

Face-context integration and trustworthiness evaluation

Marco Brambilla, Simone Mattavelli & Matteo Masi

To cite this article: Marco Brambilla, Simone Mattavelli & Matteo Masi (13 Feb 2024): Face-context integration and trustworthiness evaluation, European Review of Social Psychology, DOI: [10.1080/10463283.2024.2315915](https://doi.org/10.1080/10463283.2024.2315915)

To link to this article: <https://doi.org/10.1080/10463283.2024.2315915>



Published online: 13 Feb 2024.



Submit your article to this journal [↗](#)





View related articles [↗](#)



View Crossmark data [↗](#)



Face–context integration and trustworthiness evaluation

Marco Brambilla ^{a*}, Simone Mattavelli ^{a*} and Matteo Masi ^b

^aDepartment of Psychology, University of Milano-Bicocca, Milano, Italy; ^bDepartment of Psychology, University of Münster, Münster, Germany

ABSTRACT




Judgements of trustworthiness based on facial features have mainly been investigated by presenting faces in isolation. However, real-life situations often involve contextual cues. Here, we review our work showing that judgements of trustworthiness from faces are influenced by contextual threat. Individuals are judged as less trustworthy when their faces are surrounded by threatening, as opposed to neutral or merely negative, contexts. Delving into the mechanisms underlying face–context integration, our work reveals that the bond between trustworthiness and threat goes beyond mere stimuli congruency, suggesting that threatening contexts alter person evaluation by conveying information of adaptive significance. We propose an inferential approach to face–context integration, where faces and contexts are encoded relationally: modifying this relational encoding, via verbal or emotional cues, results in changes in face–context integration. We conclude by outlining the significance of embracing the impact of contextual cues in shaping impressions from faces.

ARTICLE HISTORY Received 30 November 2022; Accepted 5 February 2024

KEYWORDS Person perception; face; trustworthiness; threat; context

Introduction

Deciding whom to trust is a key task that people must manage throughout their life (Ames et al., 2011). Although information of that sort can be obtained from various features, one powerful source that people use to ascribe trustworthiness to others is a person's face (Jaeger et al., 2019; Oosterhof & Todorov, 2008; Todorov, 2008; Todorov et al., 2015). Indeed, a good deal of work has shown that people rapidly evaluate trustworthiness after minimal time exposure to facial cues (Todorov et al., 2009). Moreover, compared to other dimensions, judgements of trustworthiness made after

CONTACT Marco Brambilla  marco.brambilla@unimib.it; Simone Mattavelli  simone.mattavelli@unimib.it  Department of Psychology, University of Milano-Bicocca, Piazza dell'Ateneo Nuovo, 1, Milano 20126, Italy

*The first two authors contributed equally to this article and the order of names in the byline is alphabetical.

© 2024 European Association of Social Psychology

rapid exposure to facial cues show the highest correlation with judgements made in the absence of time constraints (Willis & Todorov, 2006). Similarly, people show a memory advantage for faces varying on trustworthiness than those varying on other dimensions (Rule et al., 2012). Taken together, these findings fit with the idea that the detection of trustworthiness is essential for human survival (Cosmides & Tooby, 1992). Indeed, our judgements of another person's trustworthiness are highly related to the essential decision we must make about whether they represent an opportunity or a threat (Brambilla & Leach, 2014; Brambilla et al., 2021a; Todorov et al., 2015; Willis & Todorov, 2006).

Although the human ability to accurately detect trustworthiness in others based on facial cues is generally poor (Foo et al., 2022; Jaeger et al., 2022; Rule et al., 2013), inferences of trustworthiness from facial features predict important social outcomes, including sentencing decisions (Wilson & Rule, 2015), financial leading (Duarte et al., 2012), and consumer choices (Ert et al., 2016). As trustworthiness plays a key role in many daily life decisions, defining the factors that push people to perceive humans as (un)trustworthy is key to better understanding person perception and impression formation.

Studies on how the human face impacts trustworthiness attributions have grown significantly in the last decades. These studies examined how the shape and the features of a person's face can lead people to attribute trustworthiness (for a review, Todorov et al., 2015). However, much of the research on trustworthiness inferences from faces has looked at the impact of facial features when the face is presented in isolation. For instance, experimental paradigms typically assess the attributions of trustworthiness from faces being flashed on a computer screen without any contextual information (for a discussion, see Sutherland & Young, 2022; Xie et al., 2023). However, in real life, faces are rarely encountered in isolation, that is, de-contextualised. Contextual information carries valuable information that can be integrated into the final judgement made on a person. It should be clarified that, especially within the literature of emotion recognition from faces, researchers have used the term "context" to encompass a wide variety of instances that extend beyond the target's face (see Wieser & Brosch, 2012 for a review). For instance, contextual information can include factors like target posture, which can impact the recognition of emotions conveyed by a face (Aviezer et al., 2008, 2011). Additionally, contextual information can also involve the presence of other individuals' faces, whose expressions can modulate the perceived emotional valence of surprised target faces (Neta et al., 2011). In the current review, the term "context" refers to another relevant source of information that perceivers can derive from the visual (or non-visual) scenario within which a face is displayed.

Specifically, we considered non-human cues in which faces are embedded. Thus, entities associated with human characteristics (e.g., other human faces, human bodies, etc.) were never part of the contextual stimuli employed in our studies. Prior studies have highlighted the importance of investigating this type of contextual information, as it can affect facial expression recognition and its processing at the neural level (Righart & De Gelder, 2006).

In line with this reasoning, a newly emerging perspective has shown that contextual information influences trustworthiness judgements from faces. To get a vivid idea, we might catch sight of another person walking in a park or in a narrow alleyway. It seems reasonable that running into a stranger with an untrustworthy-looking face in a poorly lit alleyway is a very different prospect from meeting them in a park on a sunny day. In this article, we review recent work from our laboratory showing that judgements of a person's trustworthiness from faces can be modified based on the nature of the surrounding contextual information.

Building on the bond between trustworthiness and threat (Brambilla et al., 2021a; Todorov et al., 2015), the present paper focuses on face-context integration on trustworthiness judgements. We define this face-context integration effect as the degree to which information external to the target individual influences attributions of trustworthiness made on the target.¹

We first review work showing that the detection of person trustworthiness is influenced by the level of threat conveyed by the visual scene in which both trustworthy- and untrustworthy-looking individuals are embedded. Specifically, we consider research showing that untrustworthy-looking individuals are more easily categorised as such when surrounded by threatening (e.g., a room with blood on the walls) rather than negative (e.g., a rundown building) or neutral (e.g., a field of grass) visual contexts.

Next, we review work showing that face-context integration occurs in a cross-modal fashion. Contextual information can be integrated into the evaluation of a person based on their faces also when the former comes from an auditory source: thus, individuals are judged as more untrustworthy when their faces are presented in temporal contingency with threatening (e.g., fire alarms) rather than negative (e.g., drill) or neutral (e.g., claps) contextual sounds.

We then discuss findings demonstrating that face-context integration is not due to a mere perceptual congruency between the dimension of threat and trustworthiness. Rather, it is likely explained by adaptive motives that force humans to spot untrustworthiness in threatening situations.

¹For all the reviewed studies, the ultimate criterion refers to the attribution of target trustworthiness (DV), which is made on facial stimuli varying on different features (IV1) and presented in contexts whose features can also vary (IV2).

Corroborating this idea, we review a series of studies showing that untrustworthy-looking faces in threatening contexts lead to more extreme dispositional attributions than trustworthy-looking faces in reassuring contexts. Moreover, such a negativity effect did not extend to another face-context binomial for which the advantage for negative emotional cues evoked by the context is of less adaptive purpose (i.e., introverted-looking faces in sad contexts).

Building on this evidence, we also consider studies showing that the bond between threat and trustworthiness in face-context integration is moderated by specific inferences made by the perceiver on the relationship between the perceptual stimuli (i.e., the face and its context). Indeed, we review evidence showing that the attribution of untrustworthiness to a social target based on their facial characteristics is amplified when faces are presented in threatening contexts ascribable to human actions. Thus, we discuss findings showing that faces in context are not passively processed by the perceiver and face-context integration is encoded relationally.

We then review a final set of studies examining the integration between contextual information and non-morphological facial features, namely emotional expressions. We emphasise that when facial features convey information that can be directly linked to the contextual scene, such as facial emotions do with internal states prompted by external (contextual) cues, then the faces and contexts influence in concert the attribution of person trustworthiness.

We conclude our review by showing the importance of embracing the role played by environmental cues in shaping impression formation from faces and by outlining a trajectory for future research.

Inferring trustworthiness from faces

People spontaneously infer a wide range of characteristics from facial appearance (Uleman et al., 1996). For instance, a person's age (Wright & Stroud, 2002), race (Richeson & Trawalter, 2005), and sex (Macrae & Martin, 2007) are perceived in a fraction of a second. People also infer personality traits from facial appearance; an important class of inferences concerns judgements of trustworthiness (Jaeger et al., 2019; Slepian et al., 2012; Todorov et al., 2008, 2015). Indeed, people start discriminating trustworthiness right after being exposed to a facial cue and more rapidly than other personality traits. As a case in point, Willis and Todorov (2006) presented participants with unfamiliar faces for 100, 500, or 1,000 ms and asked for their impressions of the targets on various trait dimensions, such as trustworthiness, aggressiveness, and likeability. A hundred-millisecond exposure was more than sufficient for participants to form specific impressions of trustworthiness. Further studies also

showed that judgements of trustworthiness from faces are an excellent approximation of the general evaluation of the target person. As such, Oosterhof and Todorov (2008) identified the most frequent trait dimensions used to describe emotionally neutral faces (see also Todorov & Oh, 2021). Judgements were submitted to a principal component analysis to identify the underlying structure of participants' judgements. The first component accounted for 63% of the variance and reflected the evaluative meaning of the trait dimensions. All positive judgements (e.g., trustworthy, intelligent) showed positive loadings on this component while all negative judgements (e.g., weird, mean) had negative loadings. Indeed, research has shown that many judgements from faces are highly correlated (Lin et al., 2021). Nevertheless, trustworthiness judgements had the highest loading on the evaluation component (.94) indicating that they most closely resemble the valence evaluation underpinning many social judgements made from faces.

Trustworthiness is also a key factor that influences how people remember facial stimuli, with untrustworthy-looking faces being more easily recalled than trustworthy-looking faces (Rule et al., 2012). In a series of studies, participants viewed various faces on a computer screen and then completed a word-search puzzle for 2 min as a filler-task. Next, they were presented with the same faces plus previously unseen distractor faces in a random order with instruction to indicate whether they had seen each face in the previous section. Finally, participants rated each face on various traits (i.e., dominance, facial maturity, likeability, trustworthiness). Results showed that participants' memory performances increased as a function of (negative) variation on trustworthiness and likeability, whereas no impact of the other dimensions was observed. Taken together, findings from trait attribution to trait memory converge to the idea that people show a preferential processing of facial trustworthiness (Olivola et al., 2014; Todorov & Duchaine, 2008; Todorov et al., 2015).

Such a preferential processing lies in the link between the dimensions of trustworthiness and threat (Brambilla et al., 2021a). Indeed, defining whether an individual is (un)trustworthy is highly related to the fundamental decision of whether that person is harmful or beneficial to the self (Ames et al., 2011; Brambilla et al., 2021a; Cosmides & Tooby, 1992). Accordingly, it has been shown that the higher the perceived untrustworthiness of a social target, the more such a target is believed to pose a threat to the stability and integrity of the whole community. By contrast, highly trustworthy social targets are perceived as beneficial for group survival and cohesion (Brambilla & Leach, 2014). At the group level, untrustworthy ingroup members are perceived as threatening to the image of their group (Brambilla et al., 2013; Leach et al., 2007; Van der Toorn et al., 2015), while untrustworthy outgroup members are perceived as posing a real and a concrete danger to the

ingroup's survival possibilities and represent a threat to the group's safety (Brambilla et al., 2012, 2013; Leidner & Castano, 2012). Therefore, trustworthiness judgements from faces predict basic approach/avoidance reactions, according to the functionalist perspective that friends and foes should be approached and avoided, respectively (Slepian et al., 2012).

Besides approach/avoidance reactions, judgements of trustworthiness from faces predict important social outcomes. For instance, people invest less money with partners who look untrustworthy (Chang et al., 2010; Rezlescu et al., 2012; Stirrat & Perrett, 2010); trustworthy-looking individuals have a higher chance of being granted loans (Duarte et al., 2012); perceiving a seller as trustworthy, even from a picture, positively affects consumer's choices (Ert et al., 2016); and crowdfunding campaigns are more successful when the entrepreneur's face is trustworthy (Duan et al., 2020). Remarkably, facial trustworthiness also affects decisions about guilt in court, as defendants who have untrustworthy-looking faces are more likely to receive the death penalty (Porter et al., 2010; Wilson & Rule, 2015; cf. Kramer & Gardner, 2020).

All this evidence might point towards a real association between facial trustworthiness and actual target's behaviour. However, impressions of trustworthiness from faces have at best only limited accuracy (Foo et al., 2022; Jaeger et al., 2022; Rule et al., 2013) meaning that their everyday use might be problematic. For instance, it has been shown that inferences of trustworthiness made from the faces of corporate criminals did not differ from inferences made from the faces of noncriminal executives. Similarly, judgements of trustworthiness did not differ between the faces of military criminals and the faces of military heroes (Rule et al., 2013). Yet, despite this limited accuracy, facial impressions have critical social consequences and are hard to shift (Chang et al., 2010) even in the presence of more valid cues. Given the importance of facial impressions to everyday decisions, it is critical that scientists understand the factors that push people to judge trustworthiness from faces, even if those judgements are far from accurate. Here, we argue that trustworthiness judgements from faces are influenced by the context in which the face is embedded.

Judging trustworthiness from faces and the surrounding visual context

Most research on evaluating trustworthiness based on facial appearance has largely centred on isolated facial images, specifically faces displayed on a screen without accompanying contextual details (Sutherland & Young, 2022; Todorov et al., 2015; Xie et al., 2023). However, anyone would agree that encountering someone face-to-face often happens in rich and informative surrounding contexts. Such contextual information unlikely goes

undetected and might contribute to shape impressions. In line with this reasoning, research on emotions has shown that the visual context surrounding the face influences the perception of facial emotions (Righart & De Gelder, 2008; see also Aviezer et al., 2008; Barrett & Kensinger, 2010). In such experiments, participants are presented with a series of faces expressing various emotions (i.e., fear, happiness, and disgust) superimposed on visual contexts that could be either congruent (e.g., fearful face in a fearful scene) or incongruent (e.g., fearful face in a happy scene) with the emotional content of the face. Results revealed faster response times and higher accuracy in the categorisation of the target's emotional expression when faces were embedded in congruent visual contexts; thus, facial expressions of fear, happiness, and disgust are more easily categorised as such when presented in fearful, happy, and disgusting visual scenes (see also Aguado et al., 2019).

Congruency effects have also been found when judging ethnicity from faces (Freeman et al., 2013). Specifically, when individuals are asked to categorise faces as White or Asian, the visual background surrounding the facial stimuli influences categorisation. Indeed, categorising a face as "Asian" is more likely when an Asian face appears in a Chinese-typed rather than an American-typed scene context. Conversely, categorising a face as "American" is more likely when a White face appears in an American-typed rather than a Chinese-typed scene context.

Together, these works suggest that recognition of both dynamic (i.e., emotions) and static (i.e., ethnicity) facial features is influenced by the visual scenes in which faces are embedded. This early evidence spurred our research on the relative importance of visual background when judging traits from faces and in particular trustworthiness.

Building on (i) the importance of congruency in face-context integration (e.g., Freeman et al., 2013; Righart & De Gelder, 2008) and (ii) the inherent link between trustworthiness and threat (for reviews, Brambilla & Leach, 2014; Todorov et al., 2015), our early work tested whether visual scenes associated with threat could alter the judgements of trustworthiness from facial features (Brambilla et al., 2018). To test this idea, we employed a mouse-tracking technique that records and analyses hand movements during categorisation tasks (Freeman & Ambady, 2010; Freeman & Johnson, 2016). Such a technique goes beyond response times, and it is thought to provide more process-sensitive information. In a typical trial, participants are required to click on a "Start" button located at the bottom-centre of the screen, which is replaced by a target. Participants then must click an appropriate response button located either at the top-left or top-right of the screen. Because the mouse is moving while a categorisation response is still evolving, its movement should provide a "read-out" of how categorisation unfolds over time (Freeman & Ambady, 2011; Freeman & Johnson, 2016). In other words, this paradigm can track how various cues

drive categorisation in real time and therefore reveal potentially subtle influences of the context that can remain hidden in a more standard categorisation measure. If the visual context influences the categorisation of trustworthiness from faces, one would expect that perceivers partially integrate the response associated with the context with that associated with the face. This would be evidenced by a partial attraction in participants' mouse trajectories towards the opposite category response before clicking their final response when the facial and context information do not match. In other words, trajectories would be facilitated when facial cues and contextual cues are compatible (e.g., an untrustworthy-looking individual in a threatening scene), and would be partially attracted to the context-associated response when incompatible (e.g., a trustworthy-looking individual in a threatening scene).

In the first experiment ($N = 51$), we asked participants to categorise the trustworthiness of faces that were shown against either threatening or neutral backgrounds. Specifically, we employed 24 computer-generated identities (12 trustworthy-looking, 12 untrustworthy-looking) borrowed from a set of pictures previously validated for facial trustworthiness (Todorov et al., 2013). We also employed four threatening (i.e., a tornado, an exploding

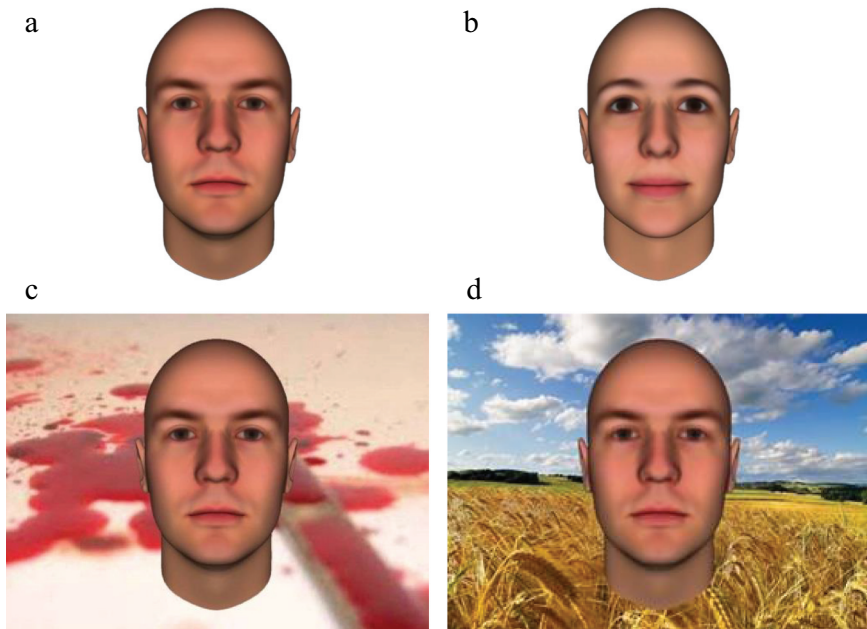


Figure 1. Examples of untrustworthy-looking (a) and trustworthy-looking (b) individuals, and of an untrustworthy-looking individual embedded in a threatening (c) or a neutral (d) context. Adapted from Brambilla et al. (2018) – experiment 1.

volcano, a bloody knife, and a fire) and four neutral (i.e., a wheat field, a grass field, a hill landscape, and a countryside landscape) visual contexts obtained from public domain websites (see [Figure 1](#)).

Participants were told that they would be presented with images of individuals in various settings and asked to categorise each person as either trustworthy or untrustworthy. Participants were instructed to make their decisions as quickly and accurately as possible, basing their judgements on their first impressions. On every trial, participants mouse-clicked a “Start” label at the bottom-centre of the screen, which was then replaced by a face-context pair in the centre of the screen. Face-context pairs were presented in randomised order, and target stimuli were categorised by clicking either the “trustworthy” or the “untrustworthy” response labels located in the top-left and top-right corners of the screen (labels’ location was counterbalanced across participants). To permit averaging and comparison across trials, we normalised trajectories into 101 time-steps and remapped leftward trajectories rightwards (inverted along the x-axis). To index the trajectories’ attraction towards the opposite category, we computed the area under the curve (AUC): the area between the observed trajectory and an idealised straight-line trajectory (Freeman & Ambady, 2010). We found that the scene context influenced the target categorisation. Indeed, untrustworthy-looking individuals elicited more direct trajectories (lower AUC) when they were embedded in threatening contexts than in neutral contexts. Conversely, trustworthy-looking individuals elicited more curved trajectories (higher AUC) when they were embedded in threatening contexts than in neutral contexts. Thus, under conditions of congruency between contextual threat and facial trustworthiness (e.g., an untrustworthy-looking individual in a threatening context), trajectories became more direct en route to the selected response. Under conditions of incongruency (e.g., a trustworthy-looking individuals in a threatening context), trajectories showed an increased attraction towards the opposite-category response associated with the context (see [Figure 2](#)).

Two additional experiments ($N = 46$ and $N = 50$) corroborated these findings in a design that offered a stringent test for the importance of face-context congruency. In such studies, we disentangled the effects of threatening scenes from negative contexts in general. Thus, we included a further experimental condition and asked participants to categorise the trustworthiness of individuals whose faces were shown against threatening, negative but unthreatening, or neutral background in a 2 (face: trustworthy-looking vs. untrustworthy-looking) \times 3 (scene context: neutral, negative, threatening) experimental design. Importantly, threatening and negative scenes were matched on valence and only differed on perceived threat. Threatening and neutral scenes were borrowed from the first experiment; negative scenes represented an abandoned factory, a forgotten toy on a bench, broken piano keys, and a degraded building. To

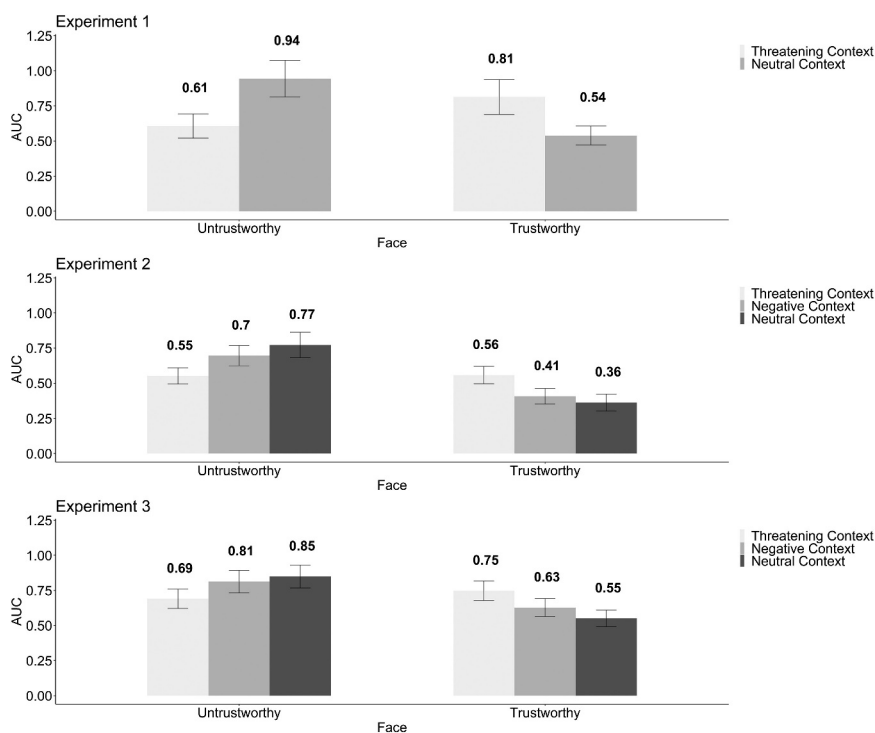


Figure 2. Untrustworthy-looking individuals are more easily categorised when embedded in threatening visual contexts. Threatening backgrounds disrupt the categorisation of trustworthy-looking individuals. Lines represent 95% confidence intervals. The mean of each condition is reported above each bar. Adapted from Brambilla et al. (2018).

increase the ecological validity of our stimuli, in the final experiment we added hairlines to the faces and embedded the facial stimuli in the visual context more naturalistically (see Figure 3).

Results from both studies confirmed that the scene in which a face is encountered alters trustworthiness categorisation. Indeed, trajectories were more direct when untrustworthy-looking individuals were shown in threatening rather than in negative and neutral scenes. Conversely, trajectories were more curved when trustworthy-looking individuals were surrounded by threatening rather than negative and neutral scenes. Trajectories did not differ between negative and neutral contexts for both trustworthy- and untrustworthy-looking individuals (see Figure 2). In sum, untrustworthy-looking individuals were more easily categorised as such when surrounded by threatening visual contexts. By contrast, threatening backgrounds disrupted the categorisation of trustworthy-looking faces. Thus, face-context integration on categorisation of facial stimuli based on their perceived

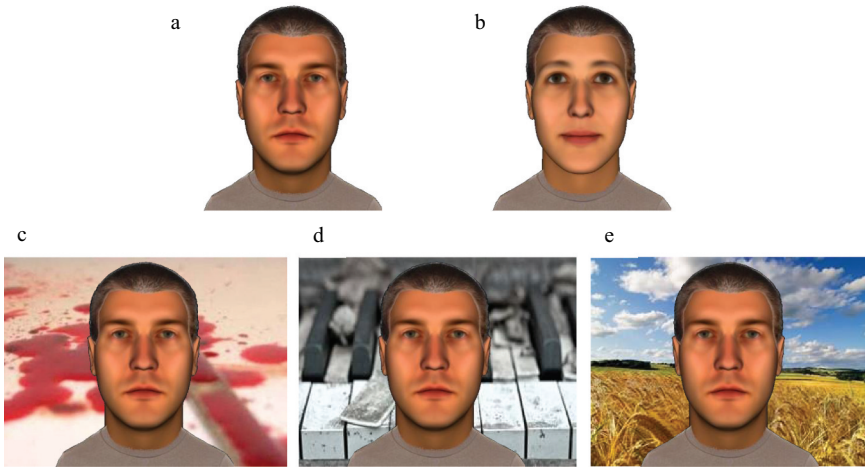


Figure 3. Examples of untrustworthy-looking (a) and trustworthy-looking (b) individuals with added hairlines, and of an untrustworthy-looking individual embedded in a threatening (c), negative but not threatening (d), or a neutral (e) context. Adapted from Brambilla et al. (2018) – experiment 3.

trustworthiness emerged when there was conceptual congruency between the domains expressed in the context (i.e., threat) and in the face (i.e., trustworthiness).

These preliminary studies speak to the malleable nature of trustworthiness such that its perception from facial cues is readily pushed around by the scene context. Indeed, visual scenes systematically altered the categorisation of a person's trustworthiness by further revealing a close connection between threat and facial (un)trustworthiness.

Judging trustworthiness from faces and the surrounding auditory context

The sight of a face is often accompanied by contextual cues that are processed either visually or under different sensory modalities. For instance, some environments are inherently characterised by threatening sounds (e.g., in neighbourhoods with high crime rates, the sound of police sirens is often heard) that may impact upon our impression of an unknown passer-by. Thus, the analysis of how the context affects trustworthiness judgements from individuals' faces should extend to contextual cues coming from different sensory modalities. Building on our findings (Brambilla et al., 2018), in another set of studies we tested whether face-context integration occurs in a cross-modal fashion and whether the evaluation of trustworthiness based on facial features is influenced by the threatening nature of the auditory

context in which the face is embedded (Brambilla et al., 2021b). To do so, we asked participants to evaluate target individuals on trustworthiness based on their faces that were surrounded by either threatening or non-threatening auditory contexts. While our early work asked participants to categorise individuals as trustworthy or untrustworthy, here we asked participants to evaluate a target person explicitly by using Likert scales. This helped us to align with the framework that theorises the dimensions of social perception (including trustworthiness) as inherently continuous (see Todorov et al., 2015).

In the first experiment ($N = 58$), we employed the same 24 computer-generated identities (12 trustworthy-looking, 12 untrustworthy-looking) used in our early work on face-context integration (see Brambilla et al., 2018, Experiment 3). Auditory stimuli (four threatening, four non-threatening) were obtained from public domain websites. Because we did not want the auditory stimuli to be somehow attributed to the facial identities, none of the selected stimuli was humanly produced (threatening sounds: ambulance siren, bombs exploding, civil defence siren, tornado; non-threatening sounds: waves on the beach, chirping birds, seagulls on the seashore, wetlands animals). Each sound was normalised in its intensity and lasted for 9 s.

Participants were asked to wear headphones and judge each person appearing on the screen on trustworthiness using a 7-point Likert scale (1 = untrustworthy to 7 = trustworthy). The experiment consisted of four blocks, two blocks for each sound condition, alternated in pairs between subjects. Each block was composed of 24 trials, one per each facial identity. Random sequences of four homogeneous sounds (all threatening vs. all non-threatening) were played contingently with the onset of faces on the computer screen. The auditory context's sequences were played throughout the entire block without interruption. No time limit was set even though participants were kindly reminded to provide their judgements as fast as possible. Results showed that untrustworthy-looking individuals were judged as more untrustworthy than trustworthy-looking ones. We also found that the auditory context altered the evaluation of target's trustworthiness from facial features: both untrustworthy-looking and trustworthy-looking individuals were judged more untrustworthy when their faces were accompanied with threatening auditory information (see Figure 4).

A second experiment ($N = 56$) corroborated these findings in a design that disentangled whether the effect we found is specific to the threat-trustworthiness congruency or indicates more general effects of negative auditory contexts. Thus, we included a further experimental condition and asked participants to evaluate target individuals on trustworthiness based on their faces that could be embedded in either threatening, negative but unthreatening, or neutral auditory contexts.

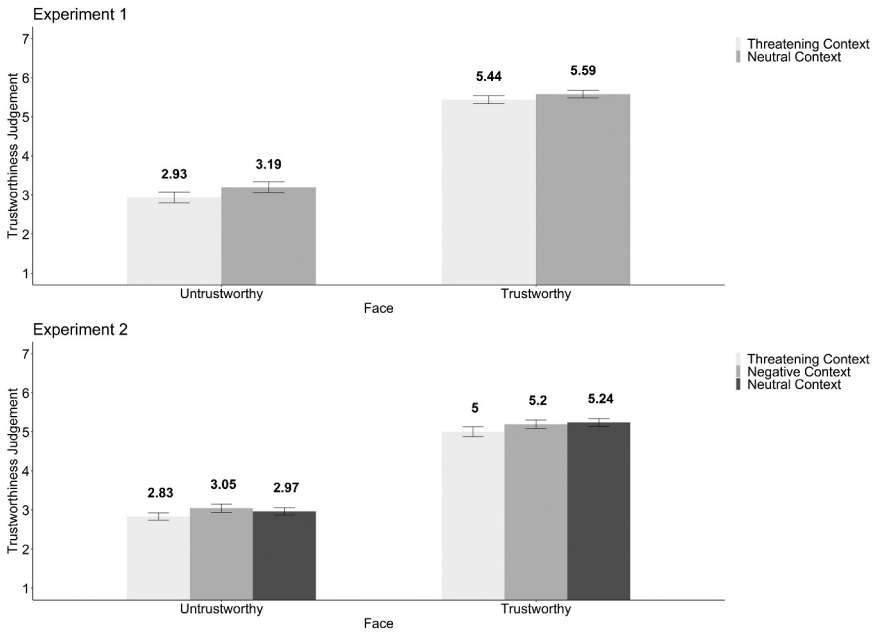


Figure 4. Individuals are judged as more untrustworthy when their faces are embedded in threatening (vs. negative vs. neutral) contextual sounds. Lines represent 95% confidence intervals. The mean of each condition is reported above each bar. Adapted from Brambilla et al. (2021a).

As in our early work, threatening and negative contexts were balanced on valence but differed on perceived threat. Threatening (i.e., traffic jam, house explosion, laser beam), negative (i.e., jackhammer, broken car, drill), and neutral sounds (i.e., pinball sounds, claps, doorbell) were either taken from online repositories (findsounds.com) or extracted from the International Affective Digitized Sounds (IADS-2, Bradley & Lang, 2007). The procedure of the experiment closely resembled the procedure used in the first experiment. The results confirmed that untrustworthiness and threat are inherently associated, as individuals were judged more untrustworthy when their faces were accompanied by threatening rather than negative or neutral contextual sounds. Moreover, we did not find any difference between neutral and negative contexts. Thus, the contextual effects were observed only when auditory stimuli conveyed threat and did not reflect a general context negativity effect.

Taken together, these latter findings add further evidence in support of the idea that judgements of trustworthiness from faces can be modified when individuals perceive the context and facial cues simultaneously. Moreover, our data show that the impact of contextual threat on trustworthiness

evaluations emerges even when considering threatening information related to different sensory modalities.

Beyond stimuli congruency: the specificity of face–context integration

The studies reviewed so far showed that the threat conveyed by contextual information (either visual or auditory) alters trustworthiness judgements based on facial features. Perhaps, the most remarkable finding is the superior impact of threatening over negative contexts in altering judgements of trustworthiness. That negative (but not threatening) and neutral contexts had a similar effect on judgements corroborates the idea that the observed face-context integration might be explained considering the conceptual congruency between the dimensions of trustworthiness and threat (see Brambilla & Leach, 2014; Brambilla et al., 2021a).

The idea that congruency matters in face-context integration is not new. A very 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). As detailed in the introduction of the article, 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 the 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. Importantly, studies investigating face-context integration in emotion recognition showed that congruency matters for both positive and negative stimuli compounds (i.e., a face and a context). Conceiving face-context integration in terms of congruency implies that, virtually, any context that has a conceptual overlap with a contingent facial disposition might ease the perception of the latter. However, judgements of another person's trustworthiness are highly related to the decisions we must make about whether they represent an opportunity or a threat (Ames et al., 2011; Brambilla & Leach, 2014; Cosmides & Tooby, 1992). To clarify this idea, the consequences of mistakenly judging a person as trustworthy, when in fact they are not, are potentially way more severe than the consequences of doing the opposite (i.e., mistakenly judging a person as untrustworthy, when in fact they are trustworthy). Under this view, the adaptive advantage of the threat-trust domain might induce higher integration (i.e., more extreme judgements on the relevant dimension) under this specific congruency.

Across three experiments (Mattavelli et al., 2022), we explored whether the integration of contextual threat in the attribution of a person's trustworthiness from facial stimuli extends beyond congruency. We tested the

specificity of face-context integration in the negative extreme of the threat-trust binomial. Our investigation followed two approaches. First, within the threat-trust domain, we tested whether the valence of the stimuli compounds could moderate the impact of congruency on face-context integration. Due to its adaptive advantage, we anticipated a stronger absolute impact of congruency on the negative extreme compared to the positive counterpart. In other words, the negative impact of a threatening context in diminishing the level of trustworthiness attributed to untrustworthy-looking individuals should be stronger than the positive impact of a reassuring context in increasing the level of trustworthiness attributed to trustworthy-looking individuals. Second, we tested the specificity of this hypothesised moderation across domains. Specifically, we compared how valence influenced face-context integration in the threat-trust binomial and in the sadness-extroversion binomial.

In the first experiment ($N = 89$), we focused on the role of valence in moderating the congruency effect. We tested the hypothesis that the evaluation of a person's trustworthiness based on their face would be influenced by the threat conveyed by the visual context in which the face is embedded. Facial and contextual stimuli were selected to belong to the same end of the threat-trust continuum. In two blocks of 72 trials each, we exposed participants to a series of faces individually presented on screen and embedded in a surrounding context. Facial stimuli were 24 identities (12 trustworthy-looking, 12 untrustworthy-looking) taken from our previous studies (Brambilla et al., 2018; Experiment 3). In each trial, targets could be either trustworthy- or untrustworthy-looking and presented either in threatening, reassuring, or neutral contexts. Four pictures (i.e., tornado, bomb explosions, volcano, blood stains) were used as threatening contexts; and other four (i.e., two beach landscapes, lake in spring, lake in autumn) as reassuring contexts (see Figure 5). Importantly, context stimuli were selected based on a pre-test that established that the two sets of stimuli did not differ from each other in absolute extremity, either on valence or on threat/reassurance. This means that any difference in face-context integration could not be attributed to context stimuli features. A grey rectangle was used as control (i.e., neutral) context stimulus. In each trial, participants rated the target on trustworthiness, using a continuous scale ranging from -3 (untrustworthy) to $+3$ (trustworthy). We were interested in the comparison between (i) the positive effect of reassuring contexts on trustworthiness attributed to trustworthy-looking targets (positive congruency) and (ii) the negative effect of threatening contexts on untrustworthiness attributed to untrustworthy-looking targets (negative congruency). Our results indicated that threatening contexts decreased the attribution of trustworthiness to untrustworthy-looking targets more than reassuring contexts increased the trustworthiness attributed to trustworthy-looking targets (see Figure 6). These findings suggest that

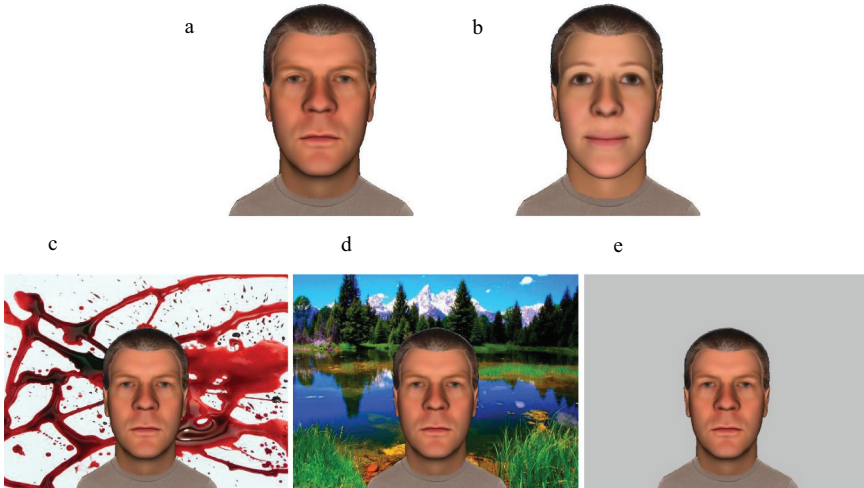


Figure 5. Examples of untrustworthy-looking (a) and trustworthy-looking (b) individuals, and of an untrustworthy-looking individual embedded in a threatening (c), reassuring (d), or a neutral (e) context. Adapted from Mattavelli et al. (2022) – experiments 1 and 2.

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.

The negativity effect found in the first experiment is compatible with an ecological-functional approach to person perception (Fiske, 1992; Gibson, 1977; McArthur & Baron, 1983). Namely, threatening scenarios are meant to put the individual in a fight or flight mode. Thus, more extreme integration on the negative end of the threat-trust binomial should be a function of the action-response (e.g., avoidance) triggered by the compound of stimuli. Alternatively, a negative effect in face-context integration might emerge regardless of what response is cued by the stimuli. A second experiment ($N = 120$) tested these alternative predictions. We confronted the role of valence in qualifying face-context integration across two different bonds, that is, the bond between trustworthiness and threat and that between extroversion and sadness. In the latter, positive congruency was obtained by presenting extrovert-looking targets in happy contexts (i.e., colourful balloons, fireworks in the sky, rainbow, colourful umbrellas), whereas negative congruency would consist of introvert-looking targets presented in sad contexts (i.e., abandoned factory, forgotten toy on a bench, broken piano keys, rainy day). The experiment consisted of two separate judgement blocks of 72 trials. One block was identical to that administered in the first experiment, with either trustworthy- or untrustworthy-looking targets surrounded by threatening, reassuring, or neutral contexts to be rated on perceived

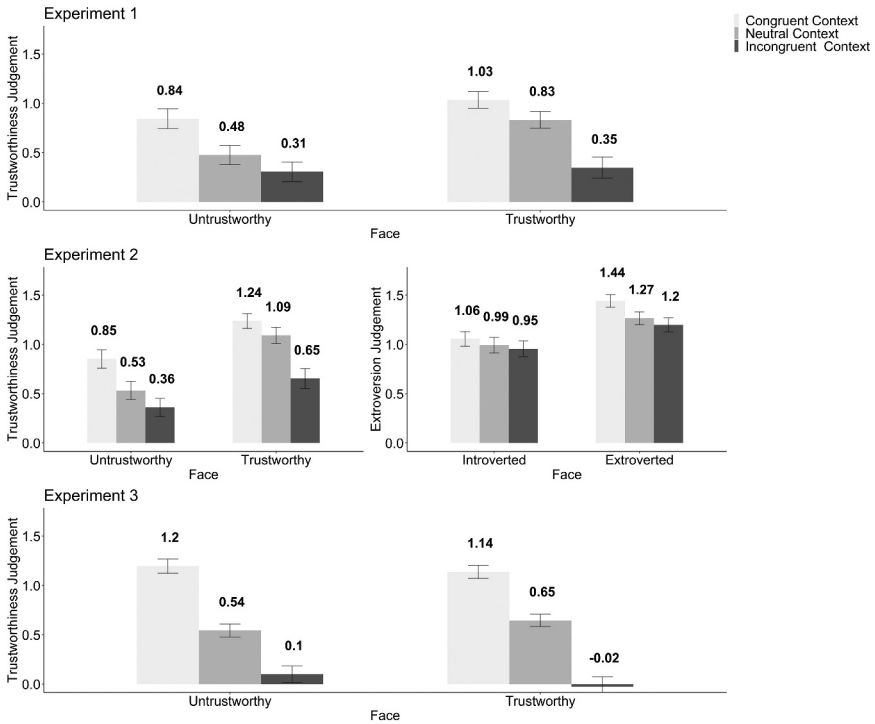


Figure 6. Face-context integration is stronger on the negative extreme of the threat-trust binominal and does not extend to another face-context binominal (sadness-extroversion). Lines represent 95% confidence intervals. The mean of each condition is reported above each bar. Adapted from Mattavelli et al. (2022).

trustworthiness ($-3 = \text{Untrustworthy}$ to $+3 = \text{Trustworthy}$). The other block consisted of introverted- or extroverted-looking faces surrounded by sad, happy, or neutral contexts (see Figure 7) and judged on perceived extroversion ($-3 = \text{Introverted}$ to $+3 = \text{Extroverted}$). Introverted- and extroverted-looking faces were borrowed from a set of pictures previously validated for facial extroversion (Todorov et al., 2013).

Results indicated a significant three-way interaction between face-context congruency, valence, and domain. Thus, the negativity effect in the face-context congruency effect was qualified by the relevant domain of judgement. In the trustworthiness domain, we replicated the results observed in the first experiment: valence moderated the effect of congruency on face context integration, such that the effect was stronger on the negative end. Instead, this was not the case in the extroversion domain (Figure 6).

This second experiment clarified that the superior face-context integration observed for negative congruency does not reflect a generalised negativity bias in face-context congruency. Rather, it was unique to the threat-

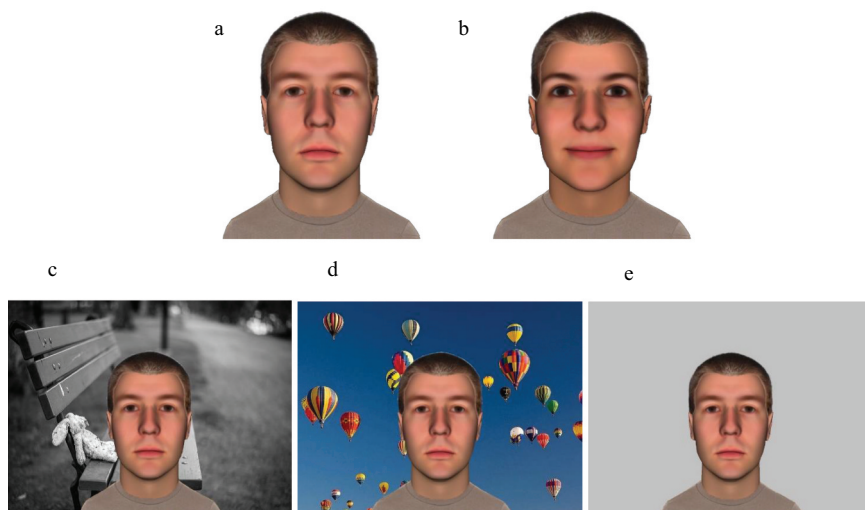


Figure 7. Examples of introverted-looking (a) and extroverted-looking (b) individuals and of an introverted-looking individual embedded in a sad (c), happy (d), or a neutral (e) context. Adapted from Mattavelli et al. (2022) – experiment 2.

trust domain. However, one element of confusion in the interpretation of the negativity effect observed in the threat-trustworthiness binomial remained. Namely, both threatening and reassuring contexts, while comparable in extremity, differed in their potential relation to the human stimuli (i.e., the faces). In fact, whereas threatening contexts such as a bomb explosion or blood stains can be easily perceived as the product of human actions, this cannot be the case with reassuring contexts such as peaceful landscapes. 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 targets were responsible for the threatening scene portrayed on the background, whereas the same inference was not possible for trustworthy-looking targets presented in reassuring scenes (we will come back to this point in the next section). A third experiment ($N=144$) was designed to address this issue. To obtain threatening and reassuring context stimuli comparable both in extremity and in their being potentially attributable to humans, we designed a learning phase at the beginning of the main experiment. In this phase, participants were exposed to eight context stimuli: four were selected to be (visually) threatening and four to be reassuring (Figure 8). None of these stimuli referred (visually) to any human action. However, each context was paired with a sentence describing either threatening (e.g., “someone molested an individual”, “someone offended an old man”) or reassuring (e.g., “someone helped an

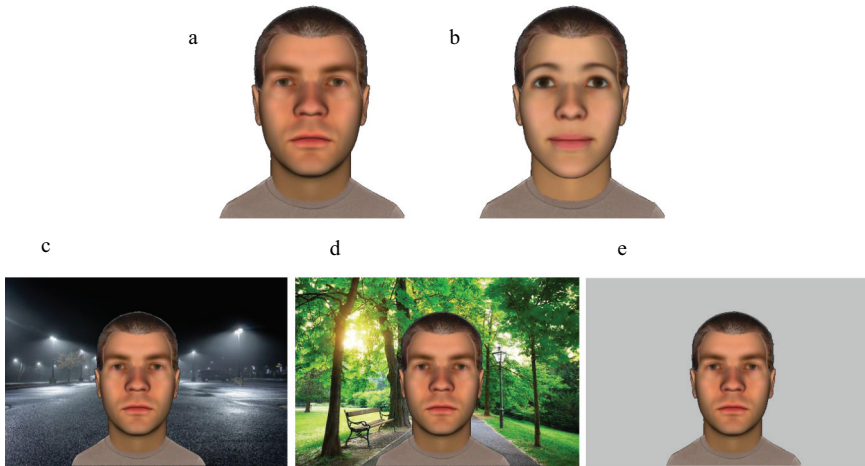


Figure 8. Examples of untrustworthy-looking (a) and trustworthy-looking (b) individuals, and of an untrustworthy-looking individual embedded in threatening (c), reassuring (d), or neutral (e) contexts. Adapted from Mattavelli et al. (2022) – experiment 3.

old lady”, “someone cheered up some strangers”) human actions, also pre-tested to be equally extreme on the threat-reassuring continuum. Throughout this learning phase, participants learned that both threatening and reassuring visual scenes were associated with a congruent human action. The rest of the experiment consisted of an evaluation task that largely mirrored that administered in the first two experiments. Confirming prior findings, we found a significant interaction between face-context congruency and valence: a stronger congruency effect emerged on the negative rather than the positive end of the trust-threat binomial (Figure 6). Thus, even after ruling-out any potential asymmetry between positive and negative congruency in terms of contexts’ human relatedness, face-context integration showed stronger on the negative end of the threat-trust binomial.

These studies showed that integrating contextual threat into the evaluation of trustworthiness from faces is not simply a matter of stimuli congruency. Within the threat-trust binomial, congruency led to more extreme judgements on the negative pole, as compared to the positive counterpart. Moreover, the higher effect of congruency for negative face-context compounds did not generalise to another domain of judgement. Taken together, these findings fit well with an ecological interpretation of face-context integration in person perception, whereby contextual threat is integrated in the evaluation of trustworthiness from targets’ faces by priming specific behavioural responses (see Gibson, 1979). For instance, a threatening scene might prime fear or avoidance responses, which ultimately influence how participants perceive and judge target stimuli presented in that specific

context. Finding a negativity bias in the impact of context on attributed trustworthiness would also be in line with a general default tendency to trust, rather than distrust, others (Katzir & Posten, 2023). Assuming a default trust assumption would imply that processing trustworthy-looking individuals in reassuring contextual scenes should simply reinforce this spontaneous assumption via two assumption-consistent cues (i.e., the face and the context). On the other hand, processing untrustworthy-looking individuals in threatening scenes should invalidate this spontaneous assumption. Thus, the stronger absolute effect observed on the negative end of the threat-trust binomial might be explained in terms of expectations-violation.

This set of studies supports the idea of an inherent link between the dimension of trustworthiness and threat (Brambilla & Leach, 2014). However, the nature of such a link remains unclear. In other words, why do people use contextual threat as a cue to infer untrustworthiness in a target person? For instance, one might hypothesise that any sort of threat might induce perceivers to ascribe untrustworthiness to a target, simply because the two dimensions are conceptually linked. Alternatively, this face-context integration effect might be qualified by the relationship that the perceiver can establish between the target individual and the surrounding context. The next set of studies delves into these alternative hypotheses.

On the nature of face–context relationship

The link between trustworthiness and threat in face-context integration can be analysed in different ways. At a pure mechanistic level of analysis, phenomena are explained as consequences of changes in the organism. Thus, one can say that contextual threat is integrated into the attribution of trustworthiness from the face because the two dimensions are associated in the mind. One might further support this idea by saying 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). Yet, this mechanistic explanation says little about *how* perceivers happen to integrate information coming from these two sources of information. At a functional level of analysis, the focus is on the environmental conditions under which a phenomenon is more or less likely to occur. Under this view, determining whether another person is worthy of trust is a function of external information suggesting that the target face might represent an opportunity of threat (Ames et al., 2011). In the studies that we have reviewed, we advanced the idea that the context could provide this sort of information. Namely, contextual threat likely informs about target's intentions. Thus, participants might use contextual information to infer that the target is dangerous or untrustworthy. This latter presupposition implies that perceivers make further intentional assumptions about individuals in

a threatening context. Logic dictates that, to make assumptions about one's intention, perceivers should somehow qualify the link between contextual threat and the target. In the next studies, we explored how face-context integration varies depending on (i) the possibility to establish meaningful relationships between the target (represented by their face) and the context and (ii) the nature of such established relationships.

The idea that face-context integration can depend on encoding relationally comes from evaluative learning research. One prominent example of learning effects is evaluative conditioning, which refers to a change in liking towards a neutral stimulus due to its repeated pairing with a valenced stimulus (De Houwer, 2007). 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). Possibly due to its reliance on pairings, evaluative conditioning has been largely interpreted as a primitive form of learning mediated by association formed between the two paired stimuli (e.g., Brinol et al., 2009). However, De Houwer (2009, 2014, 2018) proposed that evaluative conditioning is moderated by propositional reasoning. Under this view, the ultimate change in liking towards an originally neutral stimulus depends on how people qualify the nature of the relationship between paired stimuli. Supporting this argument, 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). Thus, in spite of its apparent primitive operationalisation, pairing can be conceived as a symbolic phenomenon that allows inferences about the nature of the relationship between paired stimuli (see also De Houwer & Hughes, 2016).

Just like conditioning, face-context integration relies on pairing: a face is flashed on the screen together with a contextual scene displayed in the background. Building on this parallel, we argue that if relational encoding moderates the effect of pairings on liking in evaluative conditioning, then it should moderate face-context integration on the attribution of trustworthiness. For instance, when an individual's face is presented in a threatening context, perceivers might be inclined to judge them as less trustworthy because they are meant to be the person that *actively causes* the contextual threat. Thus, the perceiver's ability 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 individual).

One way to explore the importance of relational encoding in qualifying face-context integration is via altering the nature of contextual threat. In the previous section, we argued that, beyond valence, one factor that could affect

the perceived face-context congruency lied in the potential relationship that perceivers could establish between the contextual threat and the human stimuli. Contextual threat can be sometimes attributable to the human action (e.g., a room with blood on the wall), sometimes not (e.g., an exploding volcano). If relational encoding plays a role in face-context integration, one should expect stronger effects when contextual threat can be relationally linked to the target person. Alternatively, if face-context integration reflects an unqualified link between the contextual threat and the target person, altering the meaningfulness of this potential link should not alter face-context integration.

In two initial experiments (Mattavelli, Masi, et al., 2023), we manipulated the nature of the threatening contexts to vary the extent to which the threat conveyed by such contexts could be attributed to human stimuli. These two studies were identical except for the nature of the facial stimuli: in the first experiment, we used computer-generated faces (Todorov et al., 2013), whilst in the second experiment, we used real face from the CFD (Ma et al., 2015). The design of the two studies involved a 2 (facial trustworthiness: trustworthy-looking vs untrustworthy-looking faces) \times 3 (context: human threatening vs non-human threatening vs neutral) full-within design. Participants (total $N = 156$) were presented with images of individuals surrounded by different contexts and asked to rate each person on perceived trustworthiness, using a 7-point Likert-type scale (1 = untrustworthy to 7 = trustworthy). The experiment consisted of two blocks of 72 trials each, with stimuli administered in random order. Crucial for the main research question was the selection of the context stimuli. In a pilot test, independent raters evaluated a series of context stimuli obtained from public domain websites on perceived threat (1 = not at all threatening to 7 = extremely threatening) and potential attribution to the human action (1 = not at all to 7 = extremely). We selected four human threatening contexts (i.e., rifles, gun with bullets, bloody knife, empty room with blood on the walls) and four non-human threatening contexts (i.e., rough sea, tornado, rainstorm, windstorm; see Figure 9). These two sets of stimuli were selected to differ on their potential relation with human action. Importantly, a difference in perceived threat indicated that non-human contexts were perceived as more threatening than the human ones. This double asymmetry has important implications: on the one hand, if face-context integration reflects the overall amount of threat conveyed by the context, then the effect should be stronger (lower attributed trustworthiness) for targets presented in non-human threatening contexts. On the other hand, if relational encoding matters, then one should expect a stronger integration for targets presented in human threatening contexts. In the experiments, a grey rectangle was used as neutral context to establish a baseline measure of trustworthiness attributed to targets.

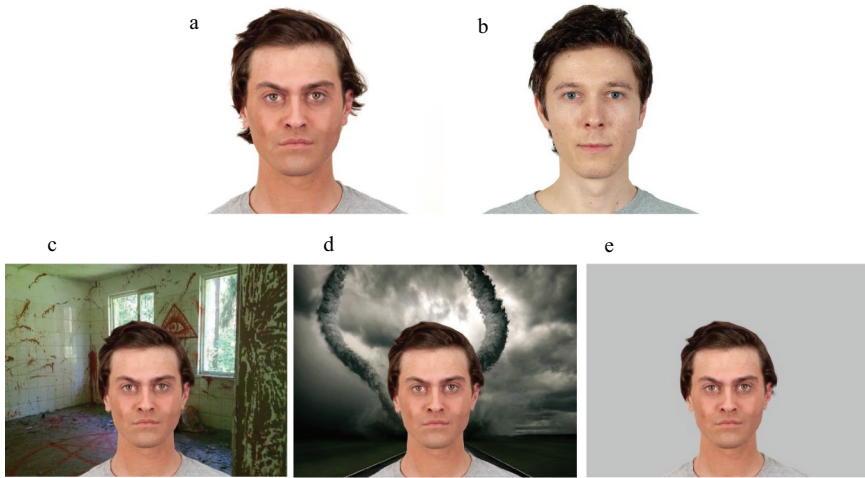


Figure 9. Examples of untrustworthy-looking (a) and trustworthy-looking (b) male individuals extracted from the CFD, and of an untrustworthy-looking individual embedded in human threatening (c), non-human threatening (d), or neutral (e) contexts. Adapted from Mattavelli, Masi, et al. (2023) – Experiment 1b.

The results were consistent across the two studies (Figure 10). Beyond the predictable effect of facial trustworthiness, with lower trustworthiness attributed to untrustworthy-looking targets than to trustworthy-looking targets, we found that human threatening contexts had higher (i.e., more negative) impact on judgements of target trustworthiness than neutral context. Comparing the non-human threatening and neutral contexts showed a lower (but still significant) effect of contextual threat. Importantly, the differential impact of human and non-human threatening context was significant such that targets were perceived as less trustworthy when embedded in human (vs. non-human) threatening contexts. Thus, although human threatening contexts were normatively rated as less threatening than the nonhuman ones, their impact on face-context integration was stronger.

In a follow-up experiment ($N = 101$), we replicated these findings and showed their generalisation across stimuli gender (Figure 11): a difference between human and non-human threatening contexts was found when participants had to rate both male and female target individuals (i.e., real faces taken from the CFD) on trustworthiness (Figure 10). Thus, across three experiments we showed that relational encoding matters more than the actual threat carried by the context stimuli in determining the attribution of (un)trustworthiness to target individuals: When the threatening scenes were potentially attributable to the individual, then the impact of the context showed stronger. The fact that human (as opposed to non-human) threatening contexts led to higher untrustworthiness does not merely prove that the

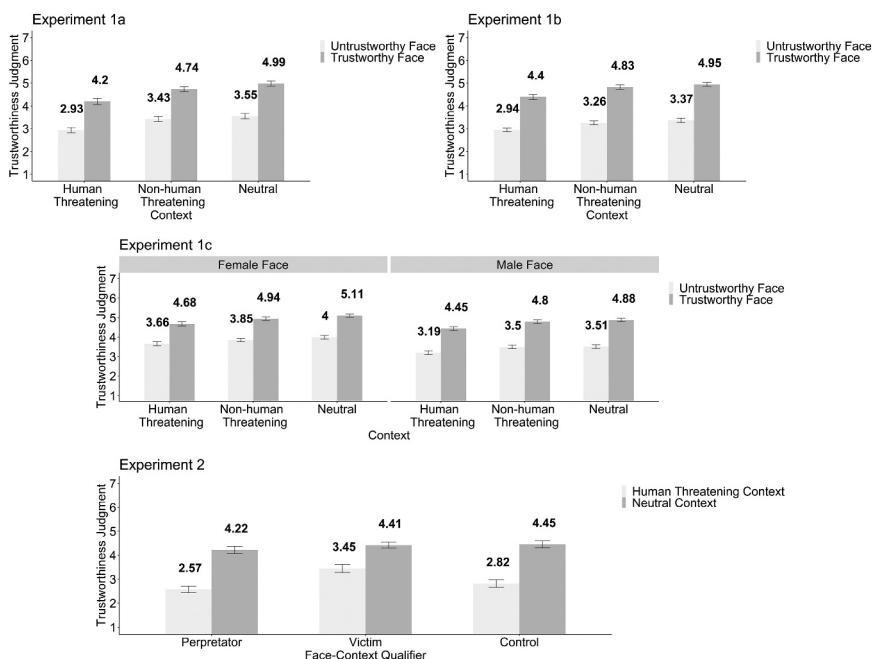


Figure 10. Individuals are judged as less trustworthy when their faces were shown in threatening contexts that were ascribable (vs. non-ascribable) to the human action (experiments 1a-1c). When instructions presented facial stimuli as belonging to the “perpetrators” of the threatening contexts, no difference with the neutral (no-instructions) condition was found in face – context integration (experiment 2). The effect was reduced when faces were presented as “victims.” lines represents 95% confidence intervals. The mean of each condition is reported above each bar. Adapted from Mattavelli, Masi, et al. (2023).

relationship between the individual and the context matters; it also suggests that perceivers encode such a relationship in a specific way.

In a final experiment ($N = 131$), we explored directly (i) how such a relationship is encoded and (ii) the consequence of altering the relational qualifier linking the target’s face to the surrounding context for their integration. Two hypotheses inspired our research questions. First, when processing a person’s face within a human threatening scenario, participants might tend to assume that the target face belongs to the person who is actively responsible for that conveyed threat (i.e., the perpetrator). Second, when additional information is available to disconfirm this default assumption, then face-context integration could be altered. We employed verbal instructions that informed participants about the nature of the relationship between the person’s face and the context, a manipulation that has been used to show the relational nature of evaluative conditioning (e.g., Fiedler & Unkelbach, 2011). In a 3 (face-context qualifier: perpetrator vs. victim vs.



Figure 11. Examples of untrustworthy-looking (a) and trustworthy-looking (b) female individuals extracted from the CFD, and of an untrustworthy-looking individual embedded in human threatening (c), non-human threatening (d), or neutral (e) contexts. Adapted from Mattavelli, Masi, et al. (2023) – experiment 1c.

control) \times 2 (context: neutral vs. human threatening) mixed design (first factor manipulated between participants) participants judged the perceived trustworthiness of target individuals represented by neutral (neither trustworthy- nor untrustworthy-looking) faces borrowed from the CFD presented in either neutral (i.e., grey rectangle) or human threatening scenarios. Crucially for the purpose of the experiment, one-third of the participants were priorly instructed that the faces appearing in threatening contexts belonged to the perpetrator; one-third of participants were informed that the very same faces belonged to the victim; one-third (control group) received no instructions. We found a main effect of the type of context, with lower attribution of trustworthiness for threatening as compared with neutral contexts. The effect of face-context qualifier was also significant. Crucially, there was a significant interaction between the two factors. Exploring this interaction revealed two important results. First, no difference emerged in the effect of the context (i.e., neutral-threatening) when comparing the perpetrator and the control condition (Figure 10). Thus, we showed that, by default, participants tended to perceive the person presented in a human threatening context as the perpetrator of the threatening scene. Second, the effect of the context on attributed trustworthiness was significantly stronger in the perpetrator than in the victim (and control) condition. Inducing participants to believe that the target face belonged to the victim of the threatening scene reduced the negative effect of the context on attributed trustworthiness.

Importantly, our relational manipulation reduced, but did not cancel, the effect of contextual threat on attributed trustworthiness: even when faces were presented as belonging to the victims of the threatening scene shown in the background, a significant (and negative) effect of the threatening (as opposed to neutral) context was found. There might be different reasons why participants maintained a tendency to judge target individuals presented as victims of the threatening context as less trustworthy than those presented in a neutral background. First, as much as instructions can rewrite a default relationship between the target and the context, one should consider that such instructions are opposed to a long history of learning, suggesting that target individuals appearing in a threatening background are likely the perpetrators, not the victims. Under such conditions, instructions might suffice to weaken, but not to cancel what has been priorly and consistently acquired via learning experiences. Second, knowing that a target is the victim of the threatening scenario does not guarantee that this person is trustworthy, especially when compared to another target presented in a neutral scenario. If one follows a relational approach to face-context integration, then it is still plausible to assume that participants found reasons to judge the victims of a threatening scene as less trustworthy than another random individual presented in a neutral scene. For instance, one might think that trustworthy people are less likely to be involved in threatening situations, regardless of their role in those situations.

Overall, this set of studies proved the relational nature of the face-context integration effect. We demonstrated that participants were more likely to integrate contextual threat in their trustworthiness attribution when such a threat was potentially ascribable to the human. It was when contextual threat justified inferences on the real nature of the person that face-context integration was maximised. In other words, face-context integration on attribution of trustworthiness is more than the sum of information conveyed by the two types of stimuli. When such information is qualified relationally, the negative impact of the threatening context on attributed untrustworthiness is maximised.

Face-context integration: the role of emotions

Mattavelli, Masi, et al. (2023) showed that relational encoding matters in face-context integration when relational information is provided explicitly via verbal instructions. Nevertheless, such explicit information is often absent in real-world encounters with individuals. Thus, it remains unknown whether people can extract relational information from other cues (such as non-verbal cues). A final set of studies delves into this intriguing question by exploring inferences drawn from facial emotions displayed by target individuals in threatening contexts.

Beyond extending the notion of relational encoding to non-verbal cues, the set of studies presented in this paragraph also sheds light on another issue raised by previous studies. In the investigation of face-context integration on person trustworthiness, the studies presented so far elucidated the significant roles of both the manipulated contextual factors and facial features. In other words, judgements about an individual's trustworthiness are influenced by (i) how trustworthy a person looks and (ii) the threat conveyed by the surrounding context. However, these two effects appear to function independently, as no conclusive evidence has yet emerged to support an interaction between contextual and facial features. This lack of interaction might indicate that facial trustworthiness and contextual threat are indeed two separate non-intersecting cues that inform perceivers about a person's trustworthiness. However, observing that faces pre-selected to appear trustworthy or untrustworthy were similarly affected by the impact of the context does not preclude the possibility that other facial features interact with contextual cues.

Previous studies examining face-context integration focused on morphological facial features (e.g., Mattavelli, Masi, et al., 2023, Experiments 1a-1c). These features encompass the structural characteristics of an individual's face, such as the arrangement of eyes, nose, mouth, and the overall facial symmetry. Remarkably, morphological features tend to remain relatively stable across time and situations. Thus, judgements on person trustworthiness based on morphological features should be less prone to alteration in response to situational information, including those activated by a visual context. Differently from morphological features, emotional expressions provide valuable insights into the momentary internal states of the target (De Gelder, 2006; Frijda, 1986). For instance, processing a smiling face often leads to the inference that the person is experiencing happiness. However, the interpretation of such expressions is deeply intertwined with the context in which they occur. The same smiling expression can signify genuine happiness in one context, while in another context, it might mask discomfort or irony. The dynamic nature of emotional interpretation highlights the nuanced relationship between emotional expressions and context thus suggesting a potential interplay between the two (for a similar argumentation, see Bublatzky et al., 2020)

To extend this understanding, we conducted two experiments that explored whether the attribution of trustworthiness from faces presented in either threatening or neutral contexts would differ based on the conveyed emotional expression – namely, happiness or fear (Mattavelli, Fiamberti, et al., 2023). Our hypothesis posited that facial emotional expressions would interact with contextual threat to modulate judgements of trustworthiness. Given that expressing happiness (as opposed to fear) in

threatening circumstances could be interpreted as a signal of malicious intent (in contrast to personal discomfort), we anticipated that trustworthiness judgements would lean more negatively towards happy faces compared to fearful faces within threatening, but not neutral, contexts. Importantly, such findings would provide stronger support for the relational encoding account of face-context integration.

In the first experiment ($N = 93$), participants saw a series of faces appearing on screen. Faces were embedded in a visual context that could be either threatening or neutral. We employed 16 face identities borrowed from the CFD (Ma et al., 2015) varying two emotional expressions, that is, fear and happiness. As threatening stimuli, we used four pictures used in Mattavelli, Masi, et al. (2023, Experiment 2) plus four other pictures of scenes pre-tested to be threatening. A grey rectangle was used in the neutral context condition (Figure 12). In each trial, participants evaluated a person's trustworthiness, using a 7-points scale (1 = untrustworthy to 7 = trustworthy). We found a main effect of the type of context, indicating that faces appearing in threatening contexts (vs. neutral contexts) were perceived as less trustworthy. There was no main effect of facial emotion. Critically, the interaction between context and emotional expression was significant. Decomposing this interaction revealed that when faces were presented in threatening



Figure 12. Examples of happy (a) and fearful (b) individuals, and of a happy (c) and fearful (d) individuals embedded in a threatening context. Adapted from Mattavelli, Fiamberti, et al. (2023) experiment 1.

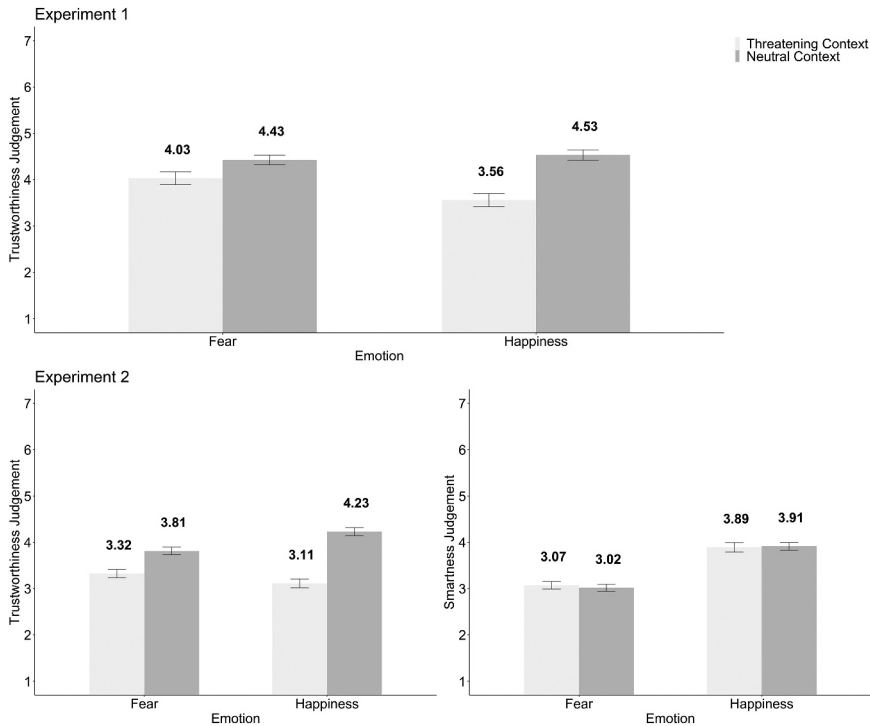


Figure 13. Individuals displaying happy faces were judged less trustworthy than those displaying fearful faces when presented in threatening contexts. The perceived trustworthiness of fearful and happy target faces was comparable when presented in neutral contexts (experiment 1). The manipulation did not impact upon smartness judgements (experiment 2). Lines represent 95% confidence intervals. The mean of each condition is reported above each bar. Adapted from Mattavelli, Fiamberti, et al. (2023).

contexts, happy individuals were judged as significantly less trustworthy than fearful individuals, whereas no difference in the perceived trustworthiness of fearful and happy individuals emerged in neutral contexts (Figure 13).

This finding is consistent with the idea that face-context integration depends on inferences regarding the role of the target individual in the surrounding context (Mattavelli, Masi, et al., 2023). Whereas fear is an emotional expression justified by contextual threat, displaying happiness in threatening circumstances is often interpreted as a signal of bad intent or cruelty, traits that are commonly associated with wrongdoers. Moreover, it shows that such inferences can be made by perceivers by using emotional information conveyed by the face. Nevertheless, an alternative explanation for the observed interaction was plausible. Namely, when faces and contexts are treated as pairs of stimuli, shared conceptual aspects between the face and the context might facilitate the cognitive processing of these pairings. In

other words, more congruent stimuli make that final stimuli pair more fluent. Previous research has demonstrated that heightened perceptual fluency leads to an amplification in stimulus evaluation (Reber et al., 1998). Within this experimental set-up, congruent face-context pairs might have elicited a more positive evaluation on the ultimate criterion, that is, trustworthiness. To clarify this amplification account, since threatening contexts should exhibit a higher degree of overlap with fearful faces as opposed to happy faces, the decreased trustworthiness attributed to individuals displaying happy faces could potentially stem from their disfluent processing (see also Righart & De Gelder, 2006). A second experiment was designed to test whether the observed effect was specific for trustworthiness or, alternatively, if the amplification in judgement generalised on an alternative dimension, that is, smartness.

In a second experiment ($N = 132$), we compared the interaction between context and emotions on two different types of attribution: trustworthiness vs. smartness. The introduction of this additional factor was justified by the hypothesis that, if the critical interaction depended solely on a fluency-driven amplification in judgement, then it should not matter whether the ultimate criterion is either relevant (i.e., trustworthiness) or irrelevant (i.e., smartness); more positive judgements for congruent stimuli should occur anyway. Conversely, an inferential account predicts that an interaction between context and emotional expression would occur only when stimuli are judged on a criterion for which the inferred relationship between the target and the context holds significance. For instance, inferring that a happy individual is the perpetrator in the surrounding threatening context holds greater significance for judgements of trustworthiness as opposed to smartness. Furthermore, in this experiment, we delved into the nature of such inferences: in a role attribution task, participants judged the likelihood with which happy vs. fearful target faces presented in threatening contexts belonged to a criminal. By correlating this role-attribution with that on both trustworthiness and smartness, we could unveil the inferential nature of face-context integration effect on person trustworthiness.

The procedure consisted of two blocks of 48 trials each. One block was formally identical to that described in Experiment 1. In the other block, participants rated each target individual – either happy or fearful in threatening or neutral contexts, on smartness (1 = not at all smart to 7 = extremely smart). The order of administration of the two blocks was counterbalanced across participants, and stimuli were administered in random order. Next, participants completed a role attribution task. In this task (16 trials), the same facial stimuli seen in the previous tasks were presented in threatening contexts only. Upon presentation of each face-context pair, participants indicated, on a 7-points Likert scale, how likely they believed that the target face belonged to a criminal (1 = not at all to 7 = definitely).

A three-way ANOVA tested whether the interaction between emotional expression and context varied as a function of the dispositional attribution required by the task (i.e., trustworthiness vs. smartness; [Figure 13](#)). We found the main effect of the type of context (i.e., more positive judgements in case of neutral contexts), facial emotion (i.e., more positive judgements for happy faces), and dispositional attribution (i.e., more positive judgements on trustworthiness). The interaction between attribution and facial emotion was significant, indicating that individuals displaying happiness were judged as smarter than those displaying fear, whereas no difference emerged on attributed trustworthiness. Also significant was the interaction between dispositional attribution and context, as well as that between facial emotion and context. Central to our research question, the three-way interaction was significant. Decomposing this interaction unveiled the replication of the interaction between facial emotion and context within the domain of trustworthiness. Specifically, participants ascribed lower trustworthiness to individuals displaying happy faces (in comparison to fearful faces) when presented within threatening contexts, while the reverse trend was observed in non-threatening contexts. Notably, the interaction between facial emotion and context did not yield statistical significance on smartness. An analysis of the impact of facial expressions on the inferred likelihood of being a criminal unveiled a significant effect: Happy individuals in threatening contexts were more likely judged as criminal than fearful individuals. Further reinforcing an inferential explanation, we found a significant correlation between the difference in face context integration on attributed probability of being a criminal in threatening contexts and the difference (i.e., fearful vs. happy faces) in face-context integration on attributed trustworthiness. Instead, no correlation emerged when we considered the difference in face context integration on attributed smartness.

Thus, this second experiment proved the specificity of the interaction between facial emotion and ruled-out a mere amplification interpretation, wherein a fluent experience with the processed stimuli led to more positive judgement on any criterion. Assessing the inferences made by participants on the role played by the target in the threatening context corroborated the idea raised by Mattavelli, Masi, et al. ([2023](#)) that this effect might be explained by relational encoding.

Overall, this set of studies reveal that attributions of trustworthiness to target individuals are moderated in concert by facial expressions and contextual information. We confirm that face-context relational encoding is key to determining whether a target person is worthy of trust. Importantly, relational encoding could change by simply altering the type of emotional expressions that target stimuli exhibited in threatening contexts. These findings speak for the active role played by perceivers in person perception and for their tendency to use and integrate available

cues to construe meaningful relationships between facial and context stimuli.

Conclusion

Summary of findings and implications

Trustworthiness inferences from others' faces are often inaccurate. Nevertheless, they impact interactions and important decisions (Ames et al., 2011). Whereas past research indicated that trustworthiness attributions stem from facial features (Jaeger et al., 2019; Slepian et al., 2012; Todorov et al., 2008, 2015), the role of contextual information in influencing such attributions has remained largely unexplored. This article reviews studies that explored the influence of context on trustworthiness attributions when individuals are assessed based on their facial appearance – namely, a face-context integration effect. In doing so, we considered both real and computer-generated faces and a wide range of contextual stimuli.

A first series of studies (Brambilla et al., 2018) showed that the visual context in which an individual's face is encountered has a significant influence on the categorisation of target individuals as either trustworthy or untrustworthy. Employing a mouse tracking technique, we found that untrustworthy-looking individuals were more easily categorised as such (i.e., more direct mouse trajectory) when surrounded by threatening visual, as opposed to both neutral and negative, contexts. By contrast, threatening backgrounds disrupted the categorisation of trustworthy-looking individuals. These findings highlight the importance of face-context integration in shaping categorisation of target individuals' trustworthiness, in conditions of conceptual congruency between the dimensions conveyed by the context and the face (i.e., threat and trustworthiness, respectively).

A second set of studies (Brambilla et al., 2021a) extended these findings by exploring face-context integration in a cross-modal fashion. Moreover, rather than on categorisation, we focused on trustworthiness judgements: participants rated individuals on trustworthiness based on their facial features by means of a Likert scale. Notably, individuals were judged as more untrustworthy when their faces were presented in threatening, rather than non-threatening, auditory contexts. Confirming earlier findings on the categorisation of target trustworthiness within visual contexts, this line of work clarified that face-context integration was unique of threatening auditory contexts, as indexed by the lack of impact of negative contexts.

These two sets of studies highlighted three important findings. Firstly, we found evidence that assessment of a target person's trustworthiness based on their facial appearance is malleable. Just like a face alone cannot provide a complete understanding of emotions in others (Wieser & Brosch, 2012),

the same holds when it comes to evaluating whether the target person is worthy of trust. Whether participants categorised a target as either trustworthy or untrustworthy or rated trustworthiness on a continuous scale, our results revealed that trustworthiness inferred from faces is vulnerable to the nature of the context. Secondly, the effect of contextual information in qualifying the level of trustworthiness attributed to a target person based on their facial cues is robust and extends across different sensory modalities (Brambilla et al., 2021a). Most studies focusing on the joint impact of facial and auditory stimuli on person perception are typically restricted to voices (for a review, see Campanella & Belin, 2007). Under such conditions, the auditory (e.g., voice) stimulus is inherently tied to targeted human identity. Using auditory stimuli that could not be attributable to the targeted identity, we showed that trustworthiness judgements of others are not just a matter of *who is seen* at but also of *what is heard*. Lastly, contextual information altered the attribution of trustworthiness from facial cues, but this was exclusive to contexts conveying threat, but not those manipulated to be negative. Hence, congruency in the affective valence of facial and contextual information could not explain face-context integration on trustworthiness judgements. Rather, our findings highlighted that it is the conceptual congruency between the dimensions of trustworthiness and threat that ultimately matters (for a review, see Brambilla & Leach, 2014; Brambilla et al., 2021a).

In a third set of studies, we delved into the specificity of congruency in affecting face-context integration (Mattavelli et al., 2022). In three experiments, we showed that the absolute (negative) impact of threatening contexts on trustworthiness attributed to untrustworthy-looking individuals showed stronger than the (positive) impact of reassuring contexts on trustworthiness attributed to trustworthy-looking individuals. These findings unveil the differential influence of threatening and reassuring contexts on trustworthiness attributions, indicating that face-context congruency cannot fully explain variations in attributed trustworthiness. Importantly, our studies clarified that the superior integration observed on the negative end of the threat-trustworthiness binomial persisted when threatening and reassuring context stimuli were selected to be equally related to human actions. Moreover, the higher effect of congruency for negative face-context compounds did not generalise to another domain of judgement, proving that this negativity bias is specifically tied to the domains of trustworthiness and threat.

In confirming that the dimensions of trustworthiness and threat are inherently linked (Brambilla & Leach, 2014; Brambilla et al., 2021a), Mattavelli et al. (2022) took one step further and unveiled that such a link is not ubiquitous along the threat-trustworthiness continuum. Our findings challenge a pure conceptual congruency account of face-context integration. Rather, they suggest that perceivers might actively

incorporate additional information to qualify the link between contextual threat and the target individual. In another set of studies (Mattavelli, Masi, et al., 2023) we tested the role of relational qualifiers. Namely, we hypothesised that relational encoding would moderate face-context integration. This was tested in two ways. First, in three separate experiments, we compared the impact of human-related (e.g., a room covered with blood) vs. non-human-related (e.g., a tornado) contextual threat and measured their differential impact on trustworthiness attributions based on an individual's face. Despite the superior (normative) threat conveyed by non-human-related context stimuli, face-context integration showed significantly stronger when the contextual threat was attributable to the human action. We interpreted this finding as indicative of a default tendency in perceivers to establish a meaningful relationship between the context and the target, with the nature of such a relationship varying according to the nature of the contextual threat. For instance, whereas a person could be seen as the victim in both human- and non-human-related threatening scenes, they could be seen as the perpetrator only in the former case. A final experiment directly tested this hypothesis. Using verbal instructions, we manipulated the nature of the relationship between the target and the context. Participants judged an individual's trustworthiness when their faces were presented in either neutral or (human-related) threatening contexts, with instructions indicating that the facial stimuli belonged to either the perpetrator or the victim of the threatening scenarios. A third (i.e., no-instructions) condition was added as a control. A significant interaction between the type of context stimuli (threatening vs. neutral) and instructions emerged. The non-significant difference in the effect of the threatening (vs. neutral) context observed when comparing the perpetrator and the control condition confirmed the hypothesised default tendency to perceive the person in a human-related threatening context as the perpetrator of the threat. Moreover, the negative impact of contextual threat on attributed trustworthiness was significantly reduced when the target was presented as the victim of the contextual threat.

This set of studies builds upon earlier findings by unveiling one underlying mechanism of face-context integration. We found that face-context integration on attribution of trustworthiness is, at least partly, relational. It was not the mere threat conveyed by the contextual stimulus that influenced the attribution of trustworthiness; rather, the extent to which such a threat justified further inferences regarding the role of the target individual. These findings confirmed prior evidence outside the realm of face perception, showing that relational reasoning is key in moderating the transfer of properties from one stimulus to another presented in spatio-temporal contingency (Fiedler & Unkelbach, 2011; Hughes et al., 2019; Moran & Bar-Anan, 2013).

While Mattavelli, Masi, et al. (2023) emphasised the importance of relational encoding when provided through verbal instructions, the final set of studies reviewed in this article explores this question in depth by investigating whether people can derive relational information also from non-verbal cues. Hence, we explored the effect of facial emotions, building on the notion that emotions can inform perceivers on how a target individual feels in a specific situation (De Gelder, 2006; Frijda, 1986). Our two studies (Mattavelli, Fiamberti, et al., 2023) showed that, just like instructions, emotions are used as valuable cues from which perceivers form inferences about the relationship between the target and the contextual threat. In a first experiment, happy individuals in threatening contexts were perceived as less trustworthy than fearful individuals. This finding underscored the role of relational encoding in face-context integration and suggested that expressing happiness in a threatening context might be perceived as a signal of malicious intent. A second experiment explored the specificity of this interaction by measuring attributions on trustworthiness and smartness. The results revealed a significant interaction between facial emotions and contextual threat only within the domain of trustworthiness, confirming the specificity of this effect. Participants' inferences about the role played by targets in threatening contexts supported the importance of relational encoding in this effect. Beyond extending the notion of relational encoding to non-verbal cues, this set of studies revealed that dynamic facial information is interpreted in the light of their interplay with contextual threat.

In summary, these findings highlight the malleability of trustworthiness attributions, qualified by contextual features. Moreover, malleable is also the integration of such contextual features into the ultimate attribution of trustworthiness. In fact, the strength of a face-context integration varied according to the type of relationship that can be established between the target and the context. This has implications also for a mechanistic understanding of face-context integration. Prior studies proposed that this integration is an inevitable process occurring at the earliest stage of stimuli processing in an automatic fashion (De Gelder et al., 2006; Meeren et al., 2005; Pourtois et al., 2000). Yet, saying that a phenomenon occurs under automaticity conditions (e.g., fast response, unawareness) does not necessarily imply that such a phenomenon is mediated by unqualified associations between concepts. In showing that face-context integration is moderated by propositional information that informs perceivers about the relationship between the target individual and the context, our studies suggest that this effect requires inferential reasoning and is not inevitable. We propose that face-context integration on attributed trustworthiness can be conceived as a consequence of an interplay between proximal and distal regularities in the environment (see De Houwer et al., 2013). Proximal regularities refer to the pairing of stimuli in the processing environment (i.e., a contextual threat

presented together with a target's face), while distal regularities are rules learned from past environments that qualify the meaning of the relationship between paired stimuli (i.e., a person that happens to be in a threatening scene is often the perpetrator).

Directions for future research

The research evidence presented in this article raises several future research possibilities. The first area that deserves further investigation pertains to the nature of the contexts in which trustworthy- and untrustworthy-looking target individuals are embedded. Individuals are often situated in contextual cues that can be processed under various sensory modalities. Our work shows that face-content integration occurs when target's faces are embedded in visual backgrounds (Brambilla et al., 2018; Mattavelli et al., 2022; Mattavelli, Fiamberti, et al., 2023; Mattavelli, Masi, et al., 2023). We also show that face-context integration occurs in a cross-modal fashion, by considering facial stimuli surrounded by auditory contexts (Brambilla et al., 2021a). While we demonstrated that face-context integration occurs regardless of whether the target's face and the context are processed in a unimodal (face-visual contexts) or a cross-modal (face-auditory contexts) fashion, unexplored is whether one of the two modalities exerts a stronger influence in biasing trustworthiness judgements. Similarly, it would be important to test how different sensory modalities integrate in affecting trustworthiness judgements. Expanding beyond visual and auditory contextual information, future research might explore the role of alternative sensory information, such as smell, on face-context integration. For instance, encountering an unfamiliar individual with the scent of burning or gunpowder is likely to evoke a significantly different perception compared to meeting them with a relaxing perfume aroma. Delving into these research questions would provide a deeper understanding of the relative contribution of specific contextual features in modulating face-context integration.

Another avenue for further research concerns a deeper analysis of the relative influence of contextual threat in face-context integration. The studies reviewed here were framed as investigations of person perception. Moreover, throughout the experimental procedures, participants received explicit instructions to judge the target person, while ignoring the context. This experimental gimmick might have downplayed the impact of context scenes in affecting trustworthiness attributions. Thus, one could ask whether face-context integration on attributed trustworthiness varies when higher emphasis is put on the context. One way to increase the salience of contextual stimuli might consist in the inclusion of an evaluation of contextual threat preceding that of the target person. Interestingly, this opens up the possibility of exploring face-context integration in a bidirectional fashion (see Hess et al., 2020 for an

example of bidirectional effect of facial emotion and the relevant context situation). Indeed, a fascinating possibility would be that face-context integration operates also in the other direction, where judgements of contextual threat are qualified by the nature of the facial stimulus (i.e., either a trustworthy- or an untrustworthy-looking). Future studies should assess whether face-context integration occurs bidirectionally and whether a change in attributed trustworthiness to target individuals presented in threatening contexts correlates with a change in attributed contextual threat.

New insights for future research stem from the role of relational encoding in qualifying face-context integration. We found that threatening contexts ascribable to human actions led to stronger face-context integration (Mattavelli, Masi, et al., 2023): Individuals were judged as less trustworthy when their faces were embedded in such contexts. Moreover, we observed that face-context integration was altered when the relationship between the face and the context was manipulated via written instructions informing that the face belonged to either the perpetrator or the victim of the surrounding context. Indeed, faces were judged as less trustworthy when they were told to belong to the perpetrator (as opposed to the victim) of the threatening surrounding context. A final set of studies (Mattavelli, Fiamberti, et al., 2023) further clarified that relational information can be actively extracted by perceivers from dynamic facial cues, such as emotional expressions, to ultimately determine trustworthiness attributions made on the target. Collectively, these findings clarified the role of relational information in face-context integration and suggest the intriguing idea of studying the influence of additional variables that might impact on the relationship that perceivers establish between the target and the context, especially when this phenomenon is examined outside a controlled experimental setting. In fact, a critical reader may imagine that face-context integration could reduce in real-life settings, where less control over other interfering variables is permitted. For instance, participants in our experiments might have paid extreme attention to the context scenes because of the way we presented our stimuli. In fact, it is possible that other cues can impact face-context integration by altering the nature of their relationship. One such cue is body movement or posture (see also Aviezer et al., 2008; Feldman- Barrett et al., 2011 for a similar argument on emotion recognition). In real-life situations, individuals are rarely processed as static entities within their context. We might see people running away or approaching a source of threat (e.g., shooting inside a shopping centre). All these cues have the potential to affect the inferences that we make about individuals processed in threatening contexts. Perhaps more importantly, all the reviewed studies employed facial stimuli that were visually juxtaposed to context stimuli of different nature (e.g., a room covered with blood). This approach, although artificial and non-natural, can effectively capture specific situations where individuals are exposed to targets presented

in context. For example, in news articles, individuals might be portrayed against a threatening backdrop. However, our method has limitations in revealing the dynamics of integration between a person and a context in more naturalistic settings. Further research could enhance our understanding by investigating how trustworthiness attributions to individuals are influenced by threatening contextual information in settings that mirror real-life scenarios, perhaps through virtual reality. This approach would yield stronger insights into the actual role of context in determining real-life decisions regarding trustworthiness inferences from faces. Additionally, these studies could also test whether other important mechanisms or individual differences, such as the ability to efficiently process multiple sources of information, may be implicated in driving face-contexts interactions in real-life scenarios.

The studies presented in the current review aimed at establishing the impact of information that is external to the individual on attributions made on the individual. This was done by employing measures that involved preestablished sets of responses. Participants categorised individuals as either trustworthy or untrustworthy via mouse movements (Brambilla et al., 2018), or rated individuals on a predefined trustworthiness scale (Brambilla et al., 2021a; Mattavelli et al., 2022; Mattavelli, Fiamberti, et al., 2023; Mattavelli, Masi, et al., 2023). Both methods constrain the ultimate criterion on which targets should be evaluated. Thus, our measures prevent any speculations on whether the effect we found at the judgement level would generalise at the perceptual level. For instance, Hassin and Trope (2000) proposed an important distinction between “reading from faces” and “reading into faces”. While “reading from faces” reflects the act of reading personality traits from facial features, “reading into faces” reflects the altered perception or attribution of facial features due to their known association with a relevant trait. For instance, one might assume that contextual threat influences what people see into the target’s face. A fascinating future perspective would be to transition from person’s trustworthiness categorisation and judgements to actual perception of trustworthiness in a person’s face. This could be done through data-driven methods that do not draw any prior assumption on the facial representation of the social target, such as the reverse correlation paradigm (Dotsch & Todorov, 2012; Dotsch et al., 2008; Mangini & Biederman, 2004). To illustrate, after seeing each face in either neutral or threatening contexts, participants could be shown with two noisy versions of the original face and asked to indicate which one appeared in a previous scene. This repeated choice task would yield two averaged classification images: one for faces in threatening contexts and another for faces in neutral contexts. Next, a separate sample of judges could evaluate these

two images on trustworthiness. If contextual threat alters the visual perception of faces, one would expect the two generated images to differ in trustworthiness (see also Rougier et al., 2021 for a similar procedure). Hence, we believe that future studies should benefit from alternative measures to better investigate the effect of face-context integration on the actual perception of trustworthiness.

Lastly, an important open question relates to the real-life implications of the current line of work. Whereas prior work has largely established that trustworthiness inferences from faces affect real-world decisions (e.g., Duarte et al., 2012; Wilson & Rule, 2015), the role of context has been poorly investigated. In our set of studies, our interest was on how contextual threat could influence perceived trustworthiness from faces. Yet, we did not explore whether contextual threat can affect perceiver's decisions. This question is especially relevant in the social media era. Social media platforms offer a fertile landscape for studying how decisions of different sorts can be influenced by the processing of individuals (represented by their faces) in context. In fact, social media platforms are used to decide whether a candidate should be recruited for a specific job position, or whether a person should be considered for a romantic appointment. Crucially, within such environments, *people* are invited to introduce themselves to others via static pictures that often display their faces with surrounding context stimuli. Exploring whether the simultaneous occurrence of a face and context stimuli perceived as either threatening (e.g., a roaring crowd at a heavy metal concert) or reassuring (e.g., a calm lake landscape) can bias perceivers' decisions would then be of high relevance. Thus, future research should consider exploring face-context integration in real-life settings to explore their impact on everyday decisions. In considering the real-life implications of face-context integration, future research should also tackle cultural differences in how faces in context shape social impressions. The studies reviewed in this article have been limited to European participants. A cross-cultural approach to studying how contextual cues influence impression formation from faces could provide a more comprehensive picture. Prior work on facial emotions, for instance, has shown that people from East Asia are more prone to incorporate contextual information when asked to evaluate the central person's facial expression, compared to their Western counterparts (Masuda et al., 2008). Building on these findings, it becomes important to understand whether the bond between (facial) trustworthiness and (contextual) threat varies cross-culturally. This work could also explore whether environmental characteristics, such as the frequency of environmental disasters or criminal acts, amplify or diminish face-content integration when people are asked to judge someone from their face.

Data availability statement

Supplementary materials, data, and analysis code are available on Open Science Framework for each study described in the article.

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.

ORCID

Marco Brambilla  <http://orcid.org/0000-0002-4774-3309>

Simone Mattavelli  <http://orcid.org/0000-0002-8934-8016>

Matteo Masi  <http://orcid.org/0000-0001-6436-1663>

References

- Aguado, L., Dieguez-Risco, T., Villalba-García, C., & Hinojosa, J. A. (2019). Double-checking emotions: Valence and emotion category in contextual integration of facial expressions of emotion. *Biological Psychology*, 146, 107723. <https://doi.org/10.1016/j.biopsycho.2019.107723>
- 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. <https://doi.org/10.1093/oxfordhb/9780195342161.001.0001>
- Aviezer, H., Bentin, S., Dudarev, V., & Hassin, R. R. (2011). The automaticity of emotional face-context integration. *Emotion*, 11(6), 1406–1414. <https://doi.org/10.1037/a0023578>
- Aviezer, H., Hassin, R. R., Ryan, J., Grady, C., Susskind, J., Anderson, A., Moscovitch, M., & 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>
- Barrett, L. F., Mesquita, B., & Gendron, M. (2011). Context in emotion perception. *Current Directions in Psychological Science*, 20(5), 286–290. <https://doi.org/10.1177/0963721411422522>
- Bradley, M. M., & Lang, P. J. (2007). The international affective digitized sounds (2nd ed., IADS-2): Affective ratings of sounds and instruction manual. Technical report B-3. University of Florida

- 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., Sacchi, S., Pagliaro, S., & Ellemers, N. (2013). Morality and intergroup relations: Threats to safety and group image predict the desire to interact with outgroup and ingroup members. *Journal of Experimental Social Psychology*, 49(5), 811–821. <https://doi.org/10.1016/j.jesp.2013.04.005>
- Brambilla, M., Sacchi, S., Rusconi, P., Cherubini, P., & Yzerbyt, V. Y. (2012). You want to give a good impression? Be honest! Moral traits dominate group impression formation. *British Journal of Social Psychology*, 51(1), 149–166. <https://doi.org/10.1111/j.2044-8309.2010.02011.x>
- Brambilla, M., Sacchi, S., Rusconi, P., & Goodwin, G. (2021a). The primacy of morality in impression development: Theory, research, and future directions. *Advances in Experimental Social Psychology*, 64, 187–262. <https://doi.org/10.1016/bs.aesp.2021.03.001>
- 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, 315–327.
- Brinol, P., Petty, R. E., & McCaslin, M. J. (2009). Changing attitudes on implicit versus explicit measures: What is the difference? In R. E. Petty, R. H. Fazio, & P. Brinol (Eds.), *Attitudes: Insights from the new implicit measures* (pp. 285–326). Psychology Press.
- Bublitzky, F., Kavcıoğlu, F., Guerra, P., Doll, S., & Junghöfer, M. (2020). Contextual information resolves uncertainty about ambiguous facial emotions: Behavioral and magnetoencephalographic correlates. *Neuroimage: Reports*, 215, 116814. <https://doi.org/10.1016/j.neuroimage.2020.116814>
- Campanella, S., & Belin, P. (2007). Integrating face and voice in person perception. *Trends in Cognitive Sciences*, 11(12), 535–543. <https://doi.org/10.1016/j.tics.2007.10.001>
- 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. *The Adapted Mind: Evolutionary Psychology and the Generation of Culture*, 163, 163–228.
- De Gelder, B. (2006). Towards the neurobiology of emotional body language. *Nature Reviews Neuroscience*, 7(3), 242–249. <https://doi.org/10.1038/nrn1872>
- De Gelder, B., Meeren, H. K., Righart, R., Van den Stock, J., Van de Riet, W. A., & Tamietto, M. (2006). Beyond the face: Exploring rapid influences of context on face processing. *Progress in Brain Research*, 155, 37–48. [https://doi.org/10.1016/S0079-6123\(06\)55003-4](https://doi.org/10.1016/S0079-6123(06)55003-4)
- De Houwer, J. (2007). A conceptual and theoretical analysis of evaluative conditioning. *The Spanish Journal of Psychology*, 10(2), 230–241. <https://doi.org/10.1017/S1138741600006491>
- 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. (2014). A propositional model of implicit evaluation. *Social and Personality Psychology Compass*, 8(7), 342–353. <https://doi.org/10.1111/spc3.12111>
- 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., Barnes-Holmes, D., & Moors, A. (2013). What is learning? On the nature and merits of a functional definition of learning. *Psychonomic Bulletin & Review*, 20(4), 631–642. <https://doi.org/10.3758/s13423-013-0386-3>
- 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>
- Dotsch, R., & Todorov, A. (2012). Reverse correlating social face perception. *Social Psychological and Personality Science*, 3(5), 562–571. <https://doi.org/10.1177/1948550611430272>
- Dotsch, R., Wigboldus, D. H., Langner, O., & Van Knippenberg, A. (2008). Ethnic out-group faces are biased in the prejudiced mind. *Psychological Science*, 19(10), 978–980. <https://doi.org/10.1111/j.1467-9280.2008.02186.x>
- Duan, Y., Hsieh, T. S., Wang, R. R., & Wang, Z. (2020). Entrepreneurs' facial trustworthiness, gender, and crowdfunding success. *Journal of Corporate Finance*, 64, 101693. <https://doi.org/10.1016/j.jcorpfin.2020.101693>
- 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), 2455–2484. <https://doi.org/10.1093/rfs/hhs071>
- Ert, E., Fleischer, A., & Magen, N. (2016). Trust and reputation in the sharing economy: The role of personal photos in Airbnb. *Tourism Management*, 55, 62–73. <https://doi.org/10.1016/j.tourman.2016.01.013>
- 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>
- Foo, Y. Z., Sutherland, C. A. M., Burton, N. S., Nakagawa, S., & Rhodes, G. (2022). Accuracy in facial trustworthiness impressions: Kernel of truth or modern physiognomy? A meta-analysis. *Personality and Social Psychology Bulletin*, 48(11), 1580–1596. <https://doi.org/10.1177/01461672211048110>
- Freeman, J. B., & Ambady, N. (2010). MouseTracker: Software for studying real-time mental processing using a computer mouse-tracking method. *Behavior Research Methods*, 42(1), 226–241. <https://doi.org/10.3758/BRM.42.1.226>
- Freeman, J. B., & Ambady, N. (2011). A dynamic interactive theory of person construal. *Psychological Review*, 118(2), 247–279. <https://doi.org/10.1037/a0022327>
- Freeman, J. B., & Johnson, K. L. (2016). More than meets the eye: Split-second social perception. *Trends in Cognitive Sciences*, 20(5), 362–374. <https://doi.org/10.1016/j.tics.2016.03.003>
- 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>
- Frijda, N. H. (1986). *The emotions*. Cambridge University Press.

- Gibson, J. J. (1977). The theory of affordances. In R. Shaw & J. Bransford (Eds.), *Perceiving, acting, and knowing* (pp. 67–82). Lawrence Erlbaum.
- Gibson, J. J. (1979). *The ecological approach to visual perception*. Houghton Mifflin.
- Hassin, R., & Trope, Y. (2000). Facing faces: Studies on the cognitive aspects of physiognomy. *Journal of Personality and Social Psychology*, 78(5), 837–852. <https://doi.org/10.1037/0022-3514.78.5.837>
- Hess, U., Dietrich, J., Kafetsios, K., Elkabetz, S., & Hareli, S. (2020). The bidirectional influence of emotion expressions and context: Emotion expressions, situational information and real-world knowledge combine to inform observers' judgments of both the emotion expressions and the situation. *Cognition and Emotion*, 34(3), 539–552. <https://doi.org/10.1080/02699931.2019.1651252>
- 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., Oud, B., Williams, T., Krumhuber, E. G., Fehr, E., & Engelmann, J. B. (2022). Can people detect the trustworthiness of strangers based on their facial appearance? *Evolution and Human Behavior*, 43(4), 296–303. <https://doi.org/10.1016/j.evolhumbehav.2022.04.004>
- Jaeger, B., Slegers, 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>
- Katzir, M., & Posten, A.-C. (2023). Are there dominant response tendencies for social reactions? Trust trumps mistrust—evidence from a dominant behavior measure (DBM). *Journal of Personality and Social Psychology*, 125(1), 57–81. <https://doi.org/10.1037/pspa0000334>
- Kramer, R. S., & Gardner, E. M. (2020). Facial trustworthiness and criminal sentencing: A comment on Wilson and rule (2015). *Psychological Reports*, 123, 1854–1868. <https://doi.org/10.1177/0033294119889582>
- 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>
- Leidner, B., & Castano, E. (2012). Morality shifting in the context of intergroup violence. *European Journal of Social Psychology*, 42(1), 82–91. <https://doi.org/10.1002/ejsp.846>
- Lin, C., Keles, U., & Adolphs, R. (2021). Four dimensions characterize comprehensive trait judgments of faces. *Nature Communications*, 12(1), 5168. <https://doi.org/10.1038/s41467-021-25500-y>
- 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>
- Macrae, C. N., & Martin, D. (2007). A boy primed sue: Feature-based processing and person construal. *European Journal of Social Psychology*, 37(5), 793–805. <https://doi.org/10.1002/ejsp.406>

- Mangini, M. C., & Biederman, I. (2004). Making the ineffable explicit: Estimating the information employed for face classifications. *Cognitive Science*, 28(2), 209–226. https://doi.org/10.1207/s15516709cog2802_4
- Masuda, T., Ellsworth, P. C., Mesquita, B., Leu, J., Tanida, S., & Van de Veerdonk, E. (2008). Placing the face in context: Cultural differences in the perception of facial emotion. *Journal of Personality and Social Psychology*, 94(3), 365–381. <https://doi.org/10.1037/0022-3514.94.3.365>
- Mattavelli, S., Fiamberti, G. C., Masi, M., & Brambilla, M. (2023). The “happy face killer” in the eyes of the beholder: Relational encoding of facial emotions in context influences trustworthiness attributions. *Journal of Experimental Social Psychology*, 109, 104517. <https://doi.org/10.1016/j.jesp.2023.104517>
- Mattavelli, S., Masi, M., & Brambilla, M. (2022). Untrusted under threat: On the superior bond between trustworthiness and threat in face-context integration. *Cognition and Emotion*, 36(7), 1273–1286. <https://doi.org/10.1080/02699931.2022.2103100>
- Mattavelli, S., Masi, M., & Brambilla, M. (2023). Not just about faces in context: Face–context relation moderates the impact of contextual threat on facial trustworthiness. *Personality and Social Psychology Bulletin*, 49(3), 376–390. <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>
- Meeren, H. K., van Heijnsbergen, C. C., & de Gelder, B. (2005). Rapid perceptual integration of facial expression and emotional body language. *Proceedings of the National Academy of Sciences*, 102(45), 16518–16523. <https://doi.org/10.1073/pnas.0507650102>
- Moran, T., & Bar-Anan, Y. (2013). The effect of object–valence relations on automatic evaluation. *Cognition and Emotion*, 27(4), 743–752. <https://doi.org/10.1080/02699931.2012.732040>
- Neta, M., Davis, F. C., & Whalen, P. J. (2011). Valence resolution of ambiguous facial expressions using an emotional oddball task. *Emotion*, 11(6), 1425–1433. <https://doi.org/10.1037/a0022993>
- Olivola, C. Y., Funk, F., & Todorov, A. (2014). Social attributions from faces bias human choices. *Trends in Cognitive Sciences*, 18(11), 566–570. <https://doi.org/10.1016/j.tics.2014.09.007>
- Oosterhof, N. N., & Todorov, A. (2008). The functional basis of face evaluation. *Proceedings of the National Academy of Sciences*, 105, 11087–11092. <https://doi.org/10.1073/pnas.0805664105>
- Porter, S., Ten Brinke, L., & Gustaw, C. (2010). Dangerous decisions: The impact of first impressions of trustworthiness on the evaluation of legal evidence and defendant culpability. *Psychology, Crime & Law*, 16(6), 477–491. <https://doi.org/10.1080/10683160902926141>
- Pourtois, G., De Gelder, B., Vroomen, J., Rossion, B., & Crommelinck, M. (2000). The time-course of intermodal binding between seeing and hearing affective information. *Neuroreport*, 11(6), 1329–1333. <https://doi.org/10.1097/00001756-200004270-00036>
- Reber, R., Winkielman, P., & Schwarz, N. (1998). Effects of perceptual fluency on affective judgments. *Psychological Science*, 9(1), 45–48. <https://doi.org/10.1111/1467-9280.00008>

- Rezlescu, C., Duchaine, B., Olivola, C. Y., Chater, N., & Rustichini, A. (2012). Unfakeable facial configurations affect strategic choices in trust games with or without information about past behavior. *Public Library of Science One*, 7(3), e34293. <https://doi.org/10.1371/journal.pone.0034293>
- Richeson, J. A., & Trawalter, S. (2005). On the categorization of admired and disliked exemplars of admired and disliked racial groups. *Journal of Personality and Social Psychology*, 89(4), 517–530. <https://doi.org/10.1037/0022-3514.89.4.517>
- Righart, R., & De Gelder, B. (2006). Context influences early perceptual analysis of faces—an electrophysiological study. *Cerebral Cortex*, 16(9), 1249–1257.
- 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>
- Rougier, M., Schmitz, M., & Yzerbyt, V. (2021). When my actions shape your looks: Experience-based properties of approach/avoidance bias the visual representation of others. *Journal of Personality and Social Psychology*, 120(5), 1146–1174. <https://doi.org/10.1037/pspa0000268>
- Rule, N. O., Krendl, A. C., Ivcevic, Z., & Ambady, N. (2013). Accuracy and consensus in judgments of trustworthiness from faces: Behavioral and neural correlates. *Journal of Personality and Social Psychology*, 104(3), 409–426. <https://doi.org/10.1037/a0031050>
- Rule, N. O., Slepian, M. L., & Ambady, N. (2012). A memory advantage for untrustworthy faces. *Cognition*, 125(2), 207–218. <https://doi.org/10.1016/j.cognition.2012.06.017>
- Slepian, M. L., Young, S. G., Rule, N. O., Weisbuch, M., & Ambady, N. (2012). Embodied impression formation: Social judgments and motor cues to approach and avoidance. *Social Cognition*, 30(2), 232–240. <https://doi.org/10.1521/soco.2012.30.2.232>
- 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., & Young, A. W. (2022). Understanding trait impressions from faces. *British Journal of Psychology*, 113(4), 1056–1078. <https://doi.org/10.1111/bjop.12583>
- Tamietto, M., Latini Corazzini, 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. (2008). Evaluating faces on trustworthiness: An extension of systems for recognition of emotions signaling approach/avoidance behaviors. *Annals of the New York Academy of Sciences*, 1124(1), 208–224. <https://doi.org/10.1196/annals.1440.012>
- 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., & Duchaine, B. (2008). Reading trustworthiness in faces without recognizing faces. *Cognitive Neuropsychology*, 25(3), 395–410. <https://doi.org/10.1080/02643290802044996>
- Todorov, A., & Oh, D. (2021). The structure and perceptual basis of social judgments from faces. *Advances in Experimental Social Psychology*, 63, 189–245. <https://doi.org/10.1016/bs.aesp.2020.11.004>

- 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>
- Todorov, A., Said, C. P., Engell, A. D., & Oosterhof, N. N. (2008). Understanding evaluation of faces on social dimensions. *Trends in Cognitive Sciences*, 12(12), 455–460. <https://doi.org/10.1016/j.tics.2008.10.001>
- Uleman, J. S., Newman, L. S., & Moskowitz, G. B. (1996). People as flexible interpreters: Evidence and issues from spontaneous trait inference. *Advances in Experimental Social Psychology*, 28, 211–279. [https://doi.org/10.1016/S0065-2601\(08\)60239-7](https://doi.org/10.1016/S0065-2601(08)60239-7)
- Van der Toorn, J., Ellemers, N., & Doosje, B. (2015). The threat of moral transgression: The impact of group membership and moral opportunity. *European Journal of Social Psychology*, 45(5), 609–622. <https://doi.org/10.1002/ejsp.2119>
- Wieser, M. J., & Brosch, T. (2012). Faces in context: A review and systematization of contextual influences on affective face processing. *Frontiers in Psychology*, 3, 471. <https://doi.org/10.3389/fpsyg.2012.00471>
- 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, J. P., & Rule, N. O. (2015). Facial trustworthiness predicts extreme criminal-sentencing outcomes. *Psychological Science*, 26(8), 1325–1331. <https://doi.org/10.1177/0956797615590992>
- 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>
- Wright, D. B., & Stroud, J. N. (2002). Age differences in lineup identification accuracy: People are better with their own age. *Law and Human Behavior*, 26(6), 641–654. <https://doi.org/10.1023/A:1020981501383>
- Xie, S. Y., Thai, S., & Hehman, E. (2023). Everyday perceiver-context influences on impression formation: No evidence of consistent effects. *Personality and Social Psychology Bulletin*, 49(6), 955–968. <https://doi.org/10.1177/01461672221085088>