RUNNING HEAD: Visual Context and Facial Trustworthiness

The Influence of Visual Context on the Evaluation of Facial Trustworthiness

Marco Brambilla, Marco Biella

University of Milano-Bicocca

Jonathan B. Freeman

New York University

Word count (excluding references and the abstract): 5673

Authors' Note: This work was supported by a Grant from the Italian Ministry of Education, University, and Research to the first author (FIRB: RBFR128CR6). The first two authors contributed equally to this article. Correspondence concerning this article should be addressed to Marco Brambilla, University of Milano-Bicocca, Department of Psychology, Piazza dell'Ateneo Nuovo, 1, 20126 – Milano (Italy). E-mail: marco.brambilla@unimib.it.

Abstract

Evaluation of facial trustworthiness is often thought to be based on facial features and relatively immune to visual context. However, we rarely encounter an isolated facial expression in the real world. In 3 Experiments using a mouse-tracking paradigm, participants were asked to categorize the trustworthiness of faces that were shown against either threatening, negative but unthreatening, or neutral scenes. Results showed that visual scenes systematically altered the categorization of facial trustworthiness. The trajectory of hand movements reflected the compatibility of facial trustworthiness and contextual threat cues of the scene. Trajectories were facilitated when facial cues and contextual cues were compatible (e.g., untrustworthy face in a threatening scene), and were partially attracted to the context-associated response when incompatible (e.g., trustworthy face in a threatening scene). Thus, the evaluation of facial trustworthiness involves dynamic updates of gradual integration of the face and the level of threat posed by the visual context.

Keywords: Trustworthiness, Face Perception; Threat; Mouse-Tracking

The Influence of Visual Context on the Evaluation of Facial Trustworthiness

Our impressions of others are often based on limited information that is spontaneously and automatically extracted from their appearance—in particular their faces (Zebrowitz, 1997; Zebrowitz & Montepare, 2008). Indeed, a growing body of research has shown that people make personality inferences from faces after minimal time exposure (Bar, Neta, & Linz, 2006; Todorov, Pakrashi, & Oosterhof, 2009; Todorov & Uleman, 2003; Willis & Todorov, 2006) and that these evaluations predict important social outcomes. For instance, inferences of dominance predict military rank attainment (Mazur, Mazur, & Keating, 1984; Mueller & Mazur 1996), while inferences of competence predict the results of political elections (Ballew & Todorov 2007; Todorov, Mandisodza, Goren, & Hall, 2005). In addition, facial dominance and competence together predict salaries of CEOs (Rule & Ambady 2008).

An important class of inferences concerns judgments of trustworthiness (Todorov, Olivola, Dotsch, & Mende-Siedlecki, 2015). Studies on economic games have shown that players are less willing to trust other players who have untrustworthy-looking faces (Chang, Doll, van't Wout, Frank, & Sanfey, 2010; Rezlescu, Duchaine, Olivola, & Chater, 2012; Stirrat & Perrett 2010) while recent experimental work reveals that defendants who have untrustworthy-looking faces are more likely to receive guilty verdicts (Porter, ten Brinke, & Gustaw, 2010; Wilson & Rule, 2015). Importantly, it has been shown that people start discriminating trustworthiness after 33 ms of exposure to a face and that the detection of trustworthiness in faces is faster than the detection of a variety of other characteristics, including competence, likeability, and dominance (Willis & Todorov, 2006; Todorov et al., 2009). In a similar vein, people show a memory advantage for faces varying on trustworthiness compared with those varying on likeability, friendliness, and dominance (Rule, Slepian, & Ambady, 2012) and facial trustworthiness predicts basic approach/avoidance responses (Slepian, Young, Rule, Weisbuch, & Ambady, 2012).

Such a preferential processing of facial trustworthiness has often been explained through a functionalist perspective. Indeed, our judgments of another person's trustworthiness are highly related to the essential decision we must make about whether they represent an opportunity or a threat (Ames, Fiske, & Todorov, 2011; Brambilla & Leach, 2014; Cosmides & Tooby, 1992). In line with this reasoning, it has been shown that perceived trustworthiness and threat are inherently linked. As such, behavioral studies have shown that the more a social target is perceived as untrustworthy, the more such a target is believed to pose a threat to the stability and integrity of the whole community. By contrast, highly trustworthy social targets are perceived as beneficial for the group survival and cohesion (Brambilla & Leach, 2014). At the group level, untrustworthy ingroup members are perceived as threatening to the image of their group (Brambilla, Sacchi, Pagliaro, & Ellemers, 2013; Leach, Ellemers, & Barreto, 2007; van der Toorn, Ellemers, & Doosje, 2015), while untrustworthy outgroup members are perceived as posing a real and a concrete danger to the ingroup's survival possibilities and represent a threat to the group's safety (Brambilla et al, 2013; Brambilla, Sacchi, Rusconi, Cherubini, & Yzerbyt, 2012; Leidner & Castano, 2012). In line with these findings, functional neuroimaging studies show that detection of trustworthiness in a face is a spontaneous, automatic process linked to activity in the amygdala (Winston, Strange, O'Doherty, & Dolan, 2002), a subcortical brain structure that tends to be implicated in the detection of potentially dangerous and threatening stimuli (Adolphs, 2010; Engell, Haxby, & Todorov, 2007; Freeman, Stolier, Ingbretsen, & Hehman, 2014; Todorov, Mende-Siedlecki, & Dotsch, 2013; Todorov, Said, Oosterhof, & Engell, 2011; Phelps & LeDoux, 2005).

In the vast majority of studies examining facial trustworthiness, faces are flashed on the computer screen, and categorization of trustworthiness quickly ensues (for a review, Todorov et al., 2015). However, faces are rarely encountered in isolation in the real world. Instead, they are typically embedded in rich contexts. For instance, we might catch sight of another person walking

in a park or waiting at the subway station. Recent studies have found that context influences the perception of facial emotions; such studies reveal facilitated response times when the emotional context of the scene and face are congruent (Aviezer, Hassin, Grady, Susskind, Anderson, Moscovitch & Bentin, 2008; Barrett & Kensinger, 2010; Righart & De Gelder, 2008). Thus, disgust, fear, and happiness are more easily recognized when faces are shown against backgrounds of natural scenes with congruent emotional significance (Righart & De Gelder, 2008). Beyond emotion recognition, contextual effects have been examined with respect to static category dimensions as well, such as ethnicity (e.g., Freeman, Ma, Han, & Ambady, 2013; Freeman, Ma, Barth, Young, Han, & Ambady, 2015). For instance, Asian categorization is more likely when an Asian face appears in a Chinese-typed rather than an American-typed scene context.

The present research sought to extend prior work by investigating whether visual context may impact the perception of trustworthiness. Indeed, while prior research has examined contextual effects with respect to emotion recognition and race categorization, hardly any experimental work has examined whether visual context influences the perception of traits, such as trustworthiness. One study found that faces were judged more trustworthy when surrounded by wealthy backgrounds (Keres & Chartier, 2016). In that study, the contextual information conveyed social status. Moreover, that study employed explicit ratings that did not permit an understanding of how facial and contextual cues were integrated during the judgment process. Here, we aimed to examine dynamics underlying the integration of facial trustworthiness and contextual cues, specifically contextual cues that convey threat. In doing so, our research is useful to broaden our understanding of the factors promoting or disrupting the processing of facial trustworthiness. Considering that prior research has shown that facial trustworthiness and the perception of threat are inherently linked (for reviews, Brambilla & Leach, 2014; Todorov et al., 2015), there is good reason to expect that visual scenes associated with threat could alter the processing of a face's trustworthiness.

To test this prediction we went beyond response times and considered a more process-sensitive methodology. Thus, we employed a mouse-tracking technique that records and analyzes hand movements during categorization tasks (Freeman & Ambady, 2010; Freeman & Johnson, 2016). Previous studies examining contextual effects suggest in some cases outcome-based measures (e.g., ratings or reaction times) may have limited sensitivity while more process-based measures such as mouse-tracking overcome this (Freeman, & Johnson, 2016; Freeman et al., 2013). As such, there are many cases where a participant's ultimate perception is not predicted to be altered by context even if the process leading up to that response would be altered considerably. In line with this reasoning, the computer mouse-tracking procedure records the position of the mouse on the x and y coordinate space, providing an online measure of the spontaneous changes across a decision process. In a typical trial, participants are required to click on a “Start” button located at the bottom-center of the screen, which is replaced by a target. Participants then must click an appropriate response button located either at the top-left or top-right of the screen. Because the mouse is moving while a categorization response is still evolving, it is able to provide a “read-out” of how categorization unfolds over time (Freeman & Ambady, 2011; Freeman & Johnson, 2016). In other words, this paradigm can track how various cues drive categorization in real time and therefore reveal potentially subtle influences of context, even when an ultimate response may not be affected.

If the visual context influences the categorization of facial trustworthiness, one would expect that perceivers partially integrate the response associated with the context with that associated with the face. This would be evidenced by a partial attraction in participants' mouse trajectories toward the opposite category response before clicking their final response when the facial and context information do not match. In other words, trajectories would be facilitated when facial cues and contextual cues are compatible (e.g., untrustworthy face in a threatening scene), and would be

partially attracted to the context-associated response when incompatible (e.g., trustworthy face in a threatening scene). We conducted three experiments to test these hypotheses.

Experiment 1

Experiment 1 was designed as a first test of our hypothesis that categorization responses of facial trustworthiness are influenced by the threatening nature of the visual context. To do so, we asked participants to categorize the trustworthiness of faces that were shown against either threatening or neutral backgrounds of natural scenes. We predicted a more direct mouse-trajectory toward the untrustworthy response button when untrustworthy faces are embedded in threatening contexts rather than in a neutral context. By contrast, we expected a more curved mouse-trajectory toward the trustworthy response button when trustworthy faces are embedded in threatening contexts rather than in a neutral context.

Method

Participants

Sample size was determined before the data collection. Specifically, an a priori power analysis was conducted for sample size estimation (using G Power 3.1; Faul, Erdfelder, Lang, & Buchner, 2007). The projected sample size needed to detect a small-to-medium effect size (f=.20; Cohen, 1988) with 80% power is N=36 for a within-subject ANOVA with 4 cells. We advertised the study on campus and all the students who responded within 4 weeks were involved in the study. Overall, we recruited 51 Italian students (36 female) aged between 19 and 75 (M_{age}=28.72, SD=12.83), with normal or corrected-to-normal vision. The sample size was comparable to those employed by previous published works on categorization of faces (Carraro, Castelli, & Negri, 2016; Freeman, 2014; Freeman et al., 2013; Righart & Gelder, 2008). In this and the subsequent studies, we report all measures, manipulations, and exclusions.

Stimuli

We employed 24 computer-generated identities (12 trustworthy, 12 untrustworthy) borrowed from a set of photos previously validated for facial trustworthiness (Todorov, Dotsch, Porter, Oosterhof, & Falvello, 2013). Specifically, trustworthy and untrustworthy faces had the highest and the lowest levels of trustworthiness, respectively. Scene context stimuli (4 neutral, 4 threatening) were obtained from public-domain websites. A pretest confirmed that the scenes were perceived as intended. In particular, independent raters (N = 26; M_{age} =23.80; SD = 2.77) were asked to indicate the extent to which each scene context was threatening using a scale ranging from 1 (*not at all*) to 7 (*extremely*). Pre-test results revealed that threatening scenes were perceived as more threatening (M =5.53, SD =1.31) than neutral scenes (M =1.16, SD =.48), $t(25)=16.61$, $p<.001$, $d=3.25$, 95% CI = [2.27, 4.23]. Importantly, scores of perceived threat were above the midpoint of the scale only for threatening scene, $t(25)=5.95$, $p<.001$, $d=1.16$, 95% CI = [.65, 1.66]. See Figure 1 for sample stimuli[1].

Procedure

Participants were told that they would be presented with images of individuals in various settings, and were asked to categorize each person as either trustworthy or untrustworthy. They were instructed to make their decisions as quickly and accurately as possible by clicking response buttons, basing their judgments on their first impressions. Participants made speeded judgments and were asked to respond within 1500 ms. On every trial, participants clicked a “Start” button at the bottom-center of the screen, which was then replaced by a face-context pair in the center of the screen. Face-context pairs were presented in randomized order, and faces were categorized by clicking a “trustworthy” or “untrustworthy” response button located in the top-left and top-right corners of the screen (counterbalanced across participants). So as to encourage mouse trajectories that are online with the actual decision process, if participants did not start moving the mouse within 250 milliseconds after the face-context pair appeared on the screen, a message advising them

to start moving the mouse earlier was displayed (Hehman, Stolier, & Freeman, 2015). Each face was presented 2 times and placed in the center location of a scene, 1 for each context type, yielding 48 trials per participant.

Results and Discussion

To permit averaging and comparison across trials, we normalized trajectories into 101 time-steps and remapped leftward trajectories rightwards (inverted along the x-axis). To index trajectories attraction toward the opposite category, we computed the maximum deviation (MD): the largest perpendicular deviation from an idealized straight line between the trajectory's start and endpoints (Freeman & Ambady, 2010). We performed a 2 (Scene Context: Neutral, Threatening) × 2 (Face: Trustworthy, Untrustworthy) within-subject ANOVA (Table 1). The main effect of scene context (F<1, p=.64) and face (F<1, p=.45) were not significant. However, there was a significant interaction between scene context and face, $F(1,50)=9.61$, p=.003, η_p^2 =.16. Specifically, untrustworthy faces elicited more direct trajectories (lower MD) when they were embedded in threatening than neutral contexts, $F(1,50)=5.09$, p=.03, η_p^2 =.09. Conversely, trustworthy faces exhibited a marginally significant tendency to elicit more deviating trajectories when they were embedded in threatening than neutral contexts, $F(1,50)=3.53$, p=.067, η_p^2 =.07.

Next, we computed the area under the curve (AUC): the area between the observed trajectory and an idealized straight-line trajectory (Freeman & Ambady, 2010),which is a related measure to MD but in some cases exhibits higher sensitivity (Hehman et al., 2015). We performed a 2 (Scene Context: Neutral, Threatening) × 2 (Face: Trustworthy, Untrustworthy) within-subject ANOVA (Table 2). The main effect of scene context (F<1, p=.74) and face (F<1, p=.40) were not significant. More importantly, the scene context × face interaction was significant, $F(1,50)=12.95$, p=.001, η_p^2 =.21. Untrustworthy faces elicited more direct trajectories (lower AUC) when they were embedded in threatening contexts than in neutral contexts, $F(1,50)=5.99$, p=.02, η_p^2 =.11.

Conversely, trustworthy faces elicited more curved trajectories (higher AUC) when they were embedded in threatening contexts than in neutral contexts, $F(1,50)=5.96$, $p=.02$, $\eta_p^2=.11$.

These findings provide initial evidence that visual context alters the processing of a face's trustworthiness. Indeed, we found that when the threatening nature of the face and context are more compatible, trajectories became more direct en route to the selected response. When they became more incompatible, trajectories showed an increased attraction toward the opposite-category response associated with the context.

Experiment 2

Experiment 2 was designed to replicate and extend the findings of Experiment 1 by investigating whether the effects we found are specific to threatening contexts or indicate more general effects of negative scene contexts. To do so, we included a further experimental condition and asked participants to categorize the trustworthiness of faces that were shown against either threatening, negative but unthreatening, or neutral backgrounds. Specifically, we predicted a more direct trajectory toward the untrustworthy response when untrustworthy faces are embedded in a threatening rather than a neutral context or negative context unrelated to threat. By contrast, we expected a more curved trajectory toward the trustworthy response when trustworthy faces are embedded in a threatening rather than a neutral context or negative context unrelated to threat, indicating a partial attraction to the untrustworthy response and an integration of facial and contextual cues.

Method

Participants

Sample size was determined before the data collection. An a priori power analysis was conducted for sample size estimation. The projected sample size needed to detect a small-to-medium effect size ($f=.20$) with 80% power is $N=28$ for a within-subject ANOVA with 6 cells. We

advertised the study on campus and all students who responded within 4 weeks and who were not involved in Experiment 1 took part to the study. Overall, we recruited 46 Italian students (33 female) aged between 19 and 49 (M_{age}=22.57, *SD*=4.78), with normal or corrected-to-normal vision. Most participants (95.7%) were right handed.

Stimuli

We used the same 24 computer-generated identities (12 trustworthy, 12 untrustworthy) of Experiment 1. Four negative scene context stimuli obtained from public-domain websites, were added to the scenes used in Experiment 1 obtaining a total of 12 scene context stimuli (4 neutral, 4 negative and 4 threatening). A pretest confirmed that the scenes were perceived as intended. The 26 independent raters who took part in the previously reported pretest also rated the extent to which negative scene contexts were threatening using a scale ranging from 1 (*not at all*) to 7 (*extremely*). Participants were further asked to indicate the valence of each scene context (4 neutral, 4 negative, 4 threatening). Thus, threatening scenes were perceived as more threatening (M = 5.53, SD = 1.31) than negative scenes (M = 2.76 SD = 1.19), $t(25)=11.08$, $p=.001$, $d=2.17$, 95% CI = [1.45, 2.88], and also more threatening than neutral scenes (M = 1.16, SD = .48), $t(25)=16.61$, $p=.001$, $d=3.25$, 95% CI = [2.27, 4.23]. Negative scenes were perceived as more threatening than neutral scenes, $t(25)=6.50$, $p=.001$, $d=1.27$, 95% CI = [.75, 1.79]. Importantly, scores of perceived threat were above the midpoint of the scale only for threatening scenes, $t(25)=5.95$, $p<.001$. Moreover, threatening and negative scenes were perceived as having the same valence, $t<1$, $p=.58$. By contrast, threatening and negative scenes were perceived as more negative than neutral scenes, $ts>12.51$, $ps<.001$. To summarize, threatening and negative scenes were comparable in terms of valence, but differed in terms of perceived threat. Neutral scenes were perceived as less negative and less threatening than the other scenes. See Figure 2 for sample stimuli.

To further exclude the possibility that the scenes were perceived as signals of (un)trustworthiness, we asked 30 Italian students not involved in the main studies (M_{age} = 22.16; *SD* = 5.69) to view each scene context and freely write down their thoughts. None of them mentioned words or concepts related to honesty, trustworthiness, or morality. More specifically, students mentioned negative concepts associated with threat (e.g., fear, danger, and risk) when viewing the threatening scenes. By contrast, students mentioned negative concepts unrelated to threat (e.g., sadness, poverty, and deterioration) when viewing the negative scenes. Students mentioned descriptive concepts (e.g., nature, green, and spring) when viewing the neutral scenes.

One concern with the mouse-tracking paradigm for trustworthiness evaluation may be that forcing subjects to make dichotomous trustworthiness decisions may bias our results or exhibit a different pattern of responses compared to continuous Likert ratings of trustworthiness. To address this issue, we recruited 100 participants from Amazon Mechanical Turk, with half of participants asked to make dichotomous trustworthiness judgments of the stimuli using the keyboard in randomized order, and the other half of participants asked to make 7-point continuous judgments of the same stimuli. Due to 8 participants not completing the task, our final sample for this task comprised of 49 participants for the dichotomous judgments and 43 participants for the continuous judgments. For each stimulus, we generated a mean for participants' dichotomous judgments (0 = untrustworthy, 1 = trustworthy), and also a mean for participants' continuous judgments (1 = untrustworthy – 7 = trustworthy). These were very strongly correlated, $r(286) = .96$, $p < .00001$. This result speaks against the possibility that forcing participants to use dichotomous responses biased the results in some manner relative to a continuous-rating assessment of facial trustworthiness.

Procedure

Following the procedure of Experiment 1, participants were told that they would be presented with images of individuals in various settings, and were asked to categorize each person as either trustworthy or untrustworthy. They were instructed to make their decisions as quickly and accurately as possible by clicking response buttons, basing their judgments on their first impressions. The mouse-tracking procedure was carried out identically as in Experiment 1. Each face was presented 3 times and placed in the center location of a scene 1 for each context type, yielding 72 trials per participant.

Results and Discussion

Following the procedure of Experiment 1, we normalized trajectories into 101 time-steps and remapped leftward trajectories rightwards (inverted along the x-axis). To index trajectories deviation toward the opposite category, we computed MD. We performed a 3 (Scene Context: Neutral, Negative, Threatening) × 2 (Face: Trustworthy, Untrustworthy) within-subject ANOVA (Table 3). The main effect of scene context was not significant, $F(2,88)=.003$, $p=.99$, $\eta_p^2=.001$. However, the main effect of face was significant, $F(1,44)=10.86$, $p=.002$, $\eta_p^2=.20$, indicating that trajectories exhibited greater deviation overall when participants evaluated untrustworthy relative to trustworthy faces. More importantly, the analysis revealed a significant interaction between scene context and face, $F(2,88)=12.38$, $p<.001$, $\eta_p^2=.22$. Specifically, untrustworthy faces elicited more direct trajectories (lower MD) when they were embedded in threatening contexts than in negative [$t(45)=2.31$, $p=.03$, $d=.34$, 95% CI = (.04, .63)] and neutral [$t(45)=2.53$, $p=.02$, $d=.37$, 95% CI = (.07, .67)] contexts. However, MD scores did not differ between neutral and negative contexts, $t(45)<1$, $p=.40$, $d=.12$. Conversely, trustworthy faces elicited more curved trajectories when they were embedded in threatening contexts than in negative [$t(44)=2.92$, $p=.006$, $d=.44$, 95% CI = (.12, .73)] and neutral [$t(44)=3.58$, $p=.001$, $d=.53$, 95% CI = (.21, .84)] contexts. However, MD scores

did not differ between neutral and negative contexts, $t(45)=1.03$, $p=.31$, $d=.16$, 95% CI = (-.13, .44).

As in Experiment 1, next we computed the related AUC measure. We performed a 3 (Scene Context: Neutral, Negative, Threatening) × 2 (Face: Trustworthy, Untrustworthy) within-subject ANOVA (Table 4). The analysis did not yield a main effect of scene context $F(2,88)=.11$, $p=.90$, $\eta_p^2=.003$. However, the main effect of face was significant, $F(1,44)=11.43$, $p=.002$, $\eta_p^2=.21$, indicating a greater curvature overall for untrustworthy relative to trustworthy faces. More importantly, the scene context ×face interaction was significant, $F(1,44)=11.36$, $p=.002$, $\eta_p^2=.21$. Untrustworthy faces elicited more direct trajectories (lower AUC) when they were embedded in threatening contexts than in negative [$t(45)=2.23$, $p=.03$, $d=.33$, 95% CI = (.03, .62)] and neutral [$t(45)=2.51$, $p=.02$, $d=.37$, 95% CI = (.06, .66)] contexts. However, AUC scores did not differ between neutral and negative contexts, $t(45)=1$, $p=.32$, $d=.15$, 95% CI = (-.14, .43). Trustworthy faces elicited more curved trajectories (higher AUC scores) when they were embedded in threatening contexts than in negative [$t(44)=2.52$, $p=.02$, $d=.38$, 95% CI = (.07, .67)] and neutral [$t(44)=3.06$, $p=.004$, $d=.46$, 95% CI = (.14, .76)] contexts. However, AUC scores did not differ between neutral and negative contexts, $t(45)<1$, $p=.35$, $d=.14$.

Taken together, the findings demonstrate that the visual context biases the categorization of facial trustworthiness. Indeed, when the threatening nature of the face and of the context were compatible, trajectories exhibited a facilitation toward the selected response. When they were incompatible, trajectories showed a partial attraction toward the opposite-category response, indicating that the context was partially integrated into the evolving evaluation. Moreover, these contextual effects were specific to the compatibility of a face's trustworthiness with the threatening nature of the scene rather a mere negative valence associated with the scene.

Experiment 3

Experiment 3 aimed at replicating and extending the findings of Experiment 2 by increasing the ecological validity of our manipulations. Indeed, in Experiment 1 and Experiment 2 we used disembodied faces without hair that floated over scenes. In Experiment 3 we added hairlines to the faces and embedded the facial stimuli in the visual contexts more naturalistically.

Method

Participants

For the recruitment of participants, we aimed at collecting as many subjects as possible over the number indicated by the power analysis of Experiment 2. We advertised the study on campus and all the students who responded within 4 weeks and that were not involved in Experiment 1 and Experiment 2 took part to the study. Overall, we recruited 50 Italian students, with normal or corrected-to-normal vision (19 male, M_{age}=22.34, SD=1.73).

Stimuli

We used the same 24 computer-generated identities (12 trustworthy, 12 untrustworthy) employed in the previous two experiments. However, the facial stimuli were modified by using Photoshop. Thus, we added hairs, necks, and shoulders to the faces in order to increase the ecological validity of our manipulations and integrate facial and contextual stimuli more naturalistically. See Figure 3 for sample stimuli.

Procedure

Following the procedure of the previous experiments, participants were told that they would be presented with images of individuals in various settings, and were asked to categorize each person as either trustworthy or untrustworthy. They were instructed to make their decisions as quickly and accurately as possible by clicking response buttons, basing their judgments on their first impressions. The mouse-tracking procedure was carried out identically as in Experiment 1 and Experiment 2. To increase the reliability of our findings, we increased the number of trials: each

face was presented 6 times and placed in the center location of a scene 2 for each context type, yielding 144 trials per participant.

Results and Discussion

We first normalized trajectories into 101 time-steps and remapped leftward trajectories rightwards (inverted along the x-axis). To index trajectories deviation toward the opposite category, we computed MD. We performed a 3 (Scene Context: Neutral, Negative, Threatening) × 2 (Face: Trustworthy, Untrustworthy) within-subject ANOVA (Table 5). In line with our hypothesis, the analysis revealed a significant interaction between scene context and face, $F(2,98)=10.82$, $p<.001$, $\eta_p^2=.18$. Specifically, untrustworthy faces elicited more direct trajectories (lower MD) when they were embedded in threatening contexts than in neutral contexts [$t(49)=3.31$, $p=.002$, $d=.47$, 95% CI = (.17, .75)] and negative contexts [$t(49)=1.75$, $p=.08$[2], $d=.24$, 95% CI = (-.03, .53)], although the latter effect reached only marginal significance. However, MD scores did not differ between neutral and negative contexts, $t(49)<1$, $p=.43$, $d=.11$. Conversely, trustworthy faces elicited more curved trajectories when they were embedded in threatening contexts than neutral [$t(49)=3.94$, $p=.001$, $d=.55$, 95% CI = (.26, .85)] and negative [$t(49)=2.65$, $p=.01$, $d=.37$, 95% CI = (.09, .66)] contexts. However, MD scores did not differ between neutral and negative contexts, $t(49)=1.40$, $p=.17$, $d=.20$, 95% CI = (-.08, .47).

We also computed the AUC measure. We performed a 3 (Scene Context: Neutral, Negative, Threatening) × 2 (Face: Trustworthy, Untrustworthy) within-subject ANOVA (Table 6). The analysis showed that the scene context × face interaction was significant, $F(1,98)=11.20$, $p=.002$, $\eta_p^2=.19$. Untrustworthy faces elicited more direct trajectories (lower AUC) when they were embedded in threatening contexts than in negative contexts [$t(49)=2.17$, $p=.035$, $d=.31$, 95% CI = (.02, .60)] and neutral contexts [$t(49)=3.42$, $p=.001$, $d=.48$, 95% CI = (.19, .77)]. However, AUC scores did not differ between neutral and negative contexts, $t(49)=.54$, $p=.60$, $d=.07$, 95% CI = (-

.20, .35). Trustworthy faces elicited more curved trajectories (higher AUC scores) when they were embedded in threatening contexts than in negative [$t(49)=2.60$, $p=.01$, $d=.37$, 95% CI = (.08, .65)] and neutral [$t(49)=4.00$, $p=.001$, $d=.57$, 95% CI = (.26, .86)] contexts. However, AUC scores did not differ between neutral and negative contexts, $t(49)=1.70$, $p=.10$, $d=.24$, 95% CI = (-.04, .52).

Taken together, the findings replicated the findings of Experiment 2 and further show that the visual context biases the categorization of facial trustworthiness.

General Discussion

Three experiments showed that the scene in which a face is encountered alters trustworthiness evaluation. By adopting a mouse-tracking technique, Experiment 1 showed that the visual context temporally influenced the evaluation of facial trustworthiness as revealed by a partial attraction in participants' mouse trajectories toward the opposite category response when the facial and contextual information were incompatible. Moreover, when compatible, the trustworthiness evaluation process was facilitated. More direct trajectories were observed when untrustworthy faces were shown in threatening rather than neutral scenes, whereas more curved trajectories were observed when trustworthy faces were shown in threatening rather than neutral scenes. Experiment 2 corroborated these findings in a design that enabled us to disentangle the effects of threatening scene contexts from negative contexts in general. Results of this study confirmed that untrustworthiness and threat are inherently associated, as trajectories were more direct when untrustworthy faces were shown in threatening rather than in negative and neutral scenes. Conversely, trajectories were more curved when trustworthy faces were surrounded by threatening rather than negative and neutral scenes. Experiment 3 corroborated these findings by using a different set of stimuli with a greater ecological validity. Thus, contextual information was represented in parallel and partially integrated into trustworthiness evaluation, even when an ultimate perception was not altered. Together, these findings provide an original contribution to the

literature on the influence of context in person perception. As such, previous studies have reported context effects in identifying facial emotions (Aviezer et al., 2008; Barrett & Kensinger, 2010; Righart & De Gelder, 2008) and social categories such as ethnicity (Freeman et al., 2015). Going beyond emotions and static category dimensions, our research shows that visual context influences the evaluation of fundamental traits as well, such as trustworthiness.

As they stand, our findings extends prior research on the factors promoting or disrupting the processing of facial trustworthiness. Extensive work has revealed that individuals detect trustworthiness in faces faster than other human traits (Willis & Todorov, 2006; for a review, Todorov et al., 2015). For instance, the amygdala may process a face's trustworthiness so rapidly that perceptual awareness is not required (Freeman et al., 2014). However, most studies in this area have examined faces without any contextual information. Thus, extending prior research our data show that judgments of facial trustworthiness can be modified when individuals perceive the background information at the same time. Our research speaks to the malleable nature of trustworthiness such that its perception is readily pushed around by scene context. The findings are also in line with prior results on impression formation and change. Indeed, it has been shown that prior knowledge regarding a target person may affect the evaluation of facial trustworthiness (Mende-Siedlecki, Cai, & Todorov, 2013). These findings reveal that extraneous information from the face (i.e., person knowledge) may affect evaluations of the face. Our findings complement these prior insights by revealing that other forms of extraneous information of the face in the form of a visual context may alter the evaluations of facial cues.

One limitation of the present work of potential concern to readers is that participants evaluated trustworthiness in a dichotomous, forced-choice design. This was chosen to be consistent with the standard mouse-tracking paradigm, but naturally one may ask whether the effects obtained may reflect some kind of artifact of the task. The pre-test data, however, which demonstrated a very

strong correlation (r = .96) between dichotomous, forced-choice responses as used here and continuous ratings of trustworthiness speak against this possibility (see Methods of Experiment 2). Nevertheless, future work could explore the generalizing of these contextual effects using different response sets or different stimuli, including the possibility of conducting mouse-tracking using a continuous scale.

Readers may also be concerned about differences between the MD and AUC measures, with occasionally weaker, marginally-significant evidence of contextual impact for the MD measure. Previous research has often found that the AUC measure tends to have higher sensitivity than the MD measure, as it incorporates the aggregated spatial attraction effect over the entire time series rather than only a single maximal point (see Hehman et al., 2015; Freeman & Ambady, 2010). As such, some of the findings we report reached only marginal significance when considering MD. However, the AUC measure yielded consistent and significant findings across the three experiments. The overall direction and pattern of results was consistent across both measures, but given AUC's higher sensitivity, it provided more statistically reliable results.

Although the focus of the present work was on the process rather than outcomes of integrating facial and contextual cues, we additionally explored how the context affected explicit categorizations of trustworthiness (see Supplementary Materials). The findings were mixed, in which incongruent context-face trials increased "incorrect" (i.e., context-associated) categorizations only in Experiments 2 and 3. Moreover, while Experiment 3 (the most ecologically valid) documented both contextual congruency and incongruency effects on categorization outcomes, Experiment 2 did not reveal a clear distinction between negative and threatening contexts. This is not especially surprising as previous mouse-tracking studies have often found that a participant's ultimate perceptual response may not be consistently altered by context or other meaningful differences even if the process leading up to that response is altered considerably (e.g., Freeman et

al., 2011; Freeman et al., 2013; Freeman, 2018). As the mouse-tracking paradigm can track how various cues drive categorization in real time, it is able to reveal potentially subtle influences of context, even when an ultimate response may not be affected.

According to a functional approach to social perception "perceiving is for doing" (Fiske, 1992) and its primary purpose is to guide people in avoiding social threats (Dunning, 2004; Heider, 1958; Zebrowitz & Collins, 1997). In this sense, people should be particularly fast in recognizing malevolent social targets (i.e., untrustworthy) especially when the context might make them able to enact their bad intentions (i.e., threatening situations). As such, one perceived, trustworthiness evaluation tends to powerfully affect social interactions by trigging a number of cognitive, affective, and behavioral effects (Todorov et al., 2015). In this sense, recognizing rapidly an untrustworthy individual under threatening circumstances might have an adaptive function. Alternatively, the effects may have arisen simply due to a domain-general property of early social perception processes' malleability to conceptually consistent information (e.g., Freeman & Johnson, 2016). Thus, any kind of context or presumably extraneous information to the initial social perception process has the potential to provide an immediate top-down constraint on perception, and the consistency or inconsistency of the information (e.g., untrustworthiness and threat being conceptually similar) can introduce predictable biases. Further research could examine whether the effects obtained generalize to other forms of conceptual consistency in contextual cues or if they are specific to threat which may suggest a more functionalist interpretation.

Moreover, our data show that threatening scenes promoted and disrupted the categorization of untrustworthy and trustworthy faces, respectively. Since we did not find any difference between negative and neutral contexts in promoting the categorization of trustworthy faces, and intriguing challenge for future research would be to test whether positive (rather than neutral) visual scenes or visual scenes priming positive moral concepts may foster the categorization of trustworthy faces.

Such studies could complement our approach and help to gain more insights on the specific conditions in which context affects trait inferences of others' faces.

References

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

Ames, D. L., Fiske, S. T., & Todorov, A. T. (2011). Impression formation: A focus on others' intents. In J. Decety & J. Cacioppo (Eds.), *The handbook of social neuroscience.*(pp. 419-433). Oxford University Press.

Aviezer, H, Hassin, RR, Ryan J, Grady C, Susskind J, Anderson A, Moscovitch M, & Bentin S. (2008). Angry, disgusted, or afraid? Studies of the malleability of emotion perception. *Psychological Science, 19*, 724–732.

Ballew, C. C., & Todorov, A. (2007). Predicting political elections from rapid and unreflective face judgments. *Proceedings of the National Academy of Sciences, 104*, 17948-17953.

Bar, M., Neta, M., & Linz, H. (2006). Very first impressions. *Emotion, 6*, 269-278.

Barrett, L. F., & Kensinger, E. A. (2010). Context is routinely encoded during emotion perception. *Psychological Science, 21*, 595-599.

Brambilla, M., & Leach, C.W. (2014). On the importance of being moral: The distinctive role of morality in social judgment. *Social Cognition, 32*, 397-408.

Brambilla, M., Sacchi, S., Pagliaro, S., & Ellemers, N. (2013). Morality and intergroup relations: Threats to safety and group image predict the desire to interact with outgroup and ingroup members. *Journal of Experimental Social Psychology, 49*, 811-821.

Brambilla, M., Sacchi, S., Rusconi, P., & Cherubini, P., Yzerbyt, V.Y. (2012). You want to give a good impression? Be honest! Moral traits dominate group impression formation. *British Journal of Social Psychology, 51*, 149-166.

Carraro, L., Castelli, L., & Negri, P. (2016). The hand in motion of liberals and conservatives reveals the differential processing of positive and negative information. Ac*ta psychologica,*

168, 78-84.

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.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hilsdale. NJ: Lawrence Earlbaum Associates, 2.

Cosmides, L., & Tooby, J. (1992). Cognitive adaptations for social exchange. In J.H. Barkow, L. Cosmides, & J. Tooby (Eds.), *The adapted mind: Evolutionary psychology and the generation of culture* (pp. 163–228). London: Oxford University Press.

Dunning, D. (2004). On the motives underlying social cognition. In M. Brewer & M. Hewstone (Eds.), *Emotion and motivation* (pp. 137–164). Oxford: Blackwell.

Engell, A. D., Haxby, J. V., & Todorov, A. (2007). Implicit trustworthiness decisions: automatic coding of face properties in the human amygdala. *Journal of cognitive neuroscience, 19*, 1508-1519.

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*, 175-191.

Fiske, S. T. (1992). Thinking is for doing: Portraits of social cognition from Daguerreotype to laserphoto. *Journal of Personality and Social Psychology, 63*, 877–889.

Freeman, J.B. (2018). Doing psychological science by hand. C*urrent Directions in Psychological Science.*

Freeman, J. B. (2014). Abrupt category shifts during real-time person perception. *Psychonomic Bulletin & Review, 21*, 85-92.

Freeman, J.B. & Ambady, N. (2010). MouseTracker: Software for studying real-time mental processing using a computer mouse-tracking method. *Behavior Research Methods, 42,* 226-241.

Freeman, J.B. & Ambady, N. (2011). A dynamic interactive theory of person construal. *Psychological Review, 118,* 247-279.

Freeman, J.B., & Johnson, K.L. (2016). More than meets the eye: Split-second social perception. *Trends in Cognitive Sciences, 20,* 362-374.

Freeman, J.B., Ma, Y., Barth, M., Young, S.G., Han, S., & Ambady, N. (2015). The neural basis of contextual influences on face categorization. *Cerebral Cortex, 25,* 415-422.

Freeman, J.B., Ma, Y., Han, S., & Ambady, N. (2013). Influences of culture and visual context on real-time social categorization. *Journal of Experimental Social Psychology, 49,* 206-210.

Freeman, J.B., Stolier, R.M., Ingbretsen, Z.A., & Hehman, E.A. (2014). Amygdala responsivity to high-level social information from unseen faces. *The Journal of Neuroscience, 34,* 10573-10581.

Hehman, E., Stolier, R. M., & Freeman, J. B. (2015). Advanced mouse-tracking analytic techniques for enhancing psychological science. *Group Processes & Intergroup Relations, 18,* 384-401.

Heider, F. (1958). *The psychology of interpersonal relations*. New York: Wiley.

Keres, A., & Chartier, C. R. (2016). The Biasing Effects of Visual Background on Perceived Facial Trustworthiness. *Psi Chi Journal of Psychological Research, 21,* 170-175.

Leach, C.W., Ellemers, N., & Barreto, M. (2007). Group virtue: The importance of morality (vs. competence and sociability) in the positive evaluation of in-groups. *Journal of Personality and Social Psychology, 93,* 234-249.

Leidner, B., & Castano, E. (2012). Morality shifting in the context of intergroup violence. *European Journal of Social Psychology, 42,* 82-91.

Mazur, A., Mazur, J., & Keating, C. (1984). Military rank attainment of a West Point class: Effects of cadets' physical features. *American Journal of Sociology*, *90*, 125-150.

Mende-Siedlecki, P., Cai, Y., & Todorov, A. (2012). The neural dynamics of updating person impressions. *Social cognitive and affective neuroscience, 8*, 623-631.

Mueller, U., & Mazur, A. (1996). Facial dominance of West Point cadets as a predictor of later military rank. *Social forces. 74*, 823-850.

Phelps, E. A., & LeDoux, J. E. (2005). Contributions of the amygdala to emotion processing: from animal models to human behavior. *Neuron, 48*, 175-187.

Porter, S., ten Brinke, L., & Gustaw, C. (2010). Dangerous decisions: The impact of first impressions of trustworthiness on the evaluation of legal evidence and defendant culpability. *Psychology, Crime & Law*, *16*, 477-491.

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.

Righart, R., & De Gelder, B. (2008). Recognition of facial expressions is influenced by emotional scene gist. *Cognitive, Affective, & Behavioral Neuroscience, 8*, 264-272.

Rule N.O., & Ambady N. (2008). The face of success: Inferences from chief executive officers' appearance predict company profits. *Psychological Science, 19*, 109-111.

Rule, N.O., Slepian, M.L., & Ambady, N. (2012). A memory advantage for untrustworthy faces. *Cognition, 125*, 207-218.

Slepian, M.L., Young, S.G., Rule, N.O., Weisbuch, M., & Ambady, N. (2012). Embodied impression formation: Social judgments and motor cues to approach and avoidance. *Social Cognition, 30*, 232-240.

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

Todorov, A., Dotsch, R., Porter, J. M., Oosterhof, N. N., & Falvello, V. B. (2013). Validation of data-driven computational models of social perception of faces. *Emotion, 13,* 724-738.

Todorov, A., Mandisodza, A. N., Goren, A., & Hall, C. C. (2005). Inferences of competence from faces predict election outcomes. *Science, 308,* 1623-1626.

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 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.

Todorov, A., Said, C. P., Oosterhof, N. N., & Engell, A. D. (2011). Task-invariant brain responses to the social value of faces. *Journal of Cognitive Neuroscience, 23,* 2766-2781.

Todorov, A., & Uleman, J. S. (2003). The efficiency of binding spontaneous trait inferences to actors' faces. *Journal of Experimental Social Psychology, 39,* 549-562.

van der Toorn, J., Ellemers, N., & Doosje, B. (2015). The threat of moral transgression: The impact of group membership and moral opportunity. *European Journal of Social Psychology, 45,* 609-622.

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

Wilson, J. P., & Rule, N. O. (2015). Facial trustworthiness predicts extreme criminal sentencing outcomes. *Psychological Science, 26,* 1325-1331.

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.

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

Zebrowitz, L., & Collins, M. (1997). Accurate social perception at zero acquaintance: The affordances of a Gibsonian approach. *Personality and Social Psychology Review, 1*, 204–223.

Zebrowitz, L.A. (1997). *Reading Faces: Window to the Soul?* Boulder, CO: Westview.

Table 1 – *Means and standard errors for MD scores as a function of face and scene context (Experiment 1)*

	Face	
Scene Context	Untrustworthy	Trustworthy
Threatening	.37(.04)a	.43 (.05)a
Neutral	.48(.05)b	.35 (.03)b

Note: Means with different subscripts in a given column are significantly different at $p<.06$. Standard errors are reported in parenthesis.

Table 2 – *Means and standard errors for AUC scores as a function of face and scene context (Experiment 1)*

	Face	
Scene Context	Untrustworthy	Trustworthy
Threatening	.61 (.09)a	.81 (.12)a
Neutral	.94 (.13)b	.54 (.07)b

Note: Means with different subscripts in a given column are significantly different at $p<.05$. Standard errors are reported in parenthesis.

Table 3 – *Means and standard errors for MD scores as a function of face and scene context (Experiment 2)*

	Face	
Scene Context	Untrustworthy	Trustworthy
Threatening	.32(.03)a	.32 (.03)a
Negative	.39 (.04)b	.24 (.03)b
Neutral	.42(.04)b	.22 (.03)b

Note: Means with different subscripts in a given column are significantly different at $p<.05$. Standard errors are reported in parenthesis.

Table 4 – *Means and standard errors for AUC scores as a function of face and scene context (Experiment 2)*

	Face	
Scene Context	Untrustworthy	Trustworthy
Threatening	.55(.06)a	.56 (.06)a
Negative	.70 (.07)b	.41 (.06)b
Neutral	.77(.09)b	.37 (.06)b

Note: Means with different subscripts in a given column are significantly different at $p<.05$. Standard errors are reported in parenthesis.

Table 5 – *Means and standard errors for MD scores as a function of face and scene context (Experiment 3)*

	Face	
Scene Context	Untrustworthy	Trustworthy
Threatening	.38(.03)a	.39 (.03)a
Negative	.42 (.03)b	.34 (.03)b
Neutral	.44(.04)b	.31 (.03)b

Note: Means with different subscripts in a given column are significantly different at $p<.05$. Standard errors are reported in parenthesis. $p=.08$ between threatening and negative contexts for untrustworthy faces.

Table 6 – *Means and standard errors for AUC scores as a function of face and scene context (Experiment 3)*

	Face	
Scene Context	Untrustworthy	Trustworthy
Threatening	.69(.07)a	.75 (.07)a
Negative	.81 (.08)b	.63 (.06)b
Neutral	.85(.08)b	.55 (.06)b

Note: Means with different subscripts in a given column are significantly different at $p<.05$. Standard errors are reported in parenthesis.

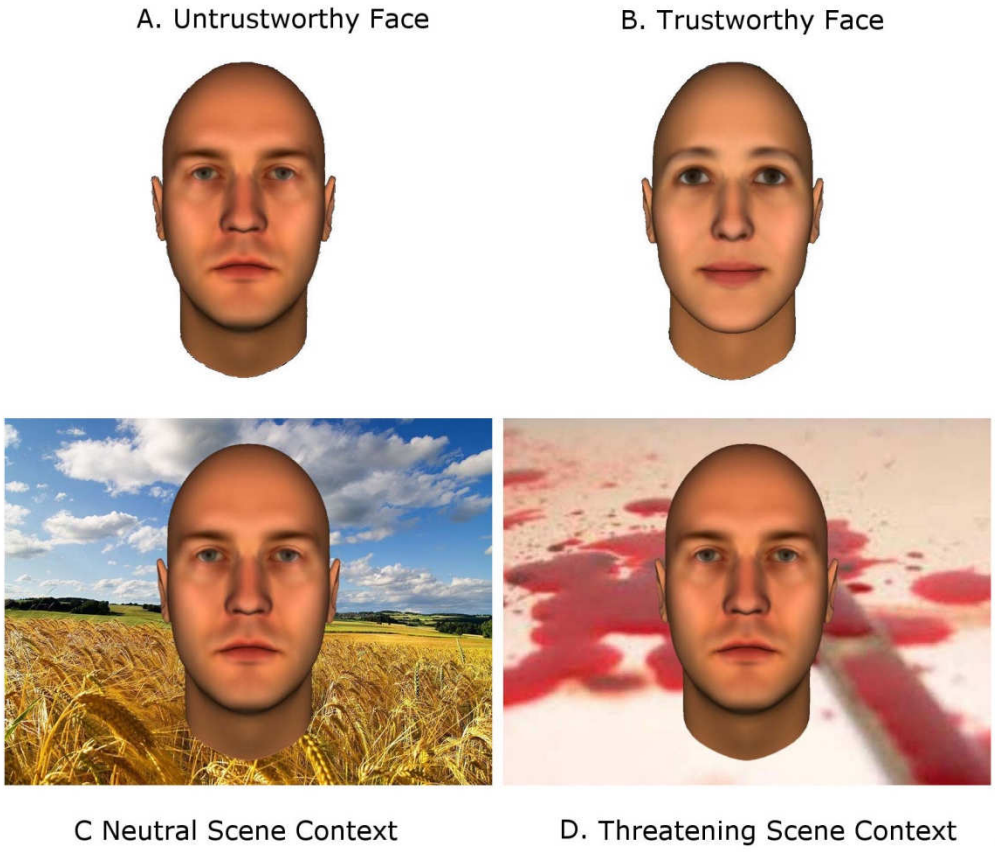

Figure 1. At the top are sample face stimuli. At the bottom are sample Neutral, and Threatening typed scene contexts (with face stimulus at the center). Experiment 1

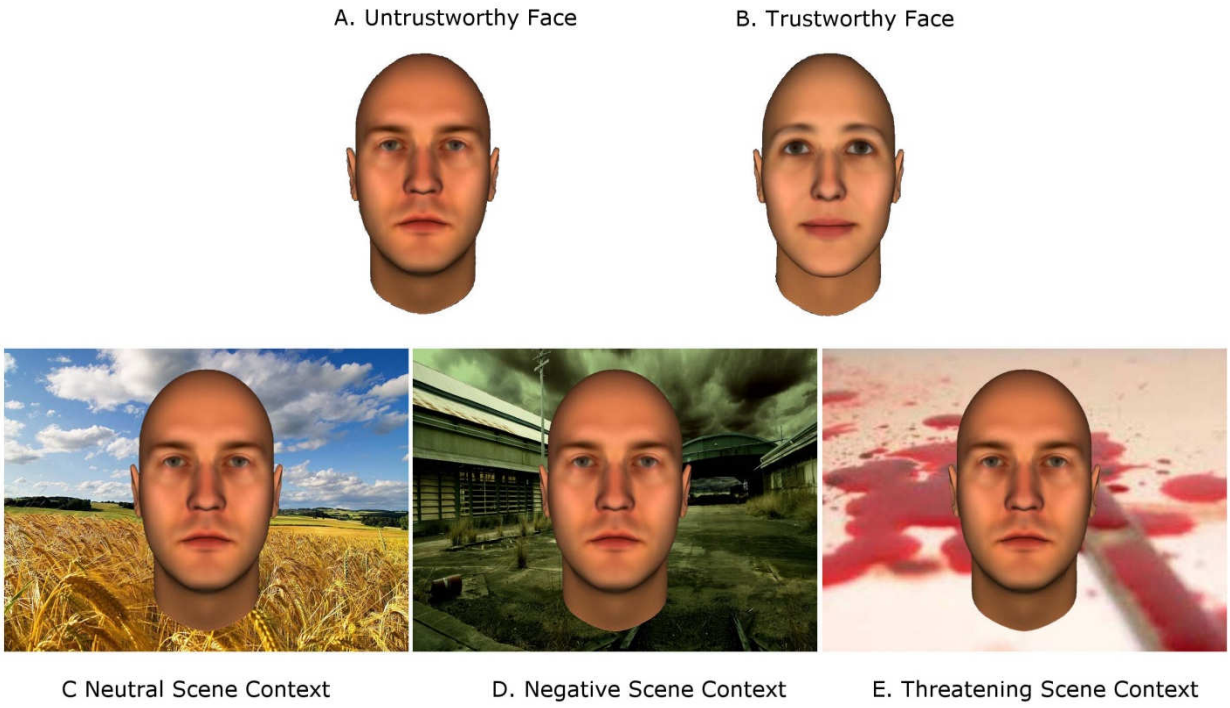

Figure 2. At the top are sample face stimuli. At the bottom are sample Neutral, Negative, and Threatening typed scene contexts (with face stimulus at the center). Experiment 2

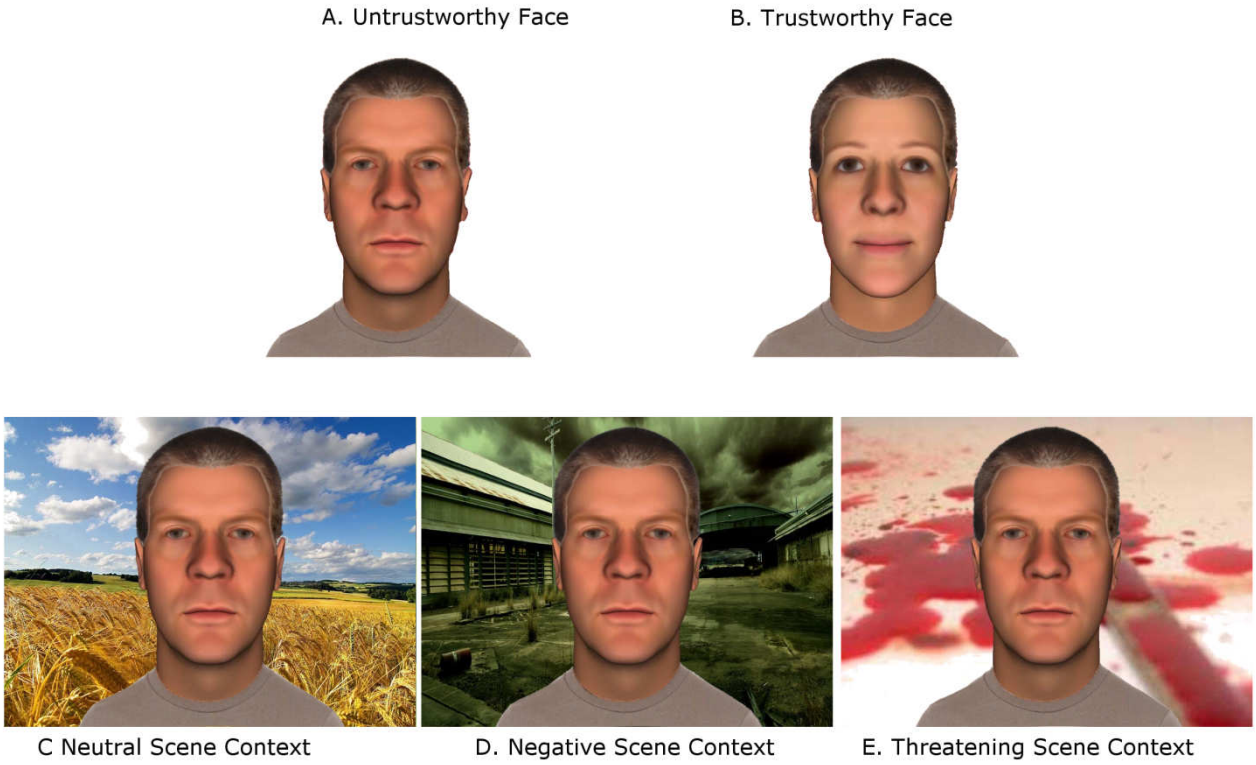

Figure 3. At the top are sample face stimuli. At the bottom are sample Neutral, Negative, and Threatening typed scene contexts (with face stimulus at the center). Experiment 3

Footnotes

[1] For the full set of scenes employed in the reported experiments, see the supplementary materials.

[2] The MD findings for untrustworthy faces appearing in threatening vs. negative contexts were mixed when looking at the individual studies: significant in Experiment 2 and marginally significant in Experiment 3. To combine the results obtained in these different studies and to increase the precision of the parameter estimates, we meta-analytically combined the results from the effect sizes reported in Experiment 2 and Experiment 3. The random effects meta-analysis (ESCI procedure; Cumming, 2012) produced the overall effect size d = .28, 95% CI [.08, .48]. This new analysis suggests that the effects we obtained on MD scores are reliable (converging with those on AUC scores).