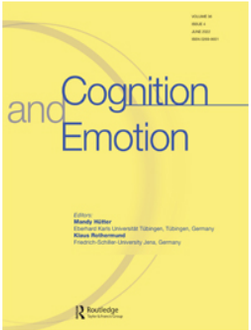

Cognition and Emotion

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/pcem20

Untrusted under threat: on the superior bond between trustworthiness and threat in face-context integration

Simone Mattavelli, Matteo Masi & Marco Brambilla

To cite this article: Simone Mattavelli, Matteo Masi & Marco Brambilla (2022): Untrusted under threat: on the superior bond between trustworthiness and threat in face-context integration, Cognition and Emotion, DOI: 10.1080/02699931.2022.2103100

To link to this article: https://doi.org/10.1080/02699931.2022.2103100

COGNITION AND EMOTION
https://doi.org/10.1080/02699931.2022.2103100

Untrusted under threat: on the superior bond between trustworthiness and threat in face-context integration

Simone Mattavelli, Matteo Masi and Marco Brambilla

Department of Psychology, University of Milano-Bicocca, Milan, Italy

ABSTRACT

The face is a powerful source to make inferences about one's trustworthiness. Recent studies demonstrated that facial trustworthiness is influenced by the level of threat conveyed by the visual scene in which faces are embedded: untrustworthy-looking faces are more likely judged as untrustworthy when shown in threatening scenes. Here, we explore whether this face-context congruency effect is specific to the negative pole of the threat-trust domain. Experiment 1 ($N = 89$) focused on the differential impact of positive vs. negative face-context congruency within the domains of threat and trust. Negative congruency (i.e. untrustworthy-looking faces in threatening contexts) led to more extreme attributions as opposed to positive congruency (i.e. trustworthy-looking faces in reassuring contexts). Experiment 2 ($N = 120$) replicated these findings by further showing their domain-specificity. The negativity bias was found in the threat-trust domain, but not when extroverted- vs. introverted-looking faces appeared in happy vs. sad context scenes. Experiment 3 ($N = 154$, pre-registered) replicated the pattern observed in Experiment 1 while controlling for the extent to which both threatening and reassuring context stimuli were related to the human action. We discussed the theoretical implications of these results for understanding how contextual information is integrated into the evaluation of facial trustworthiness.

ARTICLE HISTORY
Received 16 December 2021
Revised 13 June 2022
Accepted 13 July 2022

KEYWORDS
Face perception; context; trustworthiness; threat; congruency effect

When encountering a person for the first time, we seek cues that can tell us whether that individual deserves to be trusted (Ames et al., 2011). Information of that sort can be obtained from both behavioural and physical features. One powerful source that people use to ascribe trustworthiness to others is a person's face (see Todorov et al., 2015 for a review). A good deal of research has demonstrated that facial features that signal trustworthiness are impactful on many social outcomes. For instance, people invest less money with partners who look untrustworthy (Chang et al., 2010; Rezlescu et al., 2012; Stirrat & Perrett, 2010) while trustworthy-looking individuals have a higher chance of being granted loans (Duarte et al., 2012). Moreover, the perception of trustworthiness in others' faces proceeds more quickly than the perception of other characteristics because facial trustworthiness and threat are inherently linked (Todorov et al., 2009; Willis & Todorov, 2006). Indeed, trustworthy and untrustworthy individuals are perceived as beneficial and harmful, respectively (Todorov et al., 2015; see also Brambilla & Leach, 2014 ; Brambilla et al., 2021a). It is therefore important to unpack the factors that influence the attribution of trustworthiness to faces.

Much of the research on trustworthiness inferences from faces have looked at the impact of facial features when the face is presented in isolation (namely evaluation of trustworthiness made upon faces being flashed on the computer screen). However, there is now increasing evidence to support that context information is key to influencing how facial cues shape the impressions we form about others (Carroll & Russell, 1996; Righart & De Gelder,

CONTACT Simone Mattavelli simone.mattavelli@unimib.it Department of Psychology, University of Milan-Bicocca, 1, Piazza dell'Ateneo Nuovo, Milan 20126, Italy

Supplemental data for this article can be accessed online at https://doi.org/10.1080/02699931.2022.2103100.

2006, 2008). The context in which a face happens to be processed is key to affecting how we extract and interpret facial cues. In fact, the interplay between facial and contextual cues influences the perception of both emotions (Aviezer et al., 2008; Barrett & Kensinger, 2010; Righart & De Gelder, 2008) and ethnicity (Freeman et al., 2013). Importantly, recent research showed that context stimuli also matter when it comes to attributing facial trustworthiness. In fact, the evaluation of facial trustworthiness is influenced by the level of threat conveyed by the visual scene in which faces are embedded (Brambilla et al., 2018). Thus, untrustworthy-looking faces are more likely categorised as such when surrounded by threatening rather than both neutral and negative, but non-threatening, contexts (see also Brambilla et al., 2021b for similar findings with auditory contexts). This finding indicates that it is not just the common negative valence carried by both the face and the surrounding context, but rather the inherent link between threat and (un)trustworthiness that eases their integration.

Corroborating the idea that the dimensions of trustworthiness and threat are inherently linked (see Brambilla & Leach, 2014; Brambilla et al., 2021a), prior studies indicated that the perception of trustworthiness involves the amygdala, which is also involved in the detection of potentially threatening stimuli (Willis & Todorov, 2006). Thus, untrustworthy faces should be more likely perceived as such when surrounded by threatening (as opposed to non-threatening) contexts, because the feelings evoked by both the context and the face are congruent (e.g. threat). As a matter of fact, a similar explanation has been proposed to account for the facilitated detection of facial emotions from congruent contexts (Righart & De Gelder, 2008; Tamietto et al., 2006). For instance, Righart and De Gelder (2008) asked participants to categorise faces expressing disgust, fear, or happiness surrounded by backgrounds with either a congruent or an incongruent emotional significance (e.g. a facial expression of disgust in a context of garbage or the same expression, shown among flowers). They found an advantage in both accuracy and speed recognition for the facial expressions accompanied by congruent (vs. incongruent) scenes.

Furthermore, the dimensions of threat and (un)trustworthiness serve specific adaptive functions. In social interactions, people are primarily interested in discovering whether someone's intentions are beneficial or harmful. Thus, the processing of facial trustworthiness can be explained in ecological-functional terms, based on the idea that 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 et al., 2011; Brambilla & Leach, 2014; Cosmides & Tooby, 1992). Hence, detecting potential threats serves an adaptive need to avoid any source of danger (LeDoux, 1996). Based on an ecological account of social perception, it is by informing about actions that perception promotes individual goal attainment and species survival (Fiske, 1992; Gibson, 1977; McArthur & Baron, 1983). This adaptive advantage of the threat-trust domain has direct implications for interpreting the face-context congruency observed in Brambilla et al. (2018). In fact, not only threatening contexts are congruent with untrustworthy faces, but the action that both the context and the face trigger is one that has high adaptive value. In other words, facial untrustworthiness should be more easily perceived when the face appears in a threatening context due to the adaptive need for avoiding threat.

Importantly, conceptualising the effect in terms of general congruency implies that virtually any context that is congruent with a contingent facial disposition eases the perception of the latter. Alternatively, the adaptive advantage of the threat-trust domain would predict higher integration (i.e. more extreme dispositional judgments) under this specific congruency. These alternative hypotheses can be empirically tested within the threat-trust domain by focusing on the valence-specificity of the observed effect. Namely, if a general congruency explanation suffices to account for the effect observed by Brambilla et al. (2018), then this effect should be valence independent. In fact, just like contextual threat eases the attribution of untrustworthiness to untrustworthy-looking faces, so contextual reassurance should ease the attribution of trustworthiness to trustworthy-looking faces. In other words, under this general congruency hypothesis, no difference should be expected when comparing the face-context integration on the negative vs. positive pole of the threat-trust domain. Alternatively, if the same effect is also due to an adaptive need to spot untrustworthiness in a situation that might evoke danger (as compared to one's need to spot trustworthiness in a situation that evokes reassurance) then a stronger face-context integration should emerge when congruency is operationalised on the negative pole of the threat-trust domain.

The present work delved into this question across three experimental studies. Experiment 1 focused on the differential impact of positive vs. negative congruency of contextual and facial cues within the threat-trust domain. The human tendency to gravitate toward a vigilant monitoring of the environment that is functional to avoid threats has been often proposed to account for a positive–negative asymmetry in impression formation (Baumeister et al., 2001; Rusconi et al., 2020). If social perception consists in the integration between external cues and dispositional motives, then face-context congruency should affect judgments of facial trustworthiness with a greater extent when both are presented to be negative (threatening scene – untrustworthy face) rather than positive (reassuring scene – trustworthy face). Yet, a stronger face-context integration for untrustworthy-looking faces presented in threatening contexts does not necessarily prove that the effect is due to adaptive functions specific to the threat-trust domain. Alternatively, it might simply reflect a negativity bias in face-context integration, whereby stimuli that are congruent on the negative end of whichever domain leads to a stronger integration than stimuli congruent on the positive end. In Experiment 2 we explored whether the hypothesised effect is domain-specific. We did so by comparing the effect of positive and negative congruency within the threat-trust bond with that of another binomial where the advantage for negative emotional cues evoked by the context is of less adaptive purpose (i.e. happiness-extroversion). In fact, prior studies have shown that extroversion correlates positively with happiness (e.g. Furnham & Brewin, 1990; Williams, 1990; Wilson & Gullone, 1999). Thus, perceivers should be more likely to infer that a person is extroverted (or introverted) when presented with cues that prime happiness (or sadness). Although no prior studies have directly investigated the impact of contextual cues on the attribution of facial extroversion, here we reasoned that just like an ecological correlation might explain the effect of contextual threat-reassurance on facial trustworthiness, so contextual sadness-happiness should influence facial extroversion. Crucially, however, the impact of either negative or positive context stimuli to affect dispositional attributions made on congruent faces should vary depending on the functional value of the cue with respect to the ultimate judgment. Experiment 3 was designed to address a methodological limitation related to the context stimuli used in Experiments 1–2 to represent both threatening and reassuring scenes. Namely, we employed a preliminary conditioning procedure to ensure that both threatening and reassuring scenes were equal in their being related to the human action.

The three experiments were approved by the local ethics committee. For both the studies, we report all the manipulations and measures administered. All the materials, data, and analysis code are made available at https://osf.io/p4sfy/.

Experiment 1

Experiment 1 tested the hypothesis that the evaluation of facial trustworthiness is influenced by the threat conveyed by the visual context in which the face is embedded. We focused on the congruency between face and context (i.e. both stimuli belonging to the same end of the threat-trust continuum) and compared the effect of either positive or negative face-context congruency on the attribution of trustworthiness to faces. In a series of trials, participants were exposed to either trustworthy or untrustworthy faces embedded in contexts selected to be threatening, reassuring, or neutral. In each trial, participants rated the relevant face on trustworthiness, using a continuous scale ranging from −3 to +3. We hypothesise the congruency effect to be stronger (i.e. higher absolute variation from the baseline rate) for untrustworthy faces in threatening contexts than for trustworthy faces in reassuring contexts.

Sample size determination

We conducted a sensitivity analysis using MorePower 6.0 (Campbell & Thompson, 2012). With 89 participants, assuming a power of .80 (two-tail, $\alpha = .05$) we could detect a critical interaction effect as small as $f = .23$ ($\eta_p^2 = .05$) in a 2×3 full-within ANOVA design.

Method

Participants and Procedure

Eighty-nine Italian participants (51 females, $M_{age} = 28.77$, $SD_{age} = 10.96$) volunteered to participate in the study. We advertised the study online and included all the students who responded within a month. Participants were asked to participate in a study on face perception. Instructions informed participants that they would be presented with images of individuals located in different contexts and were

asked to rate each person on perceived trustworthiness using a 7-point Likert scale (−3 = Untrustworthy; +3 = Trustworthy). The experiment consisted of two blocks of 72 trials each with a brief pause in between. Within each block, trials consisted of the presentation of a face on the screen, surrounded by a visual context. Facial stimuli were taken from a pool of 24 pictures (12 trustworthy, 12 untrustworthy) and administered in random order without replacement. In each trial, either trustworthy- or untrustworthy-looking faces could appear in two out of four randomly extracted threatening, reassuring, or neutral contexts. Thus, each of the 24 face identities appeared six times in total, twice with each class of context. No time limit was set even though participants were kindly reminded to provide their judgments as fast as possible.

Materials

We employed 24 computer-generated identities (12 trustworthy, 12 untrustworthy) borrowed from a set of photos previously validated for facial trustworthiness (Todorov et al., 2013) and slightly modified to increase their ecological validity (Brambilla et al., 2018, 2021; see Figure 1).

Context stimuli (4 threatening, 4 reassuring) were selected via a pre-test ($N = 68$, 50 females, M_{age} = 32.32, $SD_{age} = 12.66$). Participants evaluated a series of forty stimuli on valence (*"How negative–positive is this picture?"*, on a scale ranging from 1 = negative to 7 = positive) and threat (*"How threatening-reassuring is this picture?"*, on a scale ranging from 1 = threatening to 7 = reassuring). Four stimuli (i.e. tornado, bomb explosions, volcano, blood stains) were selected to be negative and threatening (valence: $M = 1.90$, $SD = .48$; threat: $M = 1.59$, $SD = .30$) and four pictures (i.e. 2 beach landscapes, lake in spring, lake in autumn) to be positive and reassuring (valence: $M = 6.38$, $SD = .21$; threat: $M = 6.31$, $SD = .14$). Importantly, on both valence and threat, the two sets of stimuli did not differ from each other in terms of absolute extremity (valence: $p = .33$; threat: $p = .62$). A grey rectangle was used as control context.

Results

Our investigation focused on the absolute impact of congruent contexts, as opposed to both neutral and incongruent, on the attribution of trustworthiness on either trustworthy or untrustworthy faces. Therefore, the three types of contexts were coded as congruent vs. neutral vs. incongruent, based on their congruency with the focal face (e.g. congruent: threatening[reassuring] contexts with untrustworthy[trustworthy] faces; incongruent: threatening[reassuring] contexts with trustworthy[untrustworthy] faces). Neutral contexts reflected the baseline evaluation of each type of face. As we were interested in the absolute magnitude of the effects of each type of context on either trustworthy or untrustworthy faces, the

Figure 1. At the top, two sample faces employed in Experiment 1, untrustworthy on the left and trustworthy on the right. At the bottom, three samples of a face embedded into threatening, reassuring, or neutral grey background contexts (from the left).

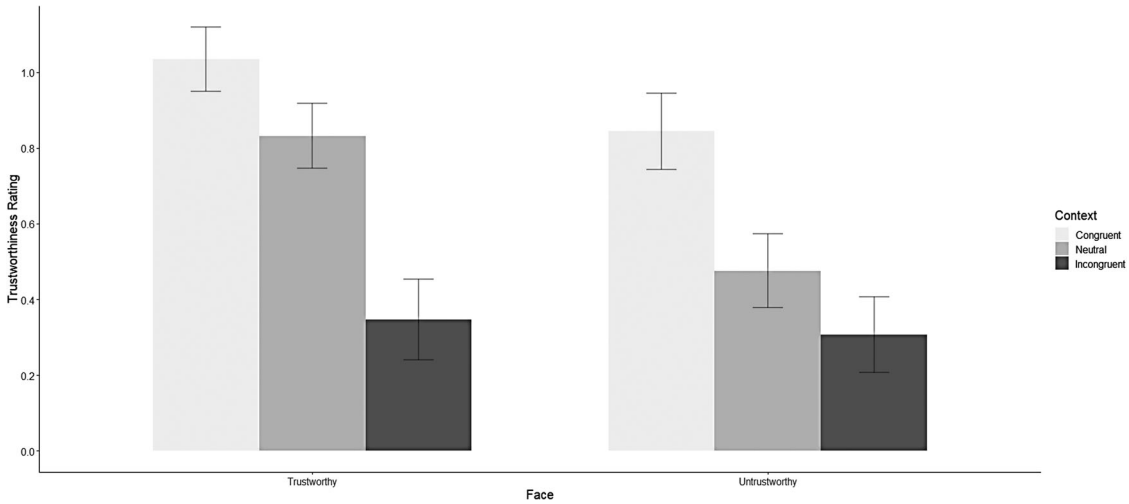

Figure 2. Results of Experiment 1.

original attribution scores received by untrustworthy faces were reversed. This means that if participants assigned −3 to an untrustworthy face, this score was coded as +3 (see Figure 2). Thus, for both trustworthy and untrustworthy faces, higher scores reflected a stronger impact of the relevant context in making the attribution of trustworthiness more extreme on the expected end of the continuum (i.e. lower trustworthiness to untrustworthy faces vs. higher trustworthiness to trustworthy faces). We conducted 2 (face: untrustworthy vs. trustworthy) × 3 (context congruency: congruent vs. neutral vs. incongruent) ANOVA.[1] Degrees of freedom and probability values have been corrected whenever the assumption of sphericity has been violated. We did not found a significant effect of face, $F(1, 88) = 2.09$, $p = .152$, $\eta_p^2 = .02$, 95% CI [.00, .10]. The effect of context congruency was significant, $F(1.24, 109.12) = 54.52$, $p < .001$, $\eta_p^2 = .38$, 95% CI [.29, .46]. Congruent contexts made the trustworthiness scores more extreme than both neutral, $t(176) = 4.86$, $p < .001$, $d = .37$, 95% CI [.21, .52], and incongruent context, $t(176) = 10.43$, $p < .001$, $d = .79$, 95% CI [.62, .95]. In line with our expectations, the interaction was significant, $F(2, 176) = 9.74$, $p < .001$, $\eta_p^2 = .10$, 95% CI [.04, .17].

We decomposed this interaction by looking at the direct contrasts between congruent, incongruent, and neutral contexts across trustworthy and untrustworthy faces. When confronted with neutral contexts, congruent contexts had significant greater impact on untrustworthy (vs. trustworthy) faces, $t(176) = 2.32$, $p = .021$, $d = .17$, 95% CI [.03, .32]. Thus, the negative impact of threatening contexts on the attribution of trustworthiness on untrustworthy faces was greater than the positive impact of reassuring contexts on trustworthy faces, $t(291) = 5.36$, $p < .001$, $d = .31$, 95% CI [.20, .43] and $t(291) = 2.95$, $p < .001$, $d = .17$, 95% CI [.06, .29], respectively. Also, congruent contexts exerted a stronger impact on untrustworthy (vs. trustworthy) faces relative to incongruent contexts, $t(176) = -2.09$, $p = .04$, $d = .16$, 95% CI [.01, .31]. Instead, comparing incongruent and neutral context yielded a stronger effect on trustworthy, relative to untrustworthy, faces, $t(176) = 4.41$, $p < .001$, $d = .33$, 95% CI [.18, .48].

These findings suggest that trustworthiness attribution is influenced differently by threatening and reassuring contexts and that face-context congruency per-se does not suffice to explain variation in attributed trustworthiness to faces. As such, we found that threatening contexts decreased the attribution of trustworthiness to untrustworthy faces more than reassuring contexts increased the trustworthiness attributed to trustworthy faces.

Experiment 2

The greater impact exerted by threatening (vs. reassuring) contexts on untrustworthy-looking (vs. trustworthy-looking) faces might reflect the adaptive function to avoid threats. Alternatively, it might simply reflect a general negativity bias whereby contextual and facial stimuli that are congruent on the negative end of the continuum lead to stronger

attribution of the relevant dispositional trait. Experiment 2 addresses these alternative predictions by testing whether the domain within which both face and context stimuli vary qualifies the direction of the effect. To this aim, we confronted the bond between trustworthiness and threat with that between extroversion and happiness. Extroversion is an inherent constituent of sociability and signals whether a person may be friendly and likable (Brambilla & Leach, 2014; Brambilla et al., 2021a; Leach et al., 2007) and should show a similar association with sad/happy scenes.[2] As such, sad contexts would be congruent to introvert faces, while happy contexts would be congruent to extrovert faces. Whereas threatening stimuli put the individual in a state of alert, sad stimuli are not meant to serve an adaptive function of fight or flight. Therefore, we expected the negativity bias registered on (un)trustworthiness and threat not to emerge on introversion and sadness.

Sample size determination

We conducted the same sensitivity analysis using MorePower 6.0 (Campbell & Thompson, 2012). In Experiment 2, with 120 participants, assuming a power of .80 (two-tail, α = .05) we could detect a critical interaction effect as small as $f = .20$ ($\eta_p^2 = .04$) in a $2 \times 3 \times 2$ full-within ANOVA design.

Method

Participants and procedure

One hundred and twenty Italian participants (81 females, $M_{age} = 32.34$, $SD_{age} = 16.12$) volunteered to participate in the study. We advertised the study online as a study on face perception. Instructions informed participants that they would be presented with images of individuals located in different contexts. The experiment consisted of two blocks of 72 trials, with a brief pause in between. One block was identical to that administered in Experiment 1, with participants exposed to either trustworthy- or untrustworthy-looking faces surrounded by threatening, reassuring, or neutral contexts. Participants' job was to rate each person on perceived trustworthiness (−3 = Untrustworthy; +3 = Trustworthy). In the other block, participants saw either introverted- or extroverted-looking faces surrounded by sad, happy, or neutral contexts. They were asked to rate each person on perceived extroversion (−3 = Introverted; +3 = Extroverted). In each block, either for trustworthy or extroversion attribution, facial stimuli were selected from a sample of 24 identities manipulated in trustworthiness (i.e. 12 trustworthy and 12 untrustworthy) or extroversion (i.e. 12 extroverted and 12 introverted) and presented for a total of three times. Once with each of the three classes of visual contexts varying in threat (i.e. threatening, reassuring, neutral) or happiness (i.e. sad, happy, neutral), respectively. The order of the two blocks was counterbalanced between participants.

Stimuli

In addition to the 24 identities manipulated in trustworthiness adopted in Experiment 1, we employed other 24 digital identities manipulated in extroversion (12 extroverted, 12 introverted, Todorov et al., 2013) which were also modified to increase their ecological validity (see Figure 3).

Context stimuli (4 threatening, 4 reassuring, 4 sad, 4 happy) were selected via a pre-test ($N = 30$, 23 females, $M_{age} = 27.97$, $SD_{age} = 6.81$). Participants evaluated sixty-two pictures obtained from public domain websites on happiness (*"How sad-happy is this picture?"*, 1 = sad – 7 = happy) and threat (*"How threatening-reassuring is this picture?"*, 1 = threatening – 7 = reassuring). Four stimuli (i.e. fire, gun with bullets, volcano, blood stains) were selected to be threatening ($M = 1.52$, $SD = .55$) and four stimuli (i.e. pile of stones, lake in spring, lake in autumn, dock outstretched towards the sea) to be reassuring ($M = 6.38$, $SD = .53$), $t\ (29) = -32.71$, $p < .001$. Other four stimuli (i.e. abandoned factory, forgotten toy on a bench, broken piano keys, bad weather conditions) were selected to be sad ($M = 1.64$, $SD = .50$) and four stimuli (i.e. colourful balloons, fireworks in the sky, rainbow, colourful umbrellas) to be happy ($M = 6.32$, $SD = .52$), $t\ (29) = -35.11$, $p < .001$. The two sets of stimuli did not differ from each other in terms of absolute extremity measured on the domain of reference (negative stimuli: $p = .36$, positive stimuli: $p = .56$). Again, a grey rectangle was used as the control context.

Results

As in Experiment 1, we computed absolute scores to reflect the impact of congruent contexts (as opposed to both incongruent and neutral) on the attribution of the focal dispositional trait on each type of face. Therefore, we proceeded by recoding the positive and negative context as either congruent

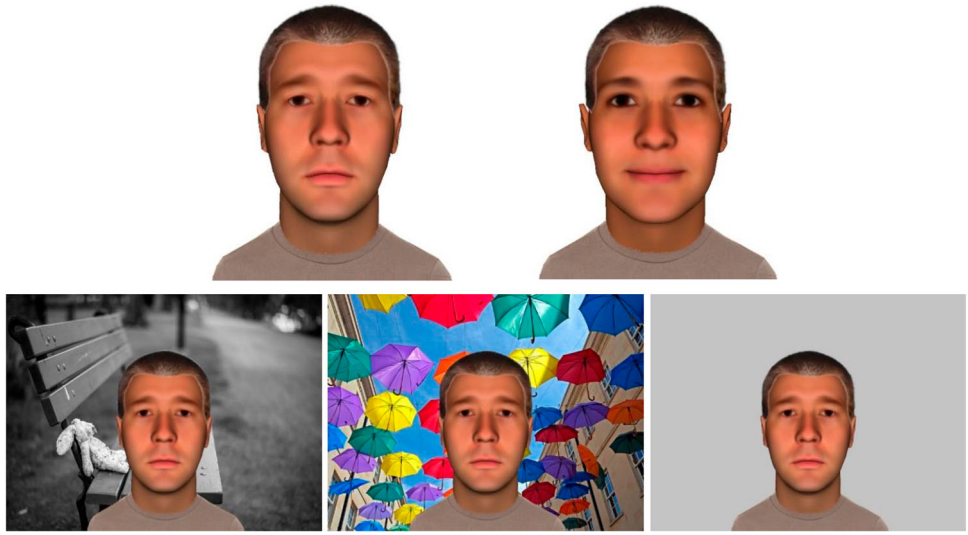

Figure 3. At the top, two sample faces employed in Experiment 2, introverted on the left and extroverted on the right. At the bottom, three samples of a face embedded into sad, happy, or neutral grey background contexts (from the left).

or incongruent. Also, scores received by negative faces on either dimension were reversed to reflect the absolute influence of the context on the trait attribution (see Figure 4).

We conducted a 2 (face: positive vs. negative) × 3 (context congruency: congruent vs. neutral vs. incongruent) × 2 (domain: trustworthiness vs. extroversion) ANOVA. We found a significant effect of face, $F(1, 119) = 16.16$, $p < .001$, $d = .37$, 95% CI [.18, .55], meaning that the attributions of the focal trait were more extreme for positive than negative faces. We also found the effect of context congruency, $F(1.26, 149.94) = 53.31$, $p < .001$, $\eta_p^2 = .31$, 95% CI [.23, .28], suggesting that within congruent contexts scores were more extreme than in neutral, $t(238) = 5.12$, $p < .001$, $d = .33$, 95% CI [.20, .46], and incongruent contexts, $t(238) = 10.32$, $p < .001$, $d = .67$, 95% CI [.53, .81], and in the latter the scores less extreme than in neutral contexts, $t(238) = -5.20$, $p < .001$, $d = .34$, 95% CI [.21, .47]. We also found a main effect of domain, $F(1, 119) = 47.32$, $p < .001$, $d = .63$, 95% CI [.43, .83], showing that extroversion scores were overall more extreme than trustworthiness scores. The interaction between face and context resulted significant, $F(1.92, 228.48) = 5.95$,

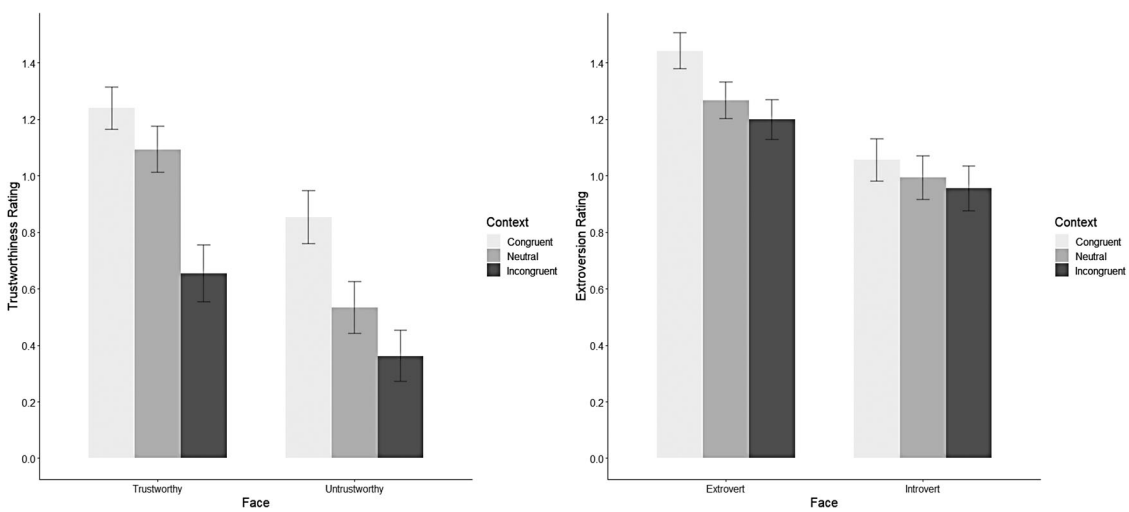

Figure 4. Results of Experiment 2.

$p = .003$, $\eta_p^2 = .05$, 95% CI [.01, .09], and so was the interaction between context and domain, $F(1.72, 204.68) = 28.21$, $p < .001$, $\eta_p^2 = .19$, 95% CI [.12, .26]. Instead, we found no interaction between face and domain, $F(1, 119) = .87$, $p = .35$. Importantly, the three-way interaction was significant, $F(1.82, 216.58) = 7.23$, $p = .001$, $\eta_p^2 = .06$, 95% CI [.02, .11]. Decomposing this interaction, we found that a negativity bias in the face-context congruency effect was qualified by the relevant domain, $t(238) = -3.56$, $p < .001$, $d = .23$, 95% CI [.10, .36]. Thus, in the trustworthiness domain we replicated the same findings observed in Experiment 1. Face-context congruency (i.e. absolute difference between congruent and neutral context) was significant for both negative, $t(799) = 6.15$, $p < .001$, $d = .22$, 95% CI [.15, .29], and positive faces, $t(799) = 2.79$, $p = .005$, $d = .10$, 95% CI [.03, .17], but the valence of the face moderated this effect significantly, $t(469) = 2.88$, $p = .004$, $d = .13$, 95% CI [.04, .22], such that the effect was stronger on the negative end. Instead, the moderating role of valence on face-context congruency in the extroversion domain showed marginal and in the opposite direction, $t(469) = -1.84$, $p = .07$, $d = .09$, 95% CI [−.01, .18]. The congruency effect was significant for positive, $t(799) = 3.34$, $p = .001$, $d = .12$, 95% CI [.05, .19], but not for negative faces, $t(799) = 1.20$, $p = .23$, $d = .04$, 95% CI [−.03, .11]. To clarify further our three-way interaction, we also looked at the difference in the congruency effect observed across the two domains by considering positive and negative face separately. We found that domain moderated significantly the congruency effect for negative faces, $t(456) = -4.03$, $p < .001$, $d = .19$, 95% CI [.10, .28], but not for positive faces, $t(456) = 0.44$, $p = .66$, $d = .02$, 95% CI [−.07, .11].

These findings suggest that face-context congruency leads to different effects depending on the relevant domain within which stimuli vary and are to be judged. Replicating Experiment 1, a negativity bias emerged on trustworthiness, such that threatening contexts influenced (negatively) the attribution of trustworthiness to congruent faces more than reassuring contexts did. This, however, was not the case when considering how sad and happy scenes impacted introverted and extroverted faces, respectively. Rather, we found a non-significant trend indicating that happy contexts influenced (positively) the attribution of extroversion to congruent faces more than sad contexts did on introversion. However, reassuring and happy contexts influenced equally the attribution of positive traits suggesting thus the absence of a positive evaluative asymmetry.

Experiment 3

In Experiment 1 we found a stronger integration when the face and the context were congruent on the negative end (i.e. threatening contexts and untrustworthy faces, as opposed to reassuring contexts and trustworthy faces). Experiment 2 replicated this finding and showed its domain-specificity. Yet, one main limitation that prevented us from drawing conclusive interpretations lay in the type of context stimuli used in our procedure. Focusing on the threat-trust domain, we realised that only threatening, but not reassuring contexts, portrayed scenes that were potentially ascribable to the human action. For instance, we used the image of a room covered with blood and a gun with bullets to represent threatening contexts, while pleasant landscapes with no human action-related objects served as reassuring contexts. Under such conditions, only in the threatening contexts, facial stimuli could be seen as responsible for the threat shown in the background. Thus, the critical effect (i.e. the difference in the impact of congruent vs. neutral contexts on trustworthy and untrustworthy faces) could reflect a mismatch in the humanity conveyed by either type of context, which might, in turn, explain the higher impact of threatening contexts in changing (negatively) the level of trustworthiness attributed to congruent facial stimuli. Corroborating this idea, Mattavelli et al. (2021) recently found that face-context integration on the attribution of trustworthiness is qualified by the relationship that the perceiver establishes between the face and the context. Namely, face-context integration is stronger when context stimuli represent scenes that are potentially ascribable to target face (e.g. a room covered with blood), than when stimuli are just threatening (e.g. a tornado).

Experiment 3 was designed to address this issue. The critical novelty introduced in this experiment was the nature of the context stimuli. Because selecting context images ascribable to the human action that could be symmetric on the threatening-reassuring continuum was no trivial task, we employed an acquisition phase at the beginning of the main experiment. In this phase, participants were exposed to eight context stimuli accurately selected via a pilot test: four were selected to be (visually) threatening and four to be reassuring. Unlike the context scenes used in prior studies, here neither threatening nor reassuring scenes referred to the human action. Rather, each image was paired with eight sentences (also pre-tested) describing either threatening or

reassuring human actions. After this phase, participants learned that both threatening and reassuring visual scenes were associated with a specific human action.

We preregistered the entire protocol of the study at OSF (https://osf.io/nd4w7)

Sample size determination

We determined our sample size based on an a-priori power analysis conducted with the pwr::pwr.f2.test() function in R. The critical effect concerned the difference in the impact of congruent vs. neutral contexts on trustworthy and untrustworthy faces. We decided to base our analysis on a small effect size $d = .15$. This effect was slightly smaller than the average effect found across our three prior experiments. At = .05 with a power = .90, the analysis suggested an overall sample of 142 participants. Because we assumed that the critical effect could be conditional on participants' ability to correctly learn whether the critical context scenes are either threatening or reassuring, we decided to slightly oversample with a 10% rate.

Method

Participants and procedure

One hundred and fifty-four participants (62 females, $M_{age} = 28.16$, $SD_{age} = 8.14$) took part in the study via Prolific Academic. Participants were asked to participate in a study about face perception. Instructions informed them that in the first phase they would see a series of images of context scenes appearing on the screen along with a sentence describing a specific situation involving human beings. This phase consisted of a single block of 32 trials, in which eight visual context stimuli were paired with eight sentences describing an action selected to be consistent with the visual nature of the context stimulus (e.g. threatening scenes with threatening actions). Each pair appeared four times with no time constraints: participants could proceed to the next trial by pressing the spacebar. After this acquisition phase, a memory test was administered. This test consisted of 8 trials, in which each context scene presented in the acquisition phase appeared individually in the centre of the screen. For each scene, participants were to indicate whether the sentence presented with that scene in the previous phase described either a threatening or a reassuring human action. Next, participants were presented with the standard procedure adopted in our previous studies. Across two consecutive blocks of 48 trials each, 12 trustworthy- and 12 untrustworthy-looking faces appeared in 4 threatening, 4 reassuring, or 4 neutral contexts (see Figure 5). Participants were asked to rate each person on perceived trustworthiness using a 7-point Likert scale (−3 = Untrustworthy; +3 = Trustworthy). Facial stimuli appeared in random order without replacement. In each trial, either trustworthy- or untrustworthy-looking faces could appear in two out of four randomly extracted threatening,

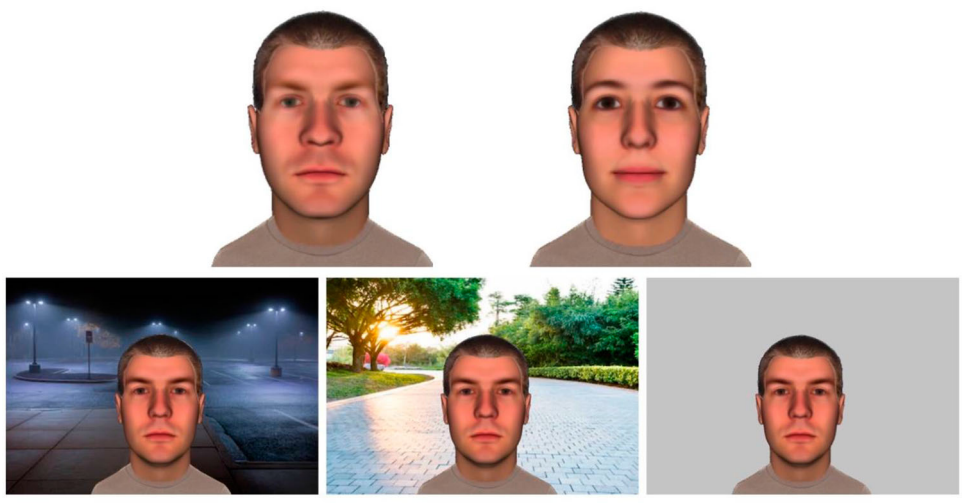

Figure 5. At the top, two sample faces employed in Experiment *3*, untrustworthy on the left and trustworthy on the right. At the bottom, three samples of a face embedded into threatening, reassuring, or neutral grey background contexts (from the left).

reassuring, or neutral contexts. Thus, each of the 24 face identities appeared four times in total, twice with each class of context. No time limit was set even though participants were kindly reminded to provide their judgments as fast as possible.

Stimuli

Context and human actions. We conducted a pilot test ($N = 43$, 30 females, $M_{age} = 32.93$, $SD_{age} = 13.47$). Participants were presented with 54 images of potential threatening and reassuring images taken from the web and 36 sentences describing threatening and reassuring scenarios. Because we wanted both images and sentences to be high in conveyed threat (reassurance) but low in reassurance (threat), we asked participants to rate each stimulus on both dimensions. To this aim, each stimulus was presented followed by the questions "How threatening is this scene (situation)?" (1 = not at all; 7 = extremely) and "How reassuring is this scene (situation)?" (1 = not at all; 7 = extremely).

The four threatening context scenes selected (i.e. two parking areas (night image), an abandoned building, a lonely street (night image)) received an average score of 4.83 ($SD = 0.83$) on threat, and of 1.69 ($SD = 0.83$) on reassurance. The four reassuring context scenes (i.e. two parks (sunny image), a new room, an elegant neighbourhood (sunny image)) received an average score of 1.57 ($SD = 1.03$) on threat, and of 4.91 ($SD = 1.62$) on reassurance. There was a significant difference between the two sets on threat and reassurance, $t(171) = 23.17$, $p < .001$ and $t(171) = -25.49$, $p < .001$. Instead the two did not differ in extremity on both their main dimension (i.e. threat for stimuli selected to be threatening, reassurance for stimuli selected to be reassuring), $t(171) = -.56$, $p = .574$, and their secondary dimension (i.e. reassurance for stimuli selected to be threatening, threat for stimuli selected to be reassuring), $t(171) = 1.20$, $p = .230$.

The four threatening human actions (i.e. "someone molested some people", "someone intimidated some tourists", "someone scared a child", "someone offended an old man") received an average score of 5.49 ($SD = 1.26$) on threat, and of 1.38 ($SD = 0.80$) on reassurance. The four reassuring actions (i.e. "someone helped an old lady", "someone offered moral support to some kids", "someone cheered up some strangers", "someone supported some guys") received an average score of 1.53 ($SD = 0.99$) on threat, and of 5.47 ($SD = 1.27$) on reassurance. There was a significant difference between the two sets of actions on threat and reassurance, $t(171) = 33.18$, $p < .001$ and $t(171) = -34.79$, $p < .001$. Instead the two did not differ in extremity on both their main dimension, $t(171) = .24$, $p = .809$, and their secondary dimension, $t(171) = -1.60$, $p = .111$.

Faces. We employed the 24 computer-generated identities (12 trustworthy, 12 untrustworthy) used in Experiment 1a-2.

Results

Only one participant showed poor memory performance after the acquisition phase (less than 75% of correct trials). In line with our pre-registered analytic plan, we adopted the same set of analyses presented in Experiment 1. The critical effect was the absolute impact of congruent contexts (as opposed to neutral contexts) on the attribution of trustworthiness on either trustworthy or untrustworthy faces.

We conducted a 2 (face: untrustworthy vs. trustworthy) × 3 (context congruency: congruent vs. neutral vs. incongruent) ANOVA (see Figure 6). Degrees of freedom and probability values have been corrected whenever the assumption of sphericity has been violated. We did not find a significant effect of face, $F(1, 153) = 0.09$, $p = .76$. The effect of context congruency was significant, $F(2, 306) = 122.24$, $p < .001$, $\eta_p^2 = .44$, 95% CI [.38, 1.00]. Congruent contexts made the trustworthiness scores more extreme than both neutral, $t(306) = 7.92$, $p < .001$, $d = .45$, 95% CI [.34, .57], and incongruent context, $t(306) = 15.64$, $p < .001$, $d = .89$, 95% CI [.76, 1.03]. In line with our expectations, the interaction was significant, $F(2, 306) = 9.33$, $p < .001$, $\eta_p^2 = .06$, 95% CI [.02, 1.00].

We decomposed this interaction by looking at the direct contrasts between congruent, incongruent, and neutral contexts across trustworthy and untrustworthy faces. Of interest for our investigation, comparing the impact of congruent vs neutral contexts on untrustworthy vs. trustworthy faces yielded a significant effect, $t(306) = 2.97$, $p = .003$, $d = .17$, 95% CI [.06, .28]. In line with prior studies, this effect indicated a stronger congruency effect on the negative rather than on the positive side of the trust-threat binomial. There was no difference in the impact of congruent vs. incongruent contexts on untrustworthy (vs. trustworthy) faces, $t(306) = -1.24$, $p = .22$, $d = .07$, 95% CI [−.04, .18]. Comparing incongruent and neutral

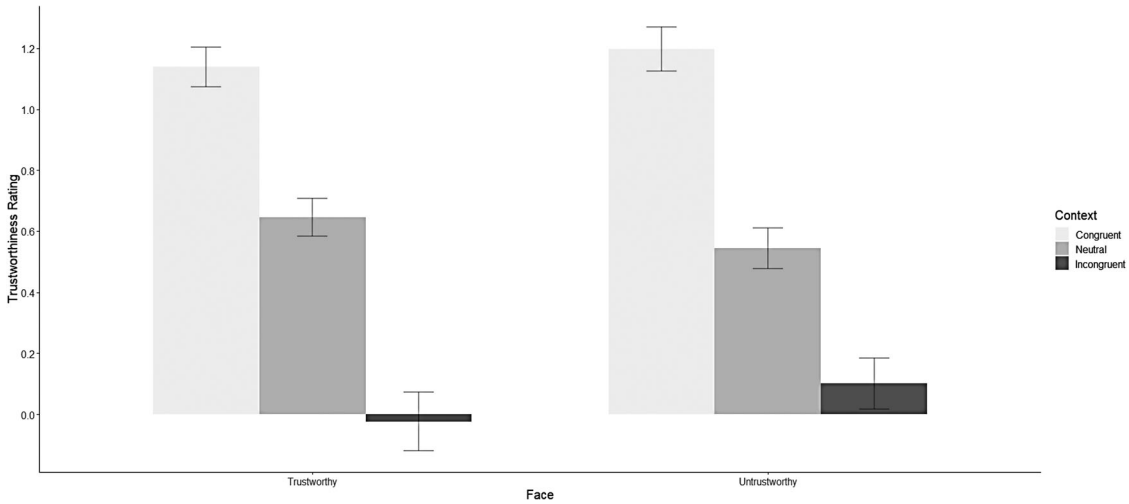

Figure 6. Results of Experiment 3.

context yielded a stronger effect on trustworthy, relative to untrustworthy, faces, $t(306) = 4.20$, $p < .001$, $d = .30$, 95% CI [.13, .35].

Importantly, the advantage in face-context integration for untrustworthy (vs, trustworthy) presented in threatening (vs. reassuring) contexts was found even in the absence of a baseline imbalance between either type of face, $t(191) = 1.07$, $p = .28$.

General discussion

Across three studies, we extended previous findings on face-context integration on the perception of human trustworthiness by investigating the conditions under which such integration is more likely to affect trustworthiness attribution. Specifically, we focused on face-context congruency. Face-context congruency occurs when the face and the context overlap with respect to (i) their signal value and (ii) their valence (Righart & De Gelder, 2008). We assessed face-context congruency as the absolute variation in the attribution of trustworthiness to facial stimuli, by comparing congruency on the positive vs. the negative extremes of the threat-trust domain. A stronger congruency effect was found in the two experiments when the context matched the face on the negative (vs. positive) extreme. Experiment 2 also proved that this negativity effect was unique of the threat-trust domain. In fact, when happy/sad scenes (in place of threatening/reassuring) were matched with extroverted/introverted faces (in place of untrustworthy/trustworthy), the superior impact of negative, over positive, face-context congruency faded away. Experiment 3 clarified that the advantage in face-context integration for the face and context stimuli belonging to the negative end of the threat-trust domain was still detectable when threatening and reassuring contexts were equally ascribable to the human action.

In unveiling how information derived from the context is integrated into the perception of facial trustworthiness, our findings converge to the idea that contextual information might elicit cues that are functional for the individual to "act upon" the embedded face. Namely, context stimuli might serve as cues that foster behavioural programmes (Darwin, 1872/1998; Panksepp, 1998; Russell & Barrett, 1999), which in turns produce consistent responses on stimuli of significant value. In the case of simultaneous processing of both contexts and faces, the context seems to foster an adaptive response to the facial stimulus. Consistent with this idea is the negativity bias observed across the three experiments for facial and context stimuli manipulated along the threat-trust domain. As far as we are aware, this is the first study to show a negativity bias in face-context integration and to suggest that such an integration might follow an adaptive path. Moreover, we showed that negativity bias in face-context congruency is not ubiquitous. Rather, it is moderated by the specific domain cued by the stimuli (and relevant for the final dispositional attribution). Thus, despite the signal value congruency between sad contexts and introverted faces, such congruency did not show stronger than the one operationalised on the positive end (happy contexts and extroverted

faces). If any, the congruency effect showed stronger on the positive end.

Experiment 3 also clarified one important element of confound in the interpretation of the stronger congruency on the negative vs. positive extreme of the threat-trust domain. Namely, in Experiments 1–2, we employed threatening contexts that could be attributed to the human action, whereas reassuring stimuli were not. Thus, the negativity effect observed in trustworthiness' attribution could have simply reflected an imbalance in the extent to which facial and contextual stimuli matched in either condition: participants might have inferred that untrustworthy-looking faces were responsible for the threatening scene portrayed on the background, whereas the same inference was not possible in for trustworthy-looking faces presented in reassuring scenes. In fact, recent research documented that face-context integration is encoded relationally by the perceiver: the stronger the relationship established between the face and the context, the higher the face-context integration (Mattavelli et al., 2021). Experiment 3 neutralised this imbalance and showed that, even when both threatening and reassuring stimuli were conditioned to be equally ascribable to the human action, a superior congruency effect on the negative end of the threat-trust domain was found.

In supporting the importance of the context and its congruency with the face, our results reconcile nicely with an ecological account of person perception (Gibson, 1977). One major pillar of Gibson's theory is that perception reflects the detection of behavioural affordances. Namely, it is the action that environmental cues suggest in relation to the target that ultimately affects our response (Zebrowitz & Montepare, 2006). To exemplify, one should take the image of an apple as the target stimulus. If the apple appears in a fruit basket, then the apple is likely to be approached. However, if the same apple appears in the hand of Snow White's stepmother, the behavioural script suggested by the relevant context might change dramatically. Likewise, context stimuli affect face perception. The ecological account also posits that our attunement to stimulus information produces overgeneralised and biased perception. Under this view, contextual visual scenes can reveal human features that are adaptive to detect and therefore more likely to be ascribed to individuals who seemingly possess similar features. In its being erroneous, overgeneralisation occurs because it is a less maladaptive error than the one that might result from underestimating or failing to respond to the stimulus (Hammond, 2007). In our studies, overgeneralisation was measured as a difference in the attribution of dispositional characteristics on a continuous scale. Future studies should explore whether these effects generalise on accuracy, recognition, and behaviour. For instance, one might wonder whether untrustworthy faces processed in threatening context are also more likely to foster physical avoidance by the perceiver.

This contribution is not free from limitations. One concerns the nature of the facial stimuli, as we employed only computer-generated faces. This was done for two main reasons. First, we wanted to be consistent with the previous studies on face-context that inspired the present investigation (Brambilla et al., 2018). Second, this specific set of computer-generated faces has the advantage to reduce familiarity with the facial stimuli, as fewer facial features are available to the perceiver to allow memory. This latter aspect was particularly relevant for the paradigm we adopted in the present set of studies, where the same face identity could be presented in distinct context scenes. However, we acknowledge that including only computer-generated faces prevent us from generalising the current findings to real-world facial stimuli. We therefore call for future investigations to address this important issue. Moreover, the studies presented here were always advertised as studies investigating face perception and, throughout the experimental procedure, participants received explicit instructions to judge the person, whereas no emphasis was put on the context. This was made to make the experimental procedure as comparable as possible with standard face perception studies. Nonetheless, we might have minimised the role of the context scenes in affecting dispositional attributions. Importantly, we found that even under such counterproductive conditions, the nature of the context scenes mattered for the dispositional attributions made on the faces. Yet, we believe that future studies should assess whether and how the role played by the context in face-context integration varies when the perceiver is not provided with cues that might orient their attention toward the face. Finally, critical readers might argue that the superior impact of congruent contexts on untrustworthy faces could be influenced by the fact that, in Experiments 1–2, untrustworthy faces were less extreme to start with (i.e. comparison between the attribution of trustworthiness in the neutral contexts). Namely, such an imbalance might imply that trustworthy faces have less room to move upward in their trustworthiness

rating (due to reassuring contexts). In acknowledging this imbalance, we highlight two important points. First, in both studies the evaluations of trustworthy faces in neutral contexts, although more extreme than their untrustworthy counterparts, were far from ceiling. This means that participants had still room to alter their evaluation of the trustworthy faces by moving towards the positive end of the trustworthiness scale. Second, and perhaps more important, the crucial effect replicated successfully in Experiment 3, where no initial imbalance between trustworthy and untrustworthy faces was found.

In sum, our findings extend previous literature on face-context integration by focusing on the impact of such integration on attribution of trustworthiness. We confirmed that a privileged bond between the domains of threat and trust exists, such that threatening contexts increase the attribution of untrustworthiness to untrustworthy faces. Significantly, we added that this face-context congruency is not observed ubiquitously on the extent of the threat-trust continuum but it is characteristic of the negative extreme. This latter finding speaks for an adaptive function of the role of the context in affecting face perception and dispositional attribution. Such an adaptive function was further confirmed by the fact that the same negativity bias was not detectable when face-context congruency was operationalised on a distinct domain. We therefore conclude that, by conveying information of adaptive significance, contextual cues bias the attribution of trustworthiness to facial stimuli by affording specific behavioural schemata that affect human responses.

Notes

1. This approach is in line with the statistical design used in prior studies on face-context integration (e.g. Brambilla et al., 2018).
2. This was also tested empirically on a separate investigation. Ninety-nine participants (48 females, *Mage* = 29.33, *SDage* = 9.74) were asked to indicate the likelihood with which they would say that a person would possess the feature Y (e.g. being introvert) knowing that they also possess the feature X (e.g. being sad). This question was asked for the four pairs of dispositional features tested in Experiment 2 (i.e. threat-untrustworthiness, reassurance-trustworthiness, sadness-introversion, happiness-extroversion). Participants answered on a scale from 0 to 100. We found that all the associations between each pair of matched features were moderately-high (threat-untrustworthiness: $M = 65.03$, $SD = 21.07$; reassurance-trustworthiness: $M = 58.36$, $SD = 19.94$; sadness-introversion: $M = 60.37$, $SD = 20.41$; happiness-extroversion: $M = 60.77$, $SD = 18.41$).

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

The author(s) reported there is no funding associated with the work featured in this article.

Data availability

All the anonymized data and materials of the studies reported in the manuscript are available. In the manuscript, we provide the link to the project on Open Science Framework. The data have not been used in prior published or in press manuscripts.

ORCID

Simone Mattavelli http://orcid.org/0000-0002-8934-8016
Matteo Masi http://orcid.org/0000-0001-6436-1663
Marco Brambilla http://orcid.org/0000-0002-4774-3309

References

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

Aviezer, H., Hassin, R. R., Ryan, J., Grady, C., Susskind, J., Anderson, A., … Bentin, S. (2008). Angry, disgusted, or afraid? Studies on the malleability of emotion perception. *Psychological Science*, *19*(7), 724–732. https://doi.org/10.1111/j.1467-9280.2008.02148.x

Barrett, L. F., & Kensinger, E. A. (2010). Context is routinely encoded during emotion perception. *Psychological Science*, *21*(4), 595–599. https://doi.org/10.1177/0956797610363547

Baumeister, R. F., Bratslavsky, E., Finkenauer, C., & Vohs, K. D. (2001). Bad is stronger than good. *Review of General Psychology*, *5*(4), 323–370. https://doi.org/10.1037/1089-2680.5.4.323

Brambilla, M., Biella, M., & Freeman, J. B. (2018). The influence of visual context on the evaluation of facial trustworthiness. *Journal of Experimental Social Psychology*, *78*, 34–42. https://doi.org/10.1016/j.jesp.2018.04.011

Brambilla, M., & Leach, C. W. (2014). On the importance of being moral: The distinctive role of morality in social judgment. *Social Cognition*, *32*(4), 397–408. https://doi.org/10.1521/soco.2014.32.4.397

Brambilla, M., Masi, M., Mattavelli, S., & Biella, M. (2021b). Faces and sounds becoming one: Cross-modal integration of facial and auditory cues in judging trustworthiness. *Social Cognition*, *39*(3), 315–327. https://doi.org/10.1521/soco.2021.39.3.315

Brambilla, M., Sacchi, S., Rusconi, P., & Goodwin, G.. (2021a). The primacy of morality in impression development: Theory,

research, and future directions. In *Advances in experimental social psychology* (Vol. 64, pp. 187–262). Academic Press. https://doi.org/10.1016/bs.aesp.2021.03.001.

Campbell, J. I., & Thompson, V. A. (2012). MorePower 6.0 for ANOVA with relational confidence intervals and Bayesian analysis. *Behavior Research Methods*, *44*(4), 1255–1265. https://doi.org/10.3758/s13428-012-0186-0

Carroll, J. M., & Russell, J. A. (1996). Do facial expressions signal specific emotions? Judging emotion from the face in context. *Journal of Personality and Social Psychology*, *70*(2), 205–218. https://doi.org/10.1037/0022-3514.70.2.205

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

Cosmides, L., & Tooby, J. (1992). Cognitive adapta- tions for social exchange. In J. Barkow, L. Cosmides, & J. Tooby (Eds.), *The adapted mind: Evolutionary Psychology and the generation of culture* (pp. 162–228). Oxford University Press.

Darwin, C. (1998). *The expression of the emotions in man and animals* (3rd ed). Oxford University Press. (Original work published 1872).

Duarte, J., & Siegel, S. (2012). Trust and credit: The role of appearance in peer-to-peer lending. *The Review of Financial Studies*, *25*(8), 2455–2484. https://doi.org/10.1093/rfs/hhs071

Fiske, S. T. (1992). Thinking is for doing: Portraits of social cognition from daguerreotype to laserphoto. *Journal of Personality and Social Psychology*, *63*(6), 877–889. https://doi.org/10.1037/0022-3514.63.6.877

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*(2), 206–210. https://doi.org/10.1016/j.jesp.2012.10.015

Furnham, A., & Brewin, C. R. (1990). Personality and happiness. *Personality and Individual Differences*, *11*(10), 1093–1096. https://doi.org/10.1016/0191-8869(90)90138-H

Gibson, J. J. (1977). The theory of affordances. In R. Shaw, & J. Bransford (Eds.), *Perceiving, acting, and knowing* (pp. 67–82). Lawrence Erlbaum.

Hammond, K. R. (2007). *Beyond rationality: The search for wisdom in a troubled time*. Oxford University Press.

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*(2), 234–249. https://doi.org/10.1037/0022-3514.93.2.234

LeDoux, J. (1996). Emotional networks and motor control: A fearful view. *Progress in Brain Research*, *107*, 437–446. https://doi.org/10.1016/S0079-6123(08)61880-4

Mattavelli, S., Masi, M., & Brambilla, M. (2021). Not just about faces in context: Face–context relation moderates the impact of contextual threat on facial trustworthiness. *Personality and Social Psychology Bulletin*. https://doi.org/10.1177/01461672211065933

McArthur, L. Z., & Baron, R. M. (1983). Toward an ecological theory of social perception. *Psychological Review*, *90*(3), 215–238. https://doi.org/10.1037/0033-295X.90.3.215

Panksepp, J. (1998). The periconscious substrates of consciousness: Affective states and the evolutionary origins of the self. *Journal of Consciousness Studies*, *5*(5-6), 566–582.

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*(3), e34293. https://doi.org/10.1371/journal.pone.0034293

Righart, R., & De Gelder, B. (2006). Context influences early perceptual analysis of faces—an electrophysiological study. *Cerebral Cortex*, *16*(9), 1249–1257. https://doi.org/10.1093/cercor/bhj066

Righart, R., & De Gelder, B. (2008). Rapid influence of emotional scenes on encoding of facial expressions: An ERP study. *Social Cognitive and Affective Neuroscience*, *3*(3), 270–278. https://doi.org/10.1093/scan/nsn021

Rusconi, P., Sacchi, S., Brambilla, M., Capellini, R., & Cherubini, P. (2020). Being honest and acting consistently: Boundary conditions of the negativity effect in the attribution of morality. *Social Cognition*, *38*(2), 146–178. https://doi.org/10.1521/soco.2020.38.2.146

Russell, J. A., & Barrett, L. F. (1999). Core affect, prototypical emotional episodes, and other things called emotion: Dissecting the elephant. *Journal of Personality and Social Psychology*, *76*(5), 805–819. https://doi.org/10.1037/0022-3514.76.5.805

Stirrat, M., & Perrett, D. I. (2010). Valid facial cues to cooperation and trust: Male facial width and trustworthiness. *Psychological Science*, *21*(3), 349–354. https://doi.org/10.1177/0956797610362647

Tamietto, M., Corazzini, L. L., de Gelder, B., & Geminiani, G. (2006). Functional asymmetry and interhemispheric cooperation in the perception of emotions from facial expressions. *Experimental Brain Research*, *171*(3), 389–404. https://doi.org/10.1007/s00221-005-0279-4

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

Todorov, A., Olivola, C. Y., Dotsch, R., & Mende-Siedlecki, P. (2015). Social attributions from faces: Determinants, consequences, accuracy, and functional significance. *Annual Review of Psychology*, *66*(1), 519–545. https://doi.org/10.1146/annurev-psych-113011-143831

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

Williams, D. G. (1990). Effects of psychoticism, extraversion, and neuroticism in current mood: A statistical review of six studies. *Personality and Individual Differences*, *11*(6), 615–630. https://doi.org/10.1016/0191-8869(90)90045-S

Willis, J., & Todorov, A. (2006). First impressions: Making up your mind after a 100-ms exposure to a face. *Psychological Science*, *17*(7), 592–598. https://doi.org/10.1111/j.1467-9280.2006.01750.x

Wilson, K., & Gullone, E. (1999). The relationship between personality and affect over the lifespan. *Personality and Individual Differences*, *27*(6), 1141–1156. https://doi.org/10.1016/S0191-8869(99)00058-6

Zebrowitz, L. A., & Montepare, J. (2006). The ecological approach to person perception: Evolutionary roots and contemporary offshoots. In M. Schaller, J. A. Simpson, & D. T. Kenrick (Eds.), *Evolution and social psychology* (pp. 81–113). Psychosocial Press.