PhD Program in Psychology, Linguistics, and Cognitive Neuroscience

DYNAMICS OF FACE-CONTEXT INTEGRATION:

HOW THREAT CUES INFLUENCE THE PROCESSING OF FACIAL TRUSTWORTHINESS

Marco Biella

Advisor: Dr. Marco Brambilla

Coordinator: Prof. Maria Teresa Guasti

DOCTORAL SCHOOL

University of Milano - Bicocca

Department of Psychology

PhD Program in Psychology, Linguistics, and Cognitive Neuroscience

Curriculum: Social, Cognitive, and Clinical Psychology

XXXI Cycle

Dynamics of Face-Context Integration:

How Threat Cues Influence the Processing of Facial Trustworthiness

PhD Candidate: **Marco Biella**

Reg. Number: 710299

Advisor: **Dr. Marco Brambilla**

Coordinator: **Prof. Maria Teresa Guasti**

ACADEMIC YEAR: 2017/2018

Table of Contents

Acknowledgements

With authorship comes merit, but merit is unfairly mapped onto authorship. My supervisor and I would not have been able to carry out half of this project without all trainee, bachelor's, and master's students who do not appear as author of this manuscript. We feel obliged to thank Alessia Macrì, Althea Frisanco, Andrea Beatrice Galeazzi, Angelica Marini, Arianna Castriota, Beatrice Leonardi, Carolina Quagliarella, Clara Caimi, Fiorella Carminati, Gloria Galimberti, Irene Sinigaglia, Johanna Baruffaldi-Preis, Luca Boiardi, Matteo Masi, Roberta Cortese, Sara Leoni, Sara Molteni, Serena Lauria, Serena Monaco for their contribution. I realized the importance of their work only now, while writing this acknowledgement. Therefore, the aforementioned list is composed by names that came to my mind and it is a non-exhaustive list.

All those students are surely grateful to our supervisor because of his guidance during their work, and I make no difference. My supervisor and I spent three years working together. We basically founded the lab from scratch, spent long hours working on our projects, and our joint venture allowed me to travel Europe as a researcher. It's such a big thing. By travelling Europe, I had the chance to meet wonderful people. People from the Social Cognition Center of Cologne who found the time to welcome me while doing science at its finest, fellow students from summer school with whom I had the chance to do "rocket science", and teacher from workshops who provided knowledge as gift.

Obviously, friends and family had a big part in this. Without their support difficult moments would have been too hard to handle and beautiful moments would have been under-celebrated. It has been a long journey and it would not have been the same without them. And then, there is Annalisa.

Abstract

Our impressions of others are often based on limited information that is spontaneously and automatically extracted from their faces. An important class of inferences concerns judgments of trustworthiness. As such, people start discriminating trustworthiness after 33ms of exposure to a face and the detection of trustworthiness in faces is faster than the detection of a variety of other characteristics, including dominance, likeability, and competence. People show a memory advantage for faces varying on trustworthiness compared with those varying on likeability, friendliness, and dominance and facial trustworthiness predicts basic approach/avoidance responses. In the vast majority of studies examining facial trustworthiness, faces are flashed on the computer screen, and categorization of trustworthiness quickly ensues. In other words, evaluation of facial trustworthiness is often thought to be based on facial features and relatively immune to context cues. However, we rarely encounter an isolated facial expression in the real world.

The present dissertation aimed at complementing and extending prior research evidence by investigating whether contextual information may impact the perception of facial trustworthiness. We conducted 8 main experiments and 5 pretests (N = 691) combining outcome based measures (i.e., ratings and explicit evaluations) with process-sensitive measures (i.e., mouse tracking). Results of each single experiment and the meta-analysis of the whole experimental data show that contextual threat information influences the evaluation of facial trustworthiness. We showed the specificity of threat information proving that the effects we found goes over and beyond negative information more in general. Thus, contextual threat information promoted the evaluation and categorization of facial untrustworthiness. By contrast, threatening contextual cues disrupted the processing of trustworthy faces. Moreover, our data suggested that such an integration occurs at a low cognitive level, in accordance with an evolutionary perspective, by showing that the integration is possible when information to be integrated lay in the same perceptual system but not when information is stored on different perceptual systems.

Taken together, our findings reveal the malleable nature of trustworthiness such that its perception is readily pushed around by scene context.

Chapter 1

General Introduction: Social Perception, Face Processing, and their Cognitive Integration with Threat Information

1.1 Introduction

A key skill needed for individual and group survival is the ability to safely navigate the environment (Fiske, 1992). Such an ability is clearly important in the scenario where our ancestors evolved, full of dangers and wild animals, and somehow it is still important nowadays. In the modern era, threats are of a different nature: the environment we have to navigate in is populated by different kind of threats. They are represented by other people we interact with and they pose a threat to our safety, to the reaching of our goals, to our image in the eyes of others (Dunbar, 2004), to our reputation and social position (Anderson, John, Keltner, & Kring, 2001) that are almost as dangerous as a wild predator back in the days. Even in ancient times this kind of threat was present and it has accompanied our evolution since then. In the past, the social nature of our species leveraged the interaction with others in order to foster survival (Fehr, Fischbacher & Gächter, 2002; Turchin, Currie, Turner & Gavrilets, 2013), but our social nature has some drawbacks concerning interaction with conspecifics. Because we have lived in groups since forever, clear group boundaries created competing groups and therefore threats (Ellemers, van Knippenberg, De Vries & Wilke, 1988; Lamont & Molnar, 2002). Our social environment navigating tools comprises specific interactions adaptions such as discriminating among others who is friend from who is foe, and in general who we can trust from who we cannot.

Social psychology has a long tradition of research on human interaction. We know very well the basic rules of human interaction such as the favoritism toward our own people (i.e., ingroup bias; Tajfel & Turner, 1979), or our natural bonding for those who are like us (Mc Pherson, Smith-Lovin & Cook, 2001). Moreover, it is well known how the individual navigate the social environment by putting instances into categories (Brewer, 1989;

Bruner, 1957; Rosch & Lloyd, 1978; Smith & Medin, 1981). Categorization is a basic and cost-effective mechanism that play a pivotal role in simplifying and making manageable the social environment (Allport, 1954; Fiske & Taylor, 1984; Lippmann, 1922; Tajfel, 1969). The contribution of the present dissertation is providing further insights regarding the latter phenomena, the ability to extract information from a social target. Specifically, we aim at understanding whether cues in the environment may impact the perception of other social agents. We aim at answering the question: "Is the same agent perceived independently from the context"? In so doing, we rely on classical and novel experimental tasks. Furthermore, we aim at exploring at which cognitive level information from the environment and from the social target are integrated.

In the following pages, we present a general review of models that explain how individuals form impressions of others and how they retrieve information needed to regulate social interaction. Such a review is mainly focused on the interpersonal level of interaction. Next, we present theories on social perception from faces, that are a primary source of social information in the environment (Todorov & Oosterhof, 2011; Todorov, Said, Engell & Oosterhof, 2008). We focus on pointing out what is widely known about face perception and which are the open issues, that we address in the following chapters.

1.2 Impression Formation and Person Perception

In psychological literature, the study of perceiving and forming impressions of others has a long tradition. An early yet prominent theory of impression formation was proposed by Solomon Asch (1946). In a series of experiments, Asch presented his experimental subjects with lists of personality characteristics. The author aimed to observe how individuals form an impression of the person described. The results of this work led to elaborate a linear model in which every characteristic impacted the final impression formed. According to this model, traits are combined and interact together as they become available until the impression is formed (Figure 1.1).

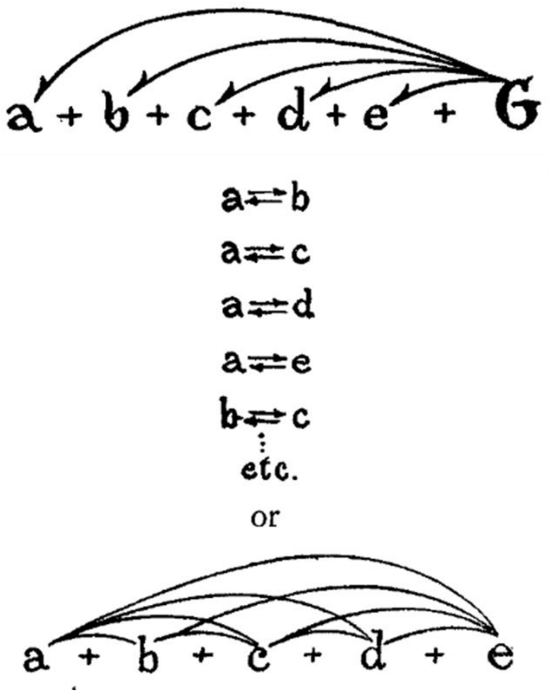

Figure 1. 1 Interactions among unitary traits in impression formation (retrieved from Asch, 1946)

From his examination, Asch theorized that forming an impression of another person is an organized process in which the perceiver needs to create "an impression of an *entire* person" (p. 284) (Figure 1.2).

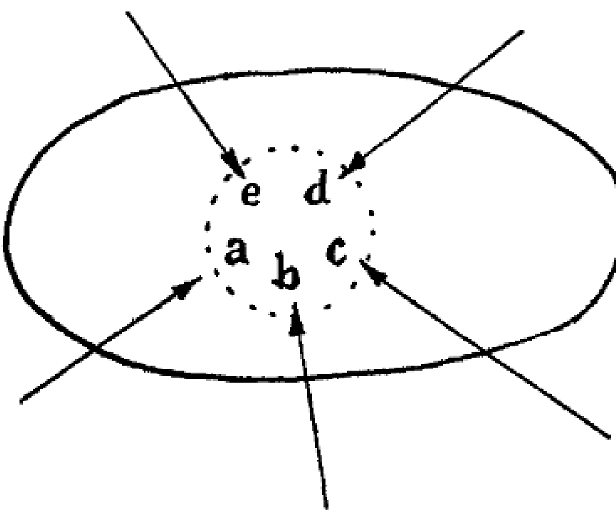

Figure 1. 2 Graphical representation of a coherent impression of a whole person (retrieved from Asch, 1946)

In his work we can find a coherence principle that guides the impression formation process. Traits interact with each other, they are interpreted according with the other characteristics in the list and influence the meaning of the following ones in order to not create contradictions. These dynamics of interaction among different pieces of information stem from the strong impact of the earlier information on the overall impression. As such, the first information provides the core impression whether the latter are used to fill the details of the general perception. Such an effect is referred to as primacy effect. Among primacy effects, we can find three type of phenomena, that is assimilation, anchoring and correction. In the first case, the information provided earlier increases the accessibility of related information which in turn confirms the former in the final impression providing a strong and coherent representation of the target (Higgins, Bargh & Lombardi, 1985; Wyer & Carlston, 1979). In the second case, such information provides a sort of impression benchmark against which new information is compared (Anderson & Lampel, 1965, Kaplan, 1969; Sherif & Hovland, 1961). In this latter case, the information is used in different ways producing opposite effects depending on the ease with which the information is processed (Martin, 1986; Schwarz & Bless, 1992). Moreover, Asch noticed that not all characteristics have the same importance in determining the final impression; they seemed to have different weights. Therefore, he divided the qualities he tested in central characteristics, those that seemed to be more basic and central, and peripheral characteristics, those that seemed secondary. This observation anticipated later studies that endeavored the primacy of some particular categories of traits over other categories.

A subsequent cognitive approach is the Information Integration Theory proposed by Anderson (1962; 1965; 1968) that refines some ideas present in Ash's work. Anderson's model postulates that a perceiver retrieves information from different sources and then integrates such information in a coherent impression before producing an explicit and overt response. In this model, all the information is acquired by the perceiver and subsequently mapped into quantitative discrete scales by means of an internal *valuation function*. After that, the quantified information is processed according to an *integration function* that produces the impression in the mind of the perceiver. Such functions work as a sort of cognitive algebra in which information from different sources is added, averaged and multiplied. A key role is played by discounting and augmentation functions that allow some different information to synergistically decrease or increase the weight of other information. Once an impression has been computed in such fashion, it can be converted in an overt response by means of a *response production function* that transforms it into a specific behavior or outcome.

These cognitive models have two main overlaps: the information as input for the impression formation and the need for coherence that a perceiver must accomplish. In both models each piece of information has its own impact on the final impression. If some information is present and is perceived by the perceiver it will be considered. The importance attributed and the specific way in which the information is processed depend on the kind of information, but the information itself won't be disregarded. Both models consider the perceiver as incoherence averse and require each process to comply with that. New information must be interpreted in light of previous information without contradiction. Therefore, some process depends on the outcome of previous or concomitant ones. Understanding how the flow of information is integrated is the general aim of the present dissertation.

1.3 Multidimensionality of Person Perception

The nature of person perception as a set of inferences of traits has been developed after the first cognitive models. The multidimensional structure of personality impression has been proposed by Rosenberg and colleagues (Rosenberg, Nelson & Vivekananthan, 1968). In their work, Rosenberg and colleagues provide a geometric representation of the set of traits that their experimental subjects used to describe some persons they knew. They used multidimensional scaling to arrange the traits used along different dimensions. The measure of the bound between two traits was obtained using the traits co-occurrences. Using this technique allowed the researchers to map different traits following the principle that closer traits are more related than further traits. Multidimensional scaling allowed to disentangle relatedness and synonymy or semantic substitutability. This means that traits that are synonyms but present different facets could lay along different dimensions. At the same time, traits that share very little semantic content but are related to the same characteristic can be placed on the same dimension. Moreover, the identification of underlying dimensions of the person perception is a subsequent step that follows the obtainment of the multidimensional configuration of traits. From their experiments, Rosenberg and colleagues found two main dimensions of person perception (Figure 1.3).

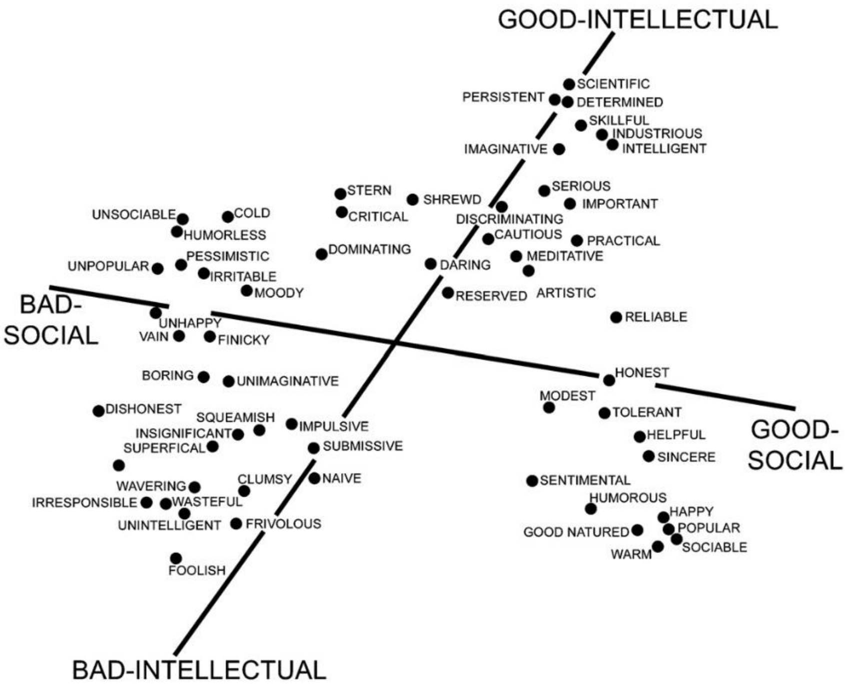

Figure 1. 3 Two-dimensional configuration of 60 traits, which shows the best-fitting axes for the properties of social desirability and intellectual desirability (retrieved from Rosenberg et al., 1968)

Such dimensions were interpreted as two bipolar continua of social desirability and intellectual desirability. Along both dimensions, traits are arranged according the bad-good polarity. Moreover, the dimensions are not perfectly orthogonal, indicating that social desirability and intellectual desirability are not totally independent. For the shake of completeness, Rosenberg and colleagues also proposed a three dimensions' solution that comprises good-bad, hard-soft, and active-passive dimensions. This solution was somehow more marginal due to the minimum increase of fit at the cost of a less succinct description of the phenomenon.

Rosenberg and colleagues noticed that, in light of their multidimensional space, antonyms used by Asch in his previous experiments could be mapped in the same configurational space. In particular, the social desirable traits could be clustered around "warm" and the social undesirable traits could be clustered around "cold". They claimed that "A strong relationship clearly exists between the two-dimensional configuration of stimulus and checklist traits and the impressions formed by Asch's subjects" (p. 292). Such finding denotes that Rosenberg multidimensional approach can explain previous theories adding a richer structure to the person perception phenomena.

1.4 Social Perception Models

Besides general cognitive models, other ones have been developed to deal with the social nature of person perception. In fact, the substrate of such models is a special kind of matter, persons. Given the intentional, agentic, or however peculiar nature of such matter, specific social perception models have been proposed.

The most widely known model of social perception is the Stereotype Content Model proposed by Susan Fiske (Fiske, Cuddy, Glick, Xu, 2002; for a review see Fiske, Cuddy, Glick, 2007). This model stems from the evolutionary need to immediately assess the intentions of the target and only later to evaluate its ability of implement such intentions. These two steps are clearly important for the survival of both groups and individuals. In fact, the evolutionary urgency to understand whether the person we are encountering intents to harm or to help us, is what makes the encounter an opportunity or a threat. Secondarily, only after we grasped the target's intentions, we can try to evaluate if he/she is able to translate its intentions into action. Consequently, the model postulates two main dimensions: warmth and competence. The first maps the first evaluation a perceiver is required to do by putting harmful intention to the low end of the warmth continuum and the good ones to the high end. The second dimension maps the subsequent process by putting individuals capable to act their intentions to the high end of the continuum and those who are less capable, to the low end. Therefore, warmth comprises characteristics such as friendliness, kindness, and trustworthiness, while competence comprises characteristics such as efficiency, competence, and capability (Abele, Cuddy, Judd, & Yzerbyt, 2008; Cuddy, Fiske, & Glick, 2008). The hierarchy of the two dimension is confirmed in later studies (Abele & Wojciszke, 2014).

The two main dimensions are suitable in order to map both social groups or other persons (Abele & Wojciszke, 2014; Russell & Fiske, 2008). The dimensions stated in the model, when crossed, create a two by two theoretical matrix composed by four cells (Figure 1.4). The model provides clear expectation of emotion and behaviors associated to each cell (Figure 1.5).

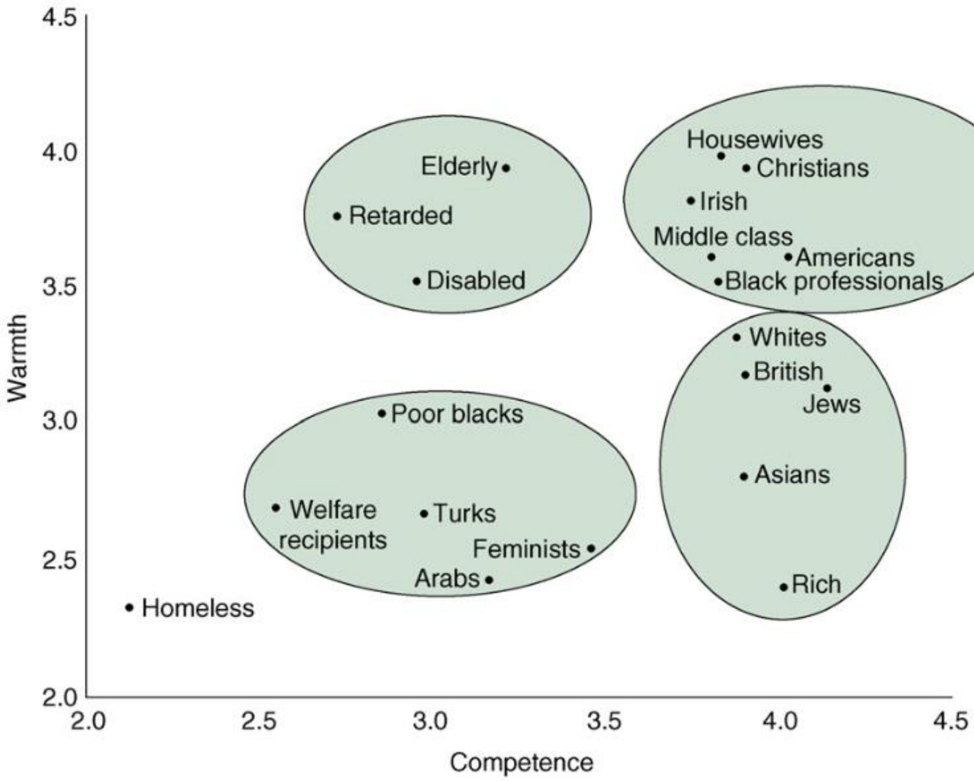

Figure 1. 4 Scatterplot of cluster analysis of social groups arranged by Warmth and Competence (retrieved from Fiske, Cuddy & Glick, 2007)

Targets in the high warmth – high competence cell are expected to evoke admiration and trigger active facilitation in the perceiver. In this situation, targets are granted high status and can exert powerful influence on the perceivers (Cuddy et al., 2009; Durante et al., 2013). In the low warmth – high competence cell, the expected emotion is envy and the behavioral reaction is passive harm. Targets in this cell are not actively harmed by perceivers that mostly avoid helping behaviors toward persons perceived as cold and competent (Fiske, Xu, Cuddy & Glick, 1999). In the high warmth – low competence cell, targets trigger pity and passive facilitation. Perceivers may isolate social targets belonging to such cells and mostly avoid harming behaviors (Cuddy, Fiske, & Glick, 2007; Cuddy, Fiske & Glick, 2008). In the low warmth – low competence cell, the emotion triggered by targets is contempt and the associated behavior is active harm. Perceivers express the highest hostility against targets perceived as low in both dimensions (Harris & Fiske, 2006).

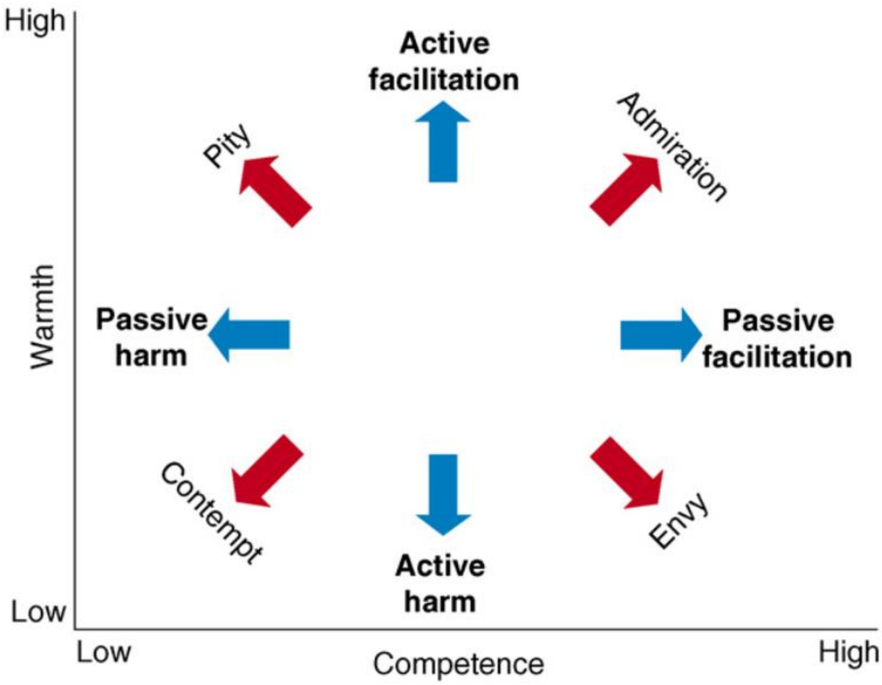

Figure 1. 5 Schematic representation of behaviors from intergroup affect and stereotypes (BIAS) map (retrieved from Fiske, Cuddy & Glick, 2007)

The model was proven to be fairly universal and not culture dependent (Durante et al., 2013). In fact, Cuddy and colleagues put a massive research effort in testing its predictions across several cultures and found substantial confirmations of the model's universality (Cuddy et al., 2009).

1.4.1 Primacy of Warmth

Although it has been shown that warmth and competence are key in order to shape social perception, research has also revealed the primacy of warmth over competence in defining the interpersonal and group relations. Indeed, from a functional perspective, person perception is a skill required in order to navigate social environment. Put differently, "social thinking is for doing" (p. 877, Fiske, 1992) and it is required in order to successfully react to the environment (Zebrowitz & Collins, 1997). Therefore, the existence of the two dimension of social perception could be explained according to the benefits they provide to individual's survival. Therefore, the dimension that provides greater and more salient information to avoid harm should be weighted more in impression formation. Such dimension is the warmth dimension.

In the literature, there are different ways of labelling warmth and competence. They may be labelled as communion and agency (Abele & Bruckmuller, 2011) and the traits associated to such dimensions could be labelled as other- or self-profitable traits respectively, because the firsts benefit the perceiver while the latter benefit the trait owner (Abele & Wojciszke, 2007). Despite different labels, the primacy of warmth over competence stays. There is experimental evidence that warmth related traits attract more attention and greater processing resources that competence related traits (Wentura, Rothermund, & Bak, 2000; Ybarra, Chan & Park, 2001) and that such traits are preferred information sources than competence related traits (De Bruin & Van Lange, 2000). Moreover, warmth related traits are inferred faster than competence related traits from ambiguous behavior descriptions (Bazinska & Wojciszke, 1996), further evidence that warmth is the preferential dimension of other-understanding and general cognitive processes. A comprehensive set of studies revealing the primary role of warmth is that by Abele and Bruckmuller (2011). The authors started defying warmth and competence, to which they refer to as communion and agency, according to their “profitability” for the trait owner or the trait perceiver. Specifically, they consider warmth as other-profitable because it informs the perceiver about attributes of the target providing key information about the safe approachability of the target itself, while competence is considered as self-profitable because the traits are the lever the owner uses in the pursuit of his or her goals. According to this dimension construal, the most informative traits for the perceiver are the ones related to the warmth dimension. Therefore, warmth should lead the perception of the target and the impression formation process in the perceiver. Abele and Bruckmuller pushed their theorizing a step further. They proposed that warmth related traits, besides being weighted more in social judgments, are also processed preferentially on earlier stages of information processing compared to competence related ones. In their experimental studies, the authors proved that warmth (i.e., communion) traits are recognized and categorized as positive or negative faster than competence related traits, they ensure faster inferences drawing from behavior, and come to mind faster when describing others.

This work stemmed from a previous one by Wojciszke and colleagues (Wojciszke, Bazinska & Jaworski, 1998). In their work, the authors based their theorizing on a goal-oriented approach to person perception (Trzebinski, 1985). They investigated impression formation processes as driven by approach/avoidance goals. In light of this assumption, warmth dimension, compared to competence, is way more informative because it can shed light on which is the target intended goal rather than the probability of goal attainment. The second is more related to competence dimension but it belongs to later stages of processing. The whole reasoning can be summarized as follows. Warmth traits inform the perceiver regarding the intentions of the target, while competence traits inform it about the ability to act such intentions. Considering that approaching someone with bad intentions is more dangerous than approaching someone with good intentions, regardless their capabilities, warmth dimension provides more insightful information for the perceiver survival. Given that capabilities are needed in order to act intentions; competence dimension plays a role in driving impression formation based on approach/avoidance goals but such inferences are relegated in subsequent stages of the impression formation process.

The aforementioned works on the comparison of warmth and competence, provide a clear hierarchy of the two dimensions. Clearly, warmth is the main dimension to regulate interaction intentions. Considering that interaction intentions rely on the impression people form on the interaction target, it is not surprising that warmth shows a primacy over competence in shaping the impression of other social agents.

1.4.2 Sub-components of Warmth: Sociability and Morality

Recent experimental work on the fundamental dimensions of social perception has shown that warmth can be divided in two sub-dimensions. Experimental literature provided evidence for the existence of morality and sociability as sub-components of the warmth dimension (Brambilla & Leach, 2014; Brambilla, Rusconi, Sacchi & Cherubini, 2011; Brambilla, Sacchi, Rusconi, Cherubini & Yzerbyt, 2012; Leach, Ellemers & Barreto, 2007). In particular, sociability can be defined as the tendency to show friendliness toward others, to interact with a welcoming attitude. On the other hand, morality is the tendency to show honesty and trustworthiness. A more detailed characterization of morality could be found in philosophers' works. For example, in Aristotele's *Nicomachean Ethics* (cited as Robinson, 1989), morality is listed as one of human virtues, but morality is a long debated topic in modern philosophy (i.e., Nietzsche, 1887/1967). However, a philosophical discussion of morality is beyond the aim of the present work. Here, we will focus on the definition of morality as one of the two subcomponents of warmth, and what MacIntyre (1984) calls "internal goods" or "goods of excellence", and Kant in 1785 (Kant & Wood, 1996) called "moral law".

It has been argued that moral principles are the most important guiding principles across cultures (Schwartz, 1992; see also Rodriguez Mosquera et al., 2002) and that such principles can exerts the function of basic cooperation devices among primates (De Waal, 1996). A strong claim about De Waal's work (1996) is made by Leach and colleagues when they said "Without a sense of morality in the group, he [De Waal, ndr] argued, it would be difficult for members to coordinate their behavior in ways that maximize benefits for themselves and the group as a whole" (p. 236, in Leach et al., 2007). In their work, Leach and colleagues (2007) stated that ingroup's morality was explicitly considered more important than competence or sociability, and that it can explain the ingroup positive evaluation better than the other two. In the same work, the authors proved that only ingroup morality can affect positive aspects of the group-level self-concept. Such impact on the positive evaluation of the ingroup is a key finding because it was tested with experimental manipulations and with both existing and synthetic (created artificially in the lab) social groups (see also Leach, Bilali, & Pagliaro, 2015). Further research in this tradition has shown that morality-based norms provide a more self-evident guideline for individual decision making than norms based on other evaluative dimensions (Ellemers, Pagliaro, Barreto, & Leach, 2008). Individuals anticipate receiving ingroup respect when adhering to morality-related norms (Pagliaro, Ellemers, & Barreto, 2011), indicating a specific concern for morality information when considering one's social identity and centrality within the group one belongs to. In sum, these findings consistently show that morality concerns play a central role in developing a group level self-concept and are crucial for maintaining a positive ingroup image (see also Ellemers & Van den Bos, 2012; Iyer, Jetten, & Haslam, 2012).

The relevance of morality has also been shown in the area of person impression and formation (Brambilla & Leach, 2014). Indeed, it has been shown that traits related to morality and sociability are differently processed in impression formation task, and that the first are preferred to the second for such kind of task (Brambilla et al. 2011). In a study by Brambilla and colleagues (2011), participant were asked to complete a trait-selection task aiming at gathering information about a target they have to form an impression on. In particular, participants were asked to judge which traits were most important in their evaluation of the target. As results, moral traits were selected more frequently than social traits participants. In a subsequent study, Brambilla and colleagues gathered evidence that, when forming an impression of a target, people tends to adopt an asymmetrically disconfirming strategy for morality trait while they adopt a symmetrical strategy for sociability or competence traits. Put differently, participants try to falsify positive moral traits and to confirm immoral traits. Considering the survival value of avoiding the approach of harmful persons, this finding indicate the greater importance of the moral dimension in the exploration of the social environment especially in uncertain situations (Hammond, 2007).

The primacy of morality over sociability and competence is clearly explained in work by Brambilla and Leach (2014). In their examination the authors framed the impression information process in interpersonal relation as a multi-componential process that, from the selection of information useful in order to form an impression of a social target, leads to a global appraisal of the formed impression of the target. Moreover, across samples from the US and Germany, the most important and desirable trait a person could possess is “trustworthy”, a clearly morality related trait (Abele & Brack, 2013; Bruckmuller & Abele, 2013; Cottrell, Neuberg & Lee, 2007). Besides the primacy of morality over sociability in traits desirability and attributed diagnosticity, Pagliaro and colleagues found that morality related information had a stronger impact on the final impression of the target than information related to non-moral characteristics (Pagliaro, Brambilla, Sacchi, D’Angelo & Ellemers, 2013). In the same work, evidence of the consequences of impression formation could be found. Specifically, the authors found that not only morality information has an impact on the formed impression that was greater than competence information, but that such morality driven emotional responses mediated the willingness to help the target in daily life.

Moreover, moral characteristics have been proven to influence actual behavior in interpersonal relations above and beyond than characteristics related to sociability. In fact, a work revealed that when interacting with someone lacking honesty, the perceiver perceived lower similarity between the self and the interaction partner, which in turn diminished the promptness to engage in behavioral synchrony (Brambilla, Sacchi, Menegatti & Moscatelli, 2016). Such effect was found only manipulating moral characteristics and not sociability traits. Specifically, the authors manipulated the characteristics of the interaction partner by providing ad-hoc descriptions. When the partner was presented manipulating moral characteristics (i.e. honesty versus dishonesty) the effect was found, while when it was presented manipulating social characteristics (i.e. friendliness versus unfriendliness) no effect was found. The key finding of this work is that moral characteristics impact the temporal coordination of interpersonal behavior by means of perceived similarity.

The aforementioned findings point out the importance of morality over sociability in person perception and its consequences in actual behavior and social life. The observation that moral characteristics are crucial in shaping impressions raises the question of which processes drive these effects. In this regard, prior work has shown that morality drives impression formation due to its key role in clarifying the intentions of others (Brambilla, Sacchi, Pagliaro, & Ellemers, 2013). As such, it has been shown that morality-related information is more socially relevant in this sense than competence- and sociability-related information as it defines whether someone represents an opportunity or a threat. Accordingly, the more a social target is perceived as immoral, the more they were perceived as posing a threat, which in turn raised a negative overall impression.

1.5 Face Perception

All the literature reviewed so far regards the process by which people form the impression of the other. Put differently, how peoples process an informational input and transform it into a coherent representation. Such input can be any kind of information.

The privileged channel by which information about others are acquired is the sight. As soon as someone is in sight, he or she is immediately categorized (Brewer, 1988; Brewer & Feinstein, 1999; Dijksterhuis, 2010; Fiske & Neuberg, 1990) and a representation is formed in the mind of the perceiver. Such a representation is derived from different information sources that are not all equal. The major source of information is the face (Macrae &

Quadflieg, 2010). Because of its “visibility and omnipresence” (Ekman, Friesen & Ellsworth, 1972; p. 1) the face dominates other sources of information. After a very brief exposure to a face, a perceiver is able to extract evolutionary valuable information that ensures the detection of possible threats (Bar, Neta & Linx, 2006). After a little longer, when just 100 ms of exposure time is reached, the perceiver has already enough information to perform reliable traits inference (Ballew & Todorov, 2007; Willis & Todorov, 2006). The speed of the information extraction process is a proof of how specialized humans are in detecting social signals. This ability is probably due evolutionary pressure and that shaped our perceptual and cognitive systems.

1.5.1 Face Space

Given the aforementioned human need to organize the social and cognitive perception, all faces encountered can be classified and organized in a perceptual space called face space. The face space is a multidimensional space where all instances lay in a certain point according to their *value* along all dimensions in the face space (Todorov et al., 2008) (Figure 1.6).

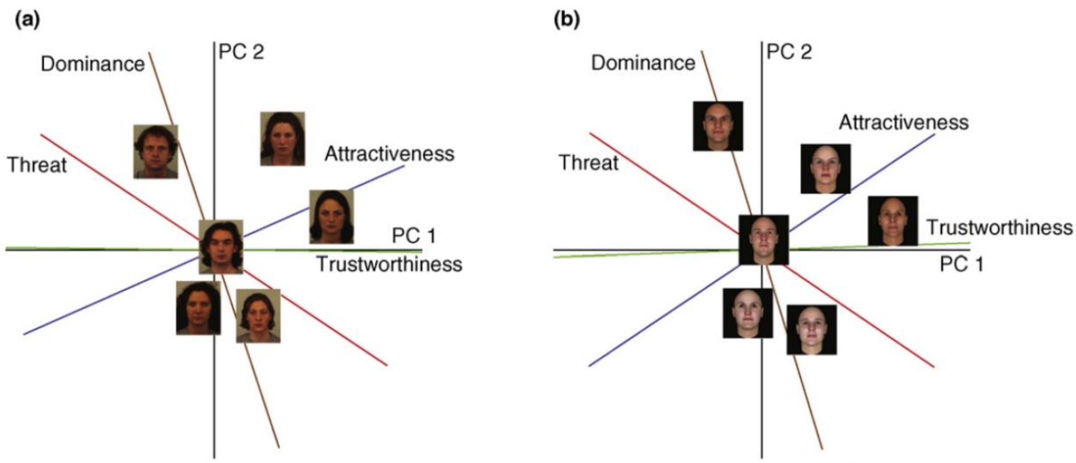

Figure 1. 6 The structure of face evaluation of real (left) and computer-generated (right) faces (retrieved from Todorov, Said, Engell & Oosterhof, 2008)

Several attempts to map the face space have been made. In a work by Todorov and colleagues (2008), the researchers tried to identify the dimension used spontaneously to characterize faces. The authors gathered trait dimensions most frequently used to describe neutral faces. Then, ratings of a database of faces have been obtained along such dimensions. From the obtained ratings, a mean on each dimension have been computed for each face. These data have been submitted to principal component analysis (PCA), a statistical procedure aiming at reduce the dimensionality of a set of correlated variables while preserving the greater amount of variance possible (Pearson, 1901; Hotelling, 1933). The outcome of the PCA is a new dataset where variables are substituted by a smaller set of principal components (PCs) that can retain enough variance but provide a parsimonious description of the same data. In the best solution obtained by Todorov and colleagues, the first two PCs accounted for more than 80% of the variance of the judgments. The variance accounted by the first two PCs was 60% and 20% respectively. The first PC shows strong positive relationship with traits such as trustworthiness and a strong negative relationship with negative traits such as aggressiveness. Therefore, it was interpreted as trustworthiness. The second PC shows a strong relationship with traits such as dominance, confidence and aggressiveness. Therefore, it was interpreted as power/dominance.

Leveraging some recent computer graphics development (Blanz & Vetter, 1999; Valentine, 1991), Todorov and colleagues created a database of faces derived by their knowledge of face perception dimensions. They acquired shape information of a set of example faces using laser scans. Then, vertices of surfaces of the scanned faces were used to describe the faces in terms of data. Such vertex positions were submitted to a PCA and a lower dimensional face space was obtained. As concluding step of the procedure, a new database of faces was obtained by gathering new face instances as linear combinations of the principal components derived in the modelling phase (Figure 1.7).

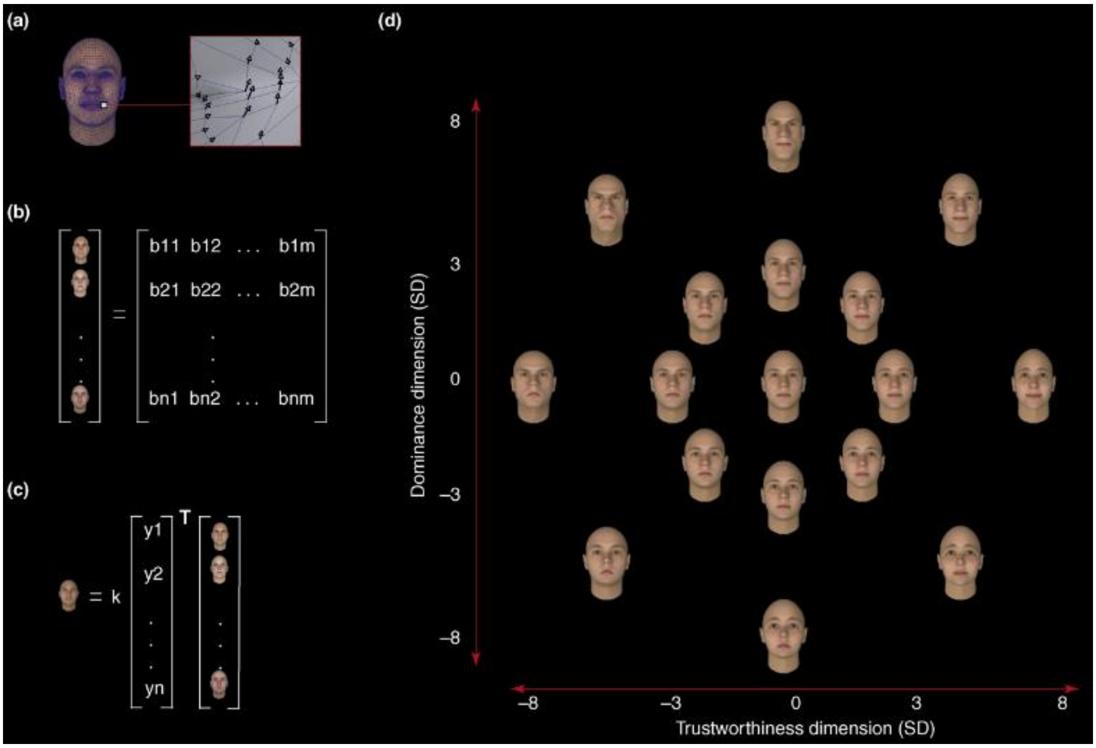

Figure 1. 7 (a) Graphical representation of features. (b) Faces are linear combinations of the 50 shape components. (c) Social judgments vector and set of randomly generated faces are multiplied. (d) 2D model of evaluation of faces. (retrieved from Todorov, Said, Engell & Oosterhof, 2008)

As conclusive validation of their computer generated database, Todorov and colleagues performed the same PCA on trait dimensions' ratings of the computer generated faces. This analysis yielded a solution that is remarkably similar to the one obtained on trait dimensions' ratings of real faces. Such similarity is an evidence of the ecological validity of the faces obtained and of the reliability of both the procedure used to produce the faces and the dimensions obtained with the PCA.

1.5.2 Universality and Reliability of the Face Space

The face space defined as a multidimensional space where every instance is placed according to its value along every dimension is more universal than it seems. For the sake of simplicity let reason considering only the first two dimensions of the face space. This simplification doesn't reduce the explanatory power of the model because the dimensions drawn from PCS are hierarchically ordered, therefore, the more dimension we consider the less increase in retained variance we have. Let's consider only trustworthiness and power/dominance, the ones found by Todorov and colleagues (2008). This representation has been empirically and independently derived by Wiggins and his colleagues (Wiggins, 1979; Wiggins, Phillips & Trapnell, 1989). The authors mapped interpersonal relationship onto two orthogonal dimensions. Such dimensions, called affiliation and dominance, seems to map perfectly onto trustworthiness and dominance dimensions of face evaluation. This kind of overlapping is not uncommon. In fact, warmth and competence from the stereotype content model (Fiske et al., 2007) can also be mapped onto the first two dimension of the face space. A theoretically sounding correspondence between warmth and trustworthiness, and between competence and dominance is easy to hypothesize. In fact, in a 2016 empirical work, Sutherland and colleagues (Sutherland, Oldmedow & Young, 2016) tested such overlapping. The authors found substantial correspondence between warmth and trustworthiness. A weaker correspondence was found between competence and dominance. Their studies involved the rating a set of real faces and, especially for warmth/trustworthiness pair, the plotted results are quite striking (Figure 1.8).

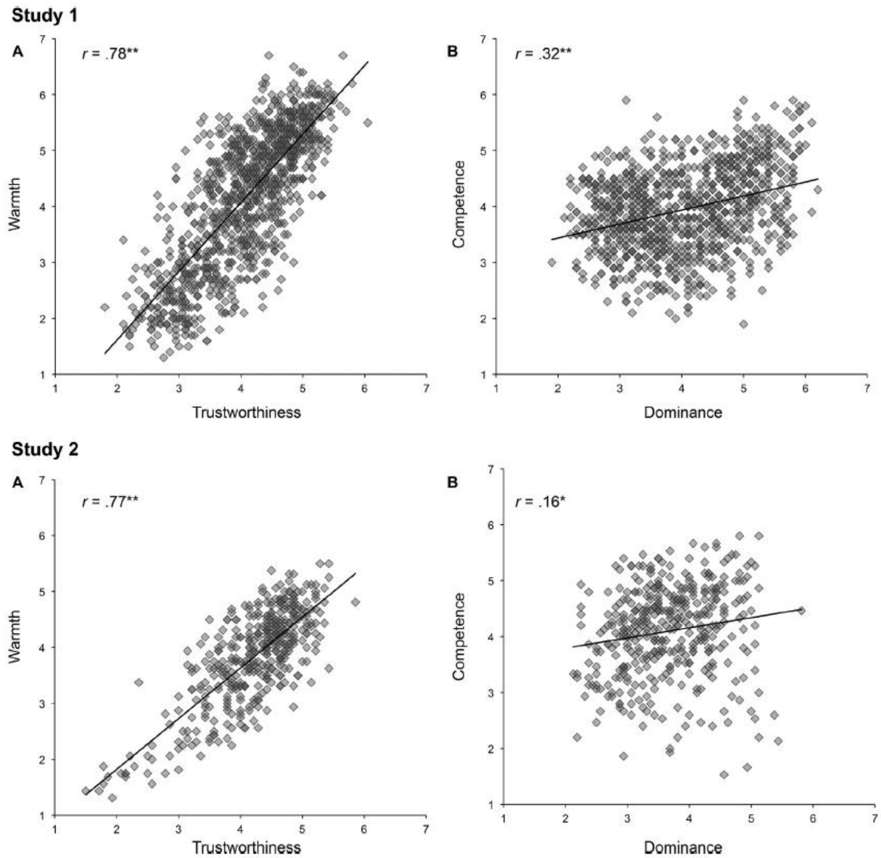

Figure 1. 8 The correlation between average trustworthiness and warmth judgements; and average competence and dominance judgements in two studies from Sutherland, Oldmeadow & Young, 2016

Further evidences of such overlapping can be found in the literature on the consequences of perceptions. In fact, how a person is perceived along trustworthiness and dominance can have real life consequences just the same mechanism happens according to how a person is perceived along warmth, especially the moral component, and competence. For instance, the perception of competence inferred from the face can predict election outcomes (Todorov, Mandisodza, Goren & Hall, 2005), the general facial appearance of a candidate can affect its success in subsequent elections (Little, Burriss, Jones & Roberts, 2007), and the physical features of a cadet can predict its success in climbing the rigid military hierarchy (Mazur, Mazur, & Keating, 1984).

Besides such outcomes, facial features can lead to severe consequences. For example, Afrocentric facial features have been proved to influence criminal sentencing (Blair, Judd,

& Chapleau, 2004), the same is true for untrustworthiness derived from facial cues in the context of both criminal sentencing (Wilson & Rule, 2015) and even capital punishment (Wilson, & Rule, 2016). An even more vicious consequences of the social perception from faces can be found in the context of self-perception. In their 2015 work, Slepian and Ames showed that being constantly perceived as untrustworthy lead the perceived to internalize the expectation of its own untrustworthiness held by the perceiver. Such internalized impression mediated the effect of appearance-based trustworthiness and actual trustworthiness (Slepian & Ames, 2015). Despite this findings, the link between perceived trustworthiness and actual trustworthiness has not been established yet. In fact, even if for example some evidence of the link between perceived honesty and clinically assessed honesty can be found (Zebrowitz, Voinescu & Collins, 1996), evidence of the link between perceived honesty and actual behaviors are weak or inconclusive (Bond, Berry & Omar, 1994). All the aforementioned research points out the important role of face perception in everyday life. Moreover, a clear overlapping between models of face perceptions, with all dimensions involved in the face space, and social perception models such as the stereotype content model emerge from the comparison of both literatures. A likely reason for such overlapping is that both (im)moral characteristics and (un)trustworthiness cues are useful danger signals for the perceiver. In fact, it has been proved that facial dynamics can indicate trustworthiness and that such dynamics rise the likelihood of interact with the target and to cooperate to a greater extent (Krumhuber, Manstead, Cosker, Marshall, Rosin & Kappas, 2007). The threat-detection based link between moral characteristics and trustworthiness can be found even at low neural level. In fact, the amygdala, a structure highly involved in the processing of negative stimuli and in signaling threat to the organism (Fox, Oler, Tromp, Fudge & Kalin, 2015), is found to react to threatening faces (Luo, Holroyd, Jones, Hendler, & Blair, 2007; Ohman, 2002). In addition, it has been shown that the amygdala reacts accordingly to the trustworthiness of the target (Santos, Almeida, Oliveiros, Castelo-Branco & Nishijo, 2016).

In the present work we rely on such overlapping between trustworthiness and warmth, and in particular with the moral component of warmth. As such, most of the studies that have shown the key role of moral characteristics in shaping person perception have used trustworthiness (and honesty) as key traits (for a review, Brambilla & Leach, 2014; see also Abele & Wojciszke, 2014). In a similar vein, studies on face perception has shown that trustworthiness facial cues are the essence of the trustworthiness dimension of facial perception (for a review, Todorov, Olivola, Dotsch, & Mende-Siedlecki, 2015). Overall, our aim was to bridge social perception literature with literature on face processing. Our key link is the similarity, in terms of perception and cognition primacy, between morality and trustworthiness. Both dimensions drive the perception of the target and provide the same evolutionary advantage to the perceiver, the ability to detect and avoid social threats.

1.5.3 Neural Correlates of Face Perception and Permeability of Face Neural Network

In the neuroscientific literature, the existence of a complete neural network dedicated to the processing of faces have been proposed (Haxby, Hoffman & Gobbini, 2000; Wiggett & Downing, 2008; Rossion, 2008). According to a work of Ishai (2008), such network involves some core and some extended structures, is distributed in the cortex, and spans from posterior visual regions to limbic structures to frontal lobes. More precisely, the core structure is composed by the fusiform gyrus (FG), the inferior occipital gyrus (IOG), and the superior temporal sulcus (STS). The FG and IOG are responsive to invariant facials structures that play a key role in identity recognition (Rossion, Caldara, Seghier, Schuller, Lazeyras, & Mayer, 2003) whereas the STS responds to gaze direction and speech related movements (Calder, Beaver, Winston, Dolan, Jenkins, Eger, & Henson, 2007; Hoffman, & Haxby, 2000; Puce, Allison, Bentin, Gore, & McCarthy, 1998). On the other hand, the extended structure involves the amygdala (AMG), the inferior frontal gyrus (IFG), and the orbitofrontal cortex (OFG). In the limbic part of the structure, the AMG process transient features such as facial expression (Breiter, Etcoff, Whalen, Kennedy, Rauch, Buckner, Strauss, Hyman & Rosen, 1996; Morris, Frith, Perrett, Rowland, Young, Calder & Dolan, 1996; Phillips, Young, Senior, Brammer, Andrew, Calder, Bullmore, Perrett, Rowland, Williams, Gray & David, 1997; Vuilleumier, Armony, Driver & Dolan, 2001) and maintain a vigilant attitude toward unfamiliar people (Gobbini & Haxby, 2007). All regions involved have reciprocal connections with the regions immediately close in the processing chain even if the strength of the connection is not always symmetrical). This configuration allows a great amount of flexibility to the network. In fact, different connections show stimulus and task dependent coupling and, at the same time, different regions show stimulus and task dependent activities (Ishai, 2008).

Since its distributed nature, and the number of general purpose regions it encompasses, this neural network is very likely permeable to information spillover from other high and low order processes. It is likely that perceiving a face by having in mind certain goals or while other processes are unfolding could result in a slightly biased perception, especially if other processes conflict or enhance the perception and categorization processes. Relying on such characteristics, the present work tries to investigate how social perception from faces and other cognitive and perceptive processes can be integrated.

1.6 Open Issues

The flourishing literature on impression formation and person perception has relied, since recent years, on textual stimuli. In the vast majority of research, persons have been presented as mere lists of traits and both impressions and perceptions have been assessed as ratings. This approach that guarantee easy experiments implementation is anyway not suited to investigate nor impression formation neither person perception. In fact, Macrae and Quadflieg (2010) noted that "despite person perception constituting a central area of inquiry in social psychology for decades, two elements have surprisingly been absent in explorations of this topic – persons and perceptions" (p. 428).

The first issue we want to address is the ecological validity of the paradigm used. We think that in order to investigate how people form an impression of others it is necessary to go over textual descriptions (Figure 1.9).

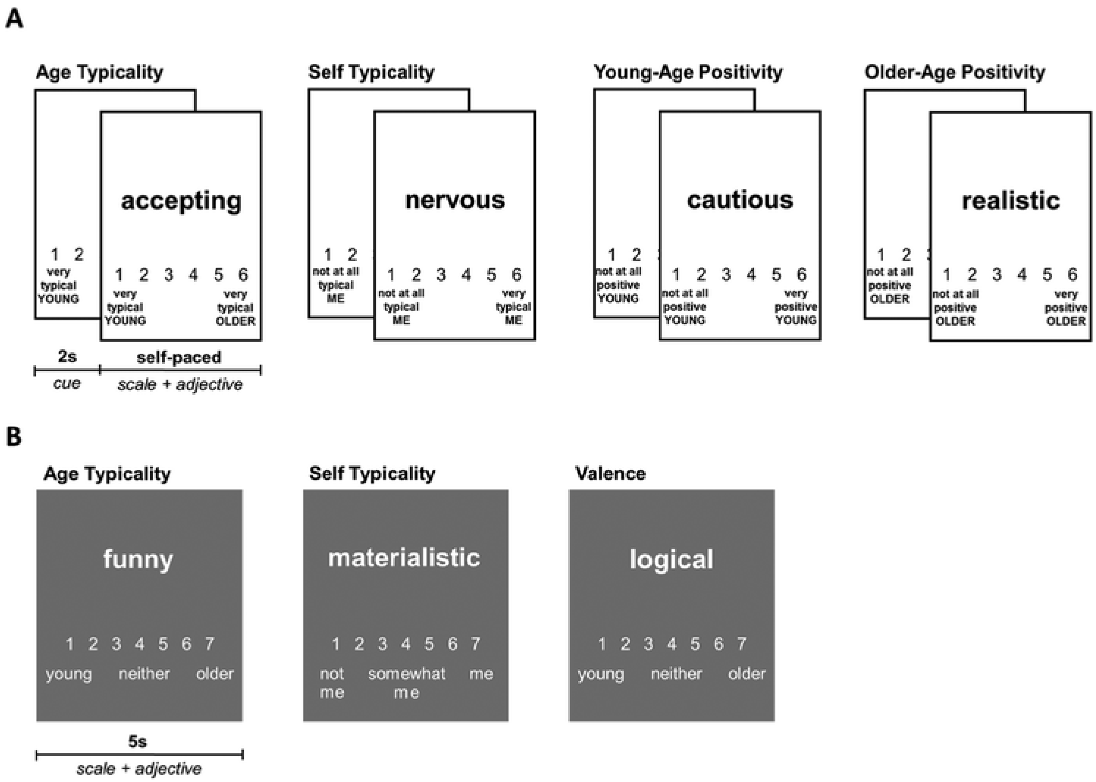

Figure 1. 9 Trial of a typical experiment presenting persons as lists of traits (retrieved from Lin, Ankudowich & Ebner, 2017)

By presenting persons as list of traits, scholars have investigated how people process list of traits and not how they perceive others. In the present work, we aim at fostering the ecological value of both stimuli and dependent measures. Across all chapters of the present work, we will test our hypotheses by relying on pictures of faces instead of lists of traits.

A second issue we are facing is the nature of the dependent variable. In scientific literature, impression of others has been assessed using explicit ratings. Such method cannot shed much light on the process of categorization of others. By using explicit ratings, only the process outcome is visible to the researcher. In the present work, mouse-tracking (Hehman, Stolier & Freeman, 2015) techniques will provide us a deep look at categorization processes while they unfold over time.

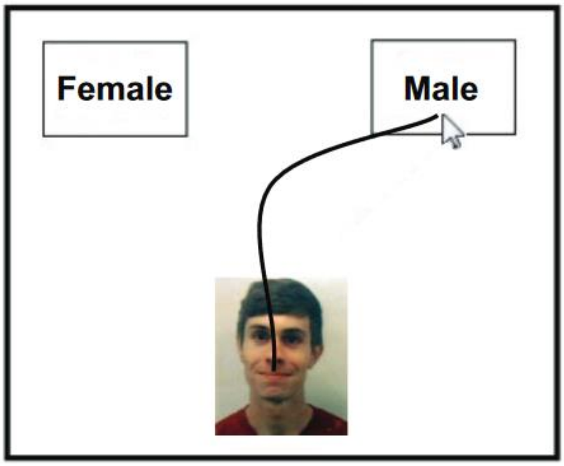

Figure 1. 10 Example of a mouse-tracking trajectory (retrieved from Heman, Stolier & Freeman, 2015)

In two choice forced choice task, such techniques allow the continuous tracking of the mouse trajectories while the process is still ongoing (Figure 1.10). Tracking hand movements while the mouse is approaching the selected response button, allows the researcher to look whether the respondents first moved toward a "wrong" response and then correct their decision or if they moved toward the selected responses from the beginning. These two options are two poles of a continuum. Therefore, researchers can have a shaded measure that points out the degree of easiness of each response (Figure 1.11).

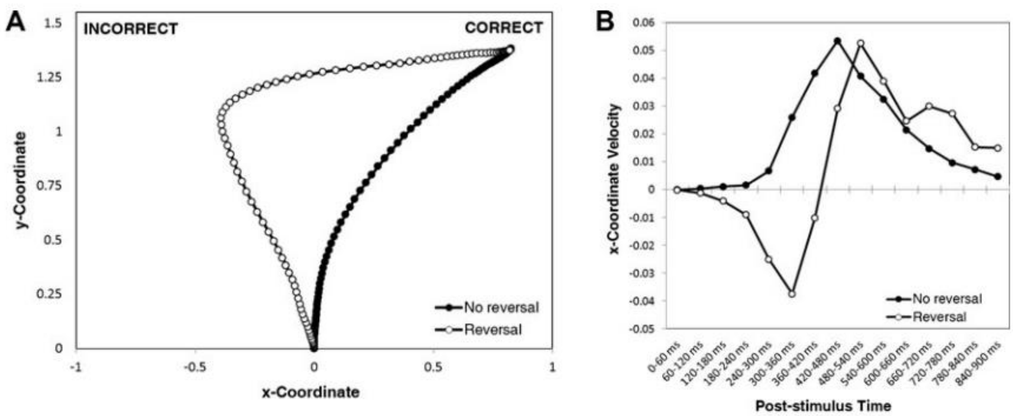

Figure 1. 11 Example of (a) mouse-tracking trajectories and (b) time course of x-axis component in mouse-tracking task (retrieved from http://www.mousetracker.org, october 2018)

Such measures are two spatial measures. The first measure is the maximum deviation (MD). This measure is operationalized as the maximum distance from the actual mouse trajectory and an ideal straight line from the starting point and the selected response. Such distance is orthogonal to the ideal straight line and indicates to what extent the participant's hand is attracted by the unselected response. Higher values of MD indicate a strong interference by the unselected response or by other co-occurring information. The second measure is the area under the curve (AUC). As MD, AUC is computed using both the actual trajectory and the ideal straight line. Since these two trajectories start and end in the same points, they delineate a close area on the computer screen. AUC is the area of the screen enclosed in these two trajectories. As for MD, AUC signals higher interference by means of high value. If the participant's hand is attracted away from the selected response, the actual trajectory will draw a wider curve resulting in higher AUC. An additional measure is the moment in time in which the MD is recorded. Such measure is called maximum distance time (MD time) and, differently from the two other spatial measures, it is a time measure. Generally, after recording the MD, the participant's hand only gets closer to the selected response button. Therefore, from this moment on, it is likely that the cognitive process under investigation has reached its end states and only a motor component is required to terminate the trial. Because of this, MD time can be considered as the amount of time required to process the stimulus without the time required to produce the response. Mouse-tracking paradigm provides one additional time measure, namely the time between the stimulus onset and the first movement of the mouse. In this work, we did not employed such a measure because our experiments are carried out under time pressure therefore, this specific measure may be distorted (Hehman et al., 2015).

Such methodological novelty can help us in assessing the dynamic process while it unfolds. All information retrieved with this technique will not be available to the researcher that rely only on the final outcome of the process (i.e. Explicit ratings). In fact, this methodology has already been used to inquiry categorization processes. For example, Freeman and colleagues (Freeman, Penner, Saperstein, Scheutz & Ambady, 2011) tested whether status cues can impact the ethnic categorization of racially ambiguous stimuli. In their work, Freeman and colleagues presented faces morphed along ethnicity from clearly withe to clearly black by adding visual cues of high and low social status, namely a business suit or a janitor suit. Results pointed out that status cues are used to resolve the competition between two contemporaneously active social categories. While performing the task, participants have the stereotype of white people and black people contemporarily active and competing against each other in guiding the final categorization. Status related cues impacted the categorization process by pushing the participant toward the stereotype-congruent response. The same year, Freeman and Ambady tested the integration between faces and voices in the domain of person categorization (Freeman & Ambady, 2011). Specifically, the authors presented facial and vocal cues in a gender categorization task while recording participants' hand movements. They manipulate gender-typicality of the face stimuli and of the vocal cues playing in the background during the categorization task. In so doing, the authors provided insights on the temporal dynamics of the integration between face and voice of the same target. In the present work, by relying on mouse-tracking techniques, we aim at obtaining accessible, process-sensitive, and real time data on which we will test our hypothesis (Stillman, Shen & Ferguson, 2018).

An additional aspect we aim to foster is the *embedded nature* of person perception, largely neglected by the literature so far. For embedded nature, we intend the fact that persons are never met in isolation but always in a rich context which convey a large amount of relevant and irrelevant information. Such aspect is still under-investigated since the vast majority of works that examined person perception, even in those using real faces, persons are presented in isolation (for a review see Todorov, Olivola, Dotsch & Mende-Siedlecki, 2015). Given the richness of information of natural contexts in which persons are always met, it is likely that the information filtering process, by which relevant information are selected and irrelevant information are discarded, could impact the perception and categorization processes. By presenting persons without any context, scholars have unduly simplified the process they were inquiring (Figure 1.12).

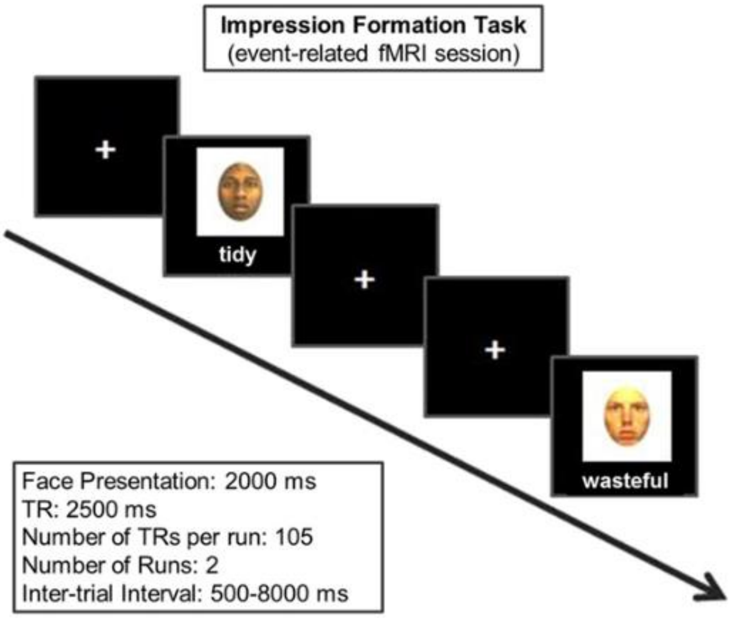

Figure 1. 12 Trial of a typical experiment presenting persons as faces without any visual context (retrieved from Li, Cardenas-Iniguez, Correll & Cloutier, 2016)

In the present work, we will try to track the influence of the context on the categorization process with a special focus on the content of the information provided by the context itself. In fact, given the evolutionary benefit the perceiver receives by correctly drawing inferences on target's morality and trustworthiness, we think that it is necessary to endeavor how threat information conveyed by the context are integrated with trustworthiness information inferred from the facial appearance of the target.

1.7 Aims of the Dissertation

Starting from the issues outlined above, the present work aims at investigating person perception using a more robust and ecologically reliable paradigm. Our main goal is to endeavor how contextual cues shapes the perception of trustworthiness of social agent. We picked trustworthiness as the social dimension of our interest because, as outlined above, it can be considered the most important dimension in terms of social perception. This general aim will be specified and developed as this work unfolds.

In the second chapter, we investigate how the information in the context in which a person is encountered impacts the trustworthiness judgement about the social target. The specific aim of this section is to understand whether it is possible that contextual cues impact the perception of the social target encountered. In so doing, the first experimental study will test whether the information embedded in the context can shape trustworthiness judgements by using classical paradigms.

In the third chapter, such paradigms will be left in favor of more insightful process-sensitive paradigms. This modification is introduced in order to obtain a more fine-grained examination of the phenomenon of interest. Here, we will investigate the ongoing categorization process when contextual information is available. The main hypothesis regards the interaction between threat information and perception of trustworthiness. Specifically, we aim at testing whether the impact of contextual cues on trustworthiness perception can be individuated in the categorization process and not only in the categorization outcome. Given the threat related nature of (un)trustworthiness (Willis & Todorov, 2006), we expect that the correct detection of an untrustworthy social target is facilitated by threat related information present in the context. Moreover, we expect this interaction to be threat specific. Therefore, we will try to disentangle threat from valence, comparing the effect of threat information with the effect of negative but non-threatening information.

In the fourth chapter, we investigate the level of the integration between threat information and judgements of trustworthiness. In order to investigate at which level such an integration takes place, we provide threat information on a different perceptual channel. While in the second and third chapters information regarding both the trustworthiness of the target and the threat embedded in the context come from visual inputs, in chapter 4 threat information is provided in the auditory channel and the information about the social target is provided in the visual channel. Such design should result in a smooth information integration if this process takes place at high level of cognitive processing. Otherwise, if the integration process takes place at low perceptual levels it should result in less or no integration between threat and trustworthiness cues.

Chapter 5 tries to summarize the findings of the previous three chapters. The findings of chapters 3 and 4 are meta-analyzed. The type of integration, visual-visual for the studies in chapter 3 and visual-auditory for the studies in chapter 4, will be taken into account. This chapter will provide a more reliable estimation of the integration of threat information with trustworthiness judgements. Moreover, the moderating role of the integration level will point out if the integration take place at low perceptual levels or at a higher cognitive levels.

Concluding, in the last chapter, all analytical results will be commented and the findings gathered in all experimental studies will be discussed.

CHAPTER 2

THREAT INFORMATION AND TRUSTWORTHINESS JUDGEMENT INTEGRATION: OUTCOME-BASED MEASURES

2.1 Introduction

In this chapter, the integration between threat information extracted by a visual context and judgements of trustworthiness about the target embedded in the context is investigated. In this chapter we consider outcome-based measures. Specifically, in the first study, such an integration was investigated by using a classical explicit rating paradigm. Untrustworthy and trustworthy faces were embedded in different types of contexts. Such contexts have been pretested in order to understand which information they convey in terms of content and to guarantee that the perception relies on consensus among perceivers. In the second study, we went beyond explicit ratings by analyzing reaction times and error rates in a categorization task. Despite trustworthiness is a continuous measure (Todorov, Baron, & Oosterhof, 2008) we introduce this modification for a main reason. Indeed, we wanted to force participant to judge targets' trustworthiness because in everyday life, judgements of trustworthiness are usually converted in discrete actions: accept a ride from someone or not, decide whether someone can keep a secret or not, trust someone or not.

In order to start probing the impact of threat information on the outcome of trustworthiness judgement and categorization processes, we developed the two experimental studies presented in this chapter and relative hypotheses. Specifically, we expect that threat information interacts with trustworthiness cues in both experiments impacting processes' outcomes. In both experiments we expect to find congruency effect (Notebaert, Gevers, Verbruggen & Liefooghe, 2006) on the interplay of the main variable we manipulated, trustworthiness and threat. Our prediction is that trustworthiness judgments in the first experiment will be pushed toward the untrustworthy side of the continuum when faces are embedded in contexts containing threat information. We expect that this effect goes above and beyond the congruency between trustworthiness and valence due the overlapping between the first and threat. Accordingly, in the second experiment, we expect threat information makes harder for participants to correctly categorize trustworthy faces. Specifically, we expect that threatening backgrounds raise overall error rates but particularly error rates in the categorization of trustworthy faces. This can be explained due to the incongruence between trustworthy cues and threat information. Once again, we expect this effect to go above and beyond a simple valence effect. As in the first experiment, we expect to find this effect even when comparing threatening and negative background that share the same valence and differentiate only regarding threat information. Since in the second experiment we collected reaction times, we can derive the same hypothesis we have for error rates and translate it on the fluidity of the categorization process as indexed by process slowing. The same pattern we expect for error rates is expected for reaction times. In fact, when the categorization task becomes harder, reaction times increases exactly as error rates.

2.2.1 Materials: Facial Stimuli

In all the experimental studies of the present work, persons are presented as faces instead of lists of traits. This is due to our aim to holistically investigate person perception and not the processing of person related traits. Therefore, we rely on a database of computer generated faces that provide us target faces along with their normative level of perceived trustworthiness (Todorov, Dotsch, Porter, Oosterhof, & Falvello, 2013). In this database, several identities are present. Each identity is obtained as a random vector of physical characteristics that defy the somatic appearance of the identity itself. Once created, each identity is morphed along a continuum that goes from 3 SD below the average trustworthiness to 3 SD above the average trustworthiness. Stopping at each integer from -3 to 3, 7 versions of the same identity are obtained (Figure 2.1).

Figure 2. 1 Example of identity morphed along trustworthiness continuum

Such identities are generated modifying only the characteristics that impact trustworthiness and keeping the other dimensions of social perception constant (i.e., dominance). This features allow us to avoid possible confounding due to other dimensions of social perception. In fact, the database we employed comprises faces morphed along trustworthiness only. Therefore, each version of each identity "scores" the same on all other dimension of social perceptions.

2.2.2 Materials: Visual backgrounds

In all experimental studies of this chapter, we manipulate the content of the information provided by the contexts. Specifically, we employed 3 types of background: neutral, negative, and threatening backgrounds. To ensure that such content is what we actually want to manipulate, we ran a pretest on a convenient sample. All contexts are drawn from public websites. The images have been downloaded and cut. All contexts resulted in 800x600 pixels jpeg files (Figure 2.2). After a selection we came up with 12 images, 4 for each context type (neutral, negative, threatening).

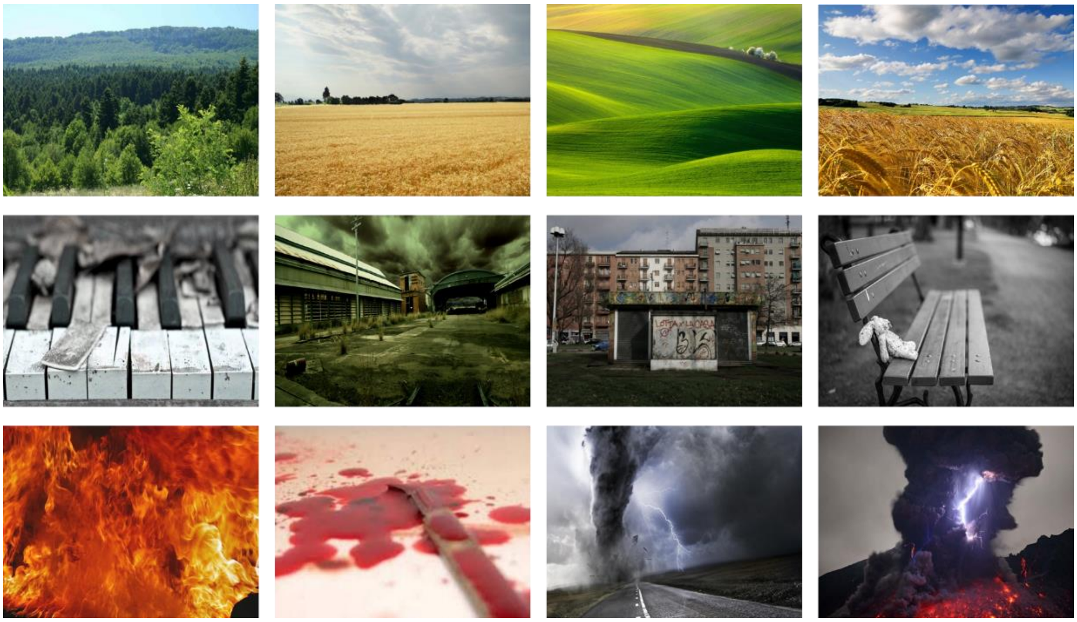

Figure 2. 2 Pretested Visual Contesxts: neutral (top), negative (middle), and threat-specific (bottom)

This final set of context have been pretested on N = 26 participants, 7 males and 19 females, ranging from 20 to 30 years old (*M* = 23.80, *SD* = 2.77). Participants saw the images one at the time in random order. They were asked to rate each context. Specifically, we asked them to indicate the extent to which each scene context was threatening using a scale ranging from 1 (*not at all*) to 7 (*extremely*), and to indicate the valence of the image on a similar scale ranging from 1 (*negative*) to 7 (*positive*). Once obtained such ratings, we computed the average valence and the average perceived threat for each type of context for each participant (Figure 2.3).

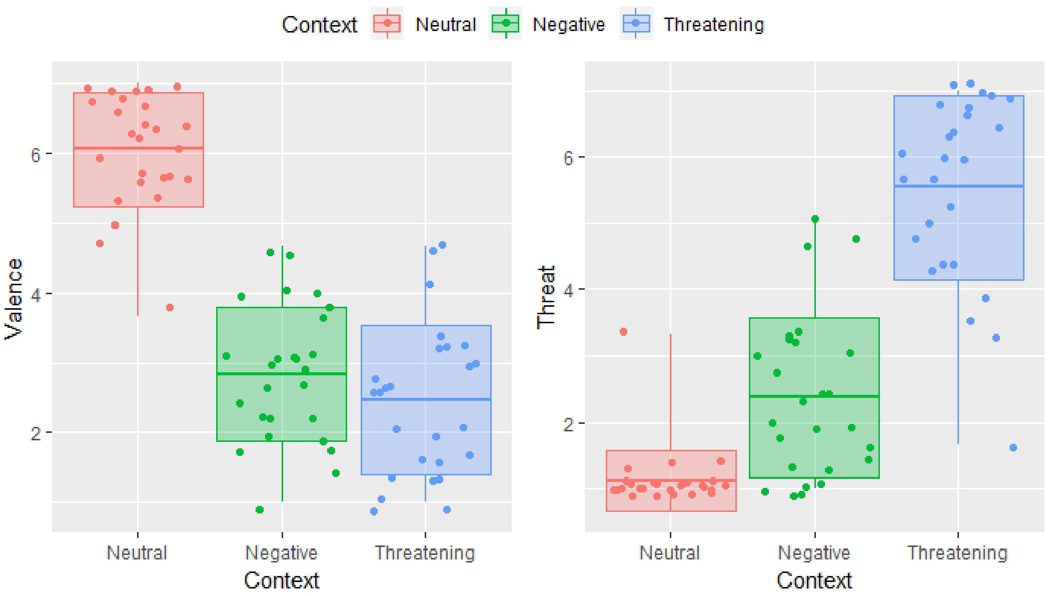

Figure 2. 3 Average valence (left) and perceived threat (right) for each type of context

Such average (Table 2.1) scores were submitted to two separate within-subjects 3 ANOVAs with the type of context (neutral, negative, threatening) as dependent variable. This analysis solves the purpose to test the differences in the perception of the images due the type of context.

	Valence	Threat
Neutral	6.06 (.82)[a]	1.16 (.48)[a]
Negative	2.85 (.96)[b]	2.76 (.19)[a]
Threatening	2.46 (1.07)[b]	5.53 (1.31)[b]

Table 2. 1 Difference in valence and threat among neutral, negative, and threatening contexts. [a, b, c] indexes indicate means that are statistically different across type of context

Regarding valence, the ANOVA revealed a significant main effect of context type, $F(2, 50) = 147.18$, $p < .001$, $\eta^2_p = .85$. Post-hoc comparisons revealed that neutral contexts (M = 6.06, SD = .82) were perceived as being less negatively valenced than both threatening (M = 2.46, SD = 1.07) and negative contexts (M = 2.85, SD = .96), $ts > 12.51$, $ps < .001$. More importantly, negative contexts and threatening contexts were perceived as sharing the same negative valence, $t(25) < 1$, $p = .58$. Such pattern is an evidence that negative and threatening contexts share the same negative valence, clearly more negative than neutral contexts.

Regarding perceived threat, the ANOVA revealed a significant main effect of context type, $F(2, 50) = 125.49$, $p < .001$, $\eta^2_p = .83$. Post-hoc comparisons revealed that threatening contexts ($M = 5.53$, $SD = 1.31$) were perceived as more threatening than both neutral ($M = 1.16$, $SD = .48$), $t(25) = 16.61$, $p < .001$, $d = 3.25$, $CI_{95\%} = [2.27, 4.23]$, and negative contexts ($M = 2.76$, $SD = 1.19$), $t(25) = 11.08$, $p = .001$, $d = 2.17$, $CI_{95\%} = [1.45, 2.88]$. This pattern of results perfectly suits our goal. In fact, negative and threatening contexts, while sharing the same negative valence, are perceived as differently threatening. Therefore, any difference found between the negative a threatening condition can be attributed to the difference in perceived threat and not in valence, disentangling the effects of the two variables.

To further exclude the possibility that the scenes were perceived as signals of (un)trustworthiness, we asked 30 Italian students not involved in the main studies ($M_{age} = 22.16$; $SD = 5.69$) to view each scene context and freely write down their thoughts. None of them mentioned words or concepts related to honesty, trustworthiness, or morality. More specifically, students mentioned negative concepts associated with threat (e.g., fear, danger, risk, etc…) when viewing the threatening scenes. By contrast, students mentioned negative concepts unrelated to threat (e.g., sadness, poverty, deterioration, etc…) when viewing the negative scenes. Students mentioned descriptive concepts (e.g., nature, green, spring, etc…) when viewing the neutral scenes. One concern with forcing subjects to make dichotomous trustworthiness decisions (see Experiment 2 and the experiments employing mouse-tracking) may bias our results or exhibit a different pattern of responses compared to continuous Likert ratings of trustworthiness. To address this issue, we recruited 100 participants from Amazon Mechanical Turk, with half of participants asked to make dichotomous trustworthiness judgments of the stimuli using the keyboard in randomized order, and the other half of participants asked to make 7-point continuous judgments of the same stimuli. Due to 8 participants not completing the task, our final sample for this task comprised of 49 participants for the dichotomous judgments and 43 participants for the continuous judgments. For each stimulus, we generated a mean for participants' dichotomous judgments (0 = untrustworthy, 1 = trustworthy), and also a mean for participants' continuous judgments (1 = untrustworthy - 7 = trustworthy). These were very strongly correlated, $r(286) = 0.96$, $p < .00001$. This result speaks against the possibility that forcing participants to use dichotomous responses biased the results in some manner relative to a continuous-rating assessment of facial trustworthiness.

2.3 Study 1: The Impact of Contextual Information on Continuous Trustworthiness Judgements

In this first experimental study, we want to test whether continuous trustworthiness judgments are impacted by the content of the information provided by the background. It is likely that persons are met in rich visual environment. For rich environment, we intend a visual scene that provides relevant and irrelevant information. Moreover, given the finite cognitive capacity of every human being, all this information must be filtered (Cowman, 2010; Luck & Vogel, 1997). Some information is retained and some other are discarded. From all these premises, we consider judging someone's trustworthiness as an information processing task. Specifically, a perceiver that draws from its visual input information about facial trustworthiness and information provided by the visual context needs to process only the trustworthiness related information. Given the strong relationship we hypothesized between trustworthiness and threat, we expect that the judgements of trustworthiness in the presence of threat information should incorporate more threatening information regarding the context because of the difficulty in separating such similar kind of information. Therefore, we expect the trustworthiness rating to be shifted toward untrustworthiness for faces embedded in threatening contexts.

Our prediction is threat specific. Considering that mere valence of the visual background can interact with facial trustworthiness and impact the judgement, we expect to find a greater effect of threatening background that goes above and beyond the effect of negative background. By sharing the same valence, negative and threatening background should exert the same effect if the effect itself is due to valence, otherwise they should exert different effects it threatening information interfere with the process of judging trustworthiness of the target.

2.3.1 Method

2.3.1.1 Participants

In order to determine the required sample size, an a priori power analysis was conducted using G Power 3.1 (Faul, Erdfelder, Lang, & Buchner, 2007). The required sample size needed to detect an effect of $\eta^2_p = .15$ with power of $(1-\beta) = .80$ (Cohen 1988) and a type 1 error of $\alpha = .05$ was N = 36 for our within-subjects experimental design.

We enrolled university students to participate in the experiment and we continued gathering data while the lab was available. 135 participants, aged between 19 and 54 years (*M* = 23.00, *SD* = 4.62), 108 females and 27 males, showed up in the lab and took part in the experiment.

2.3.1.2 Procedure

In order to test the impact of threatening information on the continuous judgements of facial trustworthiness, we elaborated a rating task in which participants are asked to rate the trustworthiness of faces embedded in the three types of pretested visual background.

We randomly selected 6 identities from the face database and considered only the extreme morphs (-3 SD, -2SD, +2SD, +3SD). Such 24 faces in total, when crossed with the 12 background we pretested, yields 288 stimuli. Given the high number of stimuli, each participants rated only a random subset. Specifically, a single participant rated each one of the 24 faces in each of the type of context and the exact context was picked randomly. This yielded 72 trials per participant.

In each trial, the participants were prompted with the stimulus and a slider below the picture. They were asked to judge how trustworthy the target was. At the extremes of the slider, we put two anchors, “*untrustworthy”* on the left end and “*trustworthy”* on the right end. The slider was divided in 6 equally spaced pieces creating a 7 point likert scale. The trial was self-paced.

Once obtained the ratings, we computed an average score for each condition for each participant. Since our experimental design is a 2 trustworthiness (trustworthy, untrustworthy) by 3 context type (neutral, negative, threatening) within-subjects design, we obtained 6 average ratings for each participant.

2.3.2 Results

The explicit ratings of trustworthiness were analyzed using a 2 trustworthiness (trustworthy, untrustworthy) by 3 context type (neutral, negative, threatening) within-subjects ANOVA. The analysis yielded a main effect of trustworthiness $F(1, 134) = 306.83$, $p < .001$, $\eta^2_p = .70$, and a main effect of context type, $F(1, 134) = 88.24$, $p < .001$, $\eta^2_p = .40$. Since untrustworthy faces ($M = 3.24$, $SE = .08$) were rated as less trustworthy than trustworthy faces ($M = 4.89$, $SE = .07$), $t(134) = 17.44$, $p < .001$, $d = 1.50$, $CI_{95\%} = [1.25, 1.75]$, the first main effect can be interpreted as a manipulation check that provides good evidence in favor of the reliability of our faces database. In the case of the second main effect, post-hoc comparisons revealed that faces embedded in different contexts were perceived differently depending on the context. In fact, faces embedded in threatening contexts ($M = 3.78$, $SE = .07$) are perceived as less trustworthy than faces embedded in both neutral ($M = 4.38$, $SE = .06$), $t(134) = 10.85$, $p < .001$, $d = .93$, $CI_{95\%} = [.73, 1.13]$, and negative contexts ($M = 4.04$, $SE = .06$), $t(134) = 6.55$, p < .001, d = .56, $CI_{95\%} = [.38, .74]$, regardless their actual trustworthiness. Moreover, faces embedded in neutral contexts were perceived as more trustworthy than faces embedded in negative contexts, $t(134) = 8.59$, $p < .001$, $d = .74$, CI*95%* = [.55, 93].

Crucially, the analysis yielded a significant interaction between trustworthiness and type of context $F(2, 268) = 20.82$, $p < .001$, $\eta^2_p = .13$. We decomposed the interaction by running post-hoc comparisons for trustworthy and untrustworthy faces separately (Table 2.2).

	Trustworthy	Untrustworthy
Neutral	5.24 (.76)^a	3.51 (1.00)^a
Negative	4.94 (.84)^b	3.14 (.98)^b
Threatening	4.50 (1.02)^c	3.06 (1.03)^c

Table 2. 2 Difference in trustworthiness rating among neutral, negative, and threatening contexts. $^{a, b, c}$ superscripts indicate means that are statistically different across type of context

For trustworthy faces (Figure 2.4), faces embedded in a threatening context ($M = 4.50$, $SD = 1.02$) were perceived as less trustworthy than faces embedded in both neutral ($M = 5.24$, $SD = .76$), $t(134) = 11.06$, $p < .001$, $d = .95$, $CI_{95\%} = [.75, 1.15]$, and negative contexts ($M = 4.94$, $SD = .84$), $t(134) = 8.09$, $p < .001$, $d = .70$, $CI_{95\%} = [.51, .88]$. Moreover, trustworthy faces embedded in neutral contexts were perceived as more trustworthy than trustworthy faces embedded in negative contexts, $t(134) = 5.96$, $p < .001$, $d = .51$, CI*95%* = [.33, .69].

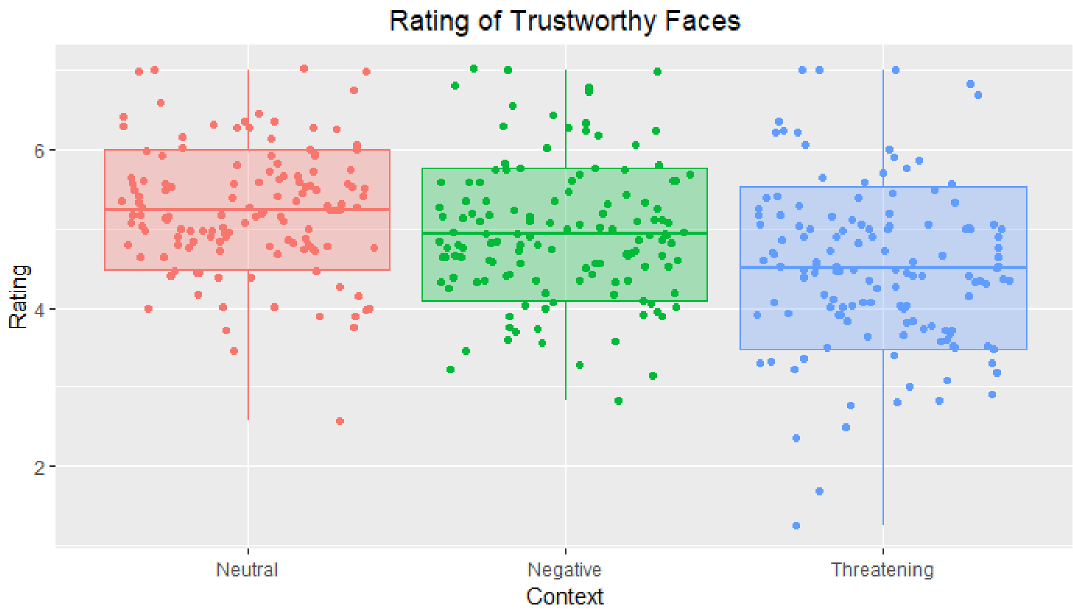

Figure 2. 4 Average trustworthy rating of trustworthy faces for each type of context (Experiment 1)

For untrustworthy faces (Figure 2.5), we found the same pattern. Faces embedded in threatening contexts (*M* = 3.06, *SD* = 1.03) were perceived as less trustworthy than faces in neutral contexts (*M* = 3.51, *SD* = 1.00), $t(134) = 8.07$, $p < .001$, $d = .69$, $CI_{95\%} = [.50, .88]$. At the same time, faces embedded in neutral contexts were perceived as more trustworthy than faces embedded in negative contexts (*M* = 3.14, *SD* = .98), $t(134) = 7.76$, $p < .001$, $d = .67$, $CI_{95\%} = [.48, .85]$. More importantly, the difference between threatening and negative (*M* = 3.14, *SD* = .98) context was statistically significant, $t(134) = 2.07$, $p = .04$, $d = .18$, $CI_{95\%} = [.01, .35]$.

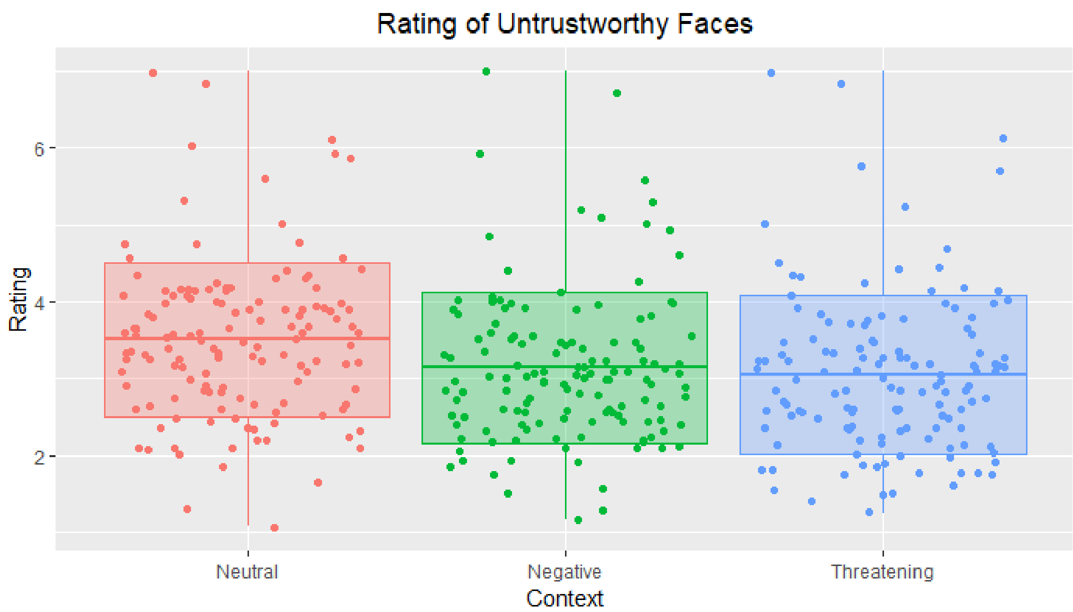

Figure 2. 5 Average trustworthy rating of untrustworthy faces for each type of context (Experiment 1)

2.3.3 Discussion

Considering the aforementioned findings, it seems that information drawn from the visual context in which the social target is encountered interfere with the trustworthiness judgement process. More precisely, valence and threat push the trustworthiness rating toward the congruent side of the judgement continuum. Interestingly, threat information seems to exert an effect independently from valence. In fact, in our experimental design, valence and threat are disentangled and the latter pushes trustworthiness rating a little bit further in the direction of the untrustworthy side of the continuum.

2.4 Study 2: The Impact of Contextual Information on the Categorization Process

In the second experiment, we tested if the results of the first experiment hold in a categorization task. We replicated Experiment 1 substituting the explicit rating task with a dichotomous categorization task. Such a modification has been introduced for two main reasons. The first is to move over the explicit rating paradigms, and the second is to begin to prepare the ground for our process-sensitive measure.

In this experiment, we relied on the same face stimuli and the same visual contexts we employed in Experiment 1. Our final stimuli are once again faces embedded in visual backgrounds. This time, by introducing the categorization task, we tested if the competition between the two categorization outcome is impacted by the information provided by the context (Freeman, Ma, Han & Ambady, 2013). In particular, we expect that threatening contexts interfere with the information drawn from the face stimulus disrupting the categorization process. Here, such as in Asch's work (1946), information drawn from the visual background interact with information drawn from the face stimulus.

Our hypothesis is that threat information drawn from threatening contexts impacts the categorization process by rising the probability of misclassifying trustworthy faces as untrustworthy. Our second hypothesis regards the fluidity of the categorization. In fact, we expect that even when correctly categorizing trustworthy faces as trustworthy, threat information present in the visual context interfere and slows down the categorization process, resulting in higher reaction times.

2.4.1 Method

2.4.1.1 Participants

In order to determine the required sample size, an a priori power analysis was conducted using G Power 3.1 (Faul, Erdfelder, Lang, & Buchner, 2007). The required sample size needed to detect an effect as the one we found in Experiment 1 ($\eta^2_p = .12$) with power of $(1-\beta) = .80$ (Cohen 1988) and a type 1 error of $\alpha = .05$ was N = 48 for our within-subjects experimental design.

To ensure a conservative sample size, we enrolled 72 Italian students, aged between 20 and 62 (*M* = 24.85, *SD* = 6.61), 54 females and 18 males. Participants took part to the experiment in exchange for course credit. The experimental procedure took part in the lab.

2.4.1.2 Procedure

Participants were told that they would be presented with images of individuals in various settings, and were asked to categorize each person as either trustworthy or untrustworthy. They were instructed to make their decisions as quickly and accurately as possible by clicking response buttons, basing their judgments on their first impressions. On every trial, participants clicked a "*Start*" button at the bottom-center of the screen, which was then replaced by a face-context pair in the center of the screen. Face-context pairs were presented in randomized order, and faces were categorized by clicking a "*trustworthy*" or "*untrustworthy*" response button located in the top-left and top-right corners of the screen (counterbalanced across participants). Participants performed the task self-paced, without timeout. However, if participants did not start moving the mouse within 250 milliseconds after the face-context pair appeared on the screen a message saying to start moving the mouse earlier appeared (for a similar procedure see Freeman, Ma, Han & Ambady, 2013). All 4 morphs of each one of the 6 identities were presented two times in each one of the 3 types of context. This permutation yielded 144 trials per participant. All faces were always embedded in the center location of a visual scene.

2.4.2 Results

Following prior research (see Righart & Gelder, 2008) we computed error rates and response times. Eight subjects had no valid trial in one or more cells and were therefore removed from the analyses.

Error rates (7.74%) were submitted to a 3 trustworthiness (trustworthy, untrustworthy) by 3 context type (neutral, negative, threatening) within-subject ANOVA. The analysis revealed a main effect of the context, $F(2,126) = 3.62$, $p = .03$, $\eta^2_p = .05$. Thus, participants made more errors when faces appeared in threatening scenes ($M = .118$, $SD = .020$) than in negative ($M = .084$, $SD = .017$), $t(126) = 2.27$, $p = .03$, $d = .20$, $CI_{95\%} = [.02, .38]$, and neutral scenes (*M* = .078, *SD* = .018), *t*(126) = 2.29, *p* = .03, *d* = .20, $CI_{95\%}$ = [.02, .38]. Errors did not differ between neutral and negative scene contexts, *t*(126) = .33, *p* = .73, *d* = .03, $CI_{95\%}$ = [-.14, .20].

The analysis also yielded a main effect of face, *F*(1,63) = 9.82, p = .003, η^2_p = .14 revealing that participants made more mistakes when confronted with trustworthy (*M* = .124, *SD* = .022) than untrustworthy faces (*M* = .062, *SD* = .016), *t*(63) = 3.15, *p* = .003, *d* = .39, $CI_{95\%}$ = [.14, .65].

More importantly, the analysis revealed an interaction effect between context and face, *F*(1,63) = 19.62, *p* = .001, η^2_p =.24. We decomposed such interaction analyzing error rates (Table 2.3) separately for trustworthy and untrustworthy faces.

	Trustworthy	Untrustworthy
Neutral	.040 (.016)[a]	.117 (.031)[a]
Negative	.119 (.026)[b]	.048 (.017)[b]
Threatening	.214 (.037)[c]	.021 (.010)[c]

Table 2. 3 Difference in error rates among neutral, negative, and threatening contexts. [a, b, c] superscripts indicate means that are statistically different across type of context

Participants were more likely to falsely categorize an untrustworthy face as trustworthy (Figure 2.6) when the face appeared in neutral context (*M* = .117, *SD* = .031) than in negative (*M* = .048, *SD* = .017), *t*(63) = 2.46, *p* = .01, *d* = .31, $CI_{95\%}$ = [.05, .56], and threatening contexts (*M* = .021, *SD* = .010), *t*(63) = 3.43, *p* < .001, *d* = .43, $CI_{95\%}$ = [.17, .68]. Moreover, participants were marginally more likely to falsely categorize an untrustworthy face as trustworthy when the face appeared in negative than in threatening scenes, *t*(63) = 1.8, *p* = .07, *d* = .22, $CI_{95\%}$ = [-.02, .47].

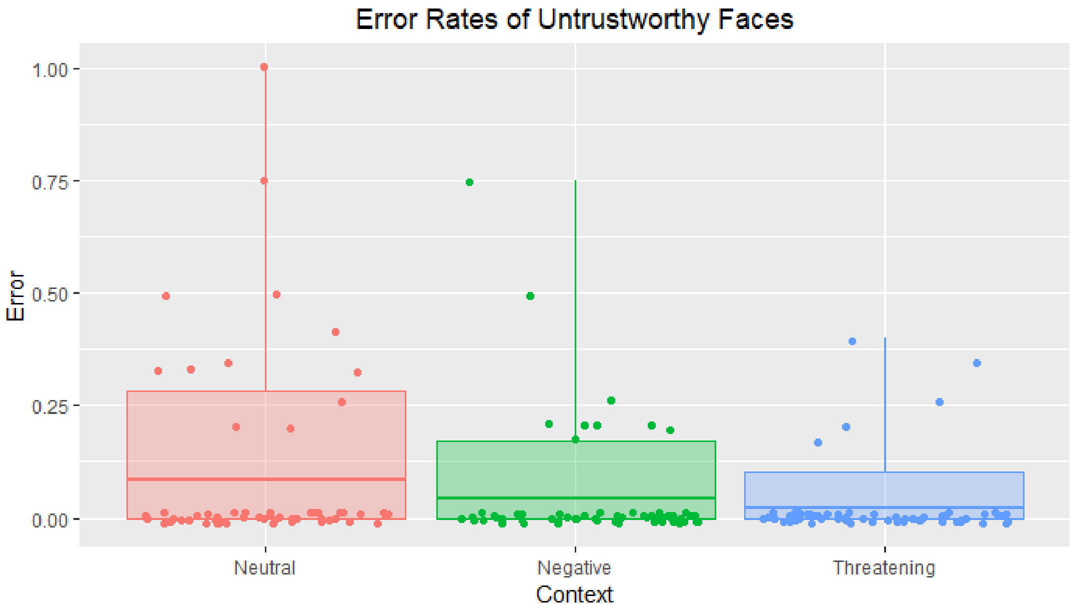

Figure 2. 6 Average error rate of untrustworthy faces for each type of context (Experiment 2)

On trustworthy faces we found the reverse pattern (Figure 2.7). Thus, participants were more likely to falsely categorize a trustworthy face as untrustworthy when the face appeared in threatening context (M = .214, SD = .037) than in negative (M = .119, SD = .026), $t(63)$ = 3.27, p = .001, d = .41, $CI_{95\%}$ = [.15, .66], and neutral contexts (M = .040, SD = .016), $t(63)$ = 4.83, p < .001, d = .60, $CI_{95\%}$ = [.33, .87]. Moreover, participants were more likely to falsely categorize a trustworthy face as untrustworthy when the face appeared in negative context scenes than in neutral contexts, $t(63)$ = 3.59, p < .001, d = .45, $CI_{95\%}$ = [.19, .70].

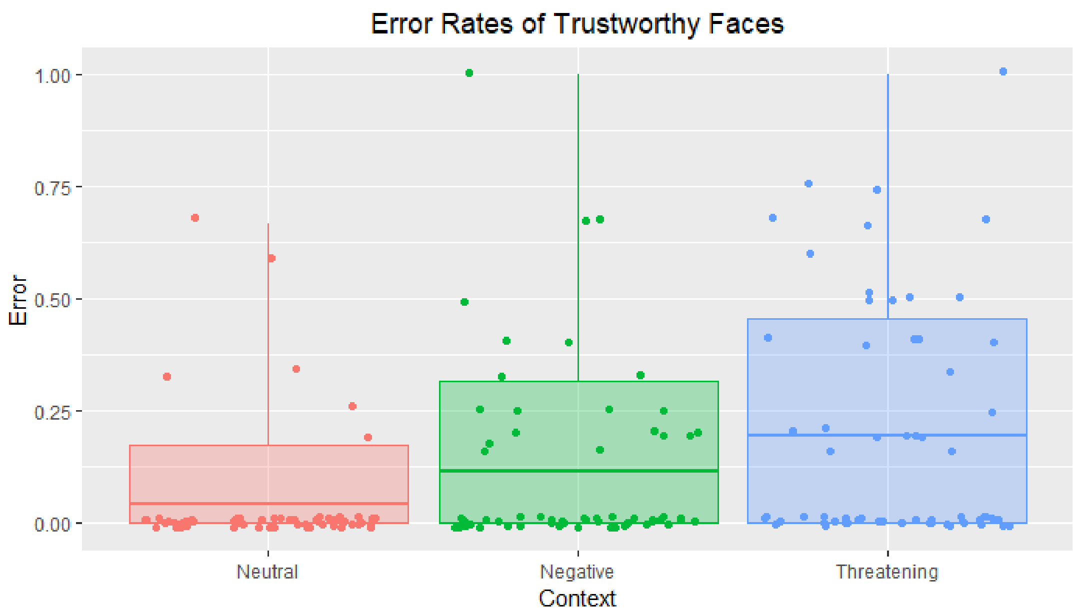

Figure 2. 7 Average error rate of trustworthy faces for each type of context (Experiment 2)

Taken together, these findings confirmed that participants were more accurate in categorizing untrustworthy faces when they were embedded in threatening contexts. By contrast, participants were less accurate in categorizing trustworthy faces when they were embedded in threatening contexts.

We next analyzed reaction times. To do so, we considered only correct responses and excluded from the analysis every trial where a trustworthy face was categorized as "untrustworthy" or where an untrustworthy face was categorized as "trustworthy". Nine subjects had no valid trial in one or more cells and were therefore removed from the analyses. We performed a 2 trustworthiness (trustworthy, untrustworthy) by 3 type of context (neutral, negative, threatening) within-subject ANOVA. The analysis yielded a main effect of context, $F(2,124) = 5.73$, $p = .02$, $\eta^2_p = .09$. Thus, participants were slower in categorizing faces when they were embedded in threatening contexts (M = 1282.624ms, SD = 17.71) than in neutral contexts (M = 1243.50ms, SD = 17.79), $t(62) = 3.85$, $p < .001$, d = .48, $CI_{95\%}$ = [.22, .74]. Reaction times did not differ between threatening and negative contexts (M = 1268.20ms, SD = 20.22), $t(62) = 1.25$, $p = .22$, $d = .16$, $CI_{95\%}$ = [-.09, .40]. However, participants were marginally slower in categorizing faces when they were embedded in negative (*M* = 1282.624ms, *SD* = 17.71) than in neutral contexts, $t(62) = 1.88$, $p = .06$, $d = .24$, $CI_{95\%} = [-.01, .49]$.

We did not find a main effect of trustworthiness ($p = .10$), although we found a significant interaction effect between the context type and trustworthiness, $F(2,124) = 16.48$, $p = .001$, $\eta^2_p = .24$. We decomposed such an interaction analyzing reaction times separately for trustworthy and untrustworthy faces (Table 2.4).

	Trustworthy	Untrustworthy
Neutral	1219.49 (20.05)^a	1267.51 (21.04)^a
Negative	1284.79 (24.30)^b	1251.61 (19.78)^a
Threatening	1327.26 (22.47)^c	1237.99 (18.63)^a

Table 2. 4 Difference in reaction times among neutral, negative, and threatening contexts. $^{a, b, c}$ superscripts indicate means that are statistically different across type of context

Reaction times did not differ when untrustworthy faces were embedded in threatening (*M* = 1237.99ms, *SD* = 18.63), negative (*M* = 1251.61ms, *SD* = 19.78) $t(62) = .97$, $p = .33$, $d = .12$, $CI_{95\%} = [-.12, .37]$, or neutral contexts (*M* = 1267.51ms, *SD* = 21.04), $t(62) = 1.72$, $p = .09$, $d = .22$, $CI_{95\%} = [-.03, .46]$. Moreover, reaction times did not differ between neutral and negative contexts, $t(62) = .93$, $p = .35$, $d = .12$, $CI_{95\%} = [-.13, .36]$ (Figure 2.8).

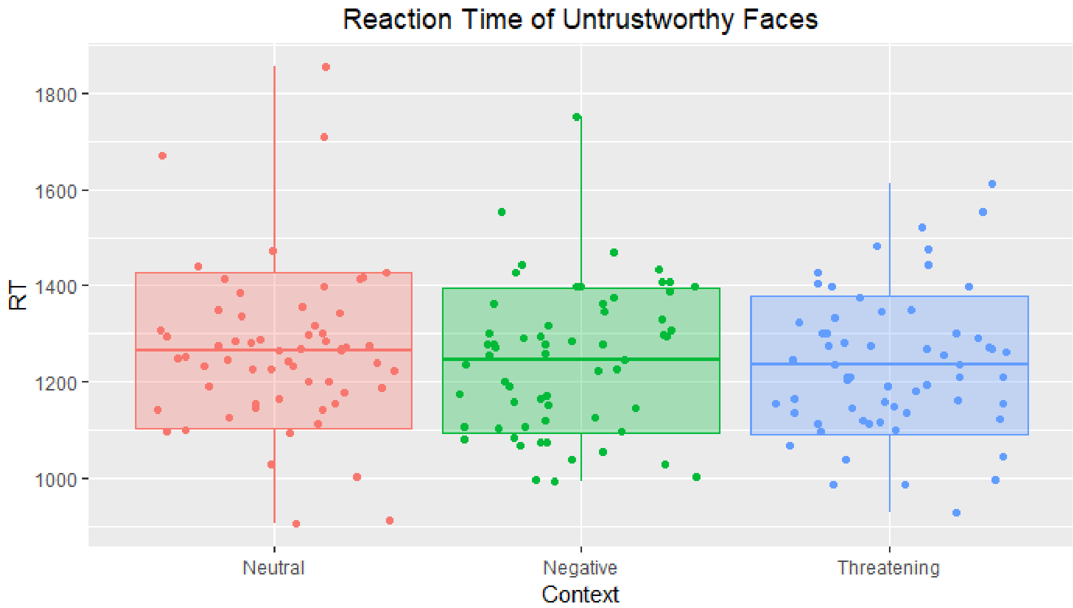

Figure 2. 8 Average reaction time of untrustworthy faces for each type of context (Experiment 2)

By contrast, participants identified trustworthy faces (Figure 2.9) slower when they were embedded in threatening contexts (*M* = 1327.26ms, *SD* = 22.47) than in neutral (*M* = 1219.49ms, *SD* = 20.05), $t(62) = 7.55$, $p < .001$, $d = .95$, $CI_{95\%} = [.65, 1.25]$, and in negative contexts (*M* = 1284.79ms, *SD* = 24.30), $t(62) = 2.31$, $p = .02$, $d = .29$, $CI_{95\%} = [.04, .54]$. Moreover, participants identified trustworthy faces slower when they were embedded in negative than in neutral backgrounds, $t(62) = 3.43$, $p = .001$, $d = .43$, $CI_{95\%} = [.17, .69]$.

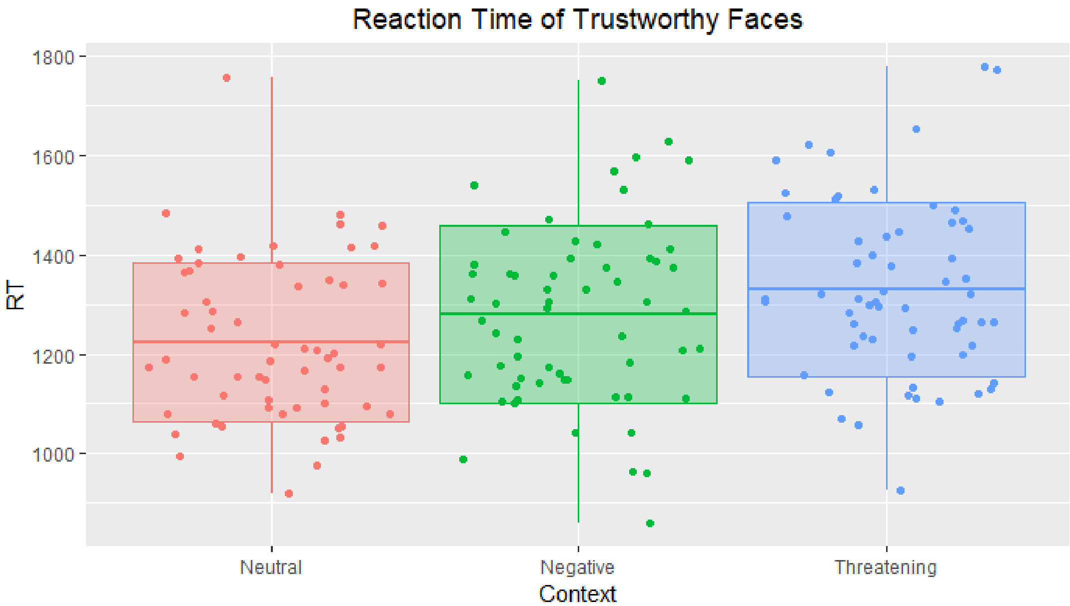

Figure 2. 9 Average reaction time of trustworthy faces for each type of context (Experiment 2)

This means that threat information slowed the correct categorization of trustworthy faces in comparison with non-threatening contexts. Such pattern of result was not found for untrustworthy faces.

2.4.3 Discussion

From the results of Experiment 2, it seems that information drawn from the visual context in which the target is embedded can impact the categorization outcome. In fact, we found main effects of context in both error rates and reaction times. Thus, threat information raises the number of categorization error and slows down the categorization process.

Crucially to our hypothesis, our results point out that trustworthiness cues and threat information drawn from the visual context interact with each other. This suggests that threat information may directly modify our impression of the perceived target or put differently, it seems that threat information is integrated with trustworthiness cues in the categorization process.

2.5 Conclusion

In two experiments, we investigated the impact of threat information on the judgement of trustworthiness and the categorization of social targets. Relying on classical outcome measure, gathered some insights about the how such information impacts the process outcome.

In both experiments, we found that threat information tends to be integrated with trustworthiness cues. Such integration induced a shift in the participants' perception. In fact, in Experiment 1, threat information drawn from contexts shifted participants' rating of the social target toward the untrustworthy side of the continuum. Ratings, the trustworthiness judgement outcome, were pushed toward the extreme of the judgement continuum that was coherent with the information provided by the context. The trend was found even for negative background pointing out that valence plays a role, but the difference between negative and threatening context was significant. Such difference obtained between two types of contexts that share the same valence, indicates that threat exerts and effect that goes beyond the effect of valence.

Similarly, in Experiment 2, we found evidence of the impact of threat information on the categorization outcome. This information drawn from threatening context induced more error in the categorization process when there was incongruence with trustworthiness cues. Moreover, in situation of incongruence, threat information slowed down the correct categorization of trustworthy targets. Both the raise in error rates and higher reaction times point out that the presence of threat information interferes with the categorization process and impact the categorization outcome. As in Experiment 1, the effects we found are added to mere valence effect. In fact, we found differences in the categorization outcomes between neutral and negative contexts. However, the difference between the two context that share the same valence but differ in terms of threat, namely negative and threatening contexts, points out that threat information exerts some unique effect on the categorization outcome. Such effect, despite going in the same direction of the effect of valence, is way beyond mere valence effect.

In both experiments, the findings suggest that threat information and trustworthiness are two types of information that share some overlapping. Clear congruence effects can be found in our data. However, the congruence between trustworthiness and threat facilitate both judgment and categorization of trustworthiness in social targets to a greater extent if compared with the congruence between valence and threat. This stable pattern of results suggests that threat information and trustworthiness provide similar insights to the perceiver. This is easy assume if we consider trustworthiness as an informative cue regard how socially threatening a target is, and the first two studies of the present work experimentally tested this assumption.

In sum, these two experiments provide initial evidence of integration between situational threat-specific information and target's trustworthiness. However, these experiments endeavored only the process' outcome. The experimental paradigms do not allow for inferences on the process dynamics. The impact of threat-specific information on trustworthiness perception and the overlapping we hypothesized between the two kind of information still need to be tested using process-sensitive experimental paradigms. This test is carried out in the next chapters.

CHAPTER 3

THREAT INFORMATION AND TRUSTWORTHINESS JUDGEMENT INTEGRATION: PROCESS-SENSITIVE MEASURES[1]

3.1 Introduction

In this chapter, the integration between threat information extracted from the visual context and judgements of trustworthiness about the target embedded in the context is investigated using a process-sensitive measure. Our main aim is to extend prior work by investigating whether visual context may impact the perception of trustworthiness.

In the previous experiments, we showed initial evidence that the visual background can impact the perception of trustworthiness. However, because of the use of explicit ratings and categorization decisions, we lack understanding of how facial and contextual cues are integrated during the judgment process. Here, we aimed to examine dynamics underlying the integration of facial trustworthiness and contextual cues, specifically contextual cues that convey threat. In doing so, our research is useful to broaden our understanding of the factors promoting or disrupting the processing of facial trustworthiness. By implying that facial trustworthiness and the perception of threat are inherently linked (for reviews, Brambilla & Leach, 2014; Todorov et al., 2015), we hypothesized that visual scenes associated with threat could alter the processing of a face's trustworthiness.

To test this prediction, we went beyond response times and considered a more process-sensitive methodology. Thus, we employed a mouse-tracking technique that records and analyzes hand movements during categorization tasks (Freeman & Ambady, 2010; Freeman & Johnson, 2016). Previous studies examining contextual effects suggest in some cases outcome-based measures (e.g., ratings or reaction times) may have limited sensitivity

[1] This chapter is based on a paper published in the *Journal of Experimental Social Psychology**

while more process-based measures such as mouse-tracking overcome this (Freeman et al., 2013; Freeman & Johnson, 2016). As such, there are many cases where a participant's ultimate perception is not predicted to be altered by context even if the process leading up to that response would be altered considerably. In line with this reasoning, the computer mouse-tracking procedure records the position of the mouse on the x and y coordinate space, providing an online measure of the spontaneous changes across a decision process. In a typical trial, participants are required to click on a “Start” button located at the bottom-center of the screen, which is replaced by a target. Participants then must click an appropriate response button located either at the top-left or top-right of the screen. Because the mouse is moving while a categorization response is still evolving, it is able to provide a “read-out” of how categorization unfolds over time (Freeman & Ambady, 2011; Freeman & Johnson, 2016). In other words, this paradigm can track how various cues drive categorization in real time and therefore reveal potentially subtle influences of context, even when an ultimate response may not be affected. If the visual context influences the categorization of facial trustworthiness, one would expect that perceivers partially integrate the response associated with the context with that associated with the face. This would be evidenced by a partial attraction in participants' mouse trajectories toward the opposite category response before clicking their final response when the facial and context information do not match. In other words, trajectories would be facilitated when facial cues and contextual cues are compatible (e.g., untrustworthy face in a threatening scene), and would be partially attracted to the context-associated response when incompatible (e.g., trustworthy face in a threatening scene). We conducted three experiments to test these hypotheses.

3.2 Study 3: The Impact of Contextual Information on the Ongoing Trustworthiness Categorization

Experiment 3 was designed as a first test of our hypothesis that categorization responses of facial trustworthiness are influenced by the threatening nature of the visual context. To do so, we asked participants to categorize the trustworthiness of faces that were shown against either threatening or neutral backgrounds. We predicted a more direct mouse-trajectory toward the untrustworthy response button when untrustworthy faces are embedded in threatening contexts rather than in a neutral context. By contrast, we expected a more curved mouse-trajectory toward the trustworthy response button when trustworthy faces are embedded in threatening contexts rather than in a neutral context.

3.2.1 Method

3.2.1.1 Participants

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

3.2.1.2 Stimuli

We employed the faces of the two previous studies. As well as for faces, scene context stimuli were the same as the experiments of chapter 2. To start with a simpler experimental design, we employed only neutral and threatening contexts (Figure 3.1).

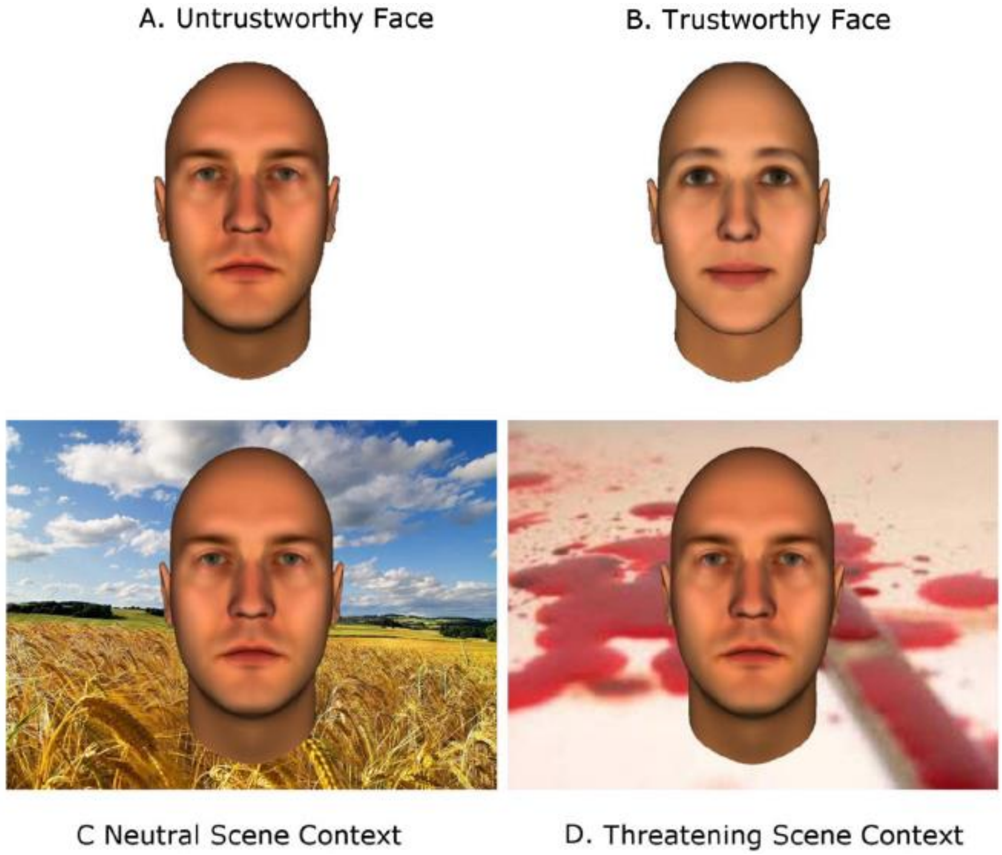

Figure 3. 1 Example of untrustworthy (a) and trustworthy (b) faces without context (upper), and of untrustworthy faces with a neutral (c) and threatening (d) contexts

3.2.1.3 Procedure

Participants were told that they would be presented with images of individuals in various settings, and were asked to categorize each person as either trustworthy or untrustworthy. They were instructed to make their decisions as quickly and accurately as possible by clicking response buttons, basing their judgments on their first impressions. Participants made speeded judgments and were asked to respond within 1500 ms. On every trial, participants clicked a “Start” button at the bottom-center of the screen, which was then replaced by a face-context pair in the center of the screen. Face-context pairs were presented in randomized order, and faces were categorized by clicking a “trustworthy” or "untrustworthy" response button located in the top-left and top-right corners of the screen (counterbalanced across participants). So as to encourage mouse trajectories that are online with the actual decision process, if participants did not start moving the mouse within 250 milliseconds after the face-context pair appeared on the screen, a message advising them to start moving the mouse earlier was displayed (Hehman, Stolier, & Freeman, 2015). Each face was presented 2 times and placed in the center location of a scene, 1 for each context type, yielding 48 trials per participant.

3.2.1.4 Results

To permit averaging and comparison across trials, we normalized trajectories into 101 time-steps and remapped leftward trajectories rightwards (inverted along the x-axis). To index trajectories attraction toward the opposite category, we computed the maximum deviation (MD): the largest perpendicular deviation from an idealized straight line between the trajectory's start and endpoints (Freeman & Ambady, 2010). We performed a 2 (Scene Context: Neutral, Threatening) × 2 (Face: Trustworthy, Untrustworthy) within-subject ANOVA. The main effect of scene context (F < 1, p = .64) and face (F < 1, p = .45) were not significant. However, there was a significant interaction between scene context and face, $F(1,50) = 9.61$, $p = .003$, $\eta^2_p = 0.16$ (Table 3.1).

	Trustworthy	Untrustworthy
Neutral	.35 (.23) [a]	.48 (.38)[a]
Threatening	.43 (.34) [b]	.37 (.28) [b]

Table 3. 1 Difference in MD. [a, b] superscripts indicate means that are statistically different across type of context

Specifically, untrustworthy faces elicited more direct trajectories (lower MD) when they were embedded in threatening (*M* = .37, *SD* = .28) than neutral (*M* = .48, *SD* = .38) contexts, $F(1,50) = 5.09$, $p = .03$, $\eta^2_p = 0.09$ (Figure 3.2).

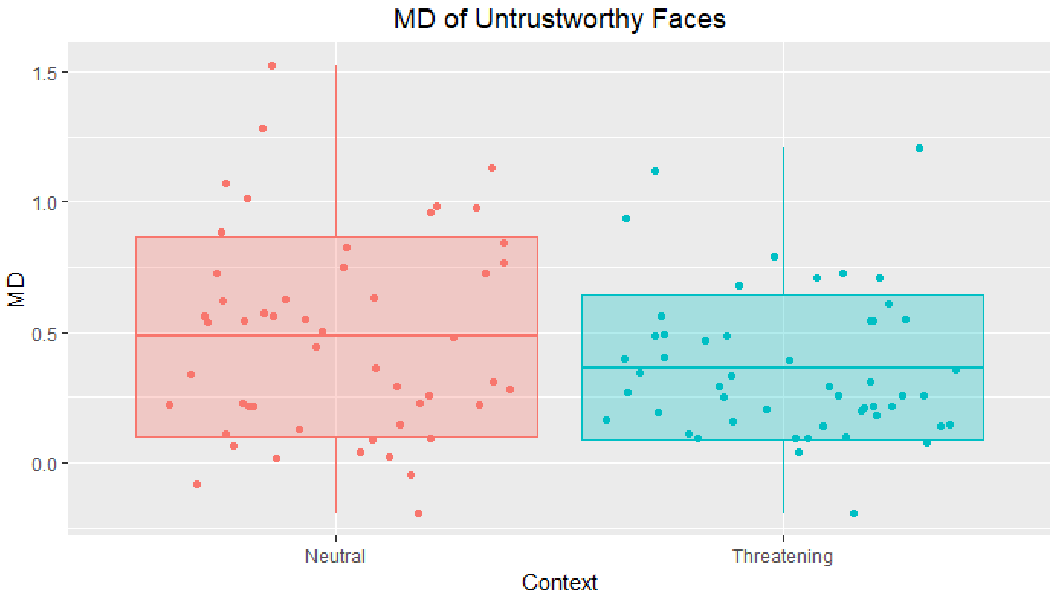

Figure 3. 2 Boxplots of average MD for untrustworthy faces for each type of context (Experiment 3)

Conversely, trustworthy faces exhibited a marginally significant tendency to elicit more deviating trajectories when they were embedded in threatening (*M* = .43, *SD* = .34) than neutral (*M* = .35, *SD* = .23) contexts, *F*(1,50) = 3.53, *p* = .067, η^2_p = 0.07 (Figure 3.3).

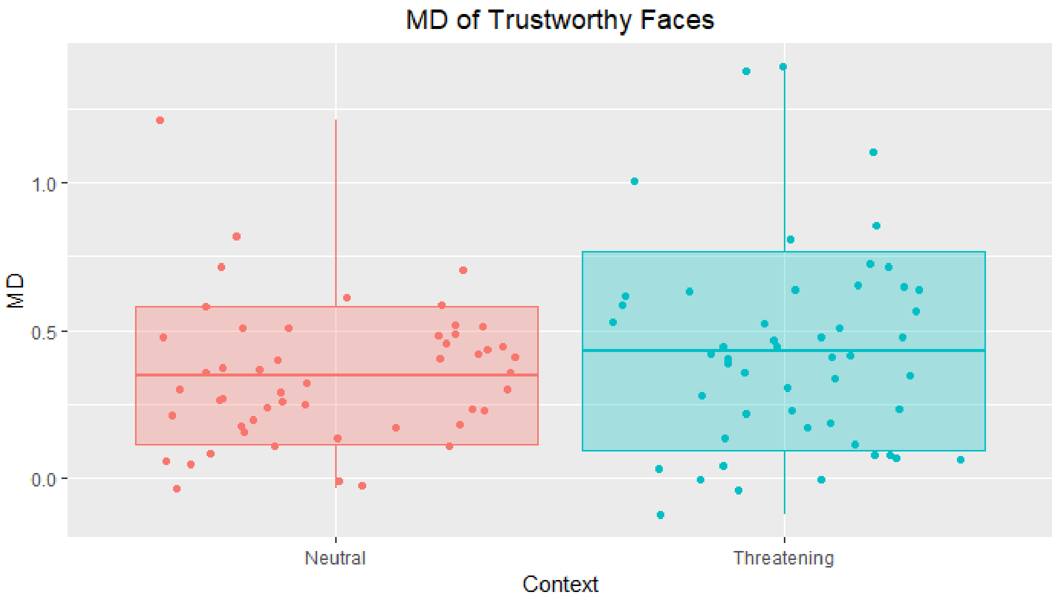

Figure 3. 3 Boxplots of average MD for trustworthy faces for each type of context (Experiment 3)

Next, we computed the area under the curve (AUC): the area between the observed trajectory and an idealized straight-line trajectory (Freeman & Ambady, 2010), which is a related measure to MD but in some cases exhibits higher sensitivity (Hehman et al., 2015). We performed a 2 (Scene Context: Neutral, Threatening) × 2 (Face: Trustworthy, Untrustworthy) within-subject ANOVA (Table 2). The main effect of scene context ($F < 1$, $p = .74$) and face ($F < 1$, $p = .40$) were not significant. More importantly, the scene context × face interaction was significant, $F(1,50) = 12.95$, $p = .001$, $\eta^2_p = 0.21$ (Table 3.2).

	Trustworthy	Untrustworthy
Neutral	.54 (.48)[a]	.94 (.93)[a]
Threatening	.81 (.88)[b]	.61 (.62)[b]

Table 3. 2 Difference in AUC. [a,b] indicates means that are statistically different across type of context

Untrustworthy faces elicited more direct trajectories (lower AUC) when they were embedded in threatening ($M = .61$, $SD = .62$) contexts than in neutral ($M = .94$, $SD = .93$) contexts, $F(1,50) = 5.99$, $p = .02$, $\eta^2_p = 0.11$ (Figure 3.4).

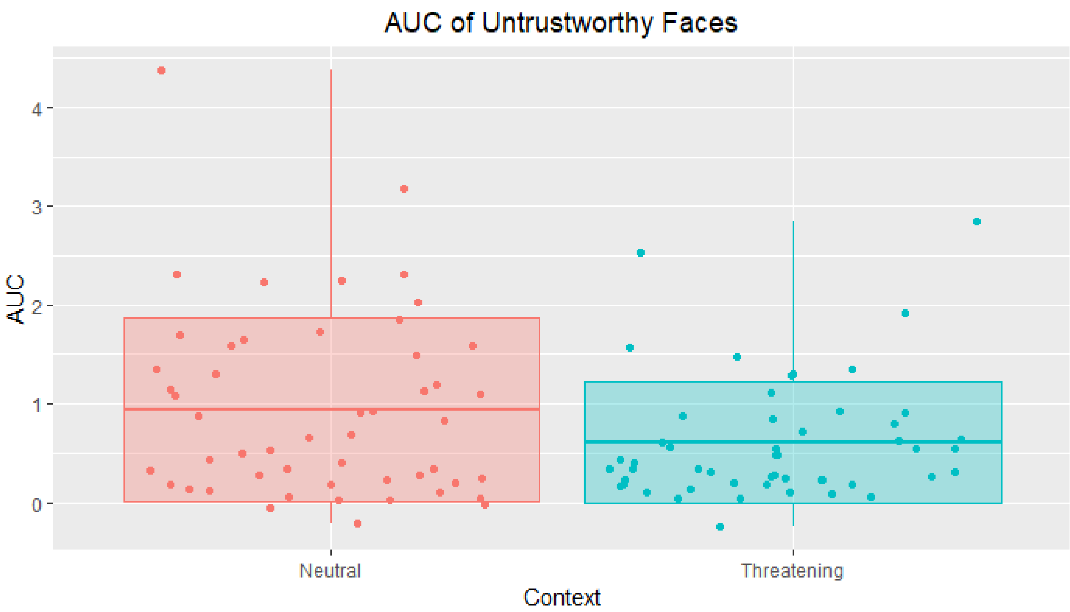

Figure 3. 4 Boxplots of average AUC for untrustworthy faces for each type of context (Experiment 3)

Conversely, trustworthy faces elicited more curved trajectories (higher AUC) when they were embedded in threatening (*M* = .81, *SD* = .88) contexts than in neutral (*M* = .54, *SD* = .48) contexts, $F(1,50) = 5.96$, $p = .02$, $\eta^2_p = 0.11$ (Figure 3.5).

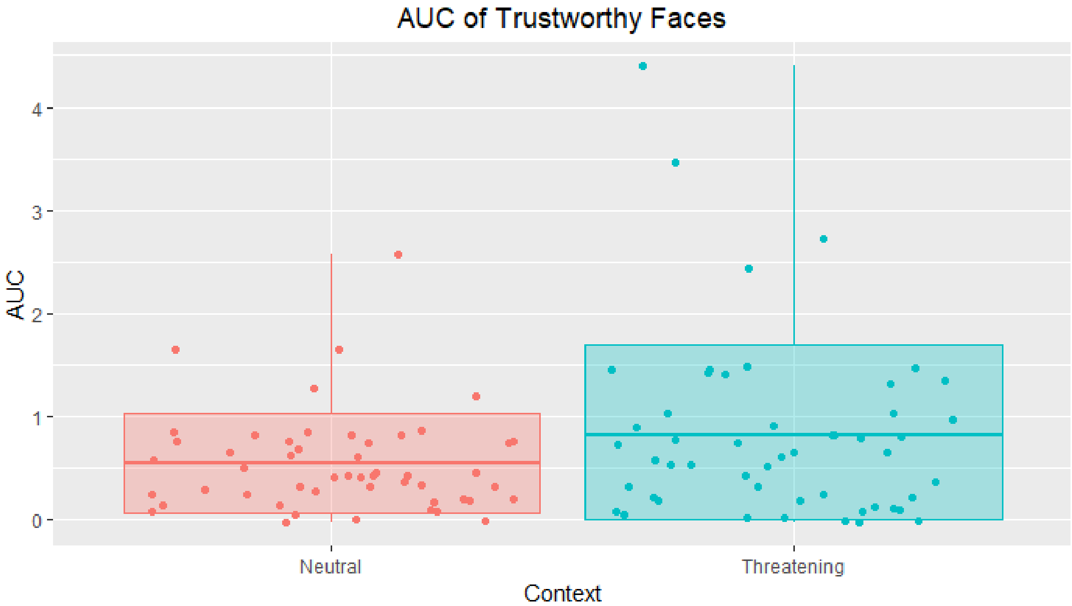

Figure 3. 5 Boxplots of average AUC for trustworthy faces for each type of context (Experiment 3)

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

3.3 Study 4: Threat Specificity on the Integration of Contextual Information and Trustworthiness Perception

Experiment 4 was designed to replicate and extend the findings of Experiment 3 by investigating whether the effects we found are specific to threatening contexts or indicate more general effects of negative scene contexts. To do so, we included a further experimental condition and asked participants to categorize the trustworthiness of faces that were shown against either threatening, negative but unthreatening, or neutral backgrounds. Specifically, we predicted a more direct trajectory toward the untrustworthy response when untrustworthy faces are embedded in a threatening rather than a neutral context or negative context unrelated to threat. By contrast, we expected a more curved trajectory toward the trustworthy response when trustworthy faces are embedded in a threatening rather than a neutral context or negative context unrelated to threat, indicating a partial attraction to the untrustworthy response and an integration of facial and contextual cues.

3.3.1. Method

3.3.1.1. Participants

Sample size was determined before the data collection. An a priori power analysis was conducted for sample size estimation. The projected sample size needed to detect a small-to-medium effect size with 80% power is N = 28 for a within-subject ANOVA. We advertised the study on campus and all students who responded within 4 weeks and who were not involved in Experiment 3 took part to the study. Overall, we recruited 46 Italian students (33 female) aged between 19 and 49 (M_{age} = 22.57, *SD* = 4.78), with normal or corrected-to-normal vision. Most participants (95.7%) were right handed.

3.3.1.2. Stimuli

We used the same 24 computer-generated identities (12 trustworthy, 12 untrustworthy) and the same visual backgrounds of Experiment 3. Four negative visual background were added, obtaining a total of 12 scene context stimuli (4 neutral, 4 negative, and 4 threatening). The negative background were the same employed in the experiments of chapter 2 (Figure 3.6).

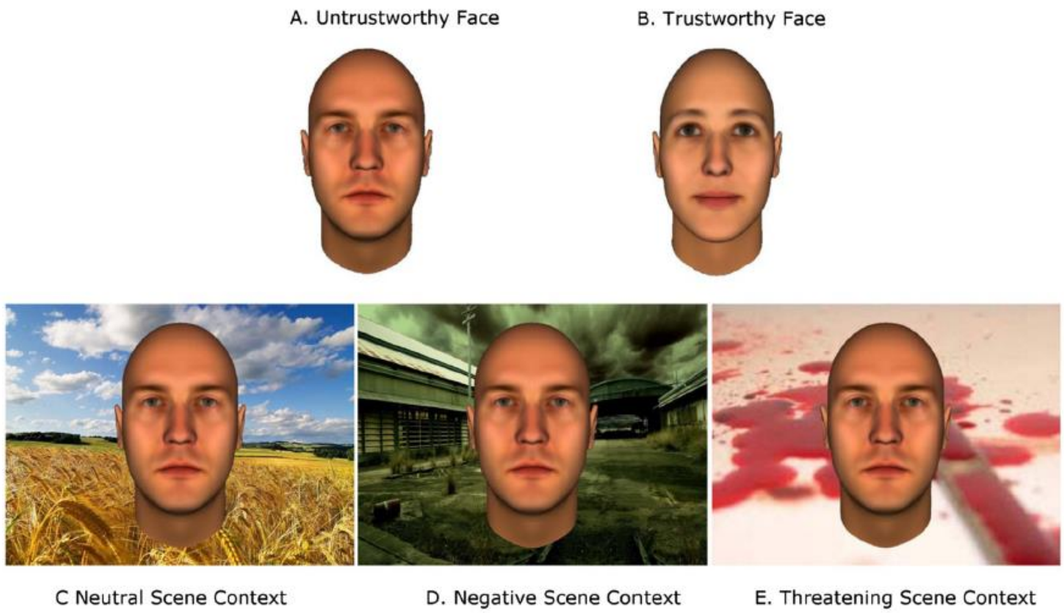

Figure 3. 6 Example of untrustworthy (a) and trustworthy (b) faces without context (upper), and of untrustworthy faces with a neutral (c), negative (d), and threatening (e) contexts

3.3.1.3. Procedure

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

3.3.2. Results

Following the procedure of Experiment 3, we normalized trajectories into 101 time-steps and remapped leftward trajectories rightwards (inverted along the x-axis). To index trajectories deviation toward the opposite category, we computed MD. We performed a 3 (Scene Context: Neutral, Negative, Threatening) × 2 (Face: Trustworthy, Untrustworthy) within-subject ANOVA. The main effect of scene context was not significant, $F(2,88) = 0.003$, $p = .99$, $\eta^2_p = 0.001$. However, the main effect of face was significant, $F(1,44) = 10.86$, $p = .002$, $\eta^2_p = 0.20$, indicating that trajectories exhibited greater deviation overall when participants evaluated untrustworthy relative to trustworthy faces. More importantly, the analysis revealed a significant interaction between scene context and face, $F(2,88) = 12.38$, $p < .001$, $\eta^2_p = 0.22$ (Table 3.3).

	Trustworthy	Untrustworthy
Neutral	.22 (.21)[a]	.42 (.28)[a]
Negative	.24 (.19)[a]	.39 (.24)[a]
Threatening	.32 (.21)[b]	.32 (.21)[b]

Table 3. 3 Difference in MD. [a, b, c] superscripts indicate means that are statistically different across type of context

Specifically, untrustworthy faces elicited more direct trajectories (lower MD) when they were embedded in threatening (M = .32, SD = .21) contexts than in negative (M = .39, SD = .24), $t(45) = 2.31$, $p = .03$, $d = 0.34$, $CI_{95\%} = [0.04, 0.63]$, and neutral (M = .42, SD = .28) $t(45) = 2.53$, $p = .02$, $d = 0.37$, $CI_{95\%} = [0.07, 0.67]$ contexts. However, MD scores did not differ between neutral and negative contexts, $t(45) < 1$, $p = .40$, $d = 0.12$ (Figure 3.7).

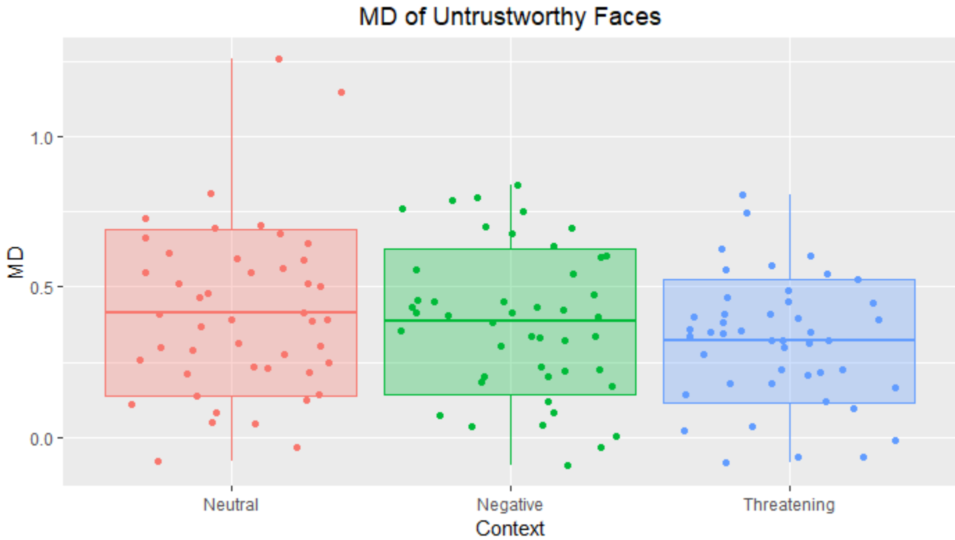

Figure 3. 7 Boxplots of average MD for untrustworthy faces for each type of context (Experiment 4)

Conversely, trustworthy faces elicited more curved trajectories when they were embedded in threatening (M = .32, SD = .21) contexts than in negative (M = .24, SD = .19), $t(44) = 2.92$, $p = .006$, $d = 0.44$, $CI_{95\%} = [0.12, 0.73]$, and neutral contexts (M = .22, SD = .21), $t(44) = 3.58$, $p = .001$, $d = 0.53$, $CI_{95\%} = [0.21, 0.84]$. However, MD scores did not differ between neutral and negative contexts, $t(45) = 1.03$, $p = .31$, $d = 0.16$, $CI_{95\%} = [-0.13, 0.44]$ (Figure 3.8).

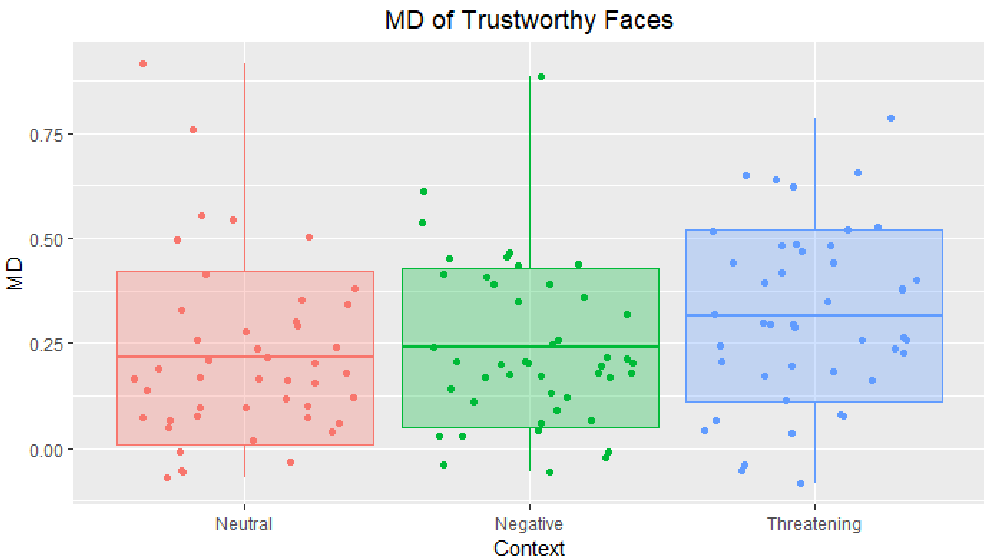

Figure 3. 8 Boxplots of average MD for trustworthy faces for each type of context (Experiment 4)

As in Experiment 3, next we computed the related AUC measure. We performed a 3 (Scene Context: Neutral, Negative, Threatening) × 2 (Face: Trustworthy, Untrustworthy) within-subject ANOVA. The analysis did not yield a main effect of scene context *F*(2,88) = 0.11, *p* = .90, η^2_p = 0.003. However, the main effect of face was significant, *F*(1,44) = 11.43, *p* = .002, η^2_p = 0.21, indicating a greater curvature overall for untrustworthy relative to trustworthy faces. More importantly, the scene context × face interaction was significant, *F*(1,44) = 11.36, *p* = .002, η^2_p = 0.21 (Table 3.4).

	Trustworthy	Untrustworthy
Neutral	.37 (.41)[a]	.79 (.61)[a]
Negative	.41 (.38)[a]	.71 (.48)[a]
Threatening	.56 (.41)[b]	.55 (.39)[b]

Table 3. 4 Difference in AUC. [a, b, c] superscripts indicate means that are statistically different across type of context

Untrustworthy faces elicited more direct trajectories (lower AUC) when they were embedded in threatening (*M* = .55, *SD* = .39) contexts than in negative (*M* = .71, *SD* = .48), *t*(45) = 2.23, *p* = .03, *d* = 0.33, $CI_{95\%}$ = [0.03, 0.62], and neutral contexts (*M* = .79, *SD* = .61), *t*(45) = 2.51, *p* = .02, *d* = 0.37, $CI_{95\%}$ = [0.06, 0.66]. However, AUC scores did not differ between neutral and negative contexts, *t*(45) = 1, *p* = .32, *d* = 0.15, $CI_{95\%}$ = [-0.14, 0.43] (Figure 3.9).

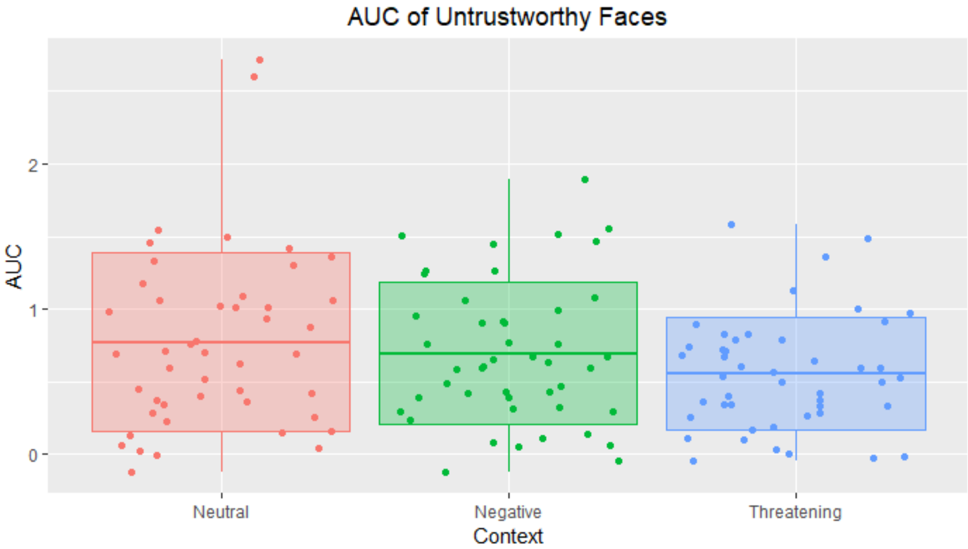

Figure 3. 9 Boxplots of average AUC for untrustworthy faces for each type of context (Experiment 4)

Trustworthy faces elicited more curved trajectories (higher AUC scores) when they were embedded in threatening (M = .56, SD = .41) contexts than in negative (M = .41, SD = .38), $t(44)$ = 2.52, p = .02, d = 0.38, $CI_{95\%}$ = [0.07, 0.67] and neutral contexts (M = .37, SD = .41), $t(44)$ = 3.06, p = .004, d = 0.46, $CI_{95\%}$ = [0.14, 0.76]. However, AUC scores did not differ between neutral and negative contexts, $t(45) < 1$, p = .35, d = 0.14 (Figure 3.10).

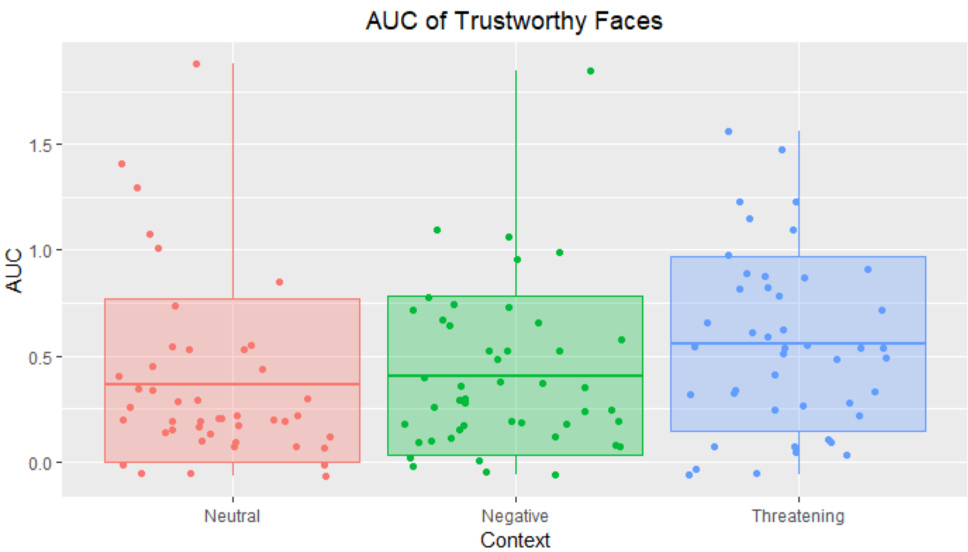

Figure 3. 10 Boxplots of average AUC for trustworthy faces for each type of context (Experiment 4)

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

3.4 Study 5: Information Integration and Threat specificity with More Realistic Stimuli

Experiment 5 aimed at replicating and extending the findings of Experiment 4 by increasing the realism of our stimuli. Indeed, in Experiment 3 and Experiment 4 we used disembodied faces without hair that floated over scenes. In Experiment 5 we added hairlines to the faces and embedded the facial stimuli in the visual contexts more naturalistically.

3.4.1 Method

3.4.1.1 Participants

For the recruitment of participants, we aimed at collecting as many subjects as possible over the number indicated by the power analysis of Experiment 4. We advertised the study on campus and all the students who responded within 4 weeks and that were not involved in Experiment 3 and Experiment 4 took part to the study. Overall, we recruited 50 Italian students, with normal or corrected-to-normal vision (19 male, M_{age} = 22.34, SD = 1.73).

3.4.1.2. Stimuli

We used the same 24 computer-generated identities (12 trustworthy, 12 untrustworthy) employed in the previous two experiments. However, the facial stimuli were modified by using Photoshop. Thus, we added hairs, necks, and shoulders to the faces in order to develop more realistic stimuli and integrate facial and contextual stimuli more naturalistically (Figure 3.11).

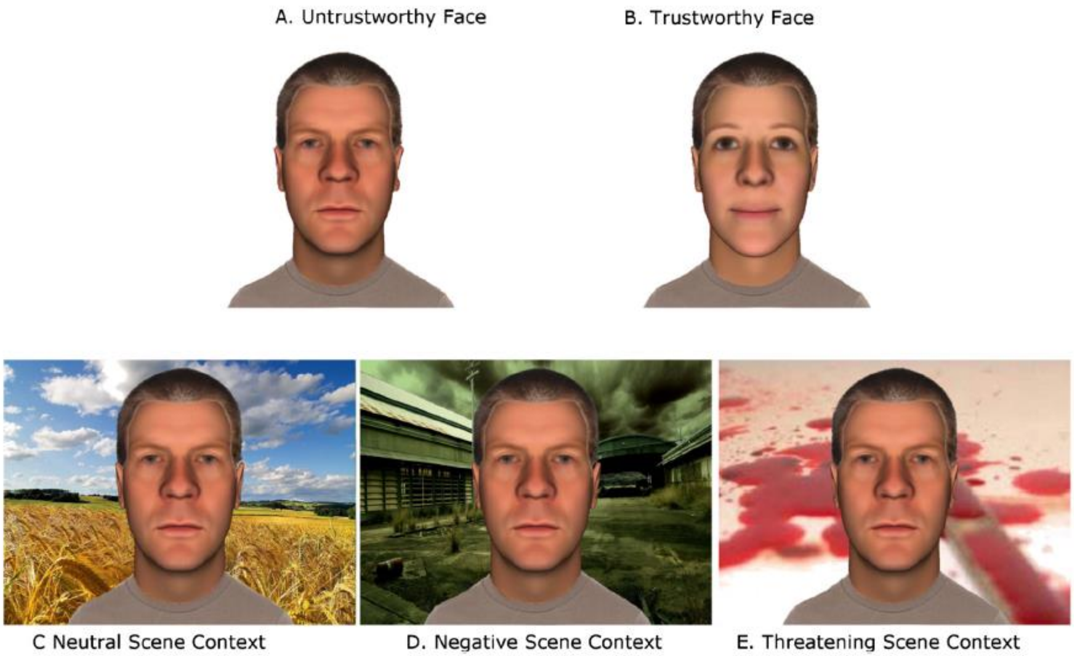

Figure 3. 11 Example of edited untrustworthy (a) and trustworthy (b) faces without context (upper), and of untrustworthy faces with a neutral (c), negative (d), and threatening (e) contexts

3.4.1.3 Procedure

Following the procedure of the previous experiments, participants were told that they would be presented with images of individuals in various settings, and were asked to categorize each person as either trustworthy or untrustworthy. They were instructed to make their decisions as quickly and accurately as possible by clicking response buttons, basing their judgments on their first impressions. The mouse-tracking procedure was carried out identically as in Experiment 3 and Experiment 4. To increase the reliability of our findings, we increased the number of trials: each face was presented 6 times and placed in the center location of a scene 2 for each context type, yielding 144 trials per participant.

3.4.2. Results

We first normalized trajectories into 101 time-steps and remapped leftward trajectories rightwards (inverted along the x-axis). To index trajectories deviation toward the opposite category, we computed MD. We performed a 3 (Scene Context: Neutral, Negative, Threatening) × 2 (Face: Trustworthy, Untrustworthy) within-subject ANOVA. In line with our hypothesis, the analysis revealed a significant interaction between scene context and face, $F(2,98) = 10.82$, $p < .001$, $\eta^2_p = 0.18$ (Table 3.5).

	Trustworthy	Untrustworthy
Neutral	.31 (.23)[a]	.42 (.25)[a]
Negative	.34 (.23)[a]	.39 (.24)[a]
Threatening	.40 (.27)[b]	.36 (.23)[b]

Table 3. 5 Difference in MD. [a, b, c] superscripts indicate means that are statistically different across type of context

Specifically, untrustworthy faces elicited more direct trajectories (lower MD) when they were embedded in threatening (*M* = .36, *SD* = .23) contexts than in neutral (*M* = .42, *SD* = .25) contexts, $t(49) = 3.31$, $p = .002$, $d = 0.47$, $CI_{95\%} = [0.17, 0.75]$, and negative (*M* = .39, *SD* = .24) contexts, $t(49) = 1.75$, $p = .08$, $d = 0.24$, $CI_{95\%} = [-0.03, 0.53]$, although the latter effect reached only marginal significance. However, MD scores did not differ between neutral and negative contexts, $t(49) < 1$, $p = .43$, $d = 0.11$ (Figure 3.12).

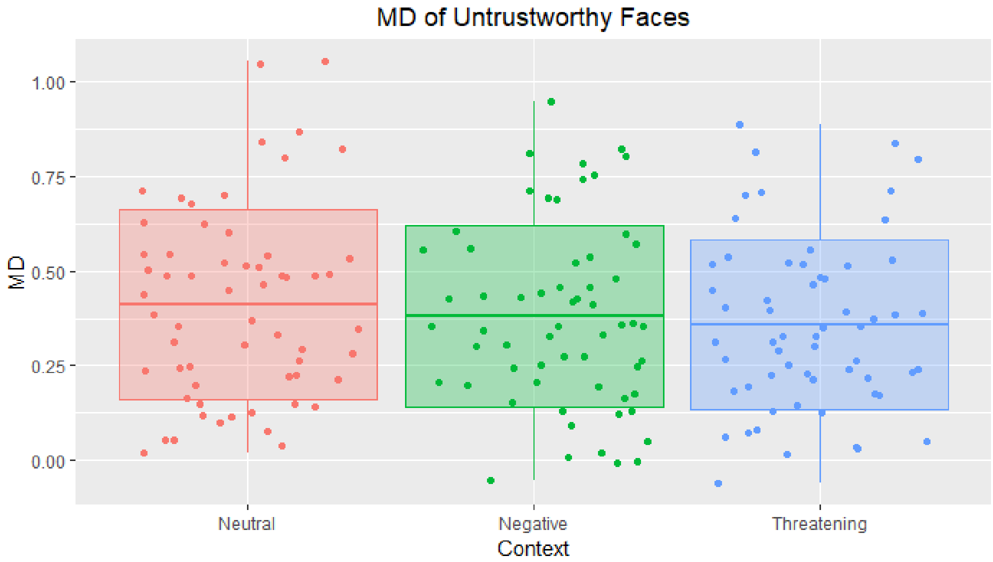

Figure 3. 12 Boxplots of average MD for untrustworthy faces for each type of context (Experiment 5)

Conversely, trustworthy faces elicited more curved trajectories when they were embedded in threatening (*M* = .40, *SD* = .27) contexts than neutral (*M* = .31, *SD* = .23), *t*(49) = 3.94, *p* = .001, *d* = 0.55, $CI_{95\%}$ = [0.26, 0.85], and negative contexts, *t*(49) = 2.65, *p* = .01, *d* = 0.37, $CI_{95\%}$ = [0.09, 0.66]. However, MD scores did not differ between neutral and negative (*M* = .34, *SD* = .23) contexts, *t*(49) = 1.40, *p* = .17, *d* = 0.20, $CI_{95\%}$ = [-0.08, 0.47] (Figure 3.13).

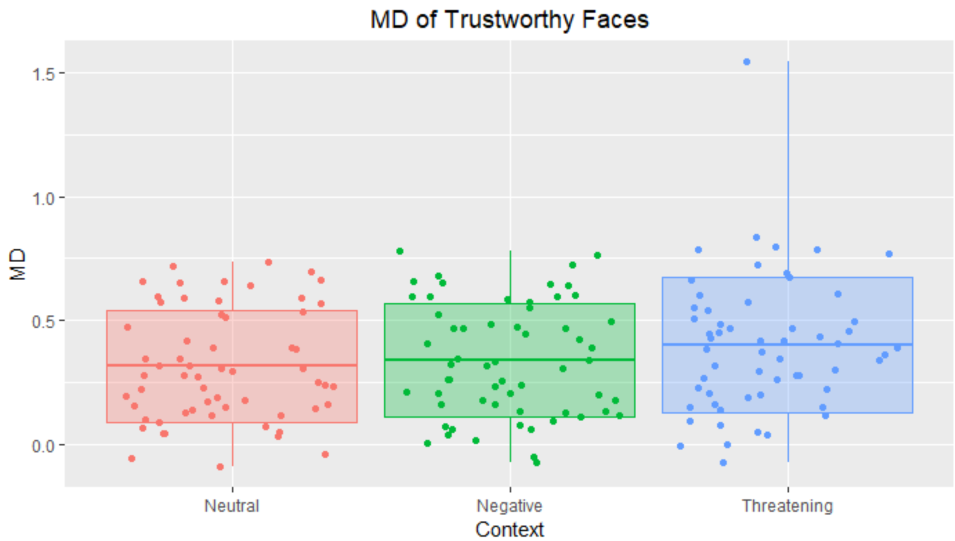

Figure 3. 13 Boxplots of average MD for trustworthy faces for each type of context (Experiment 5)

We also computed the AUC measure. We performed a 3 (Scene Context: Neutral, Negative, Threatening) × 2 (Face: Trustworthy, Untrustworthy) within-subject ANOVA. The analysis showed that the scene context × face interaction was significant, $F(1,98) = 11.20$, $p = .002$, $\eta^2_p = 0.19$ (Table 3.6).

	Trustworthy	Untrustworthy
Neutral	.56 (.45)[a]	.81 (.57)[a]
Negative	.63 (.48)[a]	.74 (.57)[a]
Threatening	.78 (.67)[b]	.65 (.48)[b]

Table 3. 6 Difference in AUC. a, b, c superscripts indicate means that are statistically different across type of context

Untrustworthy faces elicited more direct trajectories (lower AUC) when they were embedded in threatening ($M = .65$, $SD = .48$) contexts than in negative ($M = .74$, $SD = .57$) contexts, $t(49) = 2.17$, $p = .035$, $d = 0.31$, $CI_{95\%} = [0.02, 0.60]$ and neutral ($M = .81$, $SD = .57$) contexts, $t(49) = 3.42$, $p = .001$, $d = 0.48$, $CI_{95\%} = [0.19, 0.77]$. However, AUC scores did not differ between neutral and negative contexts, $t(49) = 0.54$, $p = .60$, $d = 0.07$, $CI_{95\%} = [-0.20, 0.35]$ (Figure 3.14).

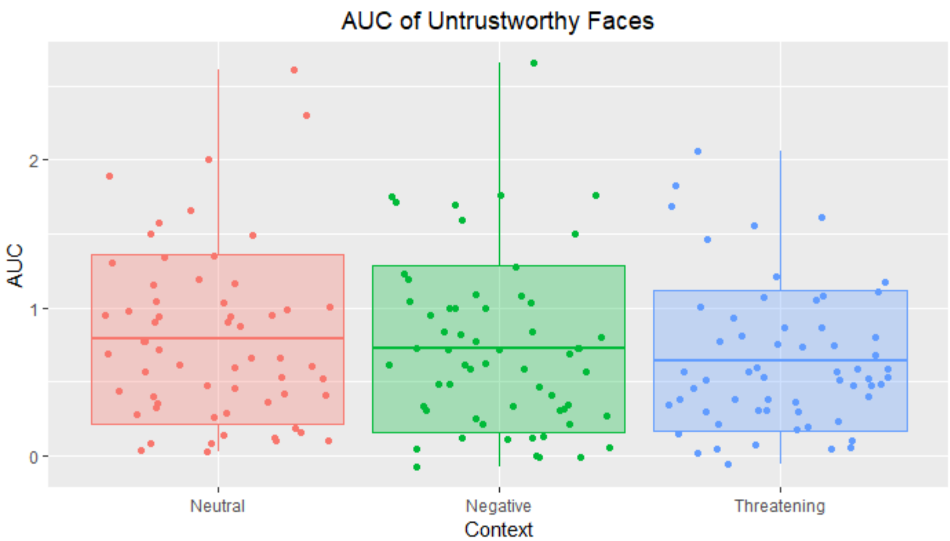

Figure 3. 14 Boxplots of average AUC for untrustworthy faces for each type of context (Experiment 5)

Trustworthy faces elicited more curved trajectories (higher AUC scores) when they were embedded in threatening (*M* = .78, *SD* = .67) contexts than in negative, $t(49) = 2.60$, $p = .01$, $d = 0.37$, $CI_{95\%} = [0.08, 0.65]$ and neutral (*M* = .56, *SD* = .45) contexts, $t(49) = 4.00$, $p = .001$, $d = 0.57$, $CI_{95\%} = [0.26, 0.86]$. However, AUC scores did not differ between neutral and negative (*M* = .63, *SD* = .48) contexts, $t(49) = 1.70$, $p = .10$, $d = 0.24$, $CI_{95\%} = [-0.04, 0.52]$ (Figure 3.15).

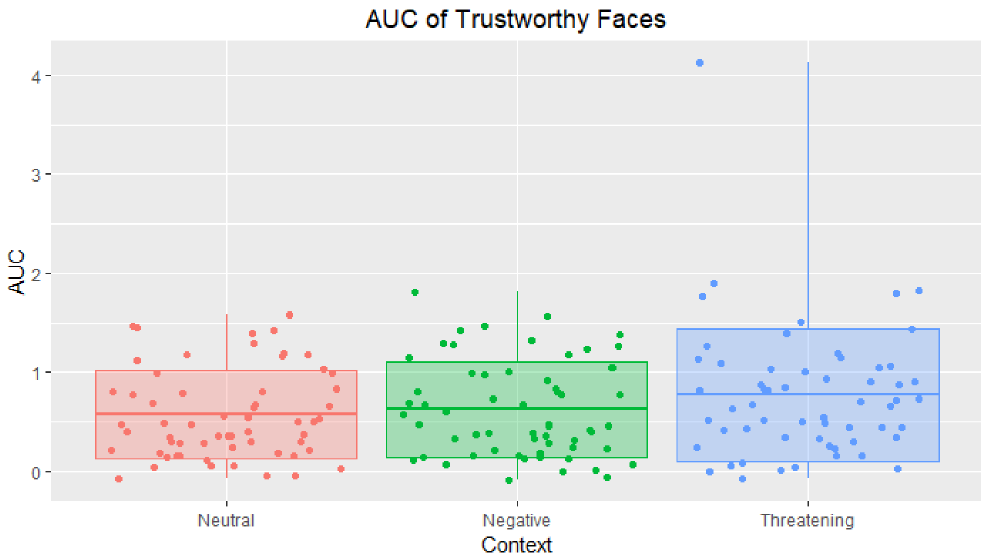

Figure 3. 15 Boxplots of average AUC for trustworthy faces for each type of context (Experiment 5)

Taken together, the findings replicated the findings of Experiment 4 and further show that the visual context biases the categorization of facial trustworthiness.

3.6 General discussion

Three experiments showed that the scene in which a face is encountered alters trustworthiness evaluation. By adopting a mouse-tracking technique, Experiment 3 showed that the visual context temporally influenced the evaluation of facial trustworthiness as revealed by a partial attraction in participants' mouse trajectories toward the opposite category response when the facial and contextual information were incompatible.

Moreover, when compatible, the trustworthiness evaluation process was facilitated. More direct trajectories were observed when untrustworthy faces were shown in threatening rather than neutral scenes, whereas more curved trajectories were observed when trustworthy faces were shown in threatening rather than neutral scenes. Experiment 4 corroborated these findings in a design that enabled us to disentangle the effects of threatening scene contexts from negative contexts in general. Results of this study confirmed that untrustworthiness and threat are inherently associated, as trajectories were more direct when untrustworthy faces were shown in threatening rather than in negative and neutral scenes. Conversely, trajectories were more curved when trustworthy faces were surrounded by threatening rather than negative and neutral scenes. Experiment 5 corroborated these findings by using a different set of more realistic stimuli. Thus, contextual information was represented in parallel and partially integrated into trustworthiness evaluation, even when an ultimate perception was not altered.

One limitation of the present work of potential concern to readers is that participants evaluated trustworthiness in a dichotomous, forced-choice design. This was chosen to be consistent with the standard mouse-tracking paradigm, but naturally one may ask whether the effects obtained may reflect some kind of artifact of the task. The pre-test data reported in chapter 2, however, which demonstrated a very strong correlation (r = 0.96) between dichotomous, forced-choice responses as used here and continuous ratings of trustworthiness speak against this possibility (see 2.2.2 Materials: Visual backgrounds in chapter 2). Nevertheless, future work could explore the generalizing of these contextual effects using different response sets or different stimuli, including the possibility of conducting mouse-tracking using a continuous scale.

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

Our data further show that threatening scenes promoted and disrupted the categorization of untrustworthy and trustworthy faces, respectively. Since we did not find any difference between negative and neutral contexts in promoting the categorization of trustworthy faces, and intriguing challenge for future research would be to test whether positive (rather than neutral) visual scenes or visual scenes priming positive moral concepts may foster the categorization of trustworthy faces. Such studies could complement our approach and help to gain more insights on the specific conditions in which context affects trait inferences of others' faces.

CHAPTER 4

THREAT INFORMATION AND TRUSTWORTHINESS JUDGEMENT INTEGRATION: MULTI-CHANNEL INTEGRATION

4.1 Introduction

In the present chapter, we aim at extending the findings on the integration of trustworthiness information with threat information. Specifically, we aim at endeavoring at which level such integration take place. There are two possible alternatives. The first is that all information are integrated at high cognitive level. In this case, information is extracted from the relative sources by the perceptual system and then integrated in order to form a unique impression. The second alternative is that information are both processed and integrated at a relatively low cognitive level. In this case the information is integrated directly by the perceptual system. The main difference between the two possible alternative is that in the first case information is easily integrated despite the sensory channel used for information extraction, while in the second case integrate information coming from different sensory channels should result in a poorer or impossible integration (Calvert, Brammer, & Iversen, 1998; Lamme & Roelfsema, 2000).

To test these two alternatives, we developed 3 experiments in which trustworthiness information and threat information are extracted by two different perceptual systems. Specifically, trustworthiness information is still presented by manipulating targets' trustworthiness while threat information is conveyed by means of auditory stimuli. Theoretically, if the integration of threat and trustworthiness information takes place at a high level of the cognitive architecture, this should lead in a slower but reliable integration. On the other hand, if such integration takes place at a low level of the cognitive architecture, this should result in the impossibility to integrate information belonging to a different perceptual system.

4.2 Study 6: Multi-Channel Integration of Environmental Sounds and Trustworthiness Cues

Experiment 6 was designed as a first investigation of the cognitive level at which threat information of facial trustworthiness are integrated. To do so, we asked participants to categorize the trustworthiness of faces presented on the screen while sounds conveying threatening (versus neutral or negative) information are played in the backgrounds. The prediction in the case of a high level integration is that untrustworthy faces paired with threatening sound should elicit a more direct mouse-trajectory toward the untrustworthy response button when compared with the same faces paired with neutral or negative sounds. By contrast, we expected a more curved mouse-trajectory toward the trustworthy response button when trustworthy faces are paired with threatening sounds in comparison with neutral or negative sounds. In the case of a low level integration, threat information conveyed by the sounds playing in the background should impact neither the categorization nor the mouse-trajectories.

4.2.1 Method

4.2.1.1 Participants

Sample size was determined before the data collection. Specifically, an a priori power analysis was conducted for sample size estimation (using G Power 3.1; Faul, Erdfelder, Lang, & Buchner, 2007). The projected sample size needed to detect a small-to-medium effect size (Cohen, 1988) with 80% power is N = 36 for a within-subject ANOVA. We advertised the study on campus and all the students who responded within 4 weeks were involved in the study. Overall, 39 participants, 20 female and 19 male, aged between 18 and 38 (M = 23.37, SD = 4.68) were enrolled in the study.

4.2.1.2 Stimuli

We employed the same faces of Experiments 3 and 4, but this time they were not embedded in any kind of visual background. Auditory stimuli used to convey threat information were drawn from the International Affective Digital Sounds (IADS) database (Bradley & Lang, 2007). The database contains a set of audio samples of different background sounds. The stimuli were edited in order to have sound samples that lasted for around 2 seconds each, and they were pretested in order to check that they were perceived as intended.

A sample of 26 Italian students was approached at the University of Milano – Bicocca, they were conducted to the lab where they provided their informed consent, and rated the materials. Specifically, each subject rated 9 stimuli: 3 neutral (i.e., Water flowing), 3 negative (i.g., Traffic jamming), 3 threatening (i.e., Bombs exploding). Participant were asked how threatening each stimulus was (1 = non-threatening – 7 = extremely threatening) and they were asked to indicate the valence of the stimulus (1 = extremely negative – 7 extremely positive). Ratings of valence and threat were averaged for each type of stimuli and for each subject (Table 4.1). Average ratings for each dimension were then submitted to separate ANOVAS with stimulus type (Neutral, Negative, Threatening) manipulated within-subjects.

	Valence	Threat
Neutral	4.19 (.77)[a]	1.72 (.88)[a]
Negative	2.26 (.69)[b]	2.32 (1.27)[a]
Threatening	2.04 (.74)[b]	4.95 (1.19)[b]

Table 4. 1 Difference in valence and threat among neutral, negative, and threatening sounds. [A,b] indicates means that are statistically different across type of sounds

The analysis revealed a main effect of stimulus type on valence $F(2, 50) = 81.37$, $p < .001$, indicating that the 3 types of context differed in terms of perceived valence. More importantly, a post hoc comparison revealed that negative ($M = 2.26$, $SD = .69$) and threatening (*M* = 2.04, *SD* = .74) sounds were not perceived as different in terms of valence, $t(25) = 1.23$, $p = .23$. Moreover, the same analysis revealed that the only stimuli to be perceived as more positive than the other two were neutral stimuli (*M* = 4.19, *SD* = .77), *ts* > 10.75, *ps* < .001. On the other hand, the analysis revealed a main effect of stimulus type on threat $F(2, 50) = 75.60$, $p < .001$, indicating that the 3 types of context differed in terms of perceived threat. More importantly, a post hoc comparison revealed that threatening (*M* = 4.95, *SD* = 1.19) sounds were perceived as more threatening than both neutral (*M* = 1.72, *SD* = .88) and negative stimuli (*M* = 2.32, *SD* = 1.27), *ts* > 2.86, *ps* < .01.

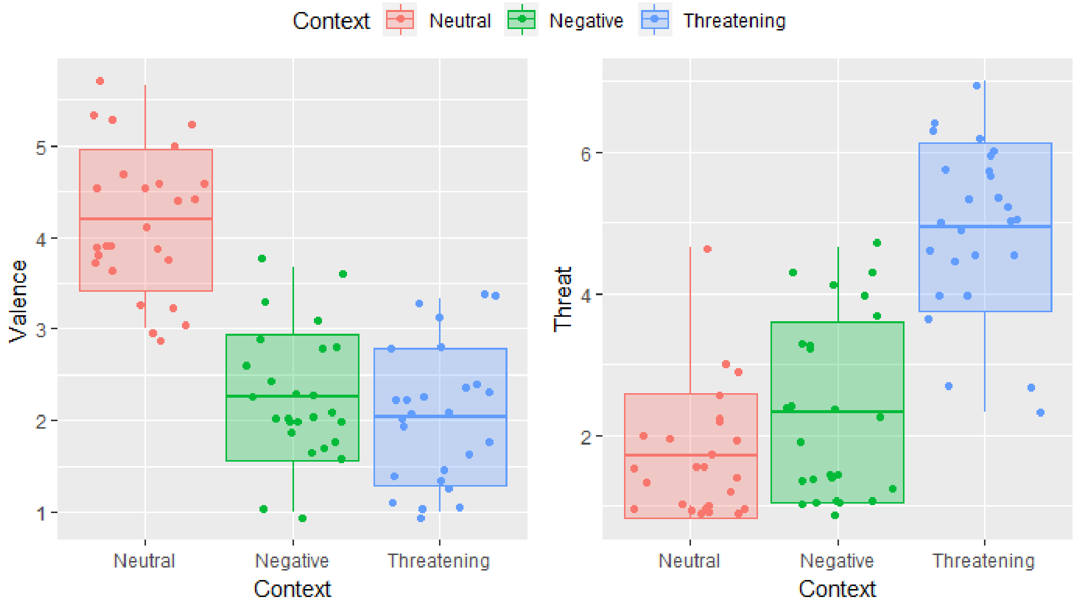

Figure 4. 1 Average valence (left) and perceived threat (right) for each type of sound

This pattern of results confirmed that our auditory stimuli were perceived as intended. Specifically, threatening stimuli were perceived as the most threatening while sharing the same valence of the negative ones (Figure 4.1).

4.2.1.3 Procedure

The procedure was largely similar to that of the previous experiments with some minor changes due to the cross-modal nature of the stimuli. As in previous experiments, participants were told that they would be presented with images of individuals while as sound was played in the background, and were asked to categorize each person as either trustworthy or untrustworthy ignoring the sound. They were asked to respond as quickly and accurately as possible by clicking response buttons, basing their judgments on their first impressions. Each trial started with the playing of the auditory stimulus followed by the face stimulus. The face was presented 1000 ms after the starting of the sound. Such modification was introduced to allow for the whole sound being played before the timeout occurs. In fact, participants were asked to respond within 1500 ms from the visual stimulus onset (2500 ms from the auditory stimulus onset). The interval between the trials was self-paced. Participants clicked a "Start" button at the bottom-center of the screen, to make the trial start. As in previous experiments, participants responded by clicking a "trustworthy" or "untrustworthy" response buttons located in the top-left and top-right corners of the screen (counterbalanced across participants). Participants were asked to start responding as early as they could. If participants did not start moving the mouse within 250 milliseconds after the face stimuli onset (1250 ms after the sound onset), a message advising them to start moving the mouse earlier was displayed (Hehman, Stolier, & Freeman, 2015). Each one of the four morphs of each identity was presented 2 times in each of the auditory condition, yielding 144 trials per participant.

4.2.1.4 Results

Following the procedure of previous experiments, we normalized trajectories into 101 time-steps and remapped leftward trajectories rightwards (inverted along the x-axis). For each trajectory, we computed the maximum deviation as a perpendicular deviation from an idealized straight line between the trajectory's start and response buttons (Freeman & Ambady, 2010). As in previous experiments, incorrect trials were excluded and the mean MD was computed for each subject in each within-subject condition.

Then, we performed a 3 (Type of sound: Neutral, Negative, Threatening) × 2 (Face: Trustworthy, Untrustworthy) within-subject ANOVA on MD (Table 4.2).

	Trustworthy	Untrustworthy
Neutral	.41 (.24)[a]	.56 (.32)[a]
Negative	.42 (.23)[a]	.49 (.31)[a]
Threatening	.40 (.24)[a]	.49 (.26)[a]

Table 4. 2 Difference in MD. [a, b, c] superscripts indicate means that are statistically different across type of sound

This analysis revealed no main effect of the type of sound, $F(2,76) = 1.26$, $p = .29$, $\eta^2_p = .03$, and no effect of face trustworthiness, $F(1,38) = 3.34$, $p = .08$, $\eta^2_p = .08$, and no face by type of sound interaction, $F(2,76) = 1.54$, $p = .22$, $\eta^2_p = .04$. For exploration purposes we reported the graphical representation of the comparison between average level of MD aggregated by type of sound for trustworthy (Figure 4.2) and untrustworthy faces (Figure 4.3).

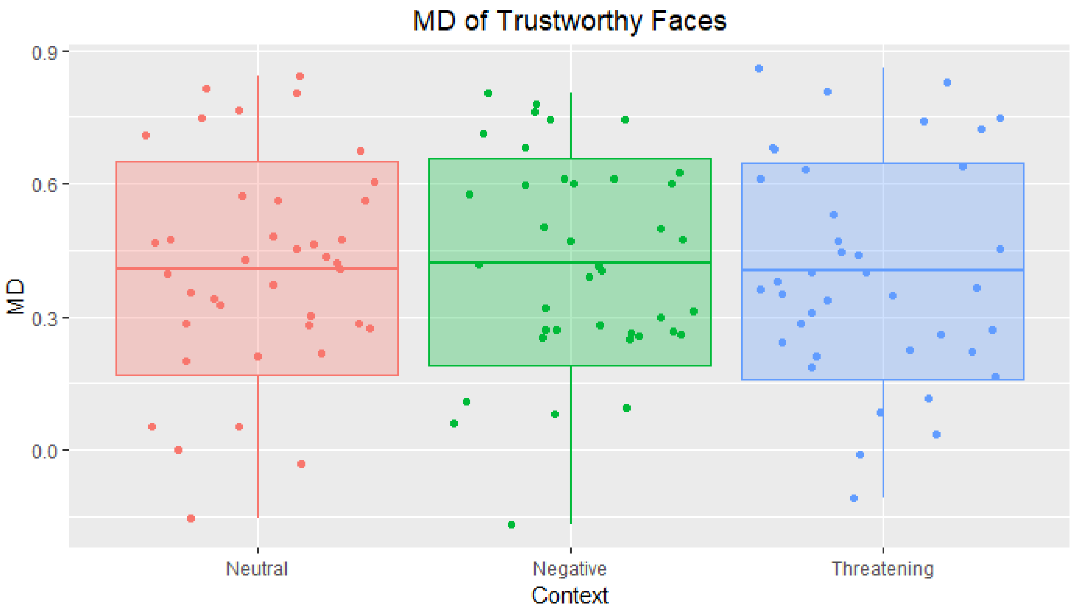

Figure 4. 2 Boxplots of average MD for trustworthy faces for each type of context (Experiment 6)

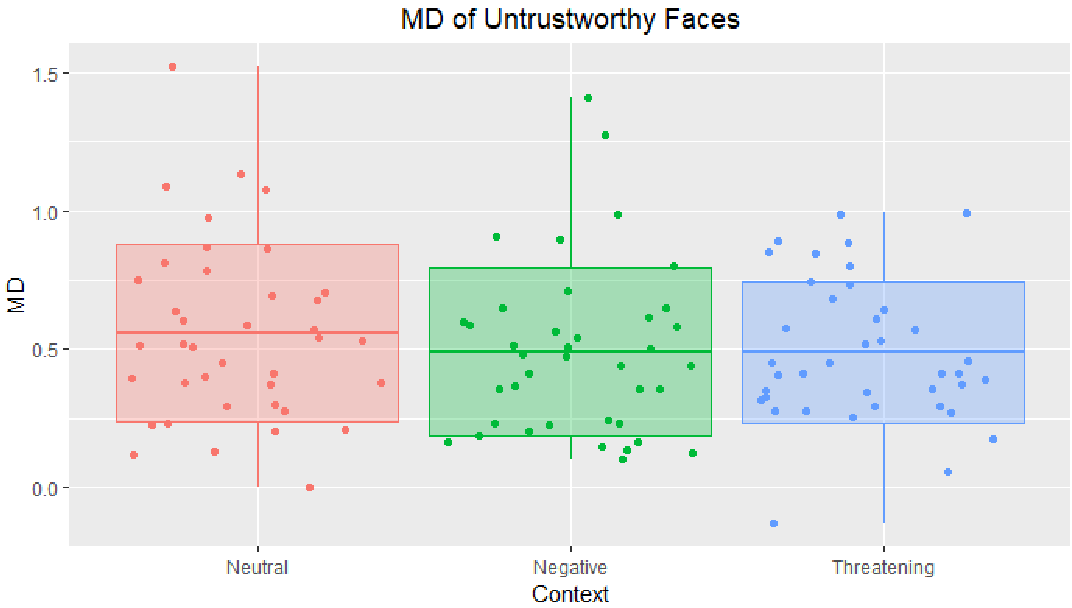

Figure 4. 3 Boxplots of average MD for untrustworthy faces for each type of context (Experiment 6)

Using the same procedure used for MD, we computed the area under the curve (AUC). The AUC was submitted to the same 3 (Type of sound: Neutral, Negative, Threatening) × 2 (Face: Trustworthy, Untrustworthy) within-subject ANOVA (Table 4.3).

	Trustworthy	Untrustworthy
Neutral	.75 (.54)[a]	1.11 (.81)[a]
Negative	.76 (.49)[a]	.97 (.76)[a]
Threatening	.74 (.53)[a]	.94 (.58)[a]

Table 4. 3 Difference in AUC. [a, b, c] superscripts indicate means that are statistically different across type of sound

The analysis revealed no main effect of the type of sound, $F(2,76) = 1.65$, $p = .20$, $\eta^2_p = .04$, and no face by type of sound interaction, $F(2,76) = 1.25$, $p = .29$, $\eta^2_p = .03$, but only a marginal effect of face trustworthiness, $F(1,38) = 3.90$, $p = .06$, $\eta^2_p = .09$. As such, trustworthy faces ($M = .75$, $SE = .08$) elicited marginally straighter trajectories than untrustworthy faces ($M = 1.00$, $SE = .10$). For exploration purposes we reported the graphical representation of the comparison between average level of AUC aggregated by type of sound for trustworthy (Figure 4.4) and untrustworthy faces (Figure 4.5).

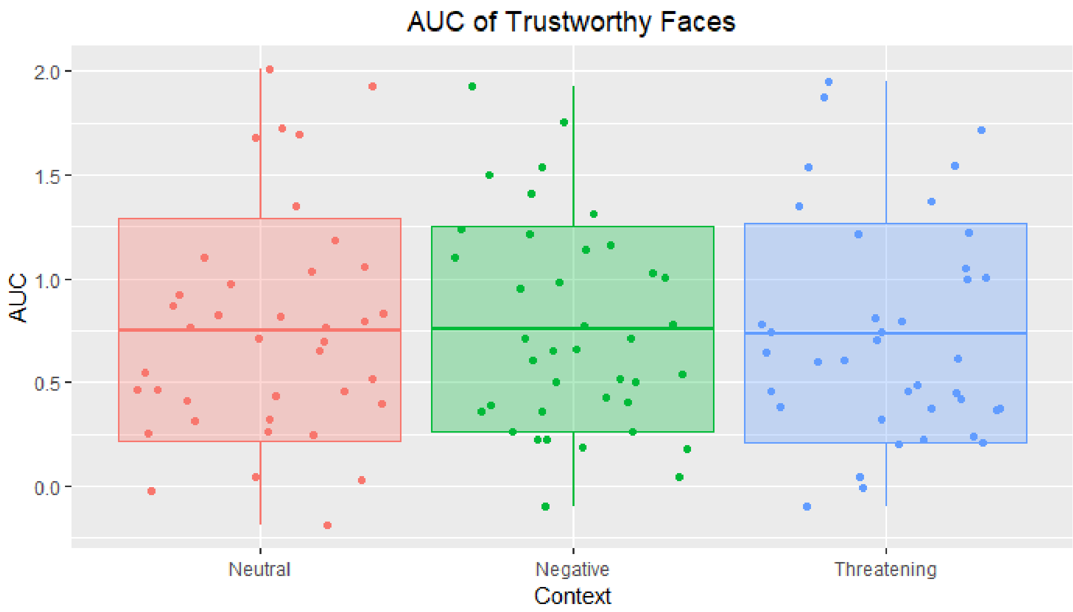

Figure 4. 4 Boxplots of average AUC for trustworthy faces for each type of context (Experiment 6)

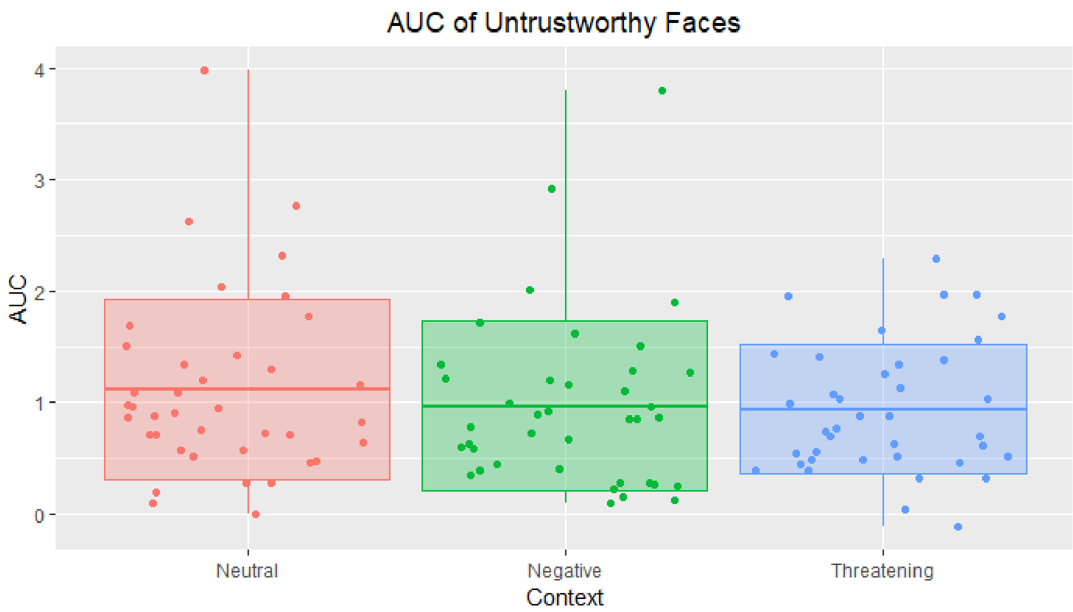

Figure 4. 5 Boxplots of average AUC for untrustworthy faces for each type of context (Experiment 6)

4.3 Study 7: Multi-Channel Integration, Stimulus Onset Asymmetry Adjustment

Experiment 6 revealed that threatening sounds do not incidentally impact the perception of trustworthiness of social targets. One reason may be that the stimuli onset asymmetry (SOA) was too large. If this was the case, the asynchrony between the image and the sound could be the culprit responsible for the lack of integration (Van Wassenhove, Grant & Poeppel, 2007). Therefore, threat information did not impact the processing of the face. To rule out this possible explanation we developed Experiment 7. We relied on design and stimuli of the previous experiment but this time, we relied on a shorter SOA. Such modification should allow for the contemporaneous processing of both auditory and visual stimuli.

A second possible explanation could be the limited sample size of Experiment 6, in which the bare minimum of participants was tested. To ensure us with enough power, we enrolled more participants than the required sample size outlined by the power analysis.

4.3.1. Method

4.3.1.1. Participants

As before, sample size was determined before the data collection. An a priori power analysis was conducted for sample size estimation (using G Power 3.1; Faul, Erdfelder, Lang, & Buchner, 2007). The projected sample size needed to detect a small-to-medium effect size (Cohen, 1988) with 80% power is N = 36 for a within-subject ANOVA. We advertised the study on campus and we enrolled an arbitrarily greater number of participants. The final sample size was N = 52.

4.3.1.2. Stimuli

Visual and auditory stimuli were the same as Experiment 6. No modifications were introduced regarding the materials. The only modification regarded the experimental procedure.

4.3.1.3. Procedure

The procedure was almost the same of Experiment 6. As before, participants were told that they would be presented with images of individuals and a sound would be played in the background. Once again, the task was to categorize each person as either trustworthy or untrustworthy ignoring the sound. As before, participants were asked to respond as quickly and accurately as possible by clicking response buttons, basing their judgments on their first impressions. Each trial started with the playing of the auditory stimulus followed by the face stimulus. This time, the face was presented only 500 ms after the starting of the sound. By reducing the SOA we aimed for the contemporaneous processing of both auditory and visual stimuli. This modification allowed for the whole sound being played before the timeout occurred, but not earlier. As before, participants were asked to respond within 1500 ms from the visual stimulus onset (2000 ms from the auditory stimulus onset). The interval between the trials was self-paced because participants had to click a “Start” button at the bottom-center of the screen in order to make the trial start. In line with experiments, participants responded by clicking a “trustworthy” or “untrustworthy” response buttons located in the top-left and top-right corners of the screen (counterbalanced across participants). Participants were asked to start responding as early as possible. If participants did not start moving the mouse within 250 milliseconds after the face stimuli onset (750 ms after the sound onset), a message advising them to start moving the mouse earlier was displayed (Hehman, Stolier, & Freeman, 2015). As in Experiment 6, each participant went through 144 trials. Specifically, each one of the four morphs of each identity was presented 2 times in each of the auditory condition.

4.3.2. Results

In order to analyze the data, we followed the procedure of previous experiments. First, we normalized trajectories into 101 time-steps and remapped leftward trajectories rightwards (inverted along the x-axis). Then, for each trajectory the maximum deviation was computed (Freeman & Ambady, 2010). Once again, incorrect trials were excluded and the mean MD was computed for each subject in each within-subject condition.

We performed a 3 (Type of sound: Neutral, Negative, Threatening) × 2 (Face: Trustworthy, Untrustworthy) within-subject ANOVA on MD (Table 4.4).

	Trustworthy	Untrustworthy
Neutral	.35 (.24)[a]	.46 (.27)[a]
Negative	.37 (.23)[a]	.45 (.29)[a]
Threatening	.36 (.21)[a]	.43 (.26)[a]

Table 4. 4 Difference in MD. [a, b, c] superscripts indicate means that are statistically different across type of sound

As in Experiment 6, analysis revealed no main effect of the type of sound, $F(1,51) = .43$, $p = .51$, $\eta^2_p = .008$, and no face by type of sound interaction, $F(2,102) = 1.06$, $p = .35$, $\eta^2_p = .02$. In contrast with the previous experiment, an effect of face trustworthiness, $F(1,51) = 4.67$, $p = .035$, $\eta^2_p = .08$, was found.

Post-hoc comparisons revealed that trustworthy faces ($M = .358$, $SE = .03$) elicited straighter trajectories than untrustworthy faces ($M = .447$, $SE = .04$), $t(52) = 2.17$, $p = .035$, $d = 0.29$, $CI_{95\%} = [0.02, 0.57]$. This is marginal to our hypothesis but in line with prior work showing the easier processing of positive information (Unkelbach, Fiedler, Bayer, Stegmüller & Danner, 2008).

Despite the lack of significance of the two-way interaction, we reported the graphical representation of the comparison between average level of MD aggregated by type of sound for trustworthy (Figure 4.6) and untrustworthy faces (Figure 4.7).

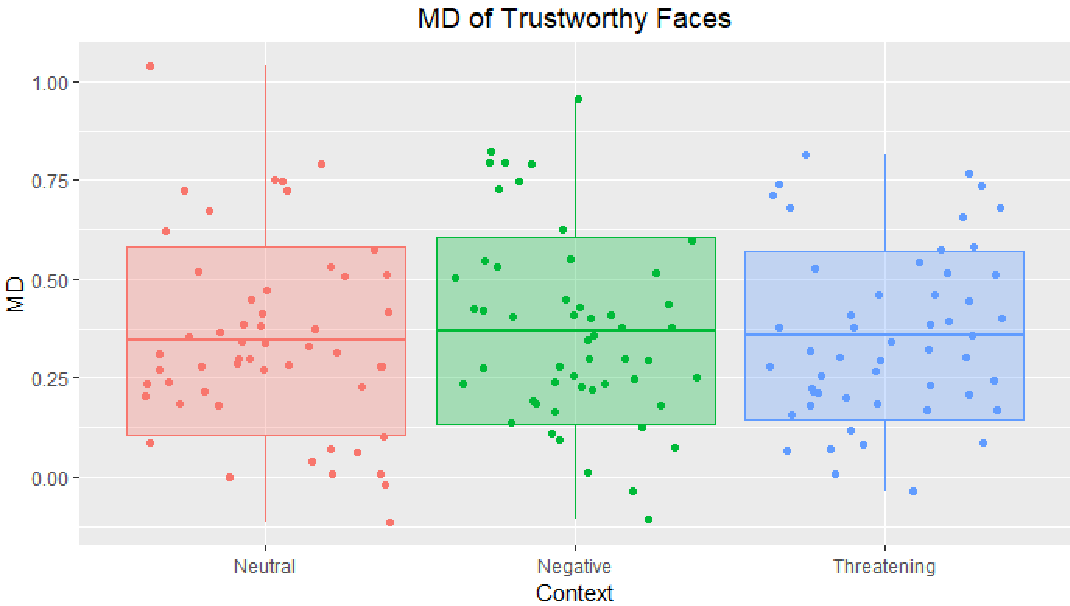

Figure 4. 6 Boxplots of average MD for trustworthy faces for each type of context (Experiment 7)

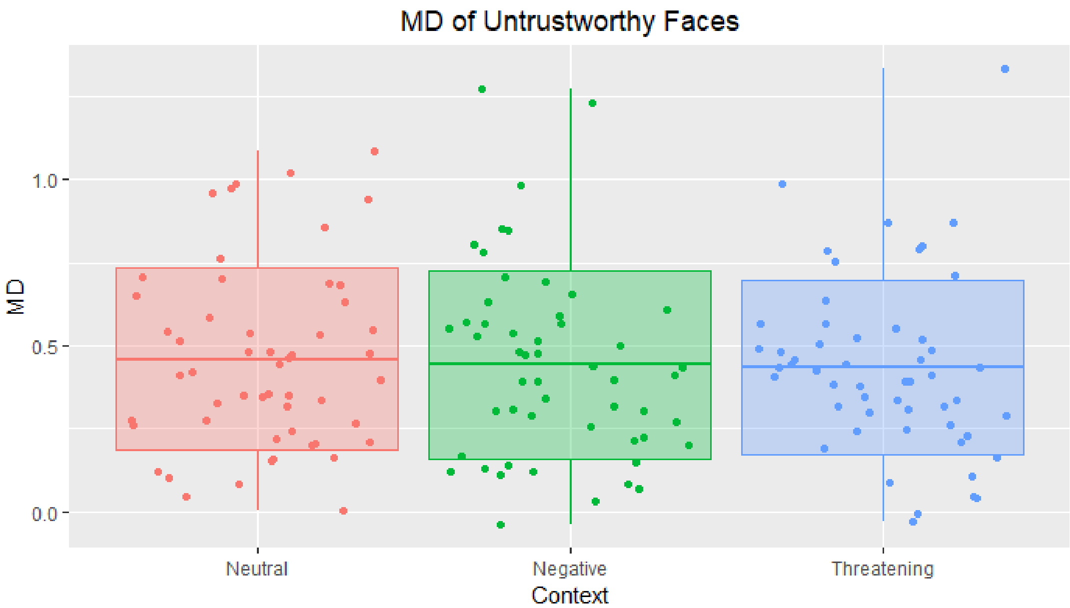

Figure 4. 7 Boxplots of average MD for untrustworthy faces for each type of context (Experiment 7)

As in the case of MD, we computed the area under the curve (AUC). AUC was submitted to the same 3 (Type of sound: Neutral, Negative, Threatening) × 2 (Face: Trustworthy, Untrustworthy) within-subject ANOVA (Table 4.5).

	Trustworthy	Untrustworthy
Neutral	.63 (.50)[a]	.92 (.66)[a]
Negative	.69 (.52)[a]	.87 (.67)[a]
Threatening	.66 (.44)[a]	.84 (.65)[a]

Table 4. 5 Difference in AUC. [a, b, c] superscripts indicate means that are statistically different across type of sound

The pattern of results is similar to MD. As in the previous experiment, the analysis revealed no main effect of the type of sound, $F(2,102) = .54$, $p = .58$, $\eta^2_p = .01$, and no face by type of sound interaction, $F(2,102) = 1.17$, $p = .31$, $\eta^2_p = .02$. As for MD, the analysis revealed an effect of face trustworthiness, $F(1,51) = 5.06$, $p = .03$, $\eta^2_p = .09$, on AUC. Post-hoc comparisons showed that trustworthy faces ($M = .66$, $SE = .06$) elicited smaller AUC than untrustworthy faces ($M = .88$, $SE = .08$), $t(51) = 2.25$, $p = .03$, $d = 0.31$, $CI_{95\%} = [0.03, 0.59]$.

For exploration purposes we reported the graphical representation of the comparison between average level of AUC aggregated by type of sound for trustworthy (Figure 4.8) and untrustworthy faces (Figure 4.9).

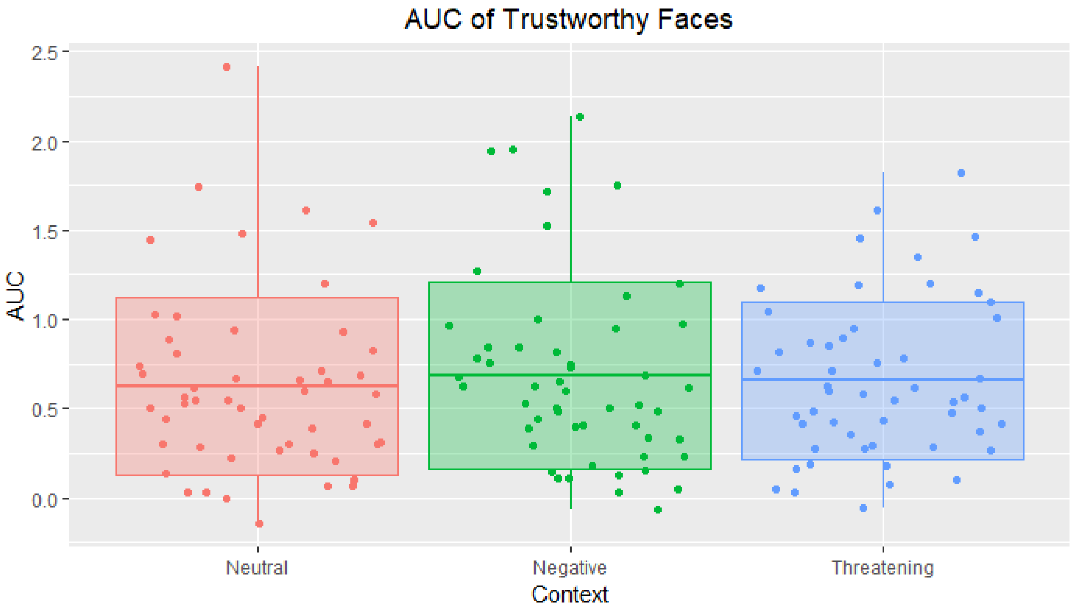

Figure 4. 8 Boxplots of average AUC for trustworthy faces for each type of context (Experiment 7)

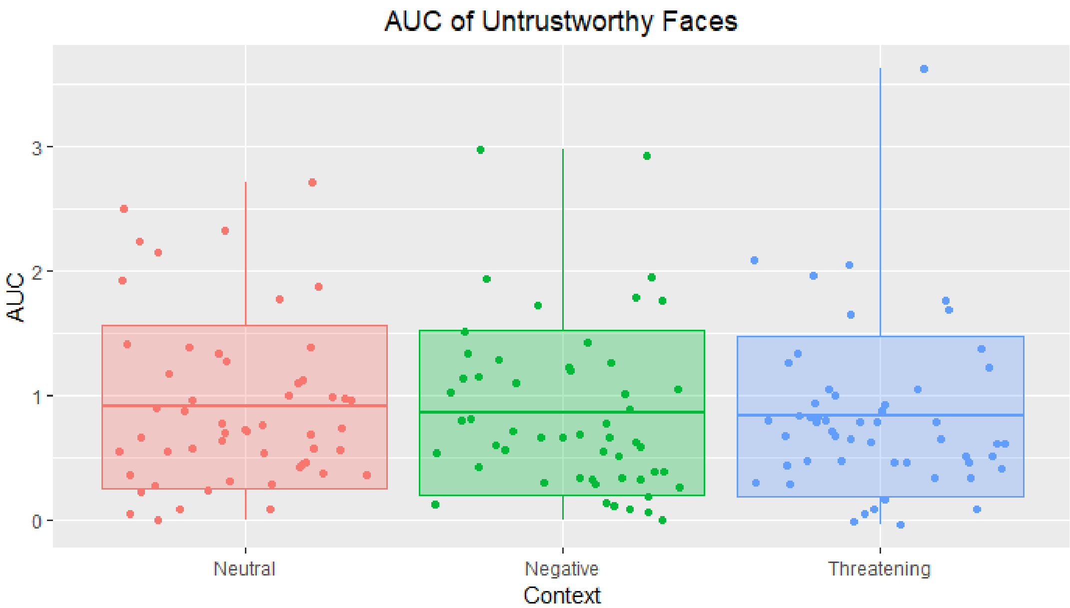

Figure 4. 9 Boxplots of average AUC for untrustworthy faces for each type of context (Experiment 7)

4.4 Study 8: Multi-Channel Integration of Human Voices and Trustworthiness Cues

Findings from experiments 6 and 7 seem to indicate that threatening sounds does not interact with the categorization of facial trustworthiness. Despite making auditory threat information and trustworthiness closer in time by reducing the stimuli onset asymmetry, still no evidence for information integration is found.

However, a remark must be made. Trustworthiness is eminently human while the sounds drawn from the IADS are environmental sounds. The auditory manipulation we employed is not a kind of information that can be brought back to human sources. For this reason, we developed a new experiment in which we manipulate threatening information in such a way that allows the perceiver to extract human-related threat information. Specifically, we created some auditory stimuli by asking a confederate to read some threatening phrases, and we employed such recordings to provide threatening information. This would give the participant the impression that the sound was associated with the voice of the facial stimuli.

With this new set of stimuli, we ran Experiment 8 that shares the design of the previous. Specifically, we still manipulate trustworthiness by showing different morphs of computer-generated identities and we employ three information conditions namely, neutral, negative, and threatening. We expect to find evidences of information integration indexed by straighter mouse trajectories when untrustworthy faces are associated with threatening sounds (i.e., in this case voices). At the same time, we expect more curved trajectories when trustworthy faces are associated with threatening sounds

4.4.1 Method

4.4.1.1 Participants

As in previous experiments, sample size was determined before the data collection. An a priori power analysis was conducted for sample size estimation (using G Power 3.1; Faul, Erdfelder, Lang, & Buchner, 2007). The projected sample size needed to detect a small-to-medium effect size (Cohen, 1988) with 80% power is N = 36 for a within-subject ANOVA. We advertised the study on campus and, as in Experiment 7, we enrolled a slightly greater number of participants. The final sample size was N = 47. Participants were all students, 35 female and 12 male, aged between 19 and 39 (*M* = 25.15, *SD* = 4.97).

4.4.1.2. Stimuli

Face stimuli are the same as previous experiments. In this experiment, we used a new set of auditory sounds. Such sounds were short phrases recorded in sample of 2 seconds. The new stimuli set was pretested in order to check that each stimulus was perceived as intended. We enrolled a sample of 22 Italian students, 15 female and 7 male, aged between 18 and 36 (*M* = 25.09, *SD* = 4.85). They were conducted to the lab where they provided their informed consent, and rated the materials. Materials comprised 12 stimuli: 4 neutral (e.g., "I'm going to have a shower"), 4 negative (e.g., "You disappointed me"), 4 threatening (e.g., "You can no longer escape me!"). Each sentence was designed to be perceived as neutral, negative or threatening according both content and prosody. For each stimulus, participants were asked how threatening it sounded (1 = non-threatening – 7 = extremely threatening) and they were asked to indicate the valence of the stimulus (1 = extremely negative – 7 extremely positive). Ratings of valence and threat were averaged for each type of stimuli and for each subject. Average ratings for each dimension were then submitted to separate ANOVAS with stimulus type (Neutral, Negative, Threatening) manipulated within-subjects (Table 4.6).

	Valence	Threat
Neutral	4.02 (.81) [a]	1.40 (.66) [a]
Negative	3.04 (.88) [b]	1.34 (.53) [a]
Threatening	2.68 (1.73) [b]	5.84 (.99) [b]

*Table 4. 6 Difference in valence and threat among neutral, negative, and threatening sounds. * indicates means that are statistically different across type of voice*

The analysis revealed a main effect of stimulus type on valence $F(2,42) = 10.24$, $p < .001$, indicating that the 3 types of context differed in terms of perceived valence. More importantly, a post hoc comparison revealed that negative (M = 3.04, SD = .88) and threatening (M = 2.68, SD = 1.73) sounds were not perceived as different in terms of valence, $t(21) = 1.17$, $p = .25$. However, both negative and threatening sounds were perceived as more negatively valenced than neutral (M = 4.02, SD = .81), $ts(21) > 3.46$, $ps < .002$. At the same time, the analysis revealed a main effect of stimulus type on threat $F(2, 42) = 311.72$, $p < .001$, indicating that the 3 types of context differed in terms of perceived threat. More importantly, a post hoc comparison revealed that threatening (M = 5.84, SD = .99) sounds were perceived as more threatening than both neutral (M = 1.40, SD = .66) and negative stimuli (M = 1.34, SD = .53), $ts > 18.19$, $ps < .001$. Interestingly, neutral and negative sentences were perceived equally threatening, $t(21) = 0.45$, $p = .66$.

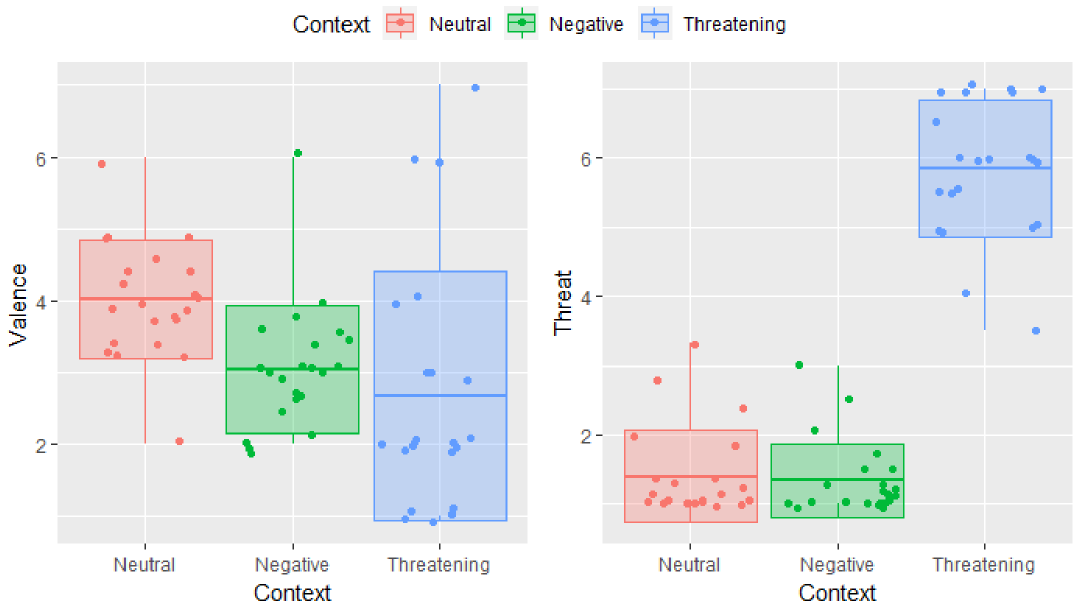

Figure 4. 10 Average valence (left) and perceived threat (right) for each type of sound (voices)

This pattern of results confirmed that the new set of auditory stimuli was perceived as intended. Specifically, threatening stimuli were perceived as the most threatening while sharing the same valence of the negative ones (Figure 4.10).

4.4.1.3 Procedure

The procedure was identical to Experiment 7. Participants were informed about the delayed presentation of a face and a sound, and their task was to categorize each person as either trustworthy or untrustworthy ignoring the sound. They were asked to respond as quickly and accurately as possible by clicking response buttons. The first stimulus was sound followed by the face stimulus. As in Experiment 7, delay between the sound and the face was 500 ms. Participants were asked to respond within 1500 ms from the visual stimulus onset (2000 ms from the auditory stimulus onset). The task was still self-paced, participants had to click a "Start" button at the bottom-center of the screen in order to make the trial start. As before, participants responded by clicking a "trustworthy" or "untrustworthy" response buttons located in the top-left and top-right corners of the screen (counterbalanced across participants). Participants were asked to start responding as early. If participants did not start moving the mouse within 250 milliseconds after the face stimuli onset (750 ms after the sound onset), a message advising them to start moving the mouse earlier was displayed (Hehman, Stolier, & Freeman, 2015). As in Experiment 7, each participant went through 144 trials. Specifically, each one of the four morphs of each identity was presented 2 times in each of the auditory condition.

4.4.2. Results

As before, we first normalized trajectories into 101 time-steps and remapped leftward trajectories rightwards. Then, for each trajectory the maximum deviation was computed (Freeman & Ambady, 2010). As before, incorrect trials were excluded and the mean MD was computed for each subject in each within-subject condition. We performed a 3 (Type of sound: Neutral, Negative, Threatening) × 2 (Face: Trustworthy, Untrustworthy) within-subject ANOVA on MD. As in Experiment 6 and 7, analysis revealed no main effect of the type of sound, $F(2,92) = 1.17$, $p = .32$, $\eta^2_p = .03$, but an effect of face trustworthiness, $F(1,46) = 4.41$, $p = .04$, $\eta^2_p = .09$. A post-hoc comparison revealed that trustworthy faces ($M = .35$, $SE = .03$) were categorized eliciting a more curved trajectory than untrustworthy faces (*M*

= .41, *SE* = .04), *t*(46) = 2.10, *p* = .04, *d* = 0.30, $CI_{95\%}$ = [0.01, 0.60]. Differently from the previous two experiments, an interaction between type of sound and face trustworthiness, *F*(2,92) = 4.11, *p* = .02, η^2_p = .08, was found (Table 4.7).

	Trustworthy	Untrustworthy
Neutral	.40 (.23)^a	.38 (.26)^a
Negative	.32 (.23)^b	.40 (.28)^{ab}
Threatening	.33 (.25)^b	.46 (.29)^b

Table 4. 7 Difference in MD. $^{a, b, c}$ superscripts indicate means that are statistically different across type of sound

We decomposed such interaction by running separate post-hoc comparisons for trustworthy and untrustworthy faces. According to such comparisons, trustworthy faces elicited a more curved trajectory when categorized while a neutral sound (*M* = .40, *SD* = .23) was playing if compared with the same face paired with a negative (*M* = .32, *SD* = .23) or a threatening (*M* = .33, *SD* = .25) sound, *ts*(45) > 2.13, *ps* < .037 (*d* = 0.33, $CI_{95\%}$ = [0.03, 0.62], and *d* = 0.31, $CI_{95\%}$ = [0.02, 0.61], respectively). However, no difference in MD was found between trustworthy faces categorized while a negative or a threatening sound was playing, *t*(45) = .27, *p* = .78, *d* = 0.04, $CI_{95\%}$ = [-0.25, 0.33] (Figure 4.11).

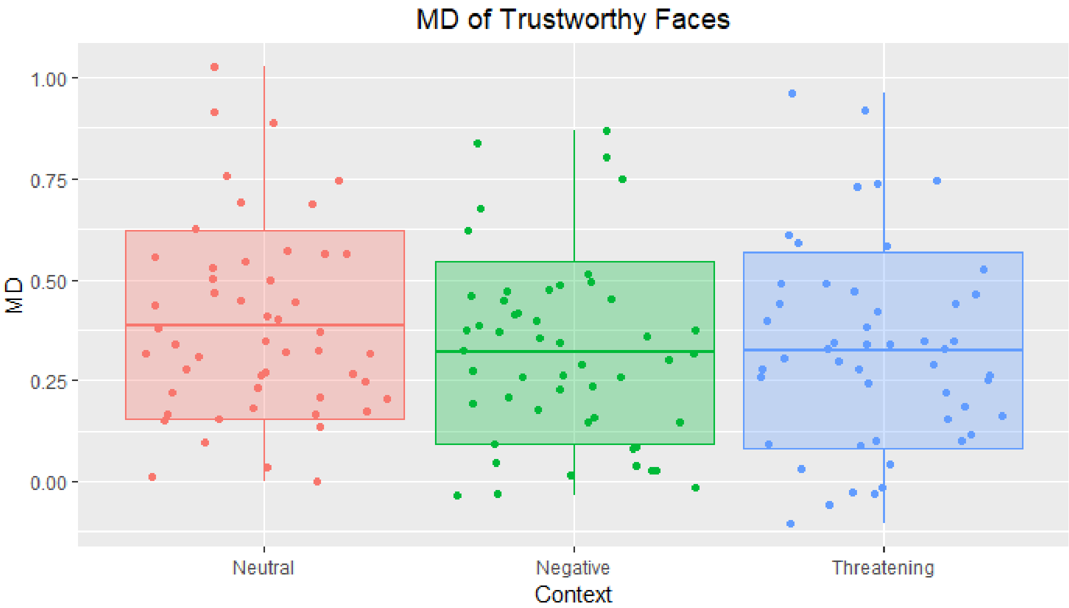

Figure 4. 11 Boxplots of average MD for trustworthy faces for each type of context (Experiment 8)

On the other hand, untrustworthy faces elicited a marginally straighter trajectory when they were categorized while a neutral sound (*M* = .38, *SD* = .26) was played in comparison with when a threatening sound (*M* = .46, *SD* = .29) was played, *t*(45) = 1.95, *p* = .057, *d* = 0.29, $CI_{95\%}$ = [-0.01, 0.58], but no difference was found between threatening and negative sound (*M* = .40, *SD* = .28), *t*(45) = 1.56, *p* = .13, *d* = 0.23, $CI_{95\%}$ = [-0.06, 0.52]. Moreover, no difference is found between neutral and negative sounds, *t*(45) = .50, *p* = .62, *d* = 0.07, $CI_{95\%}$ = [-0.21, 0.36] (Figure 4.12).

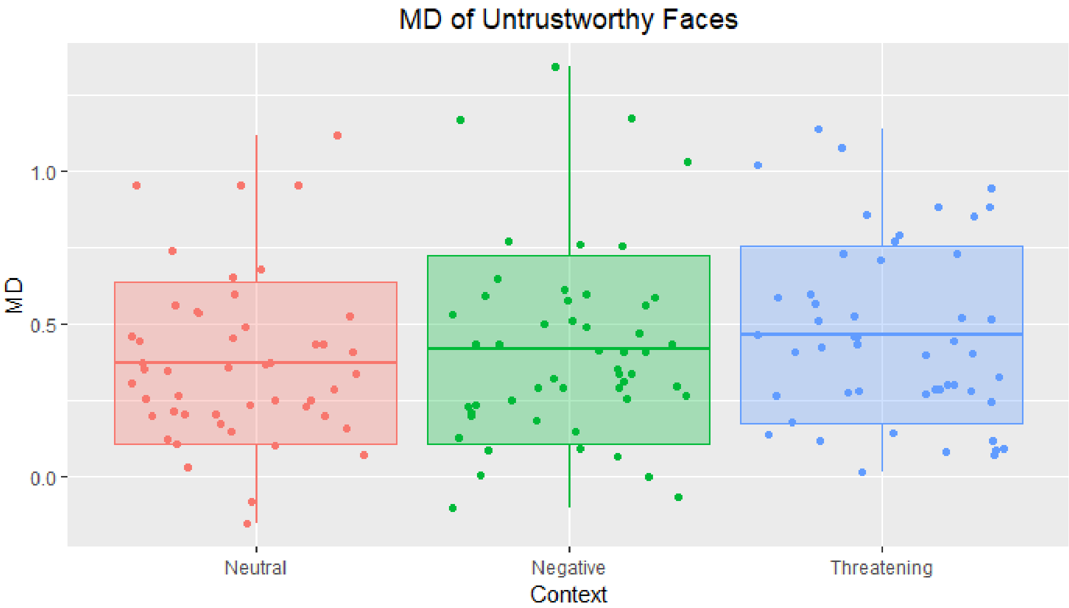

Figure 4. 12 Boxplots of average MD for untrustworthy faces for each type of context (Experiment 8)

After analyzing MD, we computed the area under the curve (AUC) using the same procedure. Then, AUC was submitted to the same 3 (Type of sound: Neutral, Negative, Threatening) × 2 (Face: Trustworthy, Untrustworthy) within-subject ANOVA (Table 4.8).

	Trustworthy	Untrustworthy
Neutral	.71 (.51)[a]	.71 (.60)[a]
Negative	.58 (.49)[a]	.71 (.65)[a]
Threatening	.55 (.51)[a]	.82 (.69)[a]

Table 4. 8 Difference in AUC. [a, b, c] superscripts indicate means that are statistically different across type of sound

The analysis revealed no main effect of the type of sound, *F*(2,92) = .59, *p* = .55, η^2_p = .01. However, we found a marginally significant main effect of face trustworthiness, *F*(1,46) = , *p* = .05, η^2_p = .82. Here, categorizing trustworthy faces (*M* = .62, *SE* = .06) elicited less wide trajectories than untrustworthy faces (*M* = .75, *SE* = .08), t(46) = 2.05, p =.05, *d* = 0.30, $CI_{95\%}$ = [0.01, 0.59]. Moreover, differently from MD, a face by type of sound interaction, *F*(2,92) = 2.73, *p* = .ns, η^2_p = .06, was not found.

For exploration purposes we reported the graphical representation of the comparison between average level of AUC aggregated by type of sound for trustworthy (Figure 4.13) and untrustworthy faces (Figure 4.14).

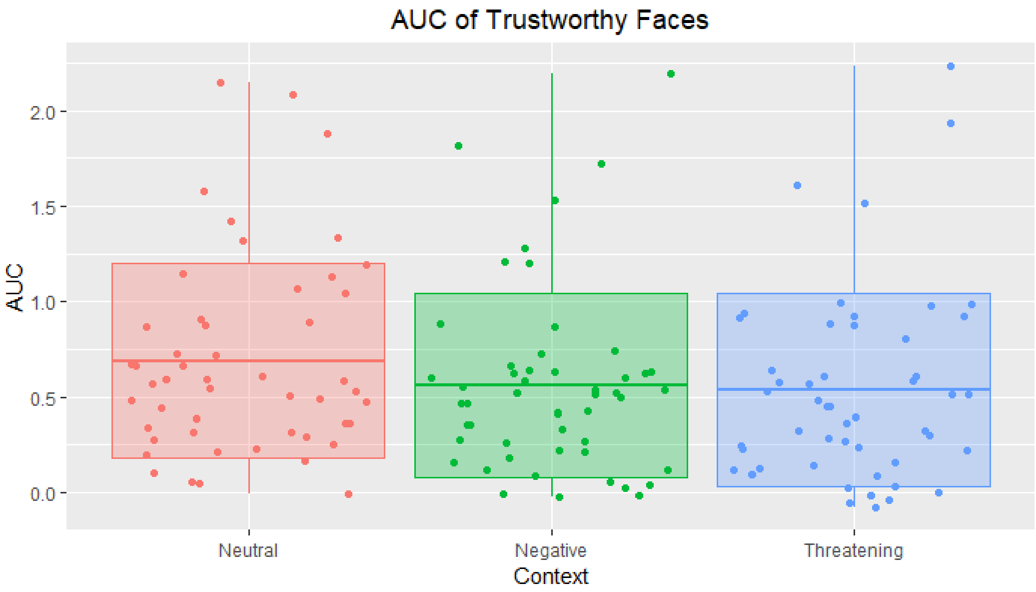

Figure 4. 13 Boxplots of average AUC for trustworthy faces for each type of context (Experiment 8)

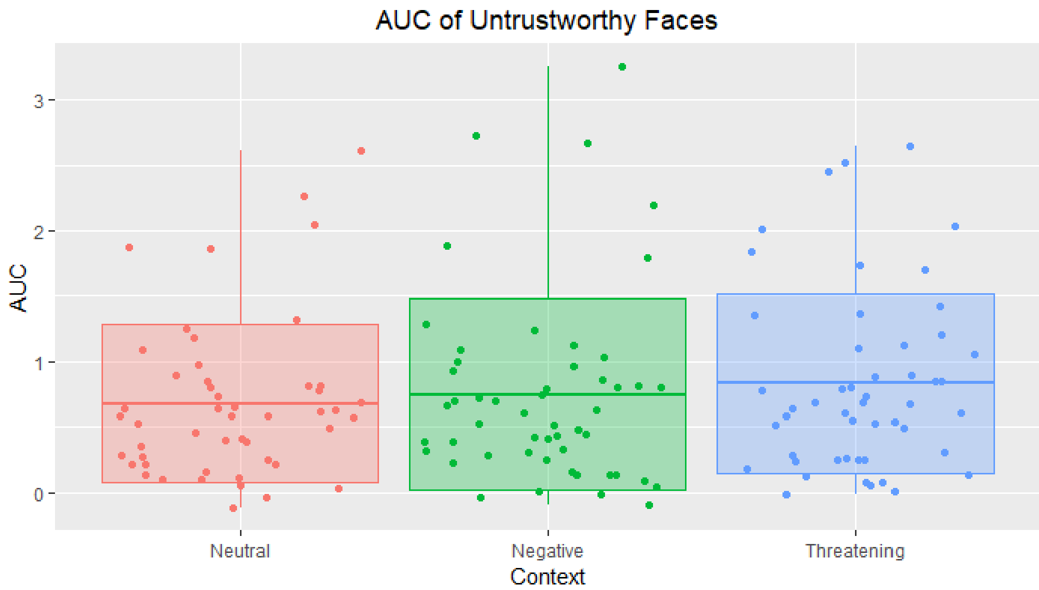

Figure 4. 14 Boxplots of average AUC for untrustworthy faces for each type of context (Experiment 8)

4.6 General discussion

The main goal of this chapter is to test whether the findings of Chapter 3 hold when information is integrated across different perceptual systems. Three studies suggested that was not the case. In Experiment 6 and 7, no interactions were found between face trustworthiness and information conveyed by the sound. In Experiment 8, our finding, pointed in the unexpected direction regarding MD of trustworthy faces. Specifically, we expected to find evidences that the threat-specific congruence with face trustworthiness should have elicited straighter trajectories for untrustworthy faces categorized while a threatening sound is played and for trustworthy faces categorized while a neutral sound is played. Such pattern of results could indicate that the threatening content of the information incidentally impacted the information about the trustworthiness of the social target, the face.

Taken together, all these findings suggest that information in the auditory and in the visual system do not interact with each other and we are left with the conclusion that such integration does not take place. However, it is reasonable to think that this lack of integration may be mostly due to the modularity of the two perceptual systems. In fact, it is plausible that information stored in the same system are more likely to interact than information stored across two different perceptual systems.

Despite these findings being uncertain, they are not discarded. In the next chapter, findings of different chapters will be summarized in order to estimate the real presence of an effect and its magnitude. Such effort will be carried out including findings of this chapter in order to avoid over-estimate the effect of interest by means of picking only findings that point in the expected direction.

CHAPTER 5

AGGREGATED FINDINGS ACROSS STUDIES: A META-ANALYSIS

5.1 Introduction

Across different experiments, we tested whether threat-related information would impact the evaluation and categorization of facial trustworthiness. Our main goal was to investigate if such kinds of information could be integrated in the mind of the perceiver. After a considerable empirical effort, we are left with some findings that point in the direction that such integration takes places. Our main focus was on investigating our hypothesis by means of process-sensitive measure such as MD and AUC, measures typical of the mouse-tracking paradigm.

The findings gathered however, should not be taken in isolation. A single study on its own is rarely enough to provide evidence for any scientific claim whether aggregating the findings could shed more light on the phenomenon of interest (Goh, Hall & Rosenthal, 2016). Moreover, it is recommended to accept the inconsistency across several findings and to handle such inconsistency in a proper way (Maner, 2014). Two aspects that will surely benefit from such recommendation are the precision of estimates (Cumming, 2014) and, in the long run, even the replicability crisis (Braver, Thoemmes & Rosenthal, 2014).

To combine our findings, we decide to aggregate our results into a meta-analysis on the measures of interest. Such techniques allow us many benefits. First, results from different studies are not interpreted in isolation, but a “bigger picture” can be sought. Second, the magnitude of the effect of interest can be reliably estimated. Such a benefit can provide a quantitative measure of how strongly information drawn from the context, both visual or auditory, distorts the processing of facial trustworthiness of the social target. Third, a meta-analysis allows us to investigate the difference in the information integration we are endeavoring due to the level at which such integration takes place. Specifically, in experiments from chapter 3, threatening information was conveyed visually, the same sensory channel used to extract information from the social target. Differently, in experiments from chapter 4, conveyed threat information by means of auditory stimuli, inducing the information integration to take place at a higher cognitive level. Different modalities, and therefore, different levels of integration, can be tested in the meta-analysis by factorizing the level of interaction and estimating differences in the variables of interest at different levels of the factor. Namely, we can compare the effect we are interested in between the two different manipulation modalities.

5.2.1 Analysis Design: Effects to be Aggregated

Our more specific hypothesis is about the specificity of threat on interfering with the processing of target's trustworthiness. We expect that, due its great overlapping with trustworthiness, threat-specific information impacts the processing of the social target above and beyond negatively valenced information. Because of that, we developed almost all studies in the present work with two control conditions, neutral and negative. In aggregating results from different studies, we aim at testing such specificity. Therefore, we decided to meta-analyze critical comparison between negative and threatening conditions.

For each study, we decomposed our interaction between trustworthy and untrustworthy faces, investigating each time our critical contrast. In fact, for each variable of interest, we computed effect sizes estimating the difference between negative and threat condition. The effect size we choose is the Cohen's d (Cohen; 1988, 1992). Such a measure, express the average difference between the means of the dependent variable across two conditions normalized by the common standard deviation. Cohen's d, as we computed, signals the magnitude of the effect by its absolute value and the direction of the effect by its sign. Specifically, since we always computed the average of the negative condition minus the average of the threat condition, a negative effect size signals that that dependent variable value is higher in threat condition than in negative condition. For example, if we consider MD as dependent variable, a negative Cohen's d is telling us that trajectories in the threat condition are more curved than trajectories in the negative condition. Effect sizes and the associated variances are computed taking into account the within-subject nature of the comparison (Lakens, 2013).

Moreover, since we are interested in process-sensitive measures, we meta-analyzed maximum deviation and area under the curve. Those are our main dependent variables. Since we are interested in such measures and in the contrast between negative and threat conditions, we meta-analyzed only results coming from experiments that provide such measure in such conditions. Therefore, we include in our meta-analysis findings from Experiment 4 through experiment 8. We conduct two separate meta-analyses for MD and AUC, and for trustworthy and untrustworthy faces. In total, we have two meta-analyses for MD and two for AUC.

5.2.2 Analysis Design: Random-Effect Model

In aggregating our results, we opted for a specific type of analysis. We ran a random-effect meta-analysis (Borenstein, Hedges, Higgins, & Rothstein, 2010). Such a technique assumes that the effect of interest is not fixed in the population, but in can vary beyond random sampling. Put differently, we assumed that effects in our studies were drawn from a distribution of effects. This choice was a conservative one because it allows an extra level of variance in the model. This framework reduces the likelihood to over-estimate the effect. However, despite this choice being a safer one, we used a random-effect model for another main reason. In fact, this choice better reflected the assumption we have about the nature of the effect. In our view, the incidental effect of threatening information on the processing of facial trustworthiness cannot be the same across all experiments because a considerable number of unidentified variables, impossible to control even in the lab, could play a role. Therefore, a random-effect model was a better choice in investigating our effect of interest.

5.2.3 Analysis Design: Investigating the Level of Integration

In order test the impact of the modality (visual vs auditory) by which the information is provided on the processing of the targets' trustworthiness, we added a variable encoding such information in the model. The result is a meta-analysis that aggregates the effect controlling for this feature of the studies and compares the visual-only studies against the visual-auditory studies.

5.3.1 Process-sensitive measures: Maximum Distance

The first dependent variable we analyzed is the maximum distance. We gathered Cohen's d from Experiment 4 to Experiment 8 along with the associated variances. We first aggregated the effects of the comparison between negative and threat condition in the categorization of trustworthy faces (Table 5.1).

	Sample Size	Cohen's d	Variance	Manipulation
Experiment 4	45	-.38	.015	visual
Experiment 5	60	-.23	.007	visual
Experiment 6	39	.08	.009	auditory
Experiment 7	52	.06	.004	auditory
Experiment 8	47	-.03	.019	auditory

Table 5. 1 Effect Sizes of the negative vs threat comparisons on MD for trustworthy faces

The negative effect sizes in experiments 4 and 5 indicate that MD was higher in the threat condition than in the negative condition. This can be taken as a signal that processing a trustworthy face in the presence of threatening information is less easy than doing the same in the presence of negative but non-threatening information. However, in experiments 6, 7, and 8, the effect sizes are close to zero.

Once gathered the results, we ran a random-effect meta-analysis in order to aggregate such effects. The analysis estimated a null effect when aggregating the effect of the auditory experiments, d = .035, se = .0594, z = .582, p = .56, $CI_{95\%}$ = [-.082, .151], but a significant effect when aggregating the effect of the visual-only experiments, d = -.329, se = .0928, z = -3.550, p = .0004, $CI_{95\%}$ = [-.511, -.147] (Figure 5.1).

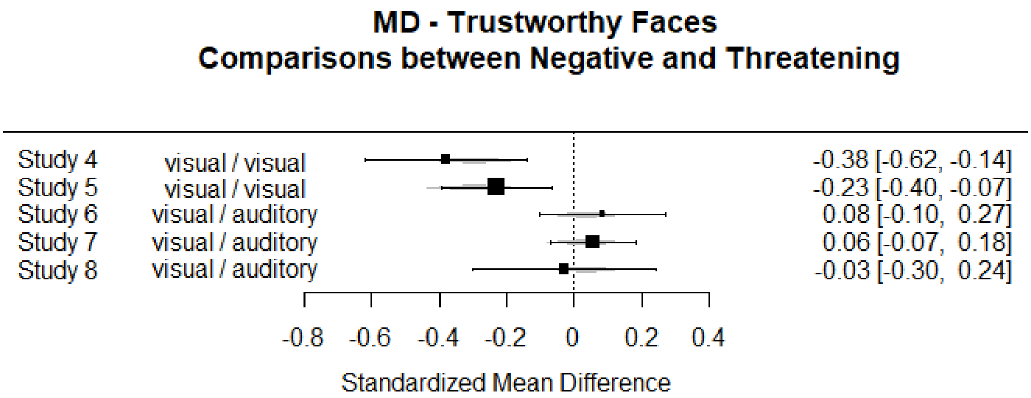

Figure 5. 1 Forest plot of standardized mean difference in MD between negative and threat conditions for trustworthy faces

Those two aggregated effects must be estimated using a random-effect model and must be kept separate due to their high heterogeneity, Q(1) = 12.60, p = .0004. Such value of heterogeneity signals that there is a substantial difference between visual-only and auditory studies.

Then, we aggregated the effects of the comparison between negative and threat condition in the categorization of untrustworthy faces (Table 5.2).

	Sample Size	Cohen's d	Variance	Manipulation
Experiment 4	45	.29	.018	visual
Experiment 5	60	.10	.009	visual
Experiment 6	39	.002	.019	auditory
Experiment 7	52	.03	.007	auditory
Experiment 8	47	-.16	.017	auditory

Table 5. 2 Effect Sizes of the negative vs threat comparisons on MD for untrustworthy faces

This time, the positive effect sizes in experiments 4 and 5 indicate that MD was lower in the threat condition than in the negative condition. This can be taken as a signal that processing an untrustworthy face in the presence of threatening information is easier than doing the same in the presence of negative but non-threatening information. This time, the effects are close to zero in experiments 6 and 7, whether the effect becomes slightly negative in Experiment 8.

After gathering the results, we ran a random-effect meta-analysis in order to aggregate such effects. As before, the analysis estimated a null effect when aggregating the effect of the auditory experiments, d = -.044, se = .0676, z = -.651, p = .51, $CI_{95\%}$ = [-.176, .088], but a significant effect when aggregating the effect of the visual-only experiments, d = .224, se = .1047, z = 2.140, p = .03, $CI_{95\%}$ = [.019, .429] (Figure 5.2).

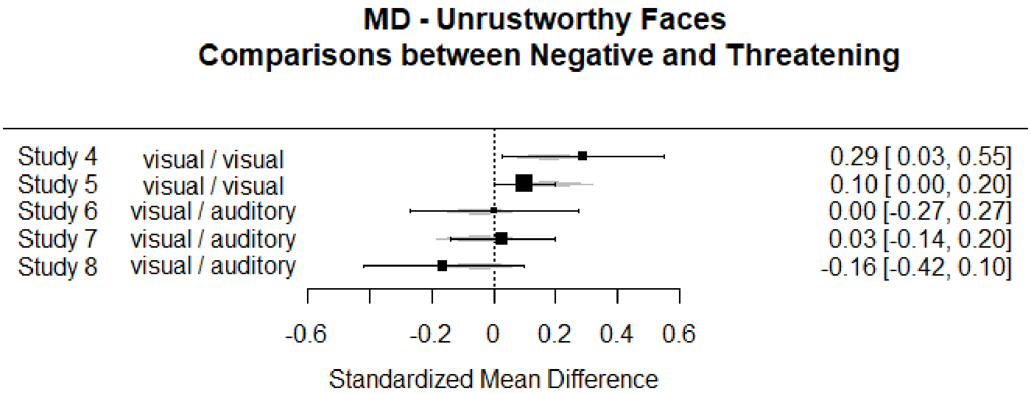

Figure 5. 2 Forest plot of standardized mean difference in MD between negative and threat conditions for untrustworthy faces

Those two aggregated effects must be estimated using a random-effect model and must be kept separate due to their high heterogeneity, Q(1) = 4.5814, *p* = .03. Such value of heterogeneity signals that there is a substantial difference between visual-only and auditory studies.

5.3.2 Process-sensitive measures: Area Under the Curve

The second dependent variable we analyzed is the Area Under the Curve. Following the same procedure used for MD, we gathered Cohen's d regarding AUC from Experiment 4 to Experiment 8 along with the associated variances. As in the previous paragraph, we first aggregated the effects of the comparison between negative and threat condition in the categorization of trustworthy faces (Table 5.3).

	Sample Size	Cohen's d	Variance	Manipulation
Experiment 4	45	-.38	.016	visual
Experiment 5	60	-.24	.008	visual
Experiment 6	39	.03	.010	auditory
Experiment 7	52	.06	.004	auditory
Experiment 8	47	.05	.025	auditory

Table 5. 3 Effect Sizes of the negative vs threat comparisons on AUC for trustworthy faces

The negative effect sizes in experiments 4 and 5 indicate that AUC was wider in the threat condition than in the negative condition. This can be taken as a signal that processing a trustworthy face in the presence of threatening information is less easy than doing the same in the presence of negative but non-threatening information. However, in experiments 6, 7, and 8, the effect sizes are close to zero.

Once gathered the results, we ran a random-effect meta-analysis in order to aggregate such effects. The analysis estimated a null effect when aggregating the effect of the auditory experiments, d = .051, se = .0658, z = .771, p = .44, $CI_{95\%}$ = [-.078, .180], but a significant effect when aggregating the effect of the visual-only experiments, d = -.349, se = .0994, z = -3.515 p = .0004, $CI_{95\%}$ = [-.544, -.154] (Figure 5.3).

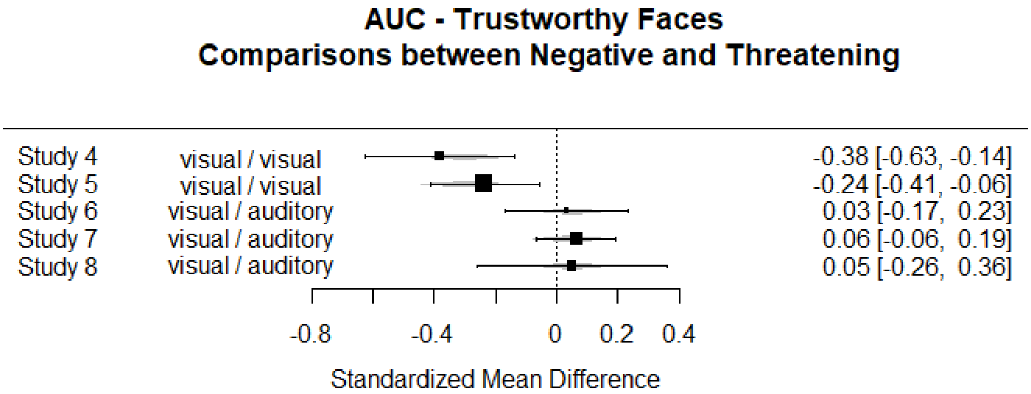

Figure 5. 3 Forest plot of standardized mean difference in AUC between negative and threat conditions for trustworthy faces

Those two aggregated effects must be estimated using a random-effect model and must be kept separate due to their high heterogeneity, Q(1) = 12.35, *p* = .0004. Such value of heterogeneity signals that there is a substantial difference between visual-only and auditory studies.

Then, we aggregated the effects of the comparison between negative and threat condition in the categorization of untrustworthy faces (Table 5.4).

	Sample Size	Cohen's d	Variance	Manipulation
Experiment 4	45	.32	.019	visual
Experiment 5	60	.15	.011	visual
Experiment 6	39	.04	.018	auditory
Experiment 7	52	.04	.009	auditory
Experiment 8	47	-.12	.018	auditory

Table 5. 4 Effect Sizes of the negative vs threat comparisons on AUC for untrustworthy faces

This time, the positive effect sizes in experiments 4 and 5 indicate that AUC was narrower in the threat condition than in the negative condition. This can be taken as a signal that processing an untrustworthy face in the presence of threatening information is easier than doing the same in the presence of negative but non-threatening information. This time, the effects are close to zero in experiments 6 and 7, whether the effect becomes slightly negative in Experiment 8.

After gathering the results, we ran a random-effect meta-analysis in order to aggregate such effects. As before, the analysis estimated a null effect when aggregating the effect of the auditory experiments, *d* = -.013, *se* = .0689, *z* = -.189, *p* = .85, $CI_{95\%}$ = [-.148, .122], but a significant effect when aggregating the effect of the visual-only experiments, *d* = .240, *se* = .1092, *z* = 2.196, *p* = .03, $CI_{95\%}$ = [.026, .454] (Figure 5.4).

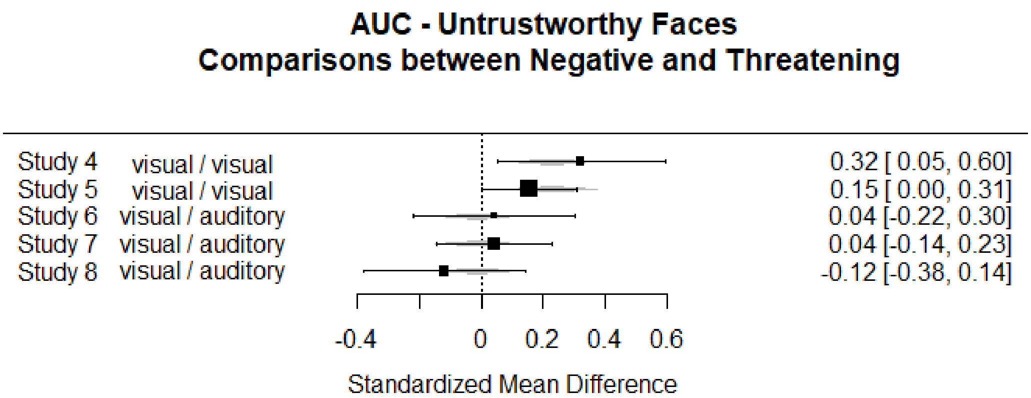

Figure 5. 4 Forest plot of standardized mean difference in AUC between negative and threat conditions for untrustworthy faces

Those two aggregated effects must be estimated using a random-effect model and must be kept separate due to their high heterogeneity, Q(1) = 4.82, *p* = .03. Such value of heterogeneity signals that there is a substantial difference between visual-only and auditory studies.

5.4 Summary

The two meta-analyses of this chapter allow for a comprehensive view of the experiments presented in the present work. According to the process-sensitive measures we employed, it seems that the combination of untrustworthiness related features and threat-specific information synergistically enhance the processing of the social target in the categorization task. At the same time, such information disrupts the efficiency of the categorization process in the presence of trustworthiness related features. However, such effects are conditional to the belonging of both information on a common perceptual system. In fact, when threat-specific information is drawn from a different perceptual system than trustworthiness and untrustworthiness features such effects vanish.

CHAPTER 6

GENERAL DISCUSSION

6.1 Summary of Main Findings

The present work intended to explore the role of contextual information in shaping the perception of facial trustworthiness. Specifically, our main aim was to test the possibility that threat-specific information attached to the social stimulus may impact the trustworthiness perception of the target. Despite a lot of work investigated trustworthiness perception (for a review, Todorov et al., 2008), the impact of contextual information is underexplored. Therefore, our experimental studies aimed at extending and complementing prior research evidence by considering the impact of contextual information upon face perception.

In Chapter 2, we explored the impact of threat information conveyed by the visual context on the perception of trustworthiness of the social target. Our data show that such information can be integrated with trustworthiness cues. Specifically, in an explicit rating task (Experiment 1), we found that threat information can shift trustworthiness perception of the target toward the untrustworthy side of the continuum. Thus, this kind of information is taken into account when elaborating a judgement of the person embedded in the threatening visual context. Moreover, even if this effect was found in the presence of negative information, the impact of threat-specific information went above and beyond the impact of negative information. Such a pattern indicates that the effect may be due the specificity of threat. In fact, threat contexts shared the same valence of negative context but only the former carried threat-specific information. The effect of the two equally negative types of context (i.e. negative and threatening conditions) points out that the threat-specific effect lays on top of a more general valence effect that, in any case, is not the only one taking place.

Experiment 2 replicated and extended the findings of Experiment 1 by employing a categorization task. Indeed, threat-specific information in the visual context resulted in a lower error rates when untrustworthy was the correct categorization. Such a result is in line with the result of Experiment 1. In both cases, the presence of threat-specific information impacted upon the outcome of the process. In Experiment 2, such outcome takes the form of categorizing the target as trustworthy or untrustworthy, but the effects highlight the congruency between trustworthiness cues and the content of threat-specific information. Here, we found the same effect of negative contexts that suggests the presence of a mere valence congruency effect. Such congruency effect however, is outperformed buy the effect of threat-specific information. In fact, threat-specific information showed an effect that goes above and beyond mere valence effect.

In Experiment 2, we substantially replicated the findings of Experiment 1. The key difference is the experimental task and the format of the trustworthiness measure. Trustworthiness is generally framed as a continuum but such a construct always translates in discrete outcomes such as trust or not trust someone, approach or avoid the encounter target and so on (Todorov et al., 2015). Because of such discrete outcome, we introduced Experiment 2. In so doing, we tested the same hypothesis of Experiment 1 by probing if the findings of that experiment can be reliable beyond the peculiarity of the construct as formalized in the literature.

In Chapter 3, we investigated the impact of threat information conveyed by the visual context on the perception of trustworthiness by examining the ongoing categorization process. In Experiment 3, we found evidence of a smoother categorization of the untrustworthy faces embedded in a threat-specific context rather than when the same face appears in other contexts. By contrast, we found a less smooth categorization when trustworthy faces appear in threatening contexts. Such findings were obtained relying on process sensitive mouse-tracking measure. By measuring MD and AUC we indexed the easiness of the categorization process while it unfolds over time.

In Experiment 4, we gathered some evidence on the impact of threat information on the perception of trustworthiness testing for the specificity of threat. By adding a negative condition, we tested whether the effect find in Experiment 3 was threat-specific or just due to mere valence. Results showed that threat-specific information had an impact that went above and beyond mere valence. In fact, the trajectories indicated a smoother integration of threat-specific information with untrustworthy cues in comparison with negative information. Such pattern of results pointed out that, on top of a mere valence effect, an effect specifically due to the congruence between untrustworthiness and threat can be found. Our explanation is that (un)trustworthiness information is inherently congruent with threat-specific information because both kinds of information signal danger to the perceiver.

In Experiment 5, we found converging results with previous findings. We replicated the pattern of results of Experiment 4 employing a new and more realistic set of stimuli. Even in this case, the categorization of trustworthy faces embedded in threatening contexts was relatively harder than the same categorization in non-threatening contexts. The categorization of untrustworthy faces doesn't mirror perfectly such results. In fact, in this case the difference between threat-specific and negative contexts was marginal for MD, but the same difference held for the comparison between neutral versus threat-specific contexts. Besides this marginal effect, AUC showed the same pattern as previous experiments and the aggregated difference in MD across similar studies between negative and threat-specific context was greater than zero.

Overall, when the experiments tested the impact of threat-specific information on trustworthiness categorization, and when information was conveyed on the same perceptual system (i.e. the visual system), we found evidence of an incidental integration of the two sources of information.

We pushed our inquiry outside the visual system, trying to investigate whether if the information integration we found could be found in higher level cognitive structure. Put differently, we tested if it is possible that information conveyed on different perceptual systems could be integrated. In Chapter 4, we provided participant with auditory contextual information, in order to assess at which level the integration between threat-specific information and trustworthiness cues takes place.

Across 3 experiments, we found null results regarding the integration under investigation. Despite different synchrony between the stimuli onset, and whether the threat information was human-made (i.e., human voices) or environmental (i.e., background sounds), we found no evidence of any incidental integration of such information with social information drawn from social targets.

In Experiment 6, no differences were found between negative and threat conditions for both trustworthy and untrustworthy faces. A possible explanation could be the asynchrony between the auditory and visual stimuli onsets. It may be that the delay between the onset of the sound and the onset of the face allowed the participants to unbound the two stimuli, making any incidental information integration easily avoidable.

Therefore, in Experiment 7, we shortened the time window between the presentation of the auditory and the visual stimuli. Such modification however, didn't impacted the results. In Experiment 7, there were no differences between trajectories in the negative and threat conditions for both trustworthy and untrustworthy faces. Here, we hypothesized that the lack of effect could be due the nature of the auditory stimuli. In fact, in Experiment 7 as well as in Experiment 6, auditory stimuli were artificial background sounds. Since such set of stimuli was not human-made, it may be that the information conveyed by the sounds is easily kept apart from the inherently social information conveyed by the social target. Experiment 8 was carried out in order to avoid such shortcoming. However, despite the introduction of a human-made set of auditory stimuli, the effect we were looking for doesn't show up.

Aggregating the effects sizes across all three multi-channel experiments indicated that no differences exists in mouse-tracking indexes between negative and threat condition for both trustworthy and untrustworthy faces when the information to be integrated are delivered on different perceptual systems. This result is in contrast with the aggregated effects when the information to be integrated are delivered on the same perceptual system. When both social information and threat information are conveyed along the visual system, the differences between negative and threat condition are different from zero for both trustworthy and untrustworthy faces. Such differences signal a facilitation effect of threat-specific information on the categorization of untrustworthy faces

6.2 Theoretical Implications

The main theoretical implications of the present work stem from our claim that untrustworthiness and threat-specific information are inherently associated. Complementing previous works revealing that trustworthiness from faces is detected faster than other human traits (Todorov et al., 2015; Willis & Todorov, 2006), our findings indicated threat as another factor able to promote or disrupt the processing of facial trustworthiness. Complementing this line of research, our findings extends the list of factors promoting or disrupting the processing of facial trustworthiness. For instance, extensive work has revealed that individuals detect trustworthiness in faces faster than other human traits (Willis & Todorov, 2006). The amygdala may process a face's trustworthiness so rapidly that perceptual awareness is not required (Freeman et al., 2014). However, most studies in this area have examined faces without any contextual information. Our work place itself in this unexplored area of the literature by extending prior research and showing that judgments of facial trustworthiness can be modified when individuals perceive the background information at the same time. Our research speaks to the situated nature of trustworthiness such that its perception is readily pushed around by scene context.

Our findings have also implication for research on impression formation. Indeed, it has been shown that prior knowledge regarding a target person may affect the evaluation of facial trustworthiness (Mende-Siedlecki, Cai, & Todorov, 2013). Extraneous information from the face (i.e., person knowledge) may impact evaluations of the face. Accordingly, our findings added visual context to the list of factors that may alter the evaluations of facial cue. Both previous knowledge and visual context can potentially trigger top-down constraints that may systematically introduce biases in social perception.

A functional approach claims that the main purpose of social perception is to guide people's behavior allowing them to approach opportunities and avoid threats (Dunning, 2004; Heider, 1958; Zebrowitz & Collins, 1997). Relying on a framework in which "social thinking is for doing" (Fiske, 1992), we place the present work in such a functional approach. According to this approach, it is plausible that, in order to survive, humans should be particularly effective in recognizing malevolent social targets (i.e., untrustworthy) especially when the context might make them able to enact their bad intentions (i.e., threatening situations). In this perspective, contextual information plays a central role because by impacting trustworthiness perception, it affects all the cognitive, affective and behavioral effects that stem from such perception (Todorov et al., 2015). Accordingly, such type of information is playing an adaptive function by enhancing the effectiveness of human's untrustworthiness perception.

Starting from this point of view, the present work has theoretical implications in some areas of inquiry in social perception. Our work provides an original contribution to the literature on the influence of context in person perception by moving beyond the investigation of emotions (Aviezer, Hassin, Ryan, Grady, Susskind, Anderson, Moscovitch & Bentin, 2008; Barrett & Kensinger, 2010) and social categories (Freeman et al., 2013; Freeman, Ma, Barth, Young., Han & Ambady, 2015). In fact, our research shows that contextual information conveyed on the same perceptual system used to draw social information can impact the evaluation and categorization of fundamental traits such as trustworthiness.

Moreover, the present work adds some insights about the difference in AUC and MD in the mouse-tracking literature. In fact, AUC may occasionally show higher sensitivity than the MD. Such difference is generally explained by the fact that AUC incorporates the aggregated spatial attraction effect over the entire time series rather than only a single maximal point (Freeman & Ambady, 2010; Hehman et al., 2015). However, in the present work, both measures are generally in accordance. In fact, when controlling for the channel that conveys contextual information, both MD and AUC measure yielded consistent findings. The overall direction and pattern of results was consistent across both measures even when statistical significance is not reached as in the case of multisensory information integration. The present work suggests that both measures are reliable enough to be used in experimental research. Beside the present work, AUC may occasionally provide more consistent results leaving the researcher with a safe choice about which measure to rely on, but in our case, MD proved to be as good as its counterpart.

6.3 Limits and Further Developments

The present work has some limitations that could be improved by future works. In our work, we developed eight experiments using static and very specific stimuli. By relying on face stimuli, we excluded the body component of social interaction. In fact, it has been shown that body posture can influence the social perception of the target (Actis-Grosso, Bossi, & Ricciardelli, 2015.; Kret, Stekelenburg, Roelofs & de Gelder, B. 2013). Further work may investigate the interplay of information drawn from face and body in the same social target. A possible implementation could be replicating our experiments by introducing full body picture stimuli. In such an implementation, different body posture could be used to manipulate threat. For example, body postures with arms extended toward the experimental subject could be perceived as more threatening than body posture in which the social target has its arm crossed on its chest. Such manipulation may shed some light on different aspects of social perception such as the overall processing of the whole social target or the integration of facial expression and intentions inferred from target posture.

Moreover, by presenting static stimuli, our work cannot speak on the continuous modulation of social perception. It is possible that some information is extracted by the dynamics of the interaction rather than the static cues present in the target. In fact, a whole literature on mimicry, an inherently dynamic mechanism, propose that trust can be built (therefore, perceived) by such a bonding process (Kret & de Dreu, 2017; Kret, Fischer & de Dreu, 2015; Kret, Tomonaga, de Dreu, Fischer & Matsuzawa, 2014). In this line of research, it has been shown that dynamic perception is correlated with trusting decision. It is plausible that a social target that mimics the subject could be perceived as more trustworthy, opening a way for a great variety of new research. It could be interesting to test whether mimicry can override the static untrustworthy perception of an untrustworthy-looking target by making it being perceived as more trustworthy due to synchronizing with the experimental subject. If this is true, we can test if the social context provided by the interaction partner itself may shift the perceiver's perception.

Since we focused on threat, we did not consider the impact of positive contextual information in driving our findings. Given a functional approach (see paragraph 6.2), positive information doesn't necessarily have to promote the detection and categorization of trustworthy cues but they may have an impact as well. It is possible that due to increased congruency between positive information and trustworthiness of the target, categorizing a trustworthy face in a positive visual context may be easier than categorizing the same stimulus in a neutral or non-positive context. In fact, even if we found evidence on the specificity of threat in impacting trustworthiness perception, our effect of interest lays on top of a more general congruence effect. However, such mere congruency effect requires direct empirical test to be claimed.

Moreover, the congruency between threat-specific information and trustworthiness is not the only congruence that may play a role in social perception. It is possible that other type of information or other characteristics of the stimuli may elicit a sort of congruence with social perception dimension. In addition, the format of contextual information is still underexplored. We provided evidences pointing out the integration of information on the visual channel but information to be integrated may take other forms. Further research investigating whether contextual information in different forms (i.e. expectations, target's stated intentions, participant's goals, ecc…) may impact how the perceiver process information drawn from the social target.

One last limitation regards the mouse-tracking paradigm. Since all our experiments employing that paradigm were carried out under time pressure, the time between the stimulus onset and the first movement cannot be considered a reliable measure. To overcome such limitation further studies that do not imply time pressure could be developed in order to gain insights on the very first reaction of experimental subjects engaged in the task of perceiving others in the environment.

References

Abele, A. E., & Brack, S. (2013). Preference for other persons' traits is dependent on the kind of social relationship. *Social Psychology*, 44, 84-94.

Abele, A.E. & Bruckmuller, S. (2011). The bigger one of the "Big Two"? Preferential processing of communal information. *Journal of Experimental Social Psychology*, 47, 935-948.

Abele, A. E., Cuddy, A. J., Judd, C. M., & Yzerbyt, V. Y. (2008). Fundamental dimensions of social judgment. *European Journal of Social Psychology*, 38, 1063-1065.

Abele, A. E., & Wojciszke, B. (2007). Agency and communion from the perspective of self vs. others. *Journal of Personality and Social Psychology*, 93, 751–763.

Abele, A. E., & Wojciszke, B. (2014). Communal and agentic content in social cognition: A Dual Perspective Model. *Advances in Experimental Social Psychology*, 50, 195-255.

Actis-Grosso, R., Bossi, F., & Ricciardelli, P. (2015). Emotion recognition through static faces and moving bodies: A comparison between typically developed adults and individuals with high level of autistic traits. *Frontiers in Psychology*, 6, 1-14.

Allport, G. W. (1954). *The nature of prejudice*. Reading, MA: Addison-Wesley.

Anderson, N. H. (1962). Application of an Additive Model to Impression Formation. *Science*, 138, 817–818

Anderson, C., John, O. P., Keltner, D., & Kring, A. M. (2001). Who attains social status? Effects of personality and physical attractiveness in social groups. *Journal of Personality and Social Psychology*, 81(1), 116-132.

Anderson, N. H. (1965). Primacy effects in personality impression formation using a generalized order effect paradigm. *Journal of Personality and Social Psychology*. 2(1): 1–9.

Anderson, N. H (1973). A Simple Model for Information Integration. In R.P. Abelson, E. Aronson, W.J. McGuire, T.M. Newcomb, M.J. Rosenberg, & P.H. Tannenbaum (Eds.), *Theories of Cognitive Consistency: A Sourcebook*. Chicago: Rand McNally.

Anderson, N. H., & Lampel, A. (1965). Effect of context on ratings of personality traits. *Psychonomic Science*,1965, 3, 433-434.

Asch, S. E. (1946). Forming impressions of personality. *The Journal of Abnormal and Social Psychology*, 41(3), 258-290.

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

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

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

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

Bazinska, R., & Wojciszke, B. (1996). Drawing inferences on moral and competencerelated traits from the same behavioral information. *Polish Psychological Bulletin*, 27, 293–299.

Blair, I. V., Judd, C. M., & Chapleau, K. M. (2004). The influence of Afrocentric facial features in criminal sentencing. *Psychological Science*, 15, 674–679.

Blanz, V., & Vetter, T. (1999). A morphable model for the synthesis of 3D faces. In Proceedings of the 26th annual conference on Computer graphics and interactive techniques, 187–194.

Bond, C. F., Berry, D. S., & Omar, A. (1994). The kernel of truth in judgments of deceptiveness. *Basic Applied Social Psychology*, 15(4), 523–534.

Bradley, M. M. & Lang, P. J. (2007). The International Affective Digitized Sounds (2nd Edition; IADS-2): Affective ratings of sounds and instruction manual. Technical report B-3. University of Florida, Gainesville, Fl.

*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.

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

Brambilla, M., Rusconi, P., Sacchi, S., & Cherubini, P. (2011). Looking for honesty: The primary role of morality (vs. sociability and competence) in information gathering. *European Journal of Social Psychology*. 41, 135-143.

Brambilla, M., Sacchi, S., Menegatti, M., & Moscatelli, S. (2016). Honesty and Dishonesty Don't Move Together: Trait Content Information Influences Behavioral Synchrony. *Journal of Nonverbal Behavior*, 40, 171–186.

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

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

Braver, S. L., Thoemmes, F. J., & Rosenthal, R. (2014). Continuously cumulating meta-analysis and replicability. *Perspectives on Psychological Science*, 9, 333–342.

Breiter, H. C., Etcoff, N. L., Whalen, P. J., Kennedy, W. A., Rauch, S. L., Buckner, R. L., Strauss, M. M., Hyman, S. E., & Rosen, B. R. (1996). Response and habituation of the human amygdala during visual processing of facial expression. *Neuron*, 17, 875–887.

Brewer, M. B. (1988). A dual process model of impression formation. In T. K. Srull & R. S. Wyer, Jr. (Eds.), *Advances in social cognition*, Vol. 1. A dual process model of impression formation (pp. 1-36). Hillsdale, NJ, US: Lawrence Erlbaum Associates, Inc.

Brewer, M. B., & Feinstein, A. S. H. (1999). Dual processes in the cognitive representation of persons and social categories. In S. Chaiken & Y. Trope (Eds.), *Dual-process theories in social psychology* (pp. 255–270). New York: The Guilford Press.

Bruckmuller, S., & Abele, A. E. (2013). The den-sity of the big two: How are agency and communion structurally represented? *Social Psychology*, 44, 63-74.

Bruner, J. S. (1957). On perceptual readiness. *Psychological Review*, 64(2), 123–152.

Calder, A. J., Beaver, J. D., Winston, J. S., Dolan, R. J., Jenkins, R., Eger, E., & Henson, R. N. A. (2007). Separate coding of different gaze directions in the superior temporal sulcus and inferior parietal lobule. *Current Biology*, 17, 20–25.

Calvert, G. A., Brammer, M. J., & Iversen, S. D. (1998). Crossmodal identification. *Trends in Cognitive Sciences*, 2(7), 247-253.

Caprariello, P. A., Cuddy, A. J. C., & Fiske, S. T. (2009). Social Structure Shapes Cultural Stereotypes and Emotions: A Causal Test of the Stereotype Content Model. *Group Processes & Intergroup Relations*, 12(2), 147-155.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hilsdale. NJ: Lawrence Earlbaum Associates2.

Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112(1), 155-159.

Cottrell, C. A., Neuberg, S. L., & Lee, N. P. (2007). What do people desire in others? A sociofunctional perspective on the importance of different valued characteristics. *Journal of Personality and Social Psychology*, 92(2), 208-231.

Cowman, N. (2010). The Magical Mystery Four: How Is Working Memory Capacity Limited, and Why? *Current Directions in Psychological Science*, 19(1), 51-57.

Cuddy, A. J. C., Fiske, S. T., Kwan, V. S. Y., Glick, P., Demoulin, S., Leyens, J., Bond, M. H., Croizet, J., Ellemers, N., Sleebos, E., Htun, T. T., Kim, H. J., Maio, G., Perry, J., Petkova, K., Todorov, V., Rodríguez-Bailón, R., Morales, E., Moya, M., Palacios, M., Smith, V., Perez, R., Vala, J., Ziegler, R. (2009). Stereotype content model across cultures: towards universal similarities and some differences. *British Journal of Social Psychology*, 48(1), 1–33.

Cuddy, A. J. C., Fiske, S. T., & Glick, P. (2007). The BIAS map: Behaviours from intergroup affect and stereotypes. *Journal of Personality and Social Psychology*, 92(4), 631-648.

Cuddy, A. J. C., Fiske, S. T., Glick, P. (2008). Warmth and Competence as Universal Dimensions of Social Perception: The Stereotype Content Model and the BIAS Map. *Advances in Experimental Social Psychology*, 40(1), 61–149.

Cumming, G. (2014). The new statistics: Why and how. *Psychological Science*, 25, 7–29.

De Bruin, E., & Van Lange, P. (2000). What people look for in others: Inferences of the perceiver and the perceived on information selection. *Personality and Social Psychology Bulletin*, 26, 206–219.

De Waal, F. (1996). *Good natured: The origins of right and wrong in humans and other animals*. Cambridge, MA: Harvard University Press.

Dijksterhuis, A. (2010). Automaticity and the unconscious. In S. T. Fiske, D. T. Gilbert, & G. Lindzey (Eds.), *Handbook of social psychology* (5th ed., Vol. 2, p. 228-267). Hoboken, NJ: John Wiley & Sons.

Dunbar, R. I. M. (2004). Gossip in evolutionary perspective. *Review of General Psychology*, 8(2), 100-110.

Durante, F., Fiske, S. T., Kervyn, N., Cuddy, A. J., Akande, A. D., Adetoun, B. E., Adewuyi, M.F., Tserere, M. M., Ramiah, A. A., Mastor, K. A., Barlow, F. K., Bonn, G., Tafarodi, R. W., Bosak, J., Cairns, E., Doherty, C., Capozza, D., Chandran, A., Chryssochoou, X., Iatridis, T., Contreras, J. M., Costa-Lopes, R., González, R., Lewis, J. I., Tushabe, G., Leyens, J. P., Mayorga, R., Rouhana, N. N., Castro, V. S., Perez, R., Rodríguez-Bailón, R., Moya, M., Morales Marente, E., Palacios Gálvez, M., Sibley, C. G., Asbrock, F., & Storari, C. C. (2008). Nations' income inequality predicts ambivalence in stereotype content: How societies mind the gap. *British Journal of Social Psychology*, 52(4), 726-746.

Ekman, P., Friesen, W. V., & Ellsworth, P. (1972). *Emotion in the Human Face*, Pergamon.

Ellemers, N., Pagliaro, S., & Barreto, M. (2013). Morality and behavioural regulation in groups: A social identity approach. *European Review of Social Psychology*, 24(1), 160-193.

Ellemers, N., Pagliaro, S., Barreto, M., & Leach, C. W. (2008). Is it better to be moral than smart? The effects of morality and competence norms on the decision to work at group status improvement. *Journal of Personality and Social Psychology*, 95, 1397-1410.

Ellemers, N., & van den Bos, K. (2012). Morality in groups: On the social-regulatory functions of right and wrong. *Social and Personality Psychology Compass*, 6, 878–889.

Ellemers, N., van Knippenberg, A., De Vries, N., & Wilke, H. (1988). Social identification and permeability of group boundaries. *European Journal of Social Psychology*, 18(6), 497-513.

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

Fehr, E., Fischbacher, U. & Gächter, S. (2002). Strong reciprocity, human cooperation, and the enforcement of social norms. *Human Nature*, 13, 1–25.

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

Fiske, S. T., Cuddy, A. J. C., Glick, P. (2007). Universal dimensions of social cognition: warmth and competence. *Trends in Cognitive Sciences*, 11(2), 77–83.

Fiske, S. T., Cuddy, A. J. C., Glick, P., Xu, J. (2002). A Model of (Often Mixed) Stereotype Content: Competence and Warmth Respectively Follow from Perceived Status and Competition. *Journal of Personality and Social Psychology*, 82(6), 878–902.

Fiske, S. T., & Neuberg, S. L. (1990). A continuum of impression formation, from category-based to individuating processes: Influences of information and motivation on attention and interpretation. *Advances in experimental social psychology*, 23, 1–74.

Fiske, S. T., & Taylor, S. E. (1984). *Social cognition*. New York: Random House.

Fiske, S. T., Xu, J., Cuddy, A. J. C., Glick, P. (1999). (Dis)respecting versus (Dis)liking: Status and Interdependence Predict Ambivalent Stereotypes of Competence and Warmth. *Journal of Social Issues*, 55(3), 473–489.

Fox, A. S., Oler, J. A., Tromp, D. P. M., Fudge, J. L., & Kalin, N. H. (2015). Extending the amygdala in theories of threat processing. *Trends in Neurosciences*, 38(5), 319-329.

Freeman, J.B. & Ambady, N. (2011). When two become one: Temporally dynamic integration of the face and voice. *Journal of Experimental Social Psychology*, 47, 259-263.

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

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

Freeman, J. B., Penner, A. M., Saperstein, A., Scheutz, M., & Ambady, N. (2011). Looking the Part: Social Status Cues Shape Race Perception. *PLOS ONE*, 6(9): e25107.

Gert, B. (1988). *A new justification of the moral rules*. Oxford: Oxford University Press.

Hammond, K. R. 2007.*Beyond rationality: The searchfor wisdom in a troubled time.* Oxford: Oxford University Press.

Gobbini, I. M., & Haxby, J. V. 2007. Neural systems for recognition of familiar faces. *Neuropsychologia*, 45, 32–41.

Goh, J. X., Hall, J. A., & Rosenthal, R. (2016). Mini Meta-Analysis of Your Own Studies: Some Arguments on Why and a Primer on How. *Social and Personality Psychology Compass*, 10(10), 535-549.

Harris, L. T., & Fiske, S. T. (2006). Dehumanizing the Lowest of the Low: Neuroimaging Responses to Extreme Out-Groups. *Psychological Science*, 17(10), 847-853.

Haxby, J. V., Hoffman, E. A., & Gobbini, M. I. (2000). The distributed human neural system for face perception. *Trends in cognitive sciences*, 4(6), 223–232.

Hehman, E., Stolier, R. M., & Freeman, J. B. (2015). Advanced mouse-tracking techniques fo enhancing psychological science. *Group Processes & Intergroup Relations*, 18(3), 384-401.

Higgins, E. T., Bargh, J. A., & Lombardi, W. J. (1985). Nature of priming effects on categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 11(1), 59-69.

Hoffman, E.A., & Haxby, J.V. (2000). Distinct representation of eye gaze and identity in the distributed human neural system for face perception. *Nature Neuroscience*, 3, 80–84.

Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology*, 24, 417–441 498–520.

Hume, D. (1896). *A treatise of human nature*. Oxford, Clarendon press.

Ishai, A. (2008). Let’s face it: it’s a cortical network. *NeuroImage*, 40(2), 415–419.

Iyer, A., Jetten, J., & Haslam, S. A. (2012). Sugaring o'er the devil: Moral superiority and group identification help individuals downplay the implications of ingroup rule-breaking. *European Journal of Social Psychology*, 42(2), 141-149.

Kant, I., & Wood, A. (1996). Groundwork of The metaphysics of morals (1785). In M. Gregor (Ed.), *Practical Philosophy* (pp. 37-108). Cambridge: Cambridge University Press.

Kaplan, M. F. (1969). Assimilation-contrast effects in impression formation. *Psychonomic Science*, 17 (4): 234-235.

Kervyn, N., Fiske, S., Yzerbyt, V. (2015). Forecasting the Primary Dimension of Social Perception. *Social Psychology*. 46(1), 36–45.

Kret, M. E., & de Dreu, C. K. W. (2017). Pupil-mimicry conditions trust in exchange partners: Moderation by oxytocin and group membership. *The Royal Society B-Biological Sciences*, 284(1850), 20162554.

Kret, M. E. Fischer, A. H. & de Dreu, C. K. W. (2015). Pupil-mimicry correlates with trust in in-group partners with dilating pupils. *Psychological Science*, 26(9), 1401-1410.

Kret, M. E., Stekelenburg, J. J., Roelofs, K., & de Gelder, B. (2013). Perception of face and body expressions using electromyography, pupillometry and gaze measures. *Frontiers in Psychology*, 4(28), 1-12.

Kret, M. E. Tomonaga, M. de Dreu, C. K. W., Fischer, A. H. & Matsuzawa, T. (2014). Role of pupil-synchronization in trust. *International Journal of Psychophysiology*, 94, 174.

Krumhuber, E., Manstead, A. S. R., Cosker, D., Marshall, D., Rosin, P. L., & Kappas, A. (2007). Facial dynamics as indicators of trustworthiness and cooperative behavior. *Emotion*, 7(4), 730-735.

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. *Frontiers in Psychology*, 4, 1-12.

Lamme, V. A. F., & Roelfsema, P. R. (2000). The distinct modes of vision offered by feedforward and recurrent processing. *Trends in Neurosciences*, 23(11), 571-579.

Lamont, M. & Molnar, V. (2002). The Study of Boundaries in the Social Sciences. *Annual Review of Sociology*, 28, 167-195.

Leach, C. W., Bilali, R., & Pagliaro, S. (2015). Groups and Morality. In M. Mikulincer and P. R. Shaver. *APA handbook of personality and social psychology*: Vol. 2. (pp. 123-149) Group Processes.

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

Lind, E. A., & Tyler, T. R. (1988). The Social Psychology of Procedural Justice. New York: Plenum Press.

Lin, T., Ankudowich, E., & Ebner, N. C. (2017). Greater perceived similarity between self and own-age others in older than young adults. *Psychology and Aging*, 32(4), 377-387.

Li, T., Cardenas-Iniguez, C., Correll, J., & Cloutier, J. (2016). The impact of motivation on race-based impression formation. *NeuroImage*, 124, 1-7.

Lippmann, W. (1922). *Public opinion*. Nueva York, Free Press.

Little, A. C., Burriss, R. P., Jones, B. C., & Roberts, S. C. (2007). Facial appearance affects voting decisions. *Evolution and Human Behavior*, 28, 18–27.

Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, 390, 279-281.

Luo, Q., Holroyd, T., Jones, M., Hendler, T., & Blair, J., (2007). Neural dynamics for facial threat processing as revealed by gamma band synchronization using MEG. *NeuroImage*, 34(2), 839-847.

MacIntyre, A. (1984). *After virtue: A study in moral theory* (2nd ed.). Notre Dame, IN: University of Notre Dame Press.

Macrae, C. N., & Quadflieg, S. (2010). Perceiving people. In S. T. Fiske, D. T. Gilbert, & G. Lindzey (Eds.), *Handbook of social psychology* (5th ed., Vol. 1, p. 428-463). Hoboken, NJ: John Wiley & Sons.

Maner, J. K. (2014). Let's put our money where our mouth is: If authors are to change their ways, reviewers (and editors) must change with them. *Perspectives on Psychological Science*, 9, 343–351.

Martin, L. L. (1986). Set/Reset: Use and Disuse of Concepts in Impression Formation. *Journal of Personality and Social Psychology*, 51(3), 493-504.

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

Mc Pherson, M., Smith-Lovin, L., & Cook, J. M. (2001). Birds of Feather Homophily in Social Networks. *Annual Review of Sociology*, 27, 415-444.

Mende-Siedlecki, P., Cai, Y., & Todorov, A. (2013). The neural dynamics of updating person impressions. *Social Cognitive and Affective Neuroscience*, 8, 623–631.

Morris, J. S., Frith, C. D., Perrett, D. I., Rowland, D., Young, A. W., Calder, A. J., & Dolan, R. J. (1996). A differential neural response in the human amygdala to fearful and happy facial expressions. *Nature*, 383, 812–815.

Newcomb, M.J. Rosenberg, & P.H. Tannenbaum (Eds.), *Theories of Cognitive Consistency: A Sourcebook*. Chicago: Rand McNally, 1968.

Nietzsche, F. (1967). *On the genealogy of morals* (W. Kaufmann & R. J. Hollingdale, Trans.). New York: Random House. (Original work published in 1887).

Notebaert, W., Gevers, W., Verbruggen, F., & Liefooghe, B. (2006). Top-down and bottom-up sequential modulations of congruency effects. *Psychonomic Bulletin & Review*, 13(1), 112-117.

Ohman, A. (2002). Automaticity and the Amygdala: Nonconscious Responses to Emotional Faces. *Current Directions in Psychological Science*, 11(2), 62-66.

Oosterhof, N. N., & Todorov, A. (2009). Shared perceptual basis of emotional expressions and trustworthiness impressions from faces. *Emotion*, 9(1), 128-133.

Pagliaro, S., Brambilla, M., Sacchi, S., D'Angelo, M., & Ellemers, N. (2013). Initial impressions determine behaviours: Morality predicts the willingness to help newcomers. *Journal of Business Ethics*, 117, 37-44.

Pagliaro, S., Ellemers, N., & Barreto, M. (2011). Sharing Moral Values: Anticipated Ingroup Respect as a Determinant of Adherence to Morality-Based (but Not Competence-Based) Group Norms. *Personality and Social Psychology Bulletin*, 37(8), 1117–1129.

Pearson, K. (1901). On lines and planes of closest fit to systems of points in space. *Philosophical Magazine*, 2, 559–572.

Phillips, M. L., Young, A. W., Senior, C., Brammer, M., Andrew, C., Calder, A. J., Bullmore, E. T., Perrett, D. I., Rowland, D., Williams, S. C. R., Gray, J. A., & David, A. S. (1997). A specific neural substrate for perceiving facial expressions of disgust. *Nature*, 389, 495–498.

Puce, A., Allison, T., Bentin, S., Gore, J. C., & McCarthy, G. (1998). Temporal cortex activation in humans viewing eye and mouth movements. *Journal of Neuroscience*, 18, 2188–2199.

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

Robinson, D. N. (1989). *Aristotle's psychology*. New York: Columbia University Press.

Rodriguez Mosquera, P. M., Manstead, A. S. R., & Fischer, A. H. (2002). Honor in the Mediterranean and Northern Europe. *Journal of Cross-Cultural Psychology*, 33(1), 16-36.

Rosch, E., & Lloyd, B. (1978). *Cognition and categorization*. Lawrence Erlbaum Associates.

Sherif, M., & Hovland, C. I. (1961). *Social judgment: Assimilation and contrast effects in communication and attitude change*. Oxford, England: Yale Univer. Press.

Robinson, D. N. (1989). *Aristotle's psychology*. New York: Columbia University Press.

Rosenberg, S., Nelson, C., & Vivekananthan, P. S. (1968). A multidimensional approach to the structure of personality impressions. *Journal of Personality and Social Psychology*, 9, 283-294.

Rossion, B., Caldara, R., Seghier, M., Schuller, A.M., Lazeyras, F., & Mayer, E. (2003). A network of occipito-temporal face-sensitive areas besides the right middle fusiform gyrus is necessary for normal face processing. *Brain*, 126, 2381–2395.

Rossion, B. (2008). Constraining the cortical face network by neuroimaging studies of acquired prosopagnosia. *NeuroImage*, 40(2), 423–426.

Russell, A. M. T., & Fiske, S. T. (2008). It's all relative: Competition and status drive interpersonal perception. *European Journal of Social Psychology*, 38, 1193-1201.

Santos, S., Almeida, I., Oliveiros, B., Castelo-Branco, M., & Nishijo, H. (2016). The Role of the Amygdala in Facial Trustworthiness Processing: A Systematic Review and Meta-Analyses of fMRI Studies. *PLOS ONE*, 11(11), e0167276.

Schwarz, N., & Bless, H. (1992). Constructing reality and its alternatives: An inclusion/exclusion model of assimilation and contrast effects in social judgment. In L. L. Martin & A. Tesser (Eds.), *The construction of social judgments* (pp. 217-245). Hillsdale, NJ, US: Lawrence Erlbaum Associates, Inc.

Stillman, P. E., Shen, X., & Ferguson, M. J. (2018). How Mouse-tracking Can Advance Social Cognitive Theory. *Trends in Cognitive Sciences*, 22(6), 531-543.

Schwartz, S. H. (1992). Universals in the content and structure of values: Theoretical advances and empirical tests in 20 countries. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 25, pp. 1–65). New York: Academic Press.

Slepian, M. L., & Ames, D. R. (2015). Internalized Imression: The Link Between Apparent Facial Trustworthiness and Deceptive Behavior Is Mediated by Targets' Expectations of How They Will Be Judged. *Psychological Science*, 27(2), 282-288.

Smith, E. E., & Medin, D. L. (1981). Categories and concepts. Harvard University Press Cambridge, MA.

Stillman, P. E., Shen, X., & Ferguson, M. J. (2018). How Mouse-tracking Can Advance Social Cognitive Theory. *Trends in Cognitive Sciences*, 22(6), 531-543.

Sutherland, C. A. M., Oldmeadow, J. A., & Young, A. W. (2016). Integrating social and facial models of person perception: Converging and diverging dimensions. *Cognition*, 157, 257-267.

Tajfel, H. (1969). Cognitive Aspects of Prejudice. *Journal of social Issues*, 25(4), 79–97.

Tajfel, H., Turner, J. C. (1979). The Social Psychology of Intergroup Relations. Brooks-Cole: Brooks-Cole. pp. 33–47.

Tooby, J., & Cosmides, L. (2010). Groups in mind: The coalitional roots of war and morality. In H. Hogh-Olesen (Ed.), *Human morality and sociality* (pp. 191–234). New York, NY: Palgrave Macmillan.

Todorov, A., Baron, S. G., & Oosterhof, N. N. (2008). Evaluating face trustworthiness: a model based approach. *Social Cognitive and Affective Neuroscience*, 3(2), 119–127.

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

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

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

Todorov, A., & Oosterhof, N. N. (2011). Modeling Social Perception of Faces. *Social Sciences*, 28(2), 117-122.

Todorov, A., Said, C. P., Engell, A. D., & Oosterhof, N. N. (2008). Understanding evaluation of faces on social dimensions. *Trends in Cognitive Science*, 12(12), 455-460.

Trzebinski, J. (1985). Action-oriented representation of implicit personalty theories. *Journal of Personality and Social Psychology*, 48(5), 1266-1278.

Turchin, P., Currie, T. E., Turner, E. A. L. & Gavrilets, S. (2013). War, space, and the evolution of Old World complex societies. *Proceeding of the National Academy of Sciences USA*, 110, 16384–16389.

Unkelbach, C., Fiedler, K., Bayer, M., Stegmüller, M., & Danner, D. (2008). Why positive information is processed faster: the density hypothesis. *Journal of Personality and Social Psychology*, 95(1), 36.

Valentine, T. (1991). A unified account of the effects of distinctiveness, inversion, and race in face recognition. *The Quarterly Journal of Experimental Psychology A*, 43, 161–204.

Van Wassenhove, V., Grant K. V., & Poeppel, D. (2007). Temporal window of integration in auditory-visual speech perception. *Neuropsychologia*, 45, 598-607.

Vuilleumier, P., Armony, J. L., Driver, J., & Dolan, R. J. (2001). Effects of attention and emotion on face processing in the human brain: an eventrelated fMRI study. *Neuron*, 30, 829–841.

Wentura, D., Rothermund, K., & Bak, P. (2000). Automatic vigilance: The attention-grabbing power of approach- and avoidance-related social information. *Journal of Personality and Social Psychology*, 78, 1024–1037.

Wiggett, A. J., & Downing, P. E. (2008). The Face Network: Overextended? (Comment on: "Let's face it: It's a cortical network" by Alumit Ishai). *NeuroImage*, 40(2), 420–422.

Wiggins, J. S., Phillips, N., & Trapnell, P. (1989). Circular reasoning about interpersonal behavior: Evidence concerning some untested assumptions underlying diagnostic classification. *Journal of Personality and Social Psychology*, 56(2), 296-305.

Wiggins, J.S. (1979) A psychological taxonomy of trait descriptive terms: the interpersonal domain. *Journal of Personality and Social Psychology*, 37, 395–412

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

Wilson, J. P., & Rule, N. O. (2015). Facial Trustworthiness Predicts Extreme Criminal-Sentencing Outcomes. *Psychological Science*, 26(8), 1325-1331.

Wilson, J. P., & Rule, N. O. (2016). Hypothetical sentencing decisions are associated with actual capital punishment outcomes: The role of facial trustworthiness. *Social Psychological and Personality Science*, 7, 331-338.

Wojciszke, B., Bazinka, R., & Jaworski, M. (1998). On the Dominance of Moral Categories in Impression Formation. *Personality and Social Psychology Bulletin*, 24(12), 1251-1263.

Ybarra, O.; Chan, E.; & Park, D. (2001). Young and Old Adults' Concerns About Morality and Competence. *Motivation and Emotion*, 25(2), 85-100.

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

Zebrowitz, L. A., Voinescu, L., & Collins, M. A., (1996). "Wide-eyed" and "crooked-faced": determinants of perceived and real honesty across the life span. *Personality and Social Psychology Bulletin*, 22(12), 1258–1269.