Empirical Research Paper

Not Just About Faces in Context: Face–Context Relation Moderates the Impact of Contextual Threat on Facial Trustworthiness

Personality and Social Psychology Bulletin
1–15

DOI: 10.1177/01461672211065933
journals.sagepub.com/home/pspb
SAGE

Simone Mattavelli[1], Matteo Masi[1], and Marco Brambilla[1]

Abstract

Recent work showed that the attribution of facial trustworthiness can be influenced by the surrounding context in which a face is embedded: contexts that convey threat make faces less trustworthy. In four studies ($N = 388$, three preregistered) we tested whether face–context integration is influenced by how faces and contexts are encoded relationally. In Experiments 1a to 1c, face–context integration was stronger when threatening stimuli were attributable to the human action. Faces were judged less trustworthy when shown in threatening contexts that were ascribable (vs. non-ascribable) to the human action. In Experiment 2, we manipulated face–context relations using instructions. When instructions presented facial stimuli as belonging to the "perpetrators" of the threatening contexts, no difference with the control (no-instructions) condition was found in face–context integration. Instead, the effect was reduced when faces were presented as "victims." We discussed the importance of considering relational reasoning when studying face–context integration.

Keywords

face perception, context, trustworthiness, relational encoding

Received June 1, 2021; revision accepted November 22, 2021

The human face is a powerful source that we use to make inferences on others' dispositions (Zebrowitz & Montepare, 2008). From someone's face, we determine whether a person is friendly, warm, or worth being trusted (Todorov et al., 2015). Most studies on face processing have been conducted by considering facial expressions without any contextual information. Indeed, faces tend to be flashed on the computer screen, and evaluations quickly ensue (for a review, Todorov et al., 2015). However, in real life, faces are rarely encountered in isolation, and the context in which they appear is often very informative. Indeed, recent research has shown that the attribution of trustworthiness to faces can be influenced by the threat conveyed by the context in which faces are embedded (Brambilla et al., 2018, 2021a; Mattavelli et al., 2021). Faces appeared more untrustworthy when they were surrounded by threatening visual contexts rather than merely negative or neutral contexts. These findings converge toward a privileged link between the dimensions of trustworthiness and threat (Brambilla et al., 2019, 2021b; Brambilla & Leach, 2014; Willis & Todorov, 2006). Unclear, however, remains the nature of the mental processes that qualify this link when it comes to face–context integration. Here, we connect face–context integration and conditioning research to shed light on the processes through which contextual threat can affect facial trustworthiness. We drew from conditioning research and borrowed the idea that humans are not passive organisms when exposed to pairings between two classes of stimuli. Rather, they impose specific encoding relationships on such pairings (De Houwer & Hughes, 2016; Fiedler & Unkelbach, 2011). Thus, we propose that the effect of face–context integration on the attribution of trustworthiness is conditional upon the nature of the relationship that can be established between the face and the context.

Inferring Trustworthiness From Faces and the Surrounding Context

When interacting with others, we seek cues that can tell us whether an individual deserves to be trusted (Ames et al., 2011). One source that people often use to ascribe trustworthiness to others is a person's face (see Todorov et al., 2015, for a review). Several facial features can convey trustworthiness.

[1]University of Milano-Bicocca, Italy

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

The shape or structure of a person's face can lead us to attribute certain traits to an individual and make assumptions about their behaviors (Oosterhof & Todorov, 2008; Sutherland et al., 2013; Vernon et al., 2014). The power of facial features to induce attributions of trustworthiness affects our intentions and behaviors. For instance, people invest less money with partners who look untrustworthy (Chang et al., 2010; Rezlescu et al., 2012; Stirrat & Perrett, 2010), whereas trustworthy-looking individuals have a higher chance of being granted loans (Duarte et al., 2012). The importance of detecting trustworthiness in others lies in the fact that (a) deciding whether a person represents either an opportunity or a threat is a decision that has a highly adaptive function, and (b) the dimensions of threat and trustworthiness are inherently linked (Ames et al., 2011; Brambilla & Leach, 2014; Cosmides & Tooby, 1992). Indeed, trustworthy and untrustworthy individuals are perceived as beneficial and harmful, respectively (Todorov et al., 2015; see also Brambilla et al., 2021b; Brambilla & Leach, 2014).

Notwithstanding the vast amount of information that a face can covey, there is now increasing evidence that contextual information influences how facial cues shape impression formation (Carroll & Russell, 1996; Righart & De Gelder, 2006, 2008). The context in which a face appears affects the way in which 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). For instance, Righart and De Gelder (2008) found that facial expressions of fear, happiness, and disgust are more easily recognized when presented in a fearful, happy, and disgusting scene, respectively. Trustworthiness does not make an exception. Indeed, the evaluation of facial trustworthiness is influenced by the level of threat conveyed by the visual scene in which faces were embedded. Thus, untrustworthy faces are more easily categorized as such when surrounded by threatening rather than negative or neutral visual contexts (Brambilla et al., 2018; see also Brambilla et al., 2021a).

On the Link Between Contextual Threat and Facial Trustworthiness

The impact of contextual threat on the attribution of facial trustworthiness speaks for the inherent bond between the two dimensions of trust and threat (Willis & Todorov, 2006). In fact, prior studies have shown that the perception of trustworthiness activates brain areas (i.e., amygdala) that are also implicated in the detection of potentially threatening stimuli (Winston et al., 2002). Moreover, according to a functional perspective of person perception, our judgments of another person's trustworthiness are highly related to the essential decisions we must make about whether they represent an opportunity or a threat (Ames et al., 2011; Brambilla & Leach, 2014; Cosmides & Tooby, 1992). This functional view can be easily applied to attribution of trustworthiness due to face–context integration, that is, threatening contexts should set the occasion for the perceiver to spot those cues that might alert an incoming threat. When such an opportunity of threat comes from facial features, that face is more likely to be judged as untrustworthy. In line with this view, Mattavelli et al. (2021) recently showed that the bond between threat and trust leads to stronger integration when both the context and the face are presented to be negative (threatening context with untrustworthy face) than positive (reassuring context with trustworthy face). Such a negativity bias fits well with an adaptive interpretation of face–context integration on the attribution of facial trustworthiness: By cueing emotions, context stimuli foster behavioral programs (Russell & Barrett, 1999) that in turn produce consistent responses on stimuli of significant value. Whereas previous findings have shown that perceived trustworthiness and threat are linked, no studies have focused directly on the nature of that link. Here, we investigate face–context integration by asking (a) what is inferred about the relationship between a threatening context and a face when the two happen to occur together, and (b) how such inferences can be manipulated to ultimately affect face–context integration.

Face–Context Integration as Stimuli-Pairing: Insights From Conditioning Research

From an operational perspective, face–context integration results from the pairing of two classes of stimuli: a face is flashed on screen and processed by the perceiver while a contextual scene is displayed on the background. Conditioning research has largely demonstrated that, when two stimuli are paired, the properties carried by one stimulus can transfer to the other (see Hofmann et al., 2010, for a review). Different accounts explain what type of mental processes could be responsible for such an effect (for an overview, see Hofmann et al., 2010). Among them, the propositional account (De Houwer, 2009, 2018; Mitchell et al., 2009) assumes that pairings are conceptually identical to verbal propositions: just as predicates relate subjects with objects in a proposition, so does pairing. However, parings differ from predicates in that the nature of the relationship between two stimuli is not overtly specified. Thus, the ultimate effect of stimuli pairing should result from an intersection between bottom-up and top-down processes: the effect is driven by the informational value conveyed by the stimuli (*bottom-up*) and by the inference that the perceiver makes about the relation between those stimuli (*top-down*). In line with this idea, the extent to which the pairing of two stimuli results in a transfer of properties from one stimulus to the other depends on the propositions that the perceiver forms about the relation between the paired stimuli. For instance,

the effect of pairing can change dramatically depending on whether participants are told that paired stimuli have either the *same* or the *opposite* meaning (e.g., Fiedler & Unkelbach, 2011; Moran & Bar-Anan, 2013) or that one stimulus either *causes* or *prevents* the other (Hughes et al., 2019). Arguably, the same reasoning might hold for faces that are presented in (paired with) contextual scenes. For instance, imagine that Bob's face is presented within a crime scene. Depending on whether Bob committed that crime or he was killed in that very same crime scene might trigger opposite responses in the perceivers. Crucially, whether one or the other relation is established between the face and the context can have an impact on the ultimate attribution of trustworthiness to Bob's face. In other words, the ability of the perceiver to construct meaning upon the relation between two classes of paired stimuli might be crucial for the property (i.e., threat) of one stimulus (i.e., the context) to affect the dispositional attribution (i.e., trustworthiness) made upon the other stimulus (i.e., the face).

The Present Research

This contribution tests the idea that the impact of face–context integration on the attribution of facial trustworthiness can be qualified by the extent to which the perceiver can form meaningful relationships between faces and their context. We hypothesize that face–context integration results from more than a mere response activated by the context (threat) and emitted toward the target face. Rather, the relationship established between the contextual threat and the face should modulate their integration. To test this prediction, we conducted four studies in which we manipulated the meaning of such a relationship in different ways. First, in Experiments 1a to 1c we altered the nature of the threatening contexts to vary the extent to which the threat conveyed by such contexts could be attributable to human stimuli. Thus, we chose threatening stimuli that were comparable in the conveyed threat, but were different with respect to their being ascribable to human actions. We tested the impact of such a manipulation on both computer-generated (Experiment 1a) and real (Experiments 1b and 1c) facial stimuli. In all the studies, the focal hypothesis concerned the differential impact of human versus nonhuman threatening contexts on the attribution of trustworthiness to facial stimuli. Thus, the impact of either type of threatening context was compared with that of a neutral context and the resulting delta was used as an estimate of their differential impact. In addition, facial stimuli were manipulated across the three studies to look either trustworthy or untrustworthy. Although we did not have a specific hypothesis on the interaction effect between context and faces, we inspected whether the hypothesized superior effect of a human threatening context was further qualified by such a facial feature. In Experiment 1c, we further tested whether the hypothesized effect could generalize across target gender. Finally, in Experiment 2, we borrowed a manipulation used in conditioning research to alter the nature of the relation between the face and the context by using explicit instructions. Such instructions were meant to inform participants about the nature of the relationship between the threatening contexts and the facial stimuli.

All studies received formal approval from the ethics committee of the University of Milano-Bicocca. We preregistered the entire protocols of Experiments 1b, 1c, and 2 on Open Science Framework (Experiment 1b: https://osf.io/q7t6u; Experiment 1c: https://osf.io/bzkca; Experiment 2: https://osf.io/2x59w). All the analysis codes are also available on Open Science Framework (https://osf.io/unbf2/). We reported all the manipulations and measures used in each study.

Experiment 1a

Prior studies investigating face–context integration on facial trustworthiness (Brambilla et al., 2018) employed a pool of heterogeneous threatening stimuli. Some threatening stimuli portrayed scenarios in which the target was potentially responsible for the event (e.g., the image of a bloody knife); some others were simply threatening but in no way attributable to the action of an individual (e.g., a tornado). Experiment 1a was designed to test the hypothesis that the evaluation of facial trustworthiness is influenced by whether the threatening visual context, in which the face is embedded, is attributable to the human action. To do so, we asked participants to rate the trustworthiness of faces that were surrounded by either human threatening, nonhuman threatening, or neutral contexts. We predicted that faces would be judged more untrustworthy when accompanied by human threatening contexts because, in that condition, participants would attribute to the facial identity the responsibility of the threatening context.

Sample Size Determination

Because no prior studies had ever investigated the effect of human versus nonhuman threatening contexts on facial trustworthiness, we adopted a conservative approach and estimated the sample size for Experiment 1a through an a priori power analysis conducted using the R function pwr.f2.test(). We opted for a medium effect size $d = .50$. At $\alpha = .05$, with a power = .95, the analysis suggested an overall sample of 63 participants. We slightly oversampled ($N = 74$) for safety reasons and then conducted a safeguard sensitivity analysis (Perugini et al., 2014) to prove that the experiment had sufficient power to detect the critical effect. We calculated the 95% confidence interval (CI) of the critical effect. Then, we used the "simr" R package to estimate the power of the study, using the 95% CI lower bounds in place of the observed effect (Green & MacLeod, 2016). This allowed us to estimate the power to detect an effect size that was 95% of the times lower than the actual effect. The

estimated (unstandardized) effect in Experiment 1a was .52 (lower bound was .44). After replacing the observed effects with the relevant lower bound, testing 74 participants yielded more than 99% power to detect the effect (see supplemental materials for sensitivity analyses on all the effects included in this study).

Method

Participants and procedure. Seventy-four Italian participants (39 females, $M_{age} = 21.92$, $SD_{age} = 2.04$) volunteered to participate in the study. The study was programmed in Qualtrics and completed online. Participants were asked to participate in a study on face perception. Instructions informed them that they would be presented with images of individuals surrounded by different contexts and were asked to rate each person on perceived trustworthiness, using a 7-point Likert-type scale (1 = *untrustworthy*, 7 = *trustworthy*). The experiment consisted of two blocks of 72 trials each, with stimuli administered in random order. No time limit was set although participants were kindly reminded to provide their judgments as fast as possible.

Stimuli. We employed 24 computer-generated identities (12 trustworthy, 12 untrustworthy) borrowed from a set of photos previously validated for facial trustworthiness (Todorov e al., 2013) and slightly modified (i.e., we added hairlines and embedded the faces in the visual context in a naturalistic way) to increase their ecological validity (Brambilla et al., 2018).

Context stimuli (four human threatening, four nonhuman threatening) were obtained from public domain websites (*human threatening*: rifles, gun with bullets, bloody knife, empty room with blood on the walls; *nonhuman threatening*: rough sea, tornado, rainstorm, windstorm; Figure 1 presents the context stimuli as they appeared in the pretest, and Figure 2 offers an example of test trials presented in Experiment 1a). A gray rectangle was used as control context. A pretest (N = 43, 30 females, $M_{age} = 32.88$, $SD_{age} = 12.27$) was conducted to control for threat and valence of both human and nonhuman threatening contexts. Independent raters evaluated all context stimuli on perceived threat (1 = *not at all threatening*, 7 = *extremely threatening*), valence (1 = *extremely negative*, 7 = *extremely positive*), and potential attribution to the human action (1 = *not at all*, 7 = *extremely*). As facial stimuli were meant to partly cover the background contexts, all context stimuli were presented with a gray silhouette of an unrecognizable individual in the center of the context image. This was done to make sure that context stimuli were evaluated based on the same portion of image that would be visible to participants in the actual experiments. Human contexts were perceived as less threatening ($M = 4.69$, $SD = 1.37$) than nonhuman contexts ($M = 5.28$, $SD = 1.22$), $t(42) = -2.34$, $p = .024$, $d = .36$, 95% CI = [.05, .66]. On valence, human contexts ($M = 2.44$, $SD = .83$) and nonhuman context ($M = 2.45$, $SD = .81$) were perfectly matched, $t(42) = -.06$, $p = .95$. Finally, human contexts were evaluated as more attributable to the action of humans ($M = 6.35$, $SD = .67$) than nonhuman contexts ($M = 2.31$, $SD = 1.08$), $t(42) = 19.41$, $p < .001$, $d = 2.96$, 95% CI = [2.26, 3.95].

Analytic Plan

Data were analyzed in a two-level generalized mixed model. The nature of the context (human vs. nonhuman vs. control), facial trustworthiness (trustworthy vs. untrustworthy), and their interaction were included as fixed factors. The individual intercept was included as random factor, as well as the intercept for the facial identities used across trials. The significant effect of the type of context on the attribution of trustworthiness was further inspected by conducting pairwise contrasts between the three conditions.

Results

Means and standard deviations are reported in Table 1 (see Figure 3 for bar graphs). A main effect of facial trustworthiness showed that participants attributed less trustworthiness to untrustworthy faces than to trustworthy faces, $b = .67$, $SE = 0.05$, $t(22) = 13.67$, $p < .001$.[1] We found a main effect of the type of context, $F(2, 10555) = 310.37$, $p < .001$. Direct contrasts showed that human threatening contexts influenced the judgment of facial trustworthiness more negatively than neutral context, $b = -.70$, $SE = 0.03$, $t(10555) = -24.00$, $p < .001$. The difference in trustworthiness attribution was significant, but weaker, when comparing nonhuman threatening contexts and neutral contexts, $b = -.18$, $SE = 0.03$, $t(10555) = -6.22$, $p < .001$. Importantly, the differential impact of human and nonhuman threatening context was significant, $\beta = -.52$, $SE = 0.03$, $t(10555) = -17.78$, $p < .001$.[2] Remarkably, the superior effect of human threatening contexts on attribution of (un) trustworthiness to embedded faces emerged although such contexts were normatively rated as less threatening than the nonhuman ones. Thus, the reported effects were not driven by context extremity. We also found a significant interaction between the nature of the context and the original trustworthiness of the facial stimuli, $F(2, 10555) = 4.62$, $p = .010$. Decomposing this interaction, we found that the difference in the impact of human threatening contexts and neutral context showed stronger for trustworthy (vs. untrustworthy) faces, $b = -.17$, $SE = 0.06$, $t(10555) = -2.97$, $p = .009$. The difference in the impact of nonhuman threatening contexts and neutral context showed a similar, although nonsignificant, trend, $b = -.12$, $SE = 0.06$, $t(10555) = -2.06$, $p = .079$. No difference was found when looking at human versus nonhuman threatening context, $b = -.05$, $SE = 0.06$,

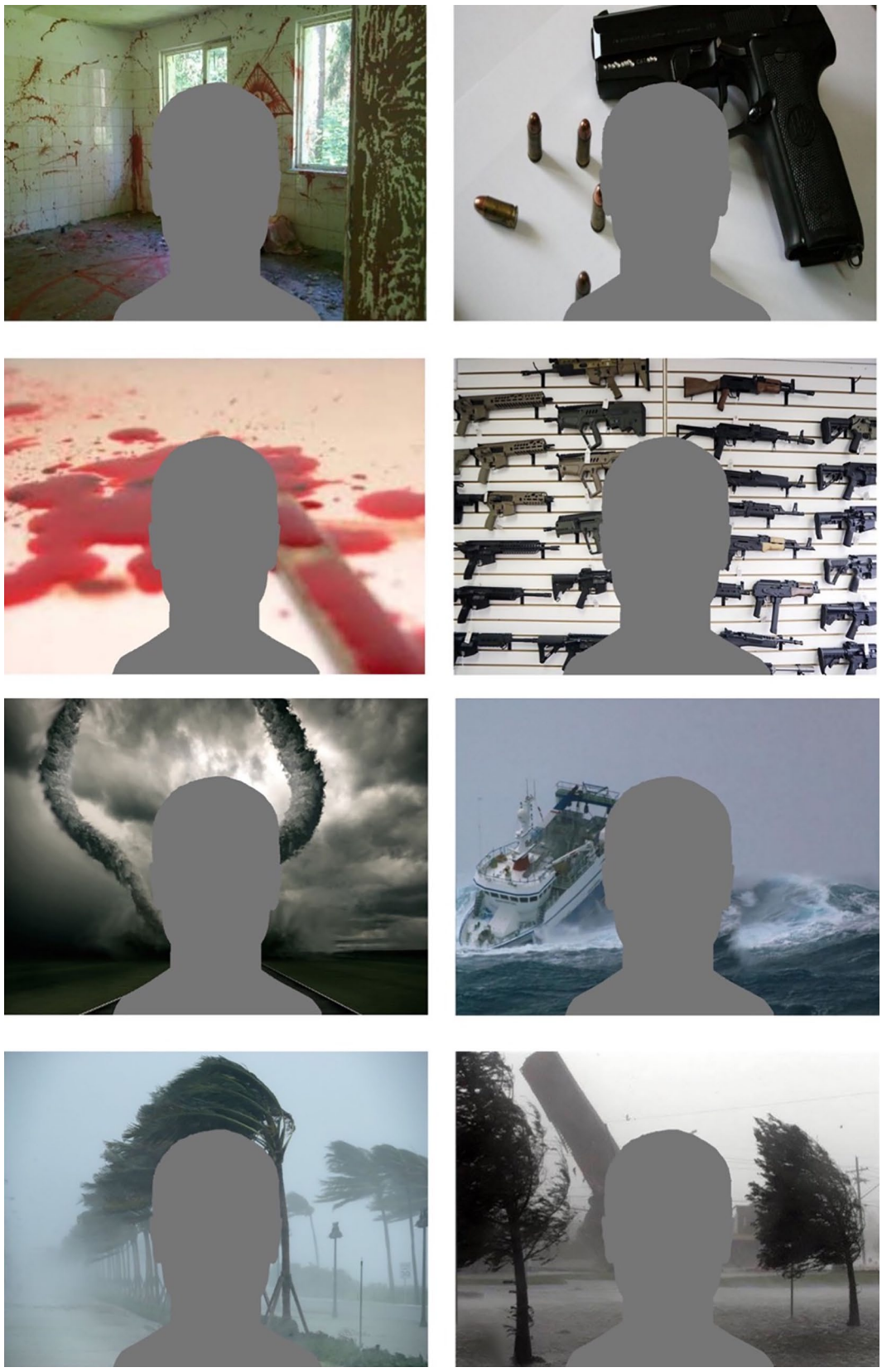

Figure 1. Human (upper panel) and nonhuman (lower panel) threatening contexts as they appeared in the pretest.

Figure 2. Example of trials presented in Experiment 1a.

Table 1. Means and Standard Deviation, Experiment 1a.

	Face			
	Untrustworthy		Trustworthy	
Context	*M*	*SD*	*M*	*SD*
Human threatening	2.93	1.40	4.20	1.56
Nonhuman threatening	3.43	1.43	4.74	1.36
Control	3.55	1.48	4.99	1.35

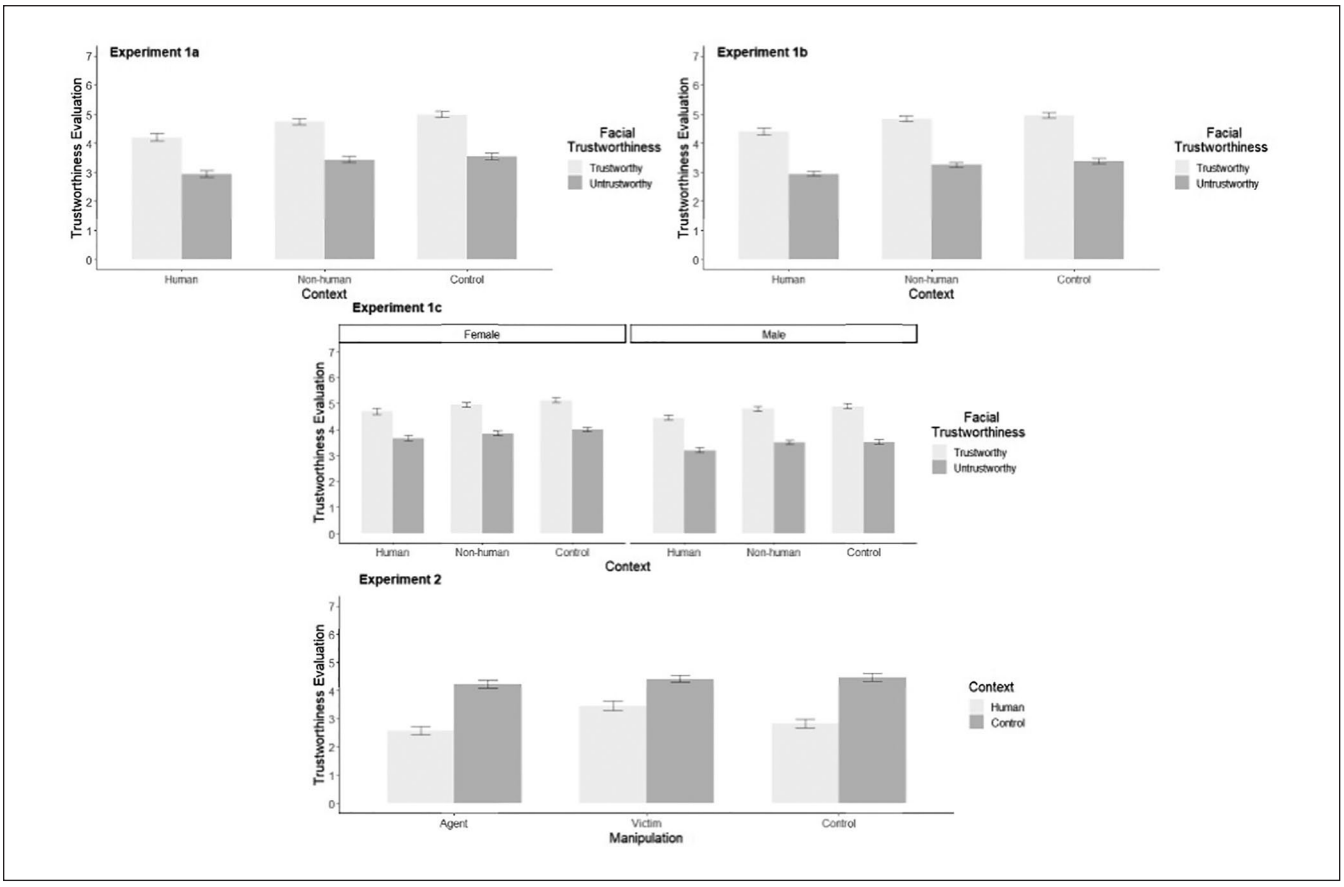

Figure 3. Bar graphs of the results for all the experiments.

Table 2. Means and Standard Deviation, Experiment 1b.

	Face			
	Untrustworthy		Trustworthy	
Context	*M*	*SD*	*M*	*SD*
Human threatening	2.94	1.39	4.40	1.62
Nonhuman threatening	3.26	1.47	4.83	1.47
Control	3.37	1.49	4.95	1.41

$t(10555) = -0.91$, $p = .365$. Thus, the superior role of human versus nonhuman threatening contexts in affecting trustworthiness judgments generalized on both trustworthy, $b = -.55$, $SE = 0.04$, $t(10555) = -13.21$, $p < .001$, and untrustworthy faces, $b = -.49$, $SE = 0.04$, $t(10555) = -11.93$, $p < .001$.[3]

Experiment 1b

Experiment 1a confirmed our hypothesis: Human threatening contexts had a higher power in influencing the attribution of trustworthiness than the nonhuman threatening contexts. We set out a second experiment that aimed at replicating the effects found in Experiment 1a, testing their generalization on real facial stimuli. The method and the procedure were identical to those of Experiment 1a.

Sample Size Determination

As for Experiment 1a, the effect of interest in Experiment 1b was the difference in the impact of human versus nonhuman threatening contexts on attribution of trustworthiness. We relied on Experiment 1a's sensitivity analysis and opted for a similar, but slightly larger, sample size in Experiment 1b. The estimated (non-standardized) critical effect was .38. As preregistered, we conducted a safeguard analysis on this critical effect. After replacing the estimated effect (.38) with the lower bound (.32), testing 82 participants yielded more than 99% power to detect the effect.

Method

Participants and procedure. Eighty-two participants (53 females, $M_{age} = 23.67$, $SD_{age} = 3.42$) volunteered to participate in the study. The procedure and the design of the experiment were identical to those of Experiment 1a, except for the type of face stimuli.

Stimuli. We employed 24 male Caucasian identities borrowed from the Chicago Face Database (Ma et al., 2015). The normative evaluation of the 12 untrustworthy faces is 2.65 ($SD = .17$), whereas the normative evaluation of the trustworthy faces is 3.74 ($SD = .11$).

Results

We adopted the same analytic plan used in Experiment 1a. We replicated the main effect of facial trustworthiness, $b = .77$, $SE = 0.09$, $t(22) = 8.13$, $p < .001$ (see Table 2 for descriptives and Figure 3 for bar graphs). In addition, a main effect of the type of context emerged, $F(2, 11699) = 158.94$, $p < .001$. Direct contrasts showed that human threatening contexts influenced the judgment of facial trustworthiness more negatively than neutral contexts, $b = -.49$, $SE = 0.03$, $t(11699) = -17.04$, $p < .001$. The difference between nonhuman threatening and neutral context was significant, $b = -.11$, $SE = 0.03$, $t(11699) = -3.96$, $p < .001$. The direct comparison between human and nonhuman threatening confirmed that the difference in their impact over facial trustworthiness was significant, $b = -.38$, $SE = 0.03$, $t(11699) = -13.07$, $p < .001$. The interaction between the nature of the context and the trustworthiness of the facial stimuli was not significant, $F(2, 11699) = 2.84$, $p = .058$. In line with Experiment 1a, human threatening contexts had stronger impact than nonhuman threatening contexts in affecting trustworthiness judgments on both trustworthy, $b = -.44$, $SE = 0.04$, $t(11699) = -10.76$, $p < .001$, and untrustworthy faces, $b = -.31$, $SE = 0.04$, $t(11699) = -7.73$, $p < .001$.

Experiment 1c

Experiments 1a and 1b showed a superior effect of human threatening contexts for non-gendered (computer-generated) faces and for real male faces. In Experiment 1c, we introduced stimulus gender (male vs. female) as an additional factor to be crossed with facial trustworthiness (trustworthy vs. untrustworthy) and the type of context (neutral vs. threatening human vs. threatening nonhuman). We explored whether the superior effect of the human threatening contexts generalized on female stimuli or, alternatively, whether gender modulated such an effect. In fact, men tend to be stereotyped as more aggressive than women (see Ellemers, 2018 for a review). Thus, if human threatening contexts have the greatest impact on trustworthiness because they suggest that the facial stimulus has an active role in the context, then such an inference might be more likely to occur with male (vs. female) facial identities.

Table 3. Means and Standard Deviation, Experiment 1c.

	Male face				Female face			
	Untrustworthy		Trustworthy		Untrustworthy		Trustworthy	
Context	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Human threatening	3.19	1.52	4.45	1.59	3.66	1.63	4.68	1.59
Nonhuman threatening	3.50	1.52	4.80	1.53	3.85	1.58	4.94	1.51
Control	3.51	1.57	4.88	1.48	4.00	1.58	5.11	1.43

Sample Size Determination

Experiment 1c was powered on the difference in the impact of human versus nonhuman threatening contexts on the attribution of trustworthiness. Because we could not estimate the size of the interaction effect between context and gender, we decided to increase the sample size and stop the data collection when 100 participants completed the study. The estimated (non-standardized) critical effect size (i.e., difference between human and nonhuman threatening contexts on the attribution of trustworthiness) observed in this study was .28. As preregistered, we conducted a sensitivity analysis on the critical effect. After replacing the observed effects with the relevant lower bound (.22), testing 101 participants yielded more than 99% power to detect the effect.

Method

We adopted a 3 (*Context*: neutral vs. human threatening, nonhuman threatening) × 2 (*Face*: trustworthy vs. untrustworthy) × 2 (*Stimuli Gender*: male vs. female) full-within design.

Participants and procedure. We recruited 101 participants (65 female, $M_{age} = 27.99$, $SD_{age} = 11.28$). Participants were asked to participate in a study on face perception. The procedure largely mirrored that of Experiments 1a and 1b, except for face stimuli gender.[4]

Stimuli. We employed 24 Caucasian identities borrowed from the Chicago Face Database (Ma et al., 2015). The normative evaluation of the six untrustworthy male faces was 2.78 ($SD = .06$), whereas for the six male trustworthy faces it was 3.80 ($SD = .07$). The normative evaluation of the six untrustworthy female faces was 2.75 ($SD = .19$), whereas for the six female trustworthy faces it was 3.82 ($SD = .06$). Context stimuli mirrored those employed in Experiments 1a and 1b.

Analytic Plan

Data were analyzed using a generalized mixed model. The nature of the context (human vs. nonhuman vs. control), facial trustworthiness (trustworthy vs. untrustworthy), gender (male vs. female), and their interaction terms were included in the model as fixed factors. The individual intercept was included as random factor, as well as the intercept for the facial identities used across trials.

Results

A main effect of facial trustworthiness showed that participants attributed less trustworthiness to untrustworthy faces as compared with trustworthy faces, $b = .60$, $SE = 0.08$, $t(20) = 7.82$, $p < .001$ (see Table 3 for descriptives and Figure 3 for bar graphs). We found a main effect of the type of context, $F(2, 14412) = 100.66$, $p < .001$. Human threatening contexts influenced the attribution of facial trustworthiness more negatively than neutral contexts, $b = -.38$, $SE = 0.03$, $t(14412) = -13.73$, $p < .001$. The differential impact of nonhuman threatening and neutral contexts was significant, $b = -.10$, $SE = 0.03$, $t(14412) = -3.77$, $p < .001$. The direct contrast between human and nonhuman threatening contexts proved that the impact of the former was stronger as compared with the latter, $b = -.28$, $SE = 0.03$, $t(14412) = -9.96$, $p < .001$. We found a significant effect of gender, $b = .16$, $SE = 0.07$, $t(20) = 7.82$, $p = .049$, with male stimuli evaluated as less trustworthy than female stimuli. Neither the interaction between face and context, $F(2, 14412) = 1.96$, $p = .141$, nor that between face and gender, $b = -.06$, $SE = 0.06$, $t(20) = -.74$, $p = .463$, were significant. Neither significant was the interaction between context and gender, $F(2, 14412) = 2.50$, $p = .082$. Indeed, the difference between human and nonhuman threatening contexts generalized on both male and female stimuli, that is, $b = -.33$, $SE = 0.04$, $t(14412) = -8.39$, $p < .001$ and $b = -.22$, $SE = 0.04$, $t(14412) = -5.70$, $p < .001$, respectively. Yet we acknowledge that the study was underpowered to detect this interactive pattern (see supplemental materials for a safeguard sensitivity analysis on this effect).

Aggregate Analysis

Data from the three studies were combined to provide evidence of the critical effect under investigation. The combined data set included 257 participants. Data were analyzed in a generalized mixed model identical to that used in Experiments 1a and 1b, except that the facial identity random intercept was

removed because the model did not converge. There was a main effect of facial trustworthiness, $b = .68$, $SE = .007$, $t(36746) = 95.91$, $p < .001$. We also found a main effect of the type of context, $F(2, 36746) = 469.99$, $p < .001$. Human threatening contexts influenced the judgment of facial trustworthiness more negatively than neutral context, $b = -.50$, $SE = .02$, $t(36746) = -29.51$, $p < .001$. The difference in trustworthiness attribution was also significant when comparing nonhuman threatening and neutral contexts, $b = -.13$, $SE = .02$, $t(36746) = -7.53$, $p < .001$. Crucial for our research question, the differential impact of human and nonhuman threatening context was significant, $b = -.37$, $SE = .02$, $t(36746) = -21.97$, $p < .001$.

We also found a significant interaction between the nature of the context and the original trustworthiness of the facial stimuli, $F(2, 36746) = 7.64$, $p < .001$. The difference in the impact of human threatening and neutral context showed stronger for trustworthy (vs. untrustworthy) faces, $b = -.13$, $SE = .03$, $t(36746) = -3.89$, $p < .001$. The difference between nonhuman threatening and neutral context was in the same direction but was not significant, $b = -.05$, $SE = .03$, $t(36746) = -1.62$, $p = .10$. Finally, the difference between human and nonhuman threatening contexts was significant and stronger for trustworthy faces, $b = -.07$, $SE = .03$, $t(36746) = -2.27$, $p = .02$. Possibly, these findings might reflect a difference in the baseline attribution of trustworthiness, inferred from evaluations given in neutral contexts, for trustworthy ($M = 4.98$, $SD = 1.41$) and untrustworthy faces ($M = 3.57$, $SD = 1.54$). In other words, the negative evaluation of untrustworthy individuals might have been closer to floor, leaving less room for change. However, because our studies were not powered on the interaction effect (see supplemental materials for a safeguard sensitivity analysis), any conclusive reasoning around this interaction would be rather premature at this stage.

Discussion

Experiments 1a to 1c showed that the nature of the threatening context impacts significantly on face–context integration: When the threatening scene was ascribable to the facial stimulus (i.e., potentially attributable to humans), then the impact of the context showed stronger. The effect replicated successfully across three experiments in which we varied both the type and the gender of the targets. Aggregated results also indicated that this effect of human threatening contexts was stronger for trustworthy faces. However, because this effect was very small and the overall study was underpowered to detect a real effect, we refrain from drawing any conclusion at this stage.

If human (vs. nonhuman) threatening contexts lead to higher untrustworthiness, then the relationship established between the face and the context is not only more meaningful, but it must be also qualified in a specific way. Specifically, we assumed that facial stimuli received less trustworthiness when presented in human threatening contexts because perceivers tended to look at them as the persons who were responsible for that action. If so, we should observe variation in the effect of face–context integration by altering the nature of the relationship between stimuli.

Experiment 2

Experiment 2 was set out to test the hypothesis that face–context integration can be conditional upon the relational qualifier that one uses to relate the face and the context. We proposed that (human) threatening contexts lead to lower attribution of facial trustworthiness because people tend to consider the individuals portrayed in those contexts as perpetrators. Throughout past experiences, perceivers learned to establish a default relation that qualifies the link between a face and a context. Here, we use verbal instructions to manipulate the relational qualifier between the face and the context and test how such qualifiers can impact the attribution of facial trustworthiness. Whereas this type of manipulation is quite typical in conditioning research (e.g., Fiedler & Unkelbach, 2011) no prior study has employed instructions to alter the nature of the relationship between contexts and faces and tested its impact on face–context integration.

Sample Size Determination

As preregistered, we adopted a sequential analysis approach (Lakens, 2014). The effect of interest was the difference in the attribution of trustworthiness in the threatening versus neutral context across conditions. We based our power analysis on two critical pairwise comparisons. First, the context effect (neutral—threatening) in the victim versus control condition. Second, the context effect (neutral—threatening) in the victim versus perpetrator condition. We aimed at a sample size that could allow us to detect two effects as small as Cohen's $f^2 = 0.039$. We used the "pwr" package in R to estimate the number of participants required in each cell, at $\alpha = .05$, and power $1\text{-}\beta = .80$ (to obtain an actual power = .80, the nominal power of each contrast was adjusted to .89). The analysis suggested 87 participants per group, leading to 261 participants. We planned and preregistered a single interim analysis with half of the suggested sample size ($N = 131$). As indicated in our preregistered protocol, we used the Pocock boundary to set the alpha level, which allowed us to stop the data collection if our interim analyses on 131 participants showed that the two critical effects were both significant at $p < .0294$. Because for both the critical contrasts the level of significance was $p < .0294$, we did not collect the remaining 130 participants.

Method

We adopted a 3 (*Face–Context Qualifier*: perpetrator vs. victim vs. control) × 2 (*Context*: neutral vs. human threatening) mixed design, with the first factor manipulated between participants. Besides the inclusion of the instruction

manipulation, there were two main variations from the design adopted in the previous studies. First, we used facial stimuli that were neutral in terms of trustworthiness. Second, we focused on the distinction between human threatening and neutral contexts and did not consider nonhuman threatening contexts. The attribution of trustworthiness to each facial identity was the dependent variable.[5]

Participants and procedure. One hundred thirty-one participants (59 females, $M_{age} = 27.99$, $SD_{age} = 11.28$) took part in the study through Prolific Academic. Participants in the three conditions underwent the same judgment phase. They were told that a series of faces would appear on screen, with each face embedded in a visual context. In each trial, participants evaluated the extent to which each face was trustworthy, using a 7-point scale (1 = *untrustworthy*, 7 = *trustworthy*). Before entering the task, participants were told that each image could include the face of an individual presented in either a crime scene or a gray background. Depending on the condition, participants were instructed that the individuals portrayed in the crime scene were either the perpetrators or the victims. Participants assigned to the control condition did not receive any instructions about the relation between the facial and context stimulus and proceeded directly to the judgment phase. Thus, for these participants, the procedure was largely comparable to that adopted in previous studies. The judgment phase consisted of two blocks of 48 trials each, with stimuli administered in random order. Participants were reminded to provide their judgments as fast as possible. In case of slow response (RTs >2,000 ms), the message "too slow" appeared on screen.

Stimuli. We employed 24 computer-generated identities from Todorov et al. (2013). Facial stimuli were selected to be neutral on trustworthiness. Similar to study 1a, facial stimuli were slightly modified to increase their ecological validity (Brambilla et al., 2021a).

Context stimuli were the same human threatening contexts used in previous studies, and the gray background serves as neutral context.

Materials. Except for the instructions administered to participants assigned to the victim and the perpetrator condition, the procedure mirrored that of Experiments 1a, 1b, and 2. Because only two types of contexts were used in this experiment, we reduced the total number of trials to 96 (divided into two separate blocks of 48 trials each). The 24 neutral faces were divided into two separate sets, to be assigned separately to either the neutral or the threatening contexts.

Instructions. The following instructions were presented to participants assigned to either the perpetrator or the victim condition:

> In this study, you will be presented with a series of faces. Each face will appear in a visual context. Some of those contexts will be neutral (grey background). Some others will portray dangerous scenes (e.g., bloody knife, arms, crime scene). Be careful: the face shown in each dangerous scene belongs to the perpetrator [victim] of the action portrayed in the background scene. For instance, the face that you will see on the next screen belongs to the perpetrator [victim] of the crime portrayed in the visual background.

Analytic Plan

Data were analyzed using generalized mixed models. The same two intercepts (i.e., individual and face identity) were included as random factors. In case of (expected) significant interaction (Context × Target–context qualifier), we conducted pairwise comparisons. We focused on the context effect (neutral—threatening) on facial trustworthiness by comparing (a) victim versus control, and (b) victim versus perpetrator. We also tested a third comparison, that is, perpetrator versus control. Then, we looked at the same comparison by considering attributions in the threatening and in the neutral contexts separately. This additional analysis was planned to test that the hypothesized impact of the manipulation at the level of the face–context relationship was due to a difference in response to facial stimuli presented in threatening, rather than in neutral, contexts. Paired sample *t* tests were conducted in each between-subject condition.

Results

Means and standard deviations are reported in Table 4 (see Figure 3 for bar graphs). We found a main effect of the type of context, $b = -1.64$, $SE = 0.06$, $t(45) = -27.17$, $p < .001$, with lower attribution of trustworthiness for threatening as compared with neutral contexts. The effect of face–context qualifier was also significant, $F(2, 128) = 5.41$, $p = .006$, as well as its interaction with the type of context, $F(2, 12420) = 106.05$, $p < .001$. The first critical contrast revealed that the effect of context (i.e., neutral—threatening) differed between the victim condition ($b = .97$, $SE = 0.06$, $z = 16.51$, $p < .001$) and the control ($b = 1.63$, $SE = 0.06$, $z = 27.43$, $p < .001$), $b = .66$, $SE = 0.05$, $z = 12.49$, $p < .001$. The second contrast further showed that the same effect of context (i.e., neutral—threatening) was stronger in the perpetrator ($b = 1.65$, $SE = 0.06$, $z = 27.17$, $p < .001$) than in the victim condition, $b = .68$, $SE = 0.05$, $z = 12.50$, $p < .001$. In essence, when faces were presented as the victims of the threatening scenes, participants were less inclined to evaluate the faces as untrustworthy. Instead, no difference emerged when comparing the perpetrator and the control condition, $b = .02$, $SE = 0.06$ $z = .32$, $p = .752$, suggesting that perceivers might have the default tendency to assume that the face belongs to the perpetrator of the threatening scene.

Importantly, the effect of context–face qualifier affected the attribution of trustworthiness appearing in threatening, but not neutral, contexts. Specifically, no significant effect of

Table 4. Means and Standard Deviation, Experiment 2.

	Context			
	Human threatening		Neutral	
Face–context relation	*M*	*SD*	*M*	*SD*
Victim	3.45	1.60	4.41	1.36
Control	2.82	1.52	4.45	1.44
Perpetrator	2.57	1.34	4.22	1.40

context–face qualifier emerged for faces in neutral contexts (all $ps > .52$). When looking at the attribution of trustworthiness to faces presented in threatening contexts, we found a significant difference between the victim condition and the control, $b = -.63$, $SE = 0.16$, $z = -3.87$, $p < .001$, and between the victim and the perpetrator condition, $b = -.88$, $SE = 0.17$, $z = -5.26$, $p < .001$. Instead, no difference emerged for faces presented in threatening contexts when comparing the perpetrator and the control condition, $b = -.25$, $SE = 0.17$, $z = -1.46$, $p = .309$.[6]

Discussion

Experiment 2 showed that face–context integration depends on the nature of the relationship that perceivers impose upon face and context stimuli. The null difference in face–context integration observed between the control condition (i.e., no-instructions) and the perpetrator condition corroborates the idea of a default relationship on facial stimuli presented in threatening contexts. Importantly, the effect of the context on the attribution of facial trustworthiness was significantly reduced when participants were induced to believe that the face belonged to the victim of the threatening scene. This emerged when the victim was compared with both the perpetrator and the control condition.

General Discussion

Processing faces in threatening contexts alters the level of trustworthiness that is attributed to the target facial stimulus. This face–context integration is so powerful that it can emerge in a few milliseconds (Brambilla et al., 2018). However, even if it can emerge so quickly, it does not imply that it is not moderated by specific higher-order reasoning imposed by the interplay between the stimuli and the perceiver (see Freeman et al., 2020). We questioned whether the response produced by the perceiver (i.e., attribution of trustworthiness) could be affected by the nature of the relationship between the face and the surrounding context. In a first set of studies (Experiments 1a to 1c), participants were presented with facial stimuli embedded in either neutral or threatening contexts. Threatening contexts varied with respect to their relationship with the human face. In spite of the weaker threat carried by the set of context stimuli that were ascribable to humans, those stimuli made computer-generated and real faces less trustworthy than (more) threatening contexts not ascribable to humans. The effect emerged on (male and female) trustworthy and untrustworthy faces. An aggregated analysis indicated that the differential impact of human versus nonhuman context was stronger on trustworthy faces, possibly due to a more positive baseline evaluation leaving higher room for observing a variation in the attribution of trustworthiness. However, future studies should explore better this interactive effect.

Whereas in this set of studies we just assumed a specific encoding schema to be key for qualifying face–context relationship, Experiment 2 proved the nature of such a relationship. Using verbal instructions, we induced participants to believe that facial stimuli belonged to either the perpetrators or the victims of the surrounding context. Such a manipulation significantly moderated face–context integration: When faces were presented as the victims of the context scenes, then the threat conveyed by the context influenced facial trustworthiness to a lesser extent than when the very same faces were the perpetrators of those contexts. Importantly, no difference was found in the effect of the context on facial trustworthiness when comparing the perpetrator condition and the no-instructions condition. Taken together, these findings suggest that the bond between threat and trust in face–context integration is not ubiquitous for any kind of threatening stimulus, and that faces in context are not just passively processed by the perceiver. Rather, the ultimate attribution of facial trustworthiness based on contextual threat is conditional upon specific conditions that determines (a) whether a meaningful relationship can be established, and (b) the nature of such a relationship.

To the best of our knowledge, this is the first empirical contribution addressing the boundaries of face–context integration. In so doing, our work combines recent findings from face–context integration with relevant insights from conditioning research. In fact, both face–context integration and conditioning are the effect of one specific type of regularity in the environment, that is, stimulus pairing. From conditioning research, we borrowed the assumption that humans can actively encode paired stimuli using specific relational operators (e.g., Fiedler & Unkelbach, 2011). When prior knowledge is applied to stimulus information in the form of self-generated propositions, conditioning is

improved. Research over the past decade has showed that conditioning effects can vary depending on how the relation between paired stimuli is generated (Fiedler & Unkelbach, 2011; Hughes et al., 2019; Moran & Bar-Anan, 2013). Here, we found that the same reasoning holds for face–context integration. Thus, this contribution shows that the effect played by contextual threat over facial trustworthiness can be better understood if one considers higher-order encoding processes that the individual imposes on face–context pairs. Under this view, face–context integration is not just about a general inherent link between threat and trust. Rather, such a link can vary in meaning, strength, or even direction.

We acknowledge that the critical conditions introduced in each experiment resulted in a mere attenuation of the effect. In Experiments 1a to 1c, the level of trustworthiness attributed to faces in nonhuman threatening context was still lower compared to the baseline. This effect is not surprising if one considers that nonhuman threatening contexts may bring a generic sense of threat that might suffice to negatively affect the attribution of trustworthiness to facial stimuli presented in such contexts. In Experiment 2, we still found a significant (and negative) effect of the threatening context in the victim condition. This effect might reflect the contrast between countervailing propositions generated from both instructions and stimuli-pairing on the relationship between faces and contexts. On one hand, instructions informed that the faces belonged to the victim of the context. On the other hand, humans have a long history of learning that tells them that stimuli that occur together often happen to be similar. As much as instructions can be powerful in affecting how stimuli that occur together are eventually processed, counteracting previously generated propositions can be tough (Zanon et al., 2014). Possibly, one way to maximize their effect might consist in presenting participants with relational qualifiers that cue how face and context are meant to be related just before each face–context trial appears on screen.

Taken together, our findings are in line with a functional interpretation of face–context integration. Based on a functional approach to social perception, the primary purpose of perceiving is to guide people in avoiding social threats (Fiske, 1992; Zebrowitz & Collins, 1997). This hypothesis finds empirical support in recent findings from face perception research, indicating that our perception of others' faces is essentially explained by the binary decision to either approach or avoid the target individual (Jones & Kramer, 2021). Similarly, threatening environments attributable to human action should offer additional reasons to avoid an individual. Specifically, humans should be more inclined to perceive individuals as untrustworthy when the context suggests that they might act malevolently because of an overprotective strategy (Hammond, 2007). Indeed, it is less costly to erroneously avoid an innocent individual rather than approaching a malevolent one. In Experiments 1a to 1c, the attributions made by the perceiver depended on whether the contextual threat was ascribable to the human action. Experiment 2 extended these findings and showed that the effect of the context was largely reduced when the target was presented as the victim of the threatening context. In line with the idea that perception is functional to action, our findings showed that face–context integration on the attribution of trustworthiness (a) is maximized when the context sets higher reasons to believe that the target is dangerous, (b) is based on a default face–context relationship (i.e., perpetrator) that perceivers impose upon the processed stimuli, and (c) is significantly reduced when such a relationship no longer justifies a sense of threat.

One important limitation relates to the generalizability of our findings based on the contextual stimuli employed. For instance, natural catastrophes (i.e., nonhuman threats) differ from scenes related to crimes (i.e., human threats) in the extent to which they activate an either personal or more general sense of threat: Whereas participants might experience the feeling of being the target of the threat conveyed by human stimuli, nonhuman stimuli are more likely to activate a threat that can damage humanity in general. In addition, compared with nonhuman stimuli, human stimuli led to a higher sense of physical threat. Future studies should replicate our findings, perhaps replacing natural catastrophes with nonhuman stimuli that can be perceived as threatening for one's personal safety (e.g., a wild dog) or using human stimuli that do not convey physical threat (e.g., a boy stealing a soda in a shop).

One question that is left unanswered here is concerned with the real-life implications of these findings. Whereas the actual power of faces to affect real-world decisions has already been established (e.g., Duarte et al., 2012; Jaeger et al., 2019; Todorov et al., 2005), less is known about the role of the context. In showing how face–context integration depends on the nature of the relationship that can be established between a face and a context, our findings confine to a controlled experimental setting. In everyday perception, faces and contexts are often processed in concert with other relevant cues. This has two main implications. First, it is easy to imagine that face–context integration could reduce in a real-life setting, where less control over other interfering variables is permitted. For instance, participants in our experiments might have placed extreme attention on the context scenes because of the way in which we presented our stimuli. Second, other cues can also impact face–context integration by altering the nature of their relationship. Take, for instance, a person's body movement. Individuals are rarely processed as static entities in context. On many occasions, we see people running away or approaching a source of threat (e.g., a riot outside a stadium). All these cues have the potential to affect the inferences that we make about individuals processed in threatening contexts. Thus, research would benefit from further studies testing face–context integration in a more naturalistic setting (perhaps through virtual reality) to provide stronger conclusions about the actual role of context in determining real-life decisions upon facial

stimuli. With that being said, we believe that our findings from Experiment 2 might have important implications with respect to how face–context integration can vary in distinct situations, that is, we showed that the contextual effect (neutral—human threatening) can be moderated by other cues. We operationalized such cues through verbal instructions informing the perceiver about the nature of the relationship between the target face and the context. However, any environmental cue might have an impact on face–context integration by informing perceivers about the nature of face–context relationship. In other words, just as instructions inform that a person is the victim of a threatening context in a direct fashion, perceiving a sense of fear (vs. anger) from either facial emotion or body movement could provide the same information indirectly.

In summary, this work offers a first empirical demonstration that attributing trustworthiness to faces presented in threatening contexts is not an inevitable effect. We proposed that the perceiver plays an active role in the processing of faces presented in threatening context, that is, the ultimate attribution of trustworthiness depends on whether meaningful relationships can be formed between the face and contextual threat as well as on the type of relational qualifier that is used to encode their link. This work opens to a new conceptualization of face–context integration that speaks for the importance of considering higher-order processes when it comes to explaining how contextual threat can bias our response to facial stimuli.

Author Note

Matteo Masi, is now affiliated to University of Surrey, Guilford, UK.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iD

Simone Mattavelli https://orcid.org/0000-0002-8934-8016

Supplemental Material

Supplemental material is available online with this article.

Notes

1. Due to the way that variance is partitioned in linear mixed models, there does not exist an agreed-upon way to calculate standardized effect sizes for individual model terms. We therefore decided not to include any standardized effect size in these analyses.
2. Note that the direct comparison between human and nonhuman threatening contexts reflects the difference in their comparison with neutral contexts. As the design of the three studies included a neutral context, estimating the difference between the two types of threatening context (i.e., human vs. nonhuman) is mathematically equivalent to estimating the overall distance between (a) the human threatening and the neutral context, and (b) the nonhuman threatening and the neutral context.
3. For all the experiments, we have also conducted the analyses using ordinal regressions, a method that is better suited when inspecting an effect on Likert-type scale data. The analyses are reported in the supplemental materials. Note that this alternative approach led to the same findings presented in the article.
4. At the end of the two blocks, participants were also asked to complete two scales that aim at assessing the extent to which a list of adjectives representative of both agency and communion are attributed to either men or women. Because the inclusion of these scales was mainly exploratory and they had no impact on the results, we did not include them in the article.
5. We conducted another study in which the relationship between faces and contexts was manipulated by an initial acquisition phase (instead of instructions). As such a manipulation showed no effect, we identified two main issues. First, the use of different facial stimuli in the acquisition and in the test phase might have prevented generalization of face–context relationship. Second, participants in the perpetrator condition showed an overall lower attribution of trustworthiness in the test phase, regardless of the nature of the context. We decided not to include this experiment in the article. However, the entire preregistered protocol and the analyses are available at https://osf.io/wp5k9/.
6. For exploratory reasons, we conducted the same analyses on reaction times. These analyses revealed a significant effect of context, $F(1, 12442) = 19.74$, $p < .001$, which indicated that participants responded faster in the control than in the two instructions conditions. Neither the effect of face–context qualifier nor the interaction were significant, that is, $F(2, 128) = .15$, $p = .860$ and $F(2, 12442) = .21$, $p = .812$, respectively.

References

Ames, D. L., Fiske, S. T., & Todorov, A. (2011). Impression formation: A focus on others' intents. In J. Decety & J. T. Cacioppo (Eds.). *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.

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

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., Carraro, L., Castelli, L., & Sacchi, S. (2019). Changing impressions: Moral character dominates impression updating. *Journal of Experimental Social Psychology*, *82*, 64–73. https://doi.org/10.1016/j.jesp.2019.01.003

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. (2021a). 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. (2021b). The primacy of morality in impression development: Theory, research, and future directions. *Advances in Experimental Social Psychology*, *64*, 187–262.

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 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). Oxford University Press.

De Houwer, J. (2009). The propositional approach to associative learning as an alternative for association formation models. *Learning & Behavior*, *37*(1), 1–20. https://doi.org/10.3758/LB.37.1.1

De Houwer, J. (2018). Propositional models of evaluative conditioning. *Social Psychological Bulletin*, *13*(3), 1–21. https://doi.org/10.5964/spb.v13i3.28046

De Houwer, J., & Hughes, S. (2016). Evaluative conditioning as a symbolic phenomenon: On the relation between evaluative conditioning, evaluative conditioning via instructions, and persuasion. *Social Cognition*, *34*(5), 480–494. https://doi.org/10.1521/soco.2016.34.5.480

Duarte, J., Siegel, S., & Young, L. (2012). Trust and credit: The role of appearance in peer-to-peer lending. *Review of Financial Studies*, *25*(8), 2455–2483. https://doi.org/10.1093/rfs/hhs071

Ellemers, N. (2018). Gender stereotypes. *Annual Review of Psychology*, *69*, 275–298. https://doi.org/10.1146/annurev-psych-122216-011719

Fiedler, K., & Unkelbach, C. (2011). Evaluative conditioning depends on higher order encoding processes. *Cognition and Emotion*, *25*(4), 639–656. https://doi.org/10.1080/02699931.2010.513497

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

Freeman, J. B., Stolier, R. M., & Brooks, J. A. (2020). Dynamic interactive theory as a domain-general account of social perception. In *Advances in experimental social psychology* (Vol. 61, pp. 237–287). Academic Press.

Green, P., & MacLeod, C. J. (2016). SIMR: An R package for power analysis of generalized linear mixed models by simulation. *Methods in Ecology and Evolution*, *7*(4), 493–498. https://doi.org/10.1111/2041-210X.12504

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

Hofmann, W., De Houwer, J., Perugini, M., Baeyens, F., & Crombez, G. (2010). Evaluative conditioning in humans: A meta-analysis. *Psychological Bulletin*, *136*(3), 390–421. https://doi.org/10.1037/a0018916

Hughes, S., Ye, Y., Van Dessel, P., & De Houwer, J. (2019). When people co-occur with good or bad events: Graded effects of relational qualifiers on evaluative conditioning. *Personality and Social Psychology Bulletin*, *45*(2), 196–208. https://doi.org/10.1177/0146167218781340

Jaeger, B., Sleegers, W. W. A., Evans, A. M., Stel, M., & van Beest, I. (2019). The effects of facial attractiveness and trustworthiness in online peer-to-peer markets. *Journal of Economic Psychology*, *75*, Article 102125. https://doi.org/10.1016/j.joep.2018.11.004

Jones, A. L., & Kramer, R. S. (2021). Facial first impressions form two clusters representing approach-avoidance. *Cognitive Psychology*, *126*, Article 101387. https://doi.org/10.1016/j.cogpsych.2021.101387

Lakens, D. (2014). Performing high-powered studies efficiently with sequential analyses. *European Journal of Social Psychology*, *44*(7), 701–710. https://doi.org/10.1002/ejsp.2023

Ma, D. S., Correll, J., & Wittenbrink, B. (2015). The Chicago face database: A free stimulus set of faces and norming data. *Behavior Research Methods*, *47*(4), 1122–1135. https://doi.org/10.3758/s13428-014-0532-5

Mattavelli, S., Masi, M., & Brambilla, M. (2021). *Untrusted under threat: On the superior bond between trustworthiness and threat in face-context integration (under review)*.

Mitchell, C. J., De Houwer, J., & Lovibond, P. F. (2009). The propositional nature of human associative learning. *Behavioral and Brain Sciences*, *32*(2), 183–246. https://doi.org/10.1017/S0140525X09000855

Moran, T., & Bar-Anan, Y. (2013). The effect of object–valence relations on automatic evaluation. *Cognition & Emotion*, *27*(4), 743–752. https://doi.org/10.1080/02699931.2012.732040

Oosterhof, N. N., & Todorov, A. (2008). The functional basis of face evaluation. *Proceedings of the National Academy of Sciences*, *105*(32), 11087–11092. https://doi.org/10.1073/pnas.0805664105

Perugini, M., Gallucci, M., & Costantini, G. (2014). Safeguard power as a protection against imprecise power estimates. *Perspectives on Psychological Science*, *9*(3), 319–332. https://doi.org/10.1177/1745691614528519

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), Article 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). Recognition of facial expressions is influenced by emotional scene gist. *Cognitive, Affective, & Behavioral Neuroscience*, *8*(3), 264–272. https://doi.org/10.3758/CABN.8.3.264

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

Sutherland, C. A., Oldmeadow, J. A., Santos, I. M., Towler, J., Burt, D. M., & Young, A. W. (2013). Social inferences from faces: Ambient images generate a three-dimensional model. *Cognition*, *127*(1), 105–118. https://doi.org/10.1016/j.cognition.2012.12.001

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., Mandisodza, A. N., Goren, A., & Hall, C. C. (2005). Inferences of competence from faces predict election outcomes. *Science*, *308*(5728), 1623–1626. https://doi.org/10.1126/science.1110589

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

Vernon, R. J., Sutherland, C. A., Young, A. W., & Hartley, T. (2014). Modeling first impressions from highly variable facial images. *Proceedings of the National Academy of Sciences*, *111*(32), E3353–E3361. https://doi.org/10.1073/pnas.1409860111

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

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*(3), 277–283. https://doi.org/10.1038/nn816

Zanon, R., De Houwer, J., Gast, A., & Smith, C. T. (2014). When does relational information influence evaluative conditioning? *Quarterly Journal of Experimental Psychology*, *67*(11), 2105–2122. https://doi.org/10.1080/17470218.2014.907324

Zebrowitz, L. A., & Collins, M. A. (1997). Accurate social perception at zero acquaintance: The affordances of a Gibsonian approach. *Personality and Social Psychology Review*, *1*(3), 204–223. https://doi.org/10.1207/s15327957pspr0103_2

Zebrowitz, L. A., & Montepare, J. M. (2008). Social psychological face perception: Why appearance matters. *Social and Personality Psychology Compass*, *2*(3), 1497–1517. https://doi.org/10.1111/j.1751-9004.2008.00109