected as the trustworthy-looking faces. The mean for the less trustworthy-looking faces, $M = -1.10$, $SD = 0.35$, was slightly more extreme than the mean for the trustworthy-looking faces, $M = 0.91$, $SD = 0.20$, but the two groups did not differ in their distance from zero, independent samples t test on distance: $t(18) = 1.44$, $p = 0.17$.

Although the less trustworthy-looking versus trustworthy-looking faces were selected based on trustworthiness ratings, the two sets of faces differed significantly on nearly all of the other trait ratings collected by Oosterhof and Todorov (2008) as well, including aggressiveness, attractiveness, caring, confidence, emotional stability, meanness, intelligence, responsibility, sociability, threateningness, unhappiness, and weirdness. The only trait dimension that the sets of faces did not differ on was dominance.

The faces, including hair, ears, and neck, were cropped out of the photographs and placed on a gray background. The images were displayed on an LCD monitor that was located approximately 57 cm away from where participants were seated. The screen had a resolution of 1024 × 768 pixels and a refresh rate of 60 Hz. At 100% size, the faces subtended 9.5° × 13.0° of visual angle.

To help rule out the possibility that any observed differences in the response to the two sets of faces were due to low-level visual confounds, the mean luminance and the standard deviation of the luminance were calculated for the faces, excluding the background. Because the photographs were in color, the luminance

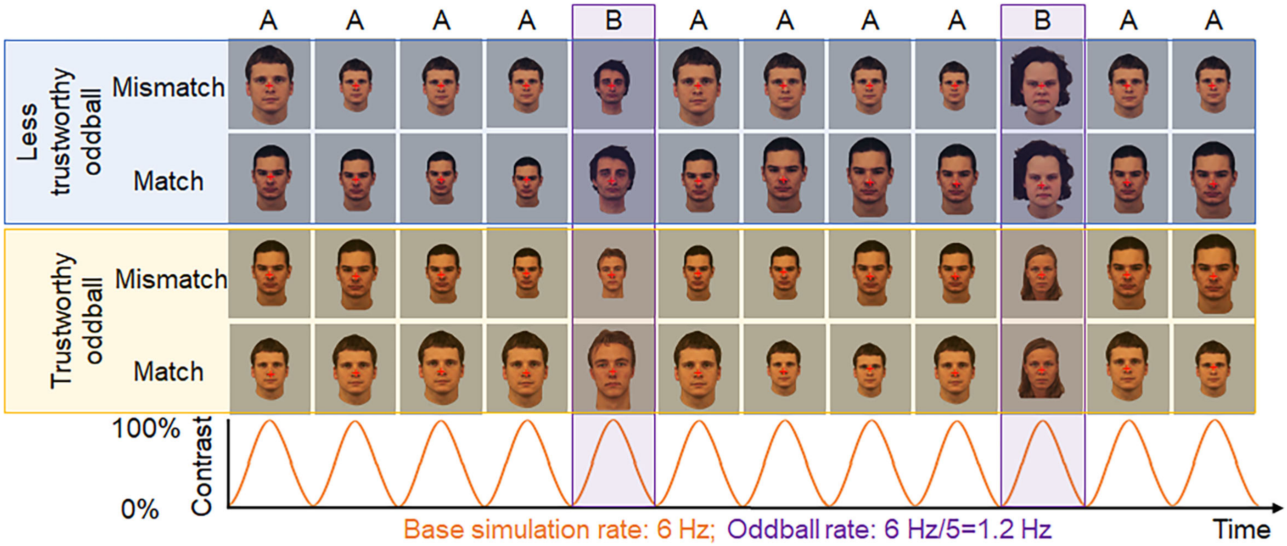

Figure 1. Experimental design. Participants viewed sequences of faces where a base face (A) was presented repeatedly at a rate of 6 Hz. Every fifth face, the base face was replaced by an oddball face (B). The oddball faces were either less trustworthy-looking (top two rows) or trustworthy-looking (bottom two rows). The base faces were also less trustworthy-looking or trustworthy-looking, such that they either matched (rows 2 and 4) or did not match (rows 1 and 3) the oddball faces in trustworthiness. This figure includes Karolinska Directed Emotional Faces set image numbers AF16NES, AF19NES, AM07NES, AM08NES, AM31NES, and AM33NES. Permission to publish these images in scientific papers can be found online at the following address http://kdef.se/home/using%20and%20publishing%20kdef%20and%20akdef.html.

was calculated by first converting the images to grayscale using Matlab's rgb2gray function. Neither the mean luminance nor the mean standard deviation of the luminance differed for the less trustworthy-looking versus trustworthy-looking sets of faces, $t < 1$ in both cases. The two sets of faces did not differ on the mean values for the red, green, or blue channels either, $t < 1$ in all cases. Finally, as a measure of the size of the cropped images, we calculated the number of pixels in each face. Again, the two sets of faces did not differ on this measure, $t < 1$.

Procedure

Stimuli were presented using PsychoPy (Peirce et al., 2019). Each trial started with a 2- to 5-second fixation cross, followed by a 60-second sequence of faces. The faces were presented at a rate of 6 Hz, which meant that each face was shown for 166.67 ms. During the presentation of each face, the contrast values of the face were sinusoidally modulated so that the face smoothly faded into and out of visibility. The contrast values went from zero at the start of the time period to full contrast during the middle of the time period, and back to zero at the end of the time period. To minimize the effect of low-level visual cues, the size of the faces was randomly varied from the presentation of one face to the next, ranging in steps of 2% from 74% to 120% of the original image size (Dzhelyova & Rossion, 2014; Liu-Shuang et al., 2014). At the end of the trial, the maximum contrast value for the face was gradually decreased from 100% to 0% so that the faces faded away. This was done to prevent abrupt eye movements at the end of the trial (Liu-Shaung et al., 2014).

At the beginning of each trial, one face was randomly selected to serve as the base face for the trial. Within the trial, the base face was presented repeatedly. Every fifth face, the base face was replaced by an oddball face of a different identity (i.e., AAAABAAAAC, etc.). The base face was either less trustworthy-looking or trustworthy-looking. For a given trial, the oddball faces were all less trustworthy-looking or all trustworthy-looking, so that they either matched the base face in their trustworthiness or did not match. This meant the experiment had a 2 (trustworthiness of base face) × 2 (match in trustworthiness between base/oddball faces) repeated measures design (Figure 1).

Neither the trustworthiness of the faces nor the periodic changes in face identity were mentioned to participants. Instead, participants were told that they would see a series of faces, but that their task was to attend to the color of a fixation cross in the middle of the screen. They were instructed to press a key whenever the color of the fixation cross changed from blue to red, which happened at eight random times during each trial. The color changes were brief, lasting only 200 ms (Liu-Shuang et al., 2014). As would be expected given the frequency of oddball faces, color changes were later found to have occurred on approximately 20% of oddball trials ($M = 19.10$, $SD = 4.60$). The frequency of the color changes during oddball trials did not differ across conditions ($F < 1$ in all cases). Immediately before the experiment, participants completed an unrelated experiment that involved the same color detection task, which meant they were familiar with it at the beginning of the experiment.

EEG Acquisition

Continuous EEG data were recorded using a Brain Vision actiChamp amplifier system (Brain Products GmbH, Munich, Germany). Data were recorded from 32 active channels mounted in a nylon cap according to the 10-20 electrode system (actiCap slim, Brain Products GmbH). One of the 32 electrodes, electrode FT9, was placed under the left eye to monitor eye blinks and therefore no data from this location are reported. To have a symmetrical electrode arrangement, data from FT10, the complementary electrode in the right hemisphere, were excluded from analysis. There was a ground electrode at Fpz. Before recording, impedances below 10 kΩ were obtained. During recording, all electrodes were referenced to electrode TP9. Data were sampled at 500 Hz.

EEG Analysis

Preprocessing

EEG data were analyzed using EEGLAB (Delorme & Makeig, 2004), in conjunction with the ERPLAB plugin (Lopez-Calderon & Luck, 2014) and Letswave (Mouraux & Iannetti, 2008). The data were first imported into EEGLAB and the subocular channel was removed. The data were then band-pass filtered between 0.1 Hz and 100 Hz using a fourth-order Butterworth filter. All channels were rereferenced to an average reference, and electrode TP9, the previous reference channel, was added back into the data set.

An average reference was used to make the results of the current experiment more directly comparable to those of previous experiments using FPVS to investigate face processing (e.g., Liu-Shuang et al., 2014). However, because the number of electrodes used in the current study is on the low end for studies employing an average reference, we also conducted a separate analysis following the same steps using a reference of Pz. The results of this analysis were extremely similar to the results of the analysis using the average reference (see Supplementary Materials).

After rereferencing, the continuous data were segmented relative to the beginning and end of each trial. The first 2 seconds of each trial were discarded to remove any transient responses due to the onset of the trial. This meant that segmentation yielded twelve 58-second segments, each of which contained 348 face presentations. For each participant, independent component analysis decomposition was run on these 12 segments and the component corresponding to eyeblinks was identified and removed. At most, a single component was removed for each participant. Channels with artifacts other than eyeblinks were estimated from neighboring channels using linear interpolation. At most, two channels were interpolated for each participant. Finally, to minimize activity that was not due to visual stimulation, the three trials for each condition for each participant were averaged together.

Frequency domain analysis

The averaged time courses for each participant were imported into Letswave. The time courses were downsampled to 250 Hz to reduce file size. A fast Fourier transform (FFT) was then applied separately to each time course, yielding amplitude spectra with a frequency resolution of 0.017 Hz (1/58).

Group-level z-scores were used to determine which harmonics to include in subsequent analyses (Liu-Shuang et al., 2014). To calculate these z-scores, the amplitude spectra for each condition were averaged together across participants. For each relevant frequency, the difference between the amplitude for that frequency and the mean of the 20 surrounding bins (10 on each side, excluding immediately adjacent bins to avoid spectral leakage) was divided by the standard deviation of the surrounding bins. Scores of greater than 3.1 ($p < 0.001$, one-tailed with signal > noise) were considered significant.

Following recent studies (e.g., Dzhelyova & Rossion, 2014; Liu-Shuang et al., 2016, Xu et al., 2017), significant responses were quantified using baseline-corrected amplitudes. For each frequency of interest, the baseline-corrected amplitude was calculated by taking the difference between the amplitude for that frequency and the mean amplitude of the 20 surrounding bins (again excluding the immediately adjacent bins). One benefit of this measure is that it allows responses to be expressed in microvolts (μV). Baseline-corrected amplitudes were calculated separately for each condition for each participant.

Responses were examined in bilateral occipitotemporal and medial occipital regions of interest (ROIs). These ROIs were defined based on previous studies (e.g., Liu-Shuang et al., 2014; Xu et al., 2017) and on the electrodes in the current cap layout. The right occipitotemporal (rOT) ROI included electrodes P8 and TP10, the left occipitotemporal (lOT) ROI included the complementary electrodes in the left hemisphere (P7 and TP9), and the medial occipital ROI included electrodes O1, O2, and Oz.

The general visual response was quantified by summing together baseline-corrected amplitudes at significant harmonics of 6 Hz. Similarly, the face individuation response was quantified by summing together baseline-corrected amplitudes at significant harmonics of 1.2 Hz. For the face individuation response, the fifth harmonic (i.e., 6 Hz) was excluded from the sum because it was confounded with the general visual response. Within the ROIs, the average baseline-corrected amplitude across the electrodes in the ROI was calculated at each harmonic before summing the amplitudes of the significant harmonics together. Mean responses were compared across conditions using repeated-measures analyses of variance.

To ensure that the method for selecting harmonics did not inadvertently bias the results toward one condition, a supplementary analysis summing together the response at the first three harmonics was conducted. This analysis yielded a pattern of significance identical to the pattern observed when summing together the significant harmonics. Given the similarity between this analysis and the main analysis, these results are not reported.

Results

Behavior

Data from one participant were excluded from further analysis due to low response accuracy (>3 SDs below the mean). Mean accuracy for the remaining participants was high ($M = 93.39$, $SD = 8.43$). Although participants' task did not involve the faces, mean accuracy was significantly greater for sequences with less trustworthy-looking ($M = 94.47$, $SD = 7.91$) versus trustworthy-looking oddball faces ($M = 92.30$, $SD = 9.39$, $F(1, 22) = 6.19$, $p = 0.02$, $\eta_p{}^2 = 0.22$). Neither the match in trustworthiness between the base face and oddball faces nor the interaction between match and the oddball type had a significant effect on accuracy (match: $F(1, 22) = 2.10$, $p = 0.16$, $\eta_p{}^2 = 0.09$; oddball type × match: $F < 1$).

The mean reaction time did not differ between conditions (overall: $M = 468.13$, $SD = 49.13$; oddball type: $F(1, 22) = 1.74$, $p = 0.20$, $\eta_p{}^2 = 0.07$; match between base and oddball face: $F < 1$; oddball type × match: $F(1, 22) = 1.14$, $p = 0.30$, $\eta_p{}^2 = 0.05$).

EEG Data

Significance of responses

There were clear responses at 1.2 Hz, 6 Hz, and their harmonics. Looking across all channels, the general visual response was significant at 6 Hz and it remained significant for all conditions until the fourth harmonic (24 Hz). Looking within the three ROIs, harmonics from 6 to 24 Hz were significant for all conditions within the lOT and rOT ROIs, and harmonics from 6 to 30 Hz were significant within the medial occipital ROI. Once again looking across all channels, the face individuation response was significant at 1.2 Hz and it remained significant for all conditions until the sixth harmonic (7.2 Hz). This was also the case in the three ROIs.

In subsequent analyses, the general visual response was quantified by summing together the response at harmonics from 6 to 24 Hz for analyses in the lOT and rOT ROIs, and the response at harmonics from 6 to 30 Hz in the medial occipital ROI. The face individuation response was quantified by summing together the response at harmonics from 1.2 to 7.2 Hz, excluding 6 Hz, within each ROI.

Response at 6 Hz and its harmonics

The general visual response at 6 Hz and its harmonics reflects the response to the appearance of the face stimuli on the background. The general visual response did not differ significantly across conditions in the lOT (oddball type: $F(1, 23) = 3.41$, $p = 0.08$, $\eta_p{}^2 = 0.13$, $F < 1$ for the other effects), medial occipital (oddball type: $F(1, 23) = 3.13$, $p = 0.09$, $\eta_p{}^2 = 0.12$; match: $F < 1$; oddball type X match: $F(1, 23) = 2.48$, $p = 0.13$, $\eta_p{}^2 = 0.10$), or rOT ROIs ($F < 1$ in all cases).

Response at 1.2 Hz and its harmonics

There was a significantly stronger face individuation response for sequences with less trustworthy-looking (lOT: $M = 0.64$, $SD = 0.48$; medial occipital: $M = 0.86$, $SD = 0.39$; rOT: $M = 0.99$ $SD = 0.32$) versus trustworthy-looking oddball faces in all three ROIs (lOT: $M = 0.52$, $SD = 0.44$, $F(1, 23) = 13.53$, $p = 0.001$, $\eta_p{}^2 = 0.37$; medial occipital: $M = 0.74$, $SD = 0.40$, $F(1, 23) = 10.92$, $p = 0.003$, $\eta_p{}^2 = 0.32$; rOT: $M = 0.86$, $SD = 0.32$, $F(1, 23) = 17.42$, $p < 0.001$, $\eta_p{}^2 = 0.43$; Figures 2 and 3). In contrast, the match in trustworthiness between the base face and the oddball faces did not have a significant effect in any of the ROIs, nor did the interaction between the oddball type and match ($F < 1$ in all cases).

To investigate the strength of the effect of oddball type, we examined whether the effect of oddball would remain significant with only half of the data. Because the factor of match provided a natural way of dividing the data, we looked separately at the simple effect of oddball type for sequences where the base face and oddball faces differed in trustworthiness versus had the same trustworthiness. When the base face and oddball faces differed in trustworthiness, there was a significantly stronger face individuation

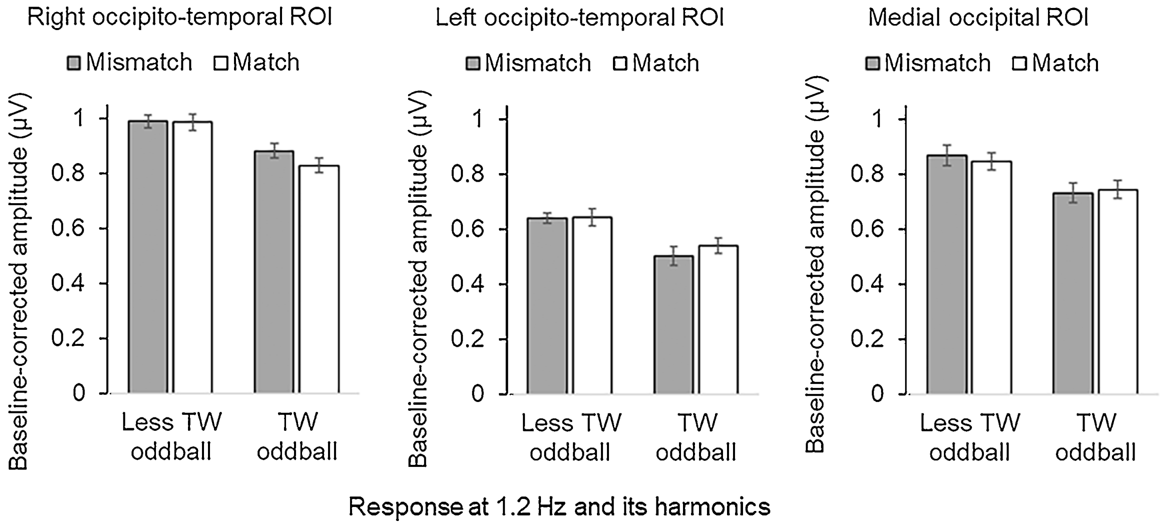

Figure 2. Response at 1.2 Hz and its harmonics in three regions of interest (ROIs). The response at 1.2 Hz and its harmonics was calculated by summing together baseline-corrected amplitudes for the significant harmonics of 1.2 Hz, with the exception of the response at 6 Hz. Mean responses are shown for sequences with less trustworthy-looking (less TW oddball) and trustworthy-looking oddball faces (TW oddball), where the base face and the oddball faces differed in trustworthiness (gray) or had the same trustworthiness (white) for A) right occipitotemporal channel locations P8 and TP10, B) left occipitotemporal channel locations P7 and TP9, and C) medial occipital channel locations O1, O2, and Oz. Error bars represent standard error of the mean.

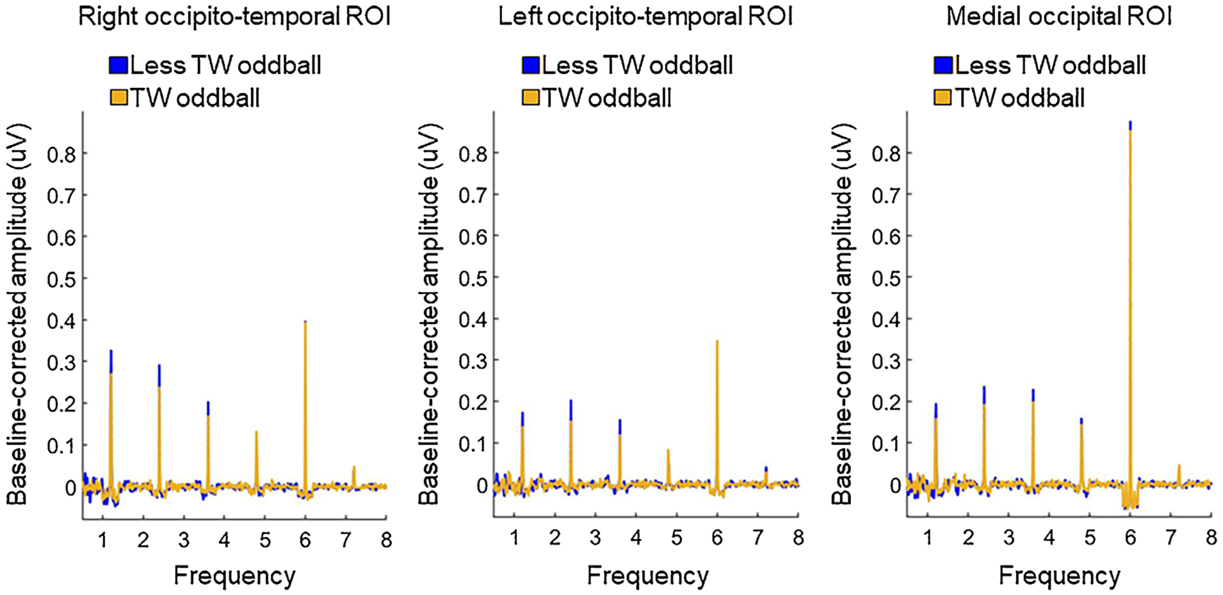

Figure 3. Frequency spectra by oddball type for the three ROIs. The graphs show the amplitude spectra for sequences with less trustworthy-looking (less TW oddball) and trustworthy-looking oddball faces (TW oddball) for each ROI, averaged across the match in trustworthiness between the base face and the oddball faces, the channels in each ROI, and all participants. The peaks in each graph represent significant responses at 1.2 Hz and its harmonics. The parts of the peaks that do not overlap (in blue) represent larger responses for sequences with less trustworthy-looking versus trustworthy-looking oddball faces. The response at 1.2 Hz and its harmonics was quantified by summing together the baseline-corrected amplitudes for the significant harmonics of 1.2 Hz, with the exception of the response at 6 Hz (see Figure 2).

response for sequences with less trustworthy-looking versus trustworthy-looking oddball faces in all three ROIs (lOT: $t(23) = 3.14$, $p = 0.005$; medial occipital: $t(23) = 2.50$, $p = 0.02$; rOT: $t(23) = 2.53$, $p = 0.02$). This was also the case for sequences where the base face and the oddball faces had the same level of trustworthiness (lOT: $t(23) = 2.15$, $p = 0.04$; medial occipital: $t(23) = 2.59$, $p = 0.02$; rOT: $t(23) = 3.33$, $p =$

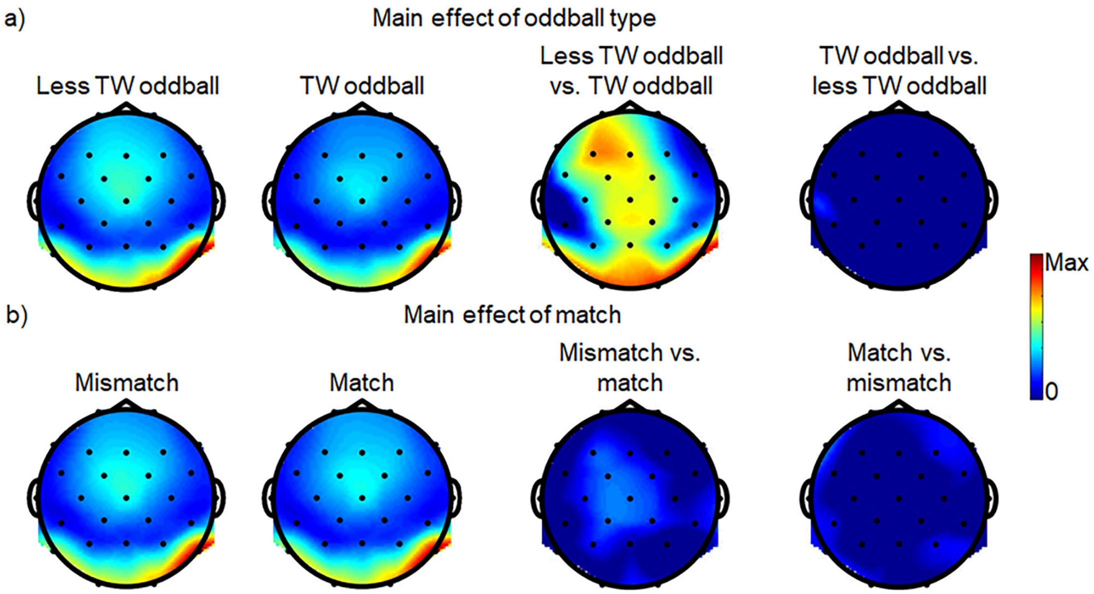

Figure 4. Scalp maps of the response at 1.2 and its harmonics by oddball type and match. The response at 1.2 Hz and its harmonics was calculated by summing together baseline-corrected amplitudes for harmonics that were significant averaging across all channels (i.e., 1.2–4.8 Hz). (a) Scalp maps for sequences with less trustworthy-looking (less TW oddball) and trustworthy-looking oddball faces (TW oddball), and the differences between the two types of sequences. (b) Scalp maps for sequences where the base face and oddball faces differed in trustworthiness (mismatch) or had the same trustworthiness (match), and the differences between the two. The color scale for the less trustworthy-looking, trustworthy-looking, mismatch, and match maps reflects a range from zero to the maximum value across all four maps. Similarly, the color scale for the difference maps reflects a range from zero to the maximum value across all four maps.

0.003). Thus, the effect of oddball type was significant with only half of the data.

Visual inspection of topographic maps revealed that the effect of oddball type was widespread. As can be seen in Figure 4a, the strongest face individuation responses were observed over right occipitotemporal sites. Comparison of the topographic maps for sequences with less trustworthy-looking versus trustworthy-looking oddball faces showed a main effect of oddball type over posterior regions of the scalp, including bilateral occipitotemporal sites and medial occipital sites, as well as more anterior regions. In contrast, no regions showed a main effect of match.

Discussion

The trustworthiness of the oddball faces had a strong influence on the face individuation response at 1.2 Hz and its harmonics. There was a stronger response to sequences with less trustworthy-looking versus trustworthy-looking oddball faces over bilateral occipitotemporal sites and medial occipital sites. Moreover, the simple effect of oddball type was significant when we split the data along the factor of match, looking separately at sequences where the base face and oddball faces differed in trustworthiness versus had the same trustworthiness. While the entire experiment consisted of 12 minutes of recorded data, this meant the effect of oddball type was significant with only 6 minutes of data.

Although the trustworthiness of the oddball faces influenced responses over right occipitotemporal electrode sites, it also influenced responses over left occipitotemporal and medial occipital sites. In fact, visual inspection of topographic maps revealed that the effect of oddball type was not limited to posterior regions of the scalp. While both lesion studies and fMRI studies have found that the amygdala plays an important role in trustworthiness judgments from faces (e.g., Adolphs et al., 1998; Engell et al., 2007; Winston et al., 2002), EEG has only limited ability to detect activity from subcortical sources and therefore it is unlikely that the observed responses directly reflect amygdala activity. However, consistent with fMRI evidence suggesting that the amygdala response to face trustworthiness modulates activity elsewhere in the brain (Todorov & Engell, 2008), the current results could instead indirectly reflect the amygdala response. More broadly, rather than speaking to the origin of the effect, the current data demonstrate that it is possible to quickly detect implicit responses to face trustworthiness using EEG.

Importantly, the differential response at 1.2 Hz and its harmonics for sequences with less trustworthy-looking versus trustworthy-looking oddball faces was observed even though participants' attention was not directed to the trustworthiness of the faces. However, although the behavioral task did not require participants to attend to the faces, accuracy was significantly higher for the sequences with the less trustworthy-looking versus trustworthy-looking oddball faces. Because fixation color changes could occur when either the base face or the oddball faces were present, it is possible that this difference in accuracy was driven by differences in attention when the oddball faces were present. At the same time, the trustworthiness of the oddball faces had only a minimal influence on the general visual response at 6 Hz. Since overall differences in attention would be expected to influence both behavior and the general visual response, this pattern of results suggests that any differences across conditions were subtle.

In contrast with the strong effect of oddball type, the match in trustworthiness between the base face and the oddball faces had very little effect on the response at 1.2 Hz and its harmonics. Although we had expected to see a stronger face individuation response for sequences of faces where the base face and oddball faces differed in trustworthiness, this was not the case. One factor that might explain the lack of effect is the number of times the base face was shown. While recent studies using FPVS (Luo et al., 2019) and ERPs (Kovacs-Balint et al., 2014) provide support for the hypothesis that the contrast between the base face and oddball faces should be important, the base face in the current study was repeated more frequently than the faces in those studies. Because repeated exposure influences stimulus evaluation (Bornstein, 1989; Zajonc, 1968), it is possible that repeated viewing of the base faces in the current study led the less trustworthy-looking and trustworthy-looking base faces to be seen as similar to each other, thereby diminishing any effect of match on the face individuation response.

Being able to quickly detect implicit responses to face trustworthiness will be useful in circumstances where it is not possible to collect explicit judgments, such as in research with infants or young children. It will also be useful for learning more about the nature of implicit neural responses to face trustworthiness. For example, it is not clear whether the size of these neural responses relates to the perceived trustworthiness of faces. Recently, Xu and colleagues (2017) found that the size of the face individuation response as measured via FPVS correlates with individual differences in face recognition ability, demonstrating that it is possible to use FPVS to index individual differences in face processing. With face evaluation, it is possible that participants who show larger trustworthiness oddball effects will also show larger differences in the perceived trustworthiness of faces. Relatedly, future work could also investigate whether different psychological states, such as a state of threat, influence implicit responses to face trustworthiness.

Although the faces in the current study were selected based on trustworthiness ratings, it is worth keeping in mind that they differed on nearly every other trait dimension examined as well. Therefore, it is not clear whether the observed effects are specifically due to face trustworthiness. Given that trait judgements tend to be highly correlated (Oosterhof & Todorov, 2008), this is likely to be the case for most other studies of face trustworthiness as well, unless the studies explicitly control for particular trait dimensions. However, it is useful to focus on face trustworthiness because face trustworthiness judgments have been found to correlate highly with a valence dimension underlying much of the variance in trait judgments from faces (Oosterhof & Todorov, 2008; Sutherland et al., 2013). This means that trustworthiness judgments capture the valence evaluation of faces and a large amount of what is shared across trait judgments.

The less trustworthy-looking and trustworthy-looking sets of faces did not differ on measures of luminance, color, or size. However, it is nonetheless possible that the effect of oddball type was driven, at least in part, by differences along a dimension unrelated to face trustworthiness. Future work could address this concern by comparing the response to faces that have been manipulated to have different levels of face trustworthiness. However, if such a study were to use a design similar to the one used in the current experiment, one challenge with this approach would be ensuring that the amount of identity change was kept constant for faces with the same versus different levels of trustworthiness.

A factor that is likely to underlie the differential response to less trustworthy-looking versus trustworthy-looking oddball faces is face typicality. Face typicality has been found to correlate with trait judgments from faces (e.g., Dotsch, Hassin, & Todorov, 2016; Sofer, Dotsch, Wigboldus, & Todorov, 2015). For example, in the set of faces used in the current study, face typicality correlates with nearly all of the trait judgments examined (see Figure 7a in Said, Dotsch, & Todorov, 2010). Researchers have investigated the role of typicality in trait judgments by manipulating face typicality, usually using artificial faces or face morphing techniques. In a behavioral study, when face typicality was manipulated by changing the position of a face in a statistical distribution of faces, this influenced trustworthiness judgments (Dotsch et al., 2016). In the brain, activity in the amygdala and the fusiform face area tracks with typicality for faces that vary in trustworthiness, as well as for faces that vary along control dimensions that have been selected to be less closely related to valance (Mattavelli, Andrews, Asghar, Towler, & Young, 2012; Said et al., 2010).

Because valence does not modulate the response in either of these regions, this has led to the proposal that the amygdala is tracking face typicality, and not valence per se (Mattavelli et al., 2012; Said et al., 2010; Todorov, Mende-Siedlecki, & Dotsch, 2013). The typicality hypothesis has the added benefit of also being able to explain the amygdala response to other face properties beyond trait judgments, such as novelty and emotion.

An open question is whether responses to face trustworthiness as measured using FPVS are similar to responses to angry versus happy expressions. Similarities in the processing of face trustworthiness and emotional expressions have previously been noted in studies using ERPs (Dzhelyova et al., 2012; Marzi et al., 2014; Rudoy & Paller, 2009; Yang et al., 2011) and in a meta-analysis of fMRI studies (Mende-Siedlecki et al., 2013). Although several studies have used FPVS to investigate responses to emotional facial expressions (Dzhelyova, Jacques, & Rossion, 2017; Gerlicher, Loon, Scholte, Lamme, & van der Leij, 2014; Zhu, Alonso-Prieto, Handy, & Barton, 2016), none of them have directly compared angry versus happy expressions. However, one of these studies did find spatially distinct responses to oddball faces with disgust, fear, and happy expressions, with more similar responses for the disgust and fear expressions (Dzhelyova et al., 2017). While the authors suggest that this similarity may be because disgust and fear are both avoidance-related emotions, the current results with face trustworthiness suggest that anger and happiness might yield dissociable responses, despite both being approach-related emotions.

In summary, participants in the current study viewed sequences of faces made up of a repeated base face and oddball faces of different identities. The trustworthiness of the oddball faces had a widespread influence on the face individuation response at 1.2 Hz and its harmonics. A stronger response to sequences with less trustworthy-looking versus trustworthy-looking oddball faces was observed after a recording time of only a few minutes, despite the fact that participants were engaged in a task that did not involve attending to the trustworthiness of the faces. These results suggest that FPVS offers an efficient means of measuring implicit responses to face evaluation.

Keywords: face individuation, face trustworthiness, fast periodic visual stimulation, implicit responses, oddball paradigm

Acknowledgments

Commercial relationships: none.
Corresponding author: Sara Verosky.

Email: sverosky@oberlin.edu.
Address: Department of Psychology, Oberlin College, 120 West Lorain, Oberlin, OH 44074.

*Equal contributions, order is alphabetical.

References

Adams, R. B., Jr., Nelson, A. J., Soto, J. A., Hess, U., & Kleck, R. E. (2012). Emotion in the neutral face: A mechanism for impression formation? *Cognition and Emotion, 26*, 431–441.

Adolphs, R., Tranel, D., & Damasio, A. R. (1998). The human amygdala in social judgment. *Nature, 393*, 470–474.

Alonso-Prieto, E., Van Belle, G., Liu-Shuang, J., Norcia, A. M., & Rossion, . (2013). The 6 Hz fundamental stimulation frequency rate for individual face discrimination in the right occipito-temporal cortex. *Neuropsychologia, 51*, 2863–2875.

Bentin, S., Allison, T., Puce, A., Perez, E., & McCarthy, G. (1996). Electrophysiological studies of face perception in humans. *Journal of Cognitive Neuroscience, 8*, 551–565.

Bornstein, R. F. (1989). Exposure and affect: overview and meta-analysis of research, 1968-1987. *Psychological Bulletin, 106*, 265–289.

Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics. *Journal of Neuroscience Methods, 134*, 9–21.

Dotsch, R., Hassin, R. R., & Todorov, A. (2016). Statistical learning shapes face evaluation. *Nature Human Behavior, 1*, Article 0001.

Dzhelyova, M., Jacques, C., & Rossion, B. (2017). At a single glance: Fast periodic visual stimulation uncovers the spatio-temporal dynamics of brief facial expression changes in the human brain. *Cerebral Cortex, 27*, 4106–4123.

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

Dzhelyova, M., & Rossion, B. (2014). The effect of parametric stimulus size variation on individual face discrimination indexed by fast periodic visual stimulation. *BMC Neuroscience, 15*, Article 87.

Engell, A. D., Haxby, J. V., & Todorov, A. (2007). Implicit trustworthiness decisions: automatic coding of faces properties in the human amygdala. *Journal of Cognitive Neuroscience, 19*, 1508–1519.

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.

Gerlicher, A. M. V., van Loon, A. M., Scholte, H. S., Lamme, V. A. F., & van der Leij, A. R. (2014). Emotional facial expressions reduce neural adaptation to facial identity. *Social Cognitive and Affective Neuroscience, 9*, 610–614.

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.

Klapper, A., Dotsch, R., van Rooij, I., & Wigboldus, D. H. J. (2016). Do we spontaneously form stable trustworthiness impressions from facial appearance? *Journal of Personality and Social Psychology, 111*, 655–664.

Kovacs-Balint, Z., Stefanics, G., Trunk, A., & Hernadi, I. (2014). Automatic detection of trustworthiness of the face: a visual mismatch negativity study. *Acta Biologica Hungarica, 65*, 1–12.

Liu-Shuang, J., Norcia, A. M., & Rossion, . (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.

Liu-Shuang, J., Torfs, K., & Rossion, B. (2016). An objective electrophysiological marker of face individualization impairment in acquired prosopagnosia with fast periodic visual stimulation. *Neuropsychologia, 83*, 100–113.

Lopez-Calderon, J., & Luck, S. J. (2014). ERPLAB: An open-source toolbox for the analysis of event-related potentials. *Frontiers in Human Neuroscience, 8*, Article 213.

Lundqvist, D., Flykt, A., & Ohman, A. (1998). *The Karolinska Directed Emotional Faces (KDEF)*. Stockholm: Department of Neurosciences Karolinska Hospital.

Luo, Q., Rossion, B., & Dzhelyova, M. (2019). A robust implicit measure of facial attractiveness discrimination. *Social Cognitive and Affective Neuroscience, 14*, 737–746.

McCarthy, G., Puce, A., Gore, J. C., & Allison, T. (1997). Face-specific processing in the human fusiform gyrus. *Journal of Cognitive Neuroscience, 9*, 605–610.

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.

Mattavelli, G., Andrews, T. J., Asghar, A. U. R., Towler, J. R., & Young, A. W. (2012). Response of face-selective brain regions to trustworthiness and gender of faces. *Neuropsychologia, 50*, 2205–2211.

Mende-Siedlecki, P., Said, C. P., & Todorov, A. (2013). The social evaluation of faces: a meta-analysis of functional neuroimaging studies. *Social Cognitive and Affective Neuroscience, 8*, 285–299.

Mouraux, A., & Iannetti, G. D. (2008). Across-trial averaging of event-related EEG responses and beyond. *Magnetic Resonance Imaging, 26*, 1041–1054.

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

Oosterhof, N. N., & Todorov, A. (2009). Shared perceptual basis of emotional expressions and trustworthiness impressions from faces. *Emotion, 9*, 128–133.

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., … Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods, 51*, 195–203.

Regan, D. (1966). Some characteristics of average steady-state and transient responses evoked by modulated light. *Electroencephalography and Clinical Neurophysiology, 20*, 238–248.

Rossion, B. (2014). Understanding individual face discrimination by means of fast periodic visual stimulation. *Experimental Brain Research, 232*, 1599–1621.

Rossion, B., & Boremanse, A. (2011). Robust sensitivity to facial identity in the right human occipito-temporal cortex as revealed by steady-state visual-evoked potentials. *Journal of Vision, 11*, 1–21.

Rudoy, J. D., & Paller, K. A. (2009). Who can you trust? Behavioral and neural differences between perceptual and memory-based influences. *Frontiers in Human Neuroscience, 3*, Article 16.

Rule, N. O., Ambady, N., & Adams, R. B., Jr.. (2009). Personality in perspective: Judgmental consistency across orientations of the face. *Perception, 38*, 1688–1699.

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.

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*, 39–47.

Sutherland, C. A. M., Oldmeadow, J. A., Santos, I. M., Towler, J., Burt, D. M., & Young, A. W. (2013). Social inference from faces: Ambient images generate a three-dimensional model. *Cognition, 127*, 105–118.

Todorov, A., & Engell, A. D. (2008). The role of the amygdala in implicit evaluation of emotionally neutral faces. *Social Cognitive and Affective Neuroscience, 3*, 303–312.

Todorov, A., Mende-Siedlecki, P., & Dotsch, R. (2013). Social judgments from faces. *Current Opinion in Neurobiology, 23*, 373–380.

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. (2009). Evaluating faces on trustworthiness after minimal time exposure. *Social Cognition, 27*, 813–833.

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

Winston, J. S., Strange, B. A., O'Doherty, J., & Dolan, R. J. (2002). Automatic and intentional brain responses during evaluation of trustworthiness of faces. *Nature Neuroscience, 5*, 277–283.

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.

Yang, D., Qi, S., Ding, C., & Song, Y. (2011). An ERP study on the time course of facial trustworthiness appraisal. *Neuroscience Letters, 496*, 147–151.

Zajonc, R. B. (1968). Attitudinal effects of mere exposure. *Journal of Personality and Social Psychology, 9*, 1–27.

Zebrowitz, L. A., Kikuchi, M., & Fellous, J. M. (2010). Facial resemblance to emotions: Group differences, impression effects, and race stereotypes. *Journal of Personality and Social Psychology, 98*, 175–189.

Zebrowitz, L. A., & Montepare, J. M. (2008). Social psychological face perception: Why appearance matters. *Social and Personality Psychology Compass, 2*, 1497–1517.

Zhu, M., Alonso-Prieto, E., Handy, T., & Barton, J. (2016). The brain frequency tuning function for facial emotion discrimination: An ssVEP study. *Journal of Vision, 16*, 1–14.