Inferring Intentions from Emotion Expressions in Social Decision Making

Abstract

In the last decade we have seen increasing experimental evidence that people make important inferences from emotion expressions about others' intentions in situations of interdependent decision making. Reverse appraisal has been proposed as one mechanism whereby people retrieve, from emotion displays, information about how others are appraising the ongoing interaction (e.g., does my counterpart find the current outcome to be goal conducive? Does s/he blame me for it?); in turn, from these appraisal attributions, people make inferences about the others' goals (e.g., is my counterpart likely to cooperate?) that shape their decision making. Here we review experimental evidence and progress that has been done in understanding this inferential mechanism and its relationship to other mechanisms for the interpersonal effects of emotion (e.g., emotional contagion and social appraisal). We discuss theoretical implications for our understanding of the role of emotion expression on human decision making, but also practical implications for the growing industry of socially intelligent machines (e.g., personal digital assistants and social robots).

1. Introduction

To engage in successful interaction, one must be able to attribute independent beliefs, desires, and intentions to others and predict their behavior (Amodio & Frith, 2006). Inferring others' mental states - or theory of mind reasoning - is especially important in mixed-motive situations, where others may or may not have compatible goals (Camerer, 2003; Rilling & Sanfey, 2011). Emotion expressions have been argued to serve important social functions, including communicating one's goals and intentions to others (Frijda & Mesquita, 1994; Keltner & Haidt, 1999; Keltner & Kring, 1998; Morris & Keltner, 2000). This information shapes decision making, for example, by helping identify whether others are likely to cooperate (Frank, 2004). The last twenty years have seen an increasing amount of experimental evidence demonstrating this interpersonal influence of emotion expressions in social decision making (for reviews see: Lerner, Valdesolo, & Kassam, 2015; van Kleef, de Dreu, & Manstead, 2010), including effects on concession-making (van Kleef, de Dreu, & Manstead, 2004, 2006), emergence of cooperation (de Melo, Carnevale, Read, & Gratch, 2014), fairness (Terada & Takeuchi, 2017; van Dijk, van Kleef, Steinel, & van Beest, 2008), trust building (Krumhuber et al., 2007), and everyday life (Parkinson & Simons, 2009). Progress has also been made in understanding the pathways by which these effects operate. Broadly speaking, emotions can serve to evoke emotions in others via mechanisms such as contagion (Niedenthal, Mermillod, Maringer, & Hess, 2010) or can serve as information, revealing the experiencer's mental state (Manstead & Fischer, 2001). Here, we focus on the latter pathway. In particular, we discuss evidence from our own research in support of the idea that people engage in a process of "reverse appraisal" (de Melo et al., 2014; Hareli & Hess, 2010) to infer the beliefs, desires and intentions of others by observing their emotional reactions to world events. We review evidence we have

found that supports this mechanism and discuss its implications for psychological theory and computational techniques.

Emotion theorists generally agree that emotions arise from ongoing, conscious or nonconscious, appraisal of events, situations, and behaviors in the environment with respect to the individual's beliefs and goals (Scherer & Moors, 2018). According to these appraisals, different emotions will be experienced with their associated patterns of physiological responses (Kreibig, 2010), action tendencies (Frijda, Kuipers, & ter Schure, 1989), and expressive behavior (Ekman, 1993). If emotions reflect one's goals through the appraisal process, it is reasonable to expect that emotion expressions reflect differentiated information about the expresser's appraisals and goals. Reverse appraisal, thus, proposes that people retrieve from emotion expressions, information about how others are appraising the ongoing interaction and, in turn, this leads to inferences about others' mental states (de Melo et al., 2014; Hareli & Hess, 2010). Reverse appraisal, therefore, is theory of mind reasoning that informs one's decision making based on the implicit information in the counterparts' emotion displays. This chapter reviews this mechanism (Section 2), supporting experimental evidence and corresponding theoretical implications (Section 3), practical applications in relevant domains (Section 4), and limitations in the current research and challenges for the future (Section 5).

2. Reverse Appraisal and Related Mechanisms

According to appraisal theories (for detailed surveys see: Ellsworth & Scherer, 2003; Scherer & Moors, 2018), people are constantly judging the events in the surrounding environment with respect to their own beliefs, desires and intentions. These judgments, or appraisals, are subjective and check constantly whether relevant events are present and if so, whether these events are beneficial or harmful to the individual's goals, who or what caused them, whether social norms have been broken and, how capable the individual is to cope with the consequences of the events. According to the pattern of appraisals that occurs, different emotions are experienced (e.g., through physiological sensations: Kreibig, 2010; Levenson, 2003; Scherer, 2001), expressed (e.g., through facial or vocal cues: Ekman, 1993; Keltner & Ekman, 2000; Scherer, Johnstone, & Klasmeyer, 2003) and corresponding action tendencies elicited (e.g., flight when experiencing fear; Frijda et al., 1989). These appraisals, moreover, can occur at different levels (e.g., sensorimotor, schematic and conceptual levels; Leventhal & Scherer, 1987) and these levels of processing can interact. This means appraisals need not occur consciously (e.g., novelty detection) and can occur more deliberately (e.g., whether someone's behavior is compatible with social norms). For appraisal theorists, thus, "emotions" consist of several components including the configuration of appraisals and their correlates in the central and peripheral nervous systems.

Though several appraisal theories have been proposed (Frijda, 1986; Lazarus, 1991; Ortony, Clore, & Collins, 1988; Reisenzein, 2009; Roseman, 2001; Scherer, 2001), there tends to be agreement on the underlying appraisal dimensions. The most basic dimension is perception of *novelty* (with respect to the current level of habituation), which tends to occur automatically, to draw attention to novel stimuli, and to help determine if adaptive action is required. Another appraisal dimension relates to *goal significance*, i.e., whether the event is relevant to the individual's goals or not. *Certainty* refers to the probability of the event actually occurring and is particularly relevant for the so-called prospective emotions (e.g., hope and fear), where both the probability of the event occurring and its consequences are in doubt. Another appraisal variable is *blameworthiness*, i.e., who or what is responsible for the event. Agency is critical to distinguish, for instance, anger (other-blame) from guilt (self-blame). A fifth appraisal variable refers to

coping potential, i.e., the evaluation of one's ability to deal with the situation. Finally, *norm compatibility* recognizes that people live in a social context and assesses how much the event conforms to society's norms. This dimension is relevant for moral emotions (e.g., guilt, compassion, and anger), whereby one assesses someone's behavior with respect to the values or norms in the social group (Haidt, 2003). Predictions tend to be consistent across theories for common emotions: joy occurs when the event is conducive to one's goals; sadness occurs when the event is obstructive to one's goals; anger occurs when the event is obstructive, caused by another agent and one has the power to control, or cope, with it; guilt occurs when the event is obstructive, caused by the self and is not norm compatible; and, so on.

If one accepts that emotional displays arise through a sequence of steps – (1) an event occurs, (2) it is appraised vis-à-vis an individual's beliefs, desires and intentions, (3) a specific pattern of appraisal elicits an internal emotion, and (4) it's external manifestation – then it becomes possible to make inferences about these earlier steps when we observe another's emotional expressions. This is the essence of reverse appraisal. Reverse appraisal theory argues that people essentially infer – as if running appraisal processes in reverse – from another's emotional expression, not only what they must feel, but the nature of the events and mental state that precipitated such an expression. These inferences can be further sharpened depending on what information is available to the observer. For example, de Melo, Carnevale, Read, and Gratch (2014) argue that when we observe someone's emotional reactions to an observable event, it supports detailed inferences about their mental state. In essence, appraisal theory can be seen as describing an equation that relates variables (events, mental state, emotions and action tendencies) to each other (Figure 1). To the extent that some of these variables are observable, it allows inferences about those that are unobservable. It is important to note that reverse appraisal may be a valid characterization of how people make inferences from emotional expressions independent as to the validity of appraisal theory itself. Appraisal theory is claimed to be a theory describing how emotions arise in people. Reverse appraisal is a theory about how people make *inferences* from another's emotion. These could be two distinct mechanisms. For example, reverse appraisal may simply reflect people's folk theories of how emotion "works" in people (Hareli, 2014; Weiner, 1987).

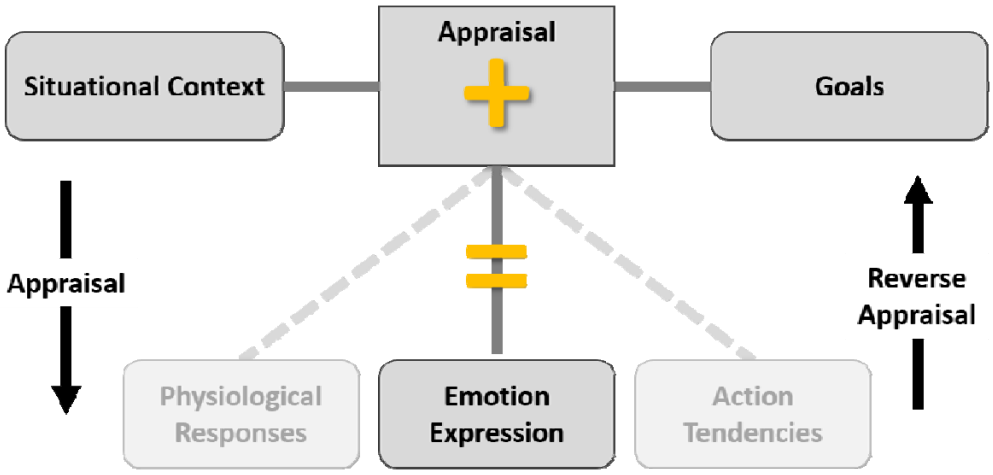

Figure 1. Appraisal and reverse appraisal theories: Appraisal theories argue people appraise, according to their goals, their situational context, which then lead to emotion expressions; complementary, in reverse appraisal, people make inferences about others' goals from information about emotion expressions and situational context.

In line with this view, Hareli and Hess have proposed that people are able to retrieve - or "reverse engineer" - appraisals from others' emotion displays to make relevant inferences (Hareli & Hess, 2010; Hess & Hareli, 2016, 2018). Whereas our research on reverse appraisal has focused primarily on how such inferences shape social decision making, Hareli and Hess have focused on how appraisals influence person perception (e.g., aggressiveness or self-confidence; Hareli & Hess, 2010) and scene perception (e.g., happy scenes are appraised less positively when matched with a disgust expression; Hess & Hareli, 2018). Together, this research emphasizes that the information people retrieve from emotion expressions shapes people's perceptions and behavior.

Reverse appraisal is a theory of mind reasoning mechanism, whereby people use the information in emotion expressions to make inferences about others' mental states. van Kleef, de Dreu, and Manstead (2010), however, rightly point out that this kind of inferential path is not the only way emotion expressions can influence people's decision making. A complementary route is the affective path (Parkinson & Simons, 2009; van Kleef et al., 2010), where others' emotional expressions directly elicit emotion in observers, and this affective experience, then, influences decision making (Damasio, 1994; Lerner et al., 2015; Loewenstein & Lerner, 2003). The prototypical would be one where others' emotion displays elicit emotions in the self that are similar (Hatfield, Cacioppo, & Rapson, 1994; Niedenthal et al., 2010) or complementary (Lanzeta & Englis, 1989). van Kleef (2016) further proposes the Emotions As Social Information (EASI) theory to integrate the various types of social effects of emotion, through the inferential path (i.e., reverse appraisal) or affective path, and describe various relevant moderators, such as the individual's epistemic motivation to process information.

Another mechanism, which is related to but different from reverse appraisal, is social appraisal (Manstead & Fischer, 2001; Parkinson, 2019).[1] This perspective also concerns the impact of others' emotional expressions on one's own emotions. Social appraisal posits that other's emotions impact one's own appraisals of the situation, which, only then, elicit emotion. For instance, Fischer, Rotteveel, Evers, and Manstead (2004) found that participants reacted more angrily to a poor grade given by an instructor when other students reacted with anger but more sadly when others reacted with sadness. Social appraisal acknowledges that other people can be the object of "regular" appraisals (e.g., regarding blameworthiness), but the point is that specific "social" appraisals occur because people care about how others react to situations. Both mechanisms claim people can retrieve appraisals from emotion displays, but social appraisal is distinct from reverse appraisal in, at least, two ways: First, social appraisal places the emphasis on own appraisals, whereas reverse appraisal emphasizes others' appraisals; second, social appraisal focuses on how others' emotional expressions lead to own emotions (via the impact of others' appraisals on own appraisals), whereas reverse appraisal focuses on how others' emotional expressions lead to inferences about others' mental states (i.e., theory of mind).

3. Experimental Evidence and Theoretical Implications

Several lines of evidence lend support to the validity of reverse appraisal theory. Scherer and Grandjean (2008) showed that people can successfully retrieve information about others' appraisals from

[1] Yet another related mechanism is *social referencing*, where others' emotional expressions are used to calibrate one's appraisals against a target object - for instance, when a toddler uses the mother's expression to decide whether to cross a fake visual cliff (Sorce, Emde, Campos, & Klinnert, 1985). However, for details on this mechanism and how it relates to social appraisal see Parkinson's chapter, in this volume.

photos of facial expressions. But, what are the behavioral consequences of these perceptions of appraisals? One line of work, advanced primarily by Hess and Hareli, examines how observed expressions shape person perception. Hareli and Hess (2010) examined hypothetical scenarios where participants were told to imagine themselves in the role of a human resources employee; they were provided with a vignette describing the candidate's narrative of a failure event in their previous job. They, then, looked at how a description of the target's emotional reactions shaped subjective impressions of the target's character. A person who reacted with anger to blame was perceived as more aggressive, and self-confident, but also as less warm and gentle than a person who reacted with sadness (Study 1). A person who reacted with a smile (Study 1) or remained neutral (Study 2) was perceived as self-confident but also as unemotional.

Hess and Hareli (2018) showed scenes from the International Affective Picture System, paired with an external observer's emotion expressions. Participants were told "the expression occurred in reaction to the scene" and were then asked to appraise the scene. The results indicated that appraisals of the scene were accordingly influenced by the expression shown – for instance, an expression of disgust led to more negative appraisals of the scene. Even though this work focuses on scene perception, it still reinforces that people were able to retrieve information about others' appraisals from their emotion expressions, which subsequently influence people's own appraisals about the situation and, presumably, influence the decisions they make.

In our research, we have explored how the inferences arising from reverse appraisal theory shape actual human decision making behavior. One line of studies used "virtual confederates" (de Melo, Carnevale, & Gratch, 2014) that could engage with participants in a number of mixed-motive experimental games (see Figure 2). Virtual confederates are computer-generated characters than can be programed to display specific behaviors and emotions (Gratch, Rickel, André, Cassell, & Petajan, 2002). In a first experiment, participants engaged in an iterated social dilemma (Dawes, 1980) with counterparts that always followed the same behavioral strategy – tit-for-tat – but expressed different patterns of emotion. The expressively cooperative counterparts would show a facial display of joy following mutual cooperation, and regret following participant exploitation (Figure 2a). Complementary, expressively competitive counterparts would show regret following mutual cooperation (since they missed an opportunity to exploit), and joy following participant exploitation (Figure 2b). As expected, participants were more likely to cooperate with the former than the latter (Figure 2c). The experiment emphasizes that, even though the exact same displays are shown, they can lead to opposite effects on cooperation and, thus, it is the information participants retrieve from emotion expressions that is critical, rather than the displays themselves.

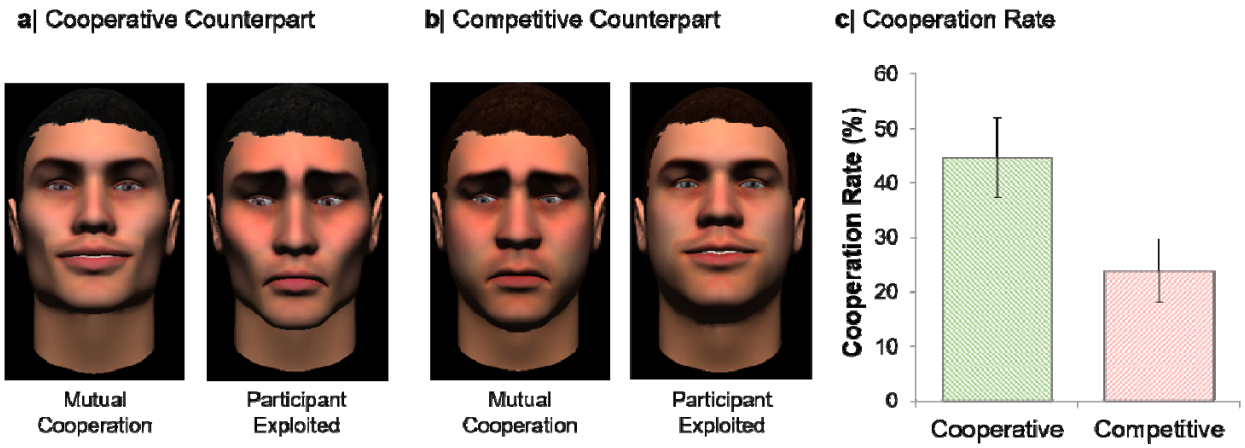

Figure 2. Context matters for the behavioral consequences of emotion expressions: **a**, the cooperative counterpart shows joy following mutual cooperation, and regret when exploiting the participant; **b**, the competitive counterpart shows regret after mutual cooperation, and joy after exploitation; **c**, participants cooperated more with cooperative than competitive counterparts, even though both followed the same strategy (tit-for-tat).

Reverse appraisal, thus, emphasizes that the context the emotion is shown is critical to understanding the behavioral consequences of emotion expressions. In contrast to earlier research suggesting that genuine smiles are an unequivocal sign of cooperation (Brown, Palameta, & Moore, 2003; Mehu, Grammer, & Dunbar, 2007; Scharlemann, Eckel, Kacelnik, & Wilson, 2001) and research suggesting that increased emotional expressivity implies others are more likely to cooperate (Boone & Buck, 2003; Schug, Matsumoto, Horita, Yamagishi, & Bonnet, 2010), reverse appraisal emphasizes that the consequences of emotion expression cannot be determined in a context-free manner. The perceived appraisals apply to the specific context – e.g., the outcome in a social dilemma – and, thus, the same emotion display can have opposite effects – e.g., increased cooperation following a smile after mutual cooperation, or decreased cooperation following a smile after exploitation.

Other researchers have also recognized the importance of context, broadly defined, for emotion perception and the consequences of emotion expressions. Avezier et al. (2008) showed that the perception of emotion in the same facial display is influenced by the body pose and situational context. Lanzetta and Englis (1989) showed that participants experienced empathy, as measured by physiological signals, when the counterpart showed a smile in a setting that was perceived to be cooperative, but counterempathy when the same smile was shown in a setting that was perceived to be competitive. More recently, Hareli and Hess (2016) have attempted to integrate research on the role of context on the meaning of emotion expressions, including the influence of situational context (as described above), the perceiver's stereotypes (e.g., Algoe, Buswell, & DeLamater, 2000), cultural display rules (e.g., Hess, Blaison, & Kafetsios, 2015), and the perceiver's motivation (e.g., Ickes & Simpson, 2004).

In a second experiment, de Melo et al. (2014) had participants engage in social dilemma scenarios where counterparts either expressed emotion in their faces or expressed the corresponding appraisal information via text (e.g., instead of a display of regret, the counterpart would say "I do not like this outcome and I blame myself for it"). Per reverse appraisal, the social effect of the textual expression of appraisals should be equivalent to the corresponding emotion expression. As expected (see Figure 3), the results revealed that the social effects on cooperation were the same if counterparts (1) expressed emotion

in their faces, (2) communicated information, via text, about appraisals, or (3) both. Moreover, these effects were not present when the counterpart showed no emotion and did not convey information about appraisals. The experiment emphasizes that it is possible to achieve the same effect with different modalities and, thus, more than the modality in itself, it is the expression modality's ability to truthfully communicate appraisals that matters for the effect.

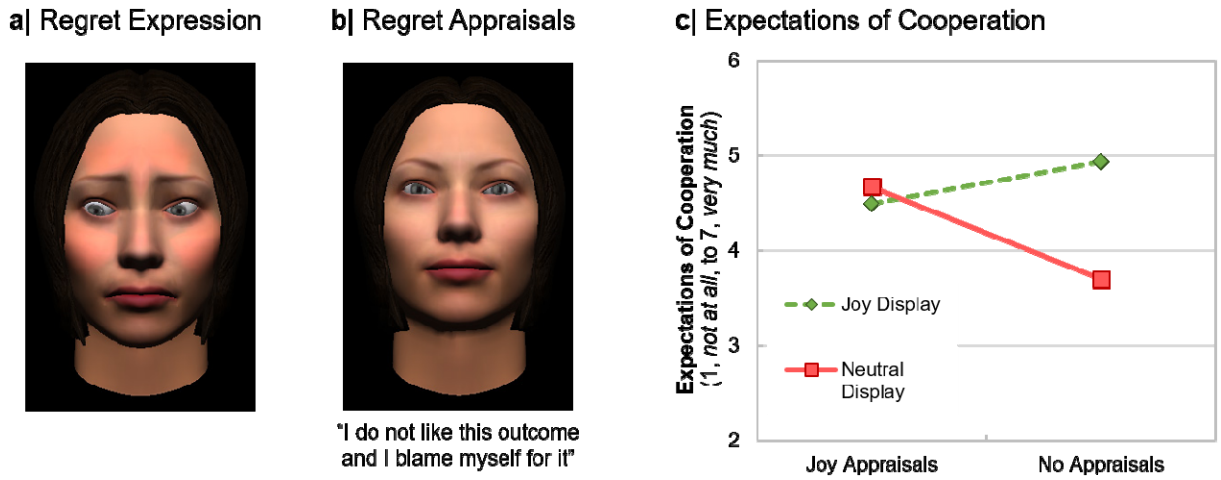

Figure 3. Appraisal expressions have the same effect as corresponding emotion displays: **a**, regret display; **b**, textual expression of the appraisals for regret; **c**, following mutual cooperation, participants expected similar levels of cooperation following displays of regret or textual expression of appraisals.

According to reverse appraisal, therefore, the emotional expression is redundant with the corresponding information about appraisals. Thus, as long as appraisals can be communicated effectively, different modalities should have the same behavioral effect in social decision making. Complementary, an expression modality is only limited as far as its ability to convey appraisals to others. In recent work, Hareli, Elkabetz, and Hess (2019) also showed that textual communication of awe appraisals alongside a neutral facial expression led to a similar effect on scene perception as a facial expression of awe. Nevertheless, much research is still needed to systematically study different communication modalities in terms of their ability to successfully encode these appraisals and, on the receiver end, decode the appraisals from the communicated signal. One exception is the work of Scherer, Mortillaro, Rotondi, Sergi, and Trzandel (2018) focusing on how appraisals are communicated through facial expressions. In a series of experiments, they showed that participants were able to retrieve specific appraisals from the components – or facial units – of the facial expression. This work suggests that (at least some) appraisals can be encoded at a more granular level than a full-blown prototypical emotion expression and, correspondingly, supports the contention that decoding of appraisals may also happen at this level of granularity.

An important factor to consider is whether the communication of appraisals is truthful. This is important because inauthentic expressions can backfire and lead to reduced trust and increased demand in mixed motive situations (Côté, Hideg, & van Kleef, 2013). It is certainly possible to provide a more detailed verbal account of one's appraisals to others, than through nonverbal means. However, economists have been quick to point out that "cheap talk" has no direct influence on the payoffs, even though in practice it can influence others' decision making (Farrell & Rabin, 1996). Complementary,

Frank (1988, 2004) argues that nonverbal cues – such as facial displays – tend to be more reliable as they are less susceptible to strategic manipulation. Research on genuine (or so-called Duchenne) smiles also indicates that some people aren't able to fake genuine facial expressions, though there are many that can (Gunnery, Hall, & Ruben, 2013). Further research is necessary to understand these kinds of tradeoffs between an expression modality's appraisal communication bandwidth and its susceptibility to strategic manipulation.

In three additional experiments, de Melo et al. (2014) presented evidence for the causal model implied by reverse appraisal: (contextualized) emotion expressions cause inferences about the expresser's appraisals, which cause inferences about the expresser's intentions. In Experiment 3, participants engaged in the iterated social dilemma with cooperative and competitive counterparts, similarly to Experiment 1; however, this time, participants were asked to report, at the end of each round, about (perceptions of) the counterparts' goal conduciveness and blameworthiness appraisals. A multiple mediation analysis of this data (Preacher & Hayes, 2008) indicated that perceptions of appraisals were mediating the effects of expressions of joy and guilt on the decision to cooperate, thus, supporting the proposed causal model. Experiment 4 extended this study by having participants engage in scenarios where each possible outcome of the social dilemma occurs, crossed with expressions of joy, sadness, regret, or anger. After receiving information about the outcome (or context) and the counterpart's expression, participants were asked about perceptions of the counterpart's appraisals and whether the counterpart was likely to cooperate in the future. The results, first, showed that people were able to retrieve appraisals from emotion displays in a way that was compatible with appraisal theories of emotion (e.g., people would perceive a counterpart that expressed anger to find the outcome to be goal obstructive and blame the participant for it). Secondly, similarly to the previous experiment, perceptions of appraisals mediated the effects of emotion expressions on expectations of cooperation, thus providing further support for the proposed causal model. However, Spencer, Zanna, and Fong (2005) argue that multiple mediation analyses are no substitute for showing experimental evidence for each link in a causal model. Thus, Experiment 5 focused on the second link in the causal model – i.e., perceptions of appraisal cause inferences about counterpart's intentions – and had participants engage in scenarios where counterparts expressed the appraisals corresponding to joy, sadness, regret, or anger, for each possible outcome of the social dilemma. The results revealed very similar effects on expectations of cooperation, as those with the corresponding expressions of emotion in the previous experiment. Therefore, since Experiment 4 provided evidence for the first causal link, and Experiment 5 for the second, according to Spencer et al. (2005) there is evidence for the reverse appraisal causal model.

This causal model, hence, implies that the effect of emotion expressions on behavior must occur through perceptions of appraisals. This inferential path contrasts to the affective path (van Kleef et al., 2010), where other emotions influence the receiver's emotions and decision making more directly, such as by a process of contagion or mimicry (Hatfield, Cacioppo, & Rapson, 1994; Niedenthal et al., 2010). Reverse appraisal, thus, implies a process based on theory of mind (ToM), i.e., inferences about others' mental states. This distinction is akin to the separation between emotional and cognitive empathy, which tend to involve different brain regions (Shamay-Tsoory, 2011). The former pertains to the ability to (empathetically) experience the other's emotion, perhaps through a process of simulation (Gallese, 2007). The latter involves ToM reasoning with the objective of understanding the others' beliefs and emotions. The ToM neural substrate is known to include the medial prefrontal cortex (mPFC), the superior temporal sulcus, the temporoparietal junction, and the temporal poles (Amodio & Frith, 2006). Research has further identified different subcomponents of the ToM network, with one important differentiation being

between cognitive ToM – focused on inferences about others' beliefs, typically measured using false belief tasks, and involving the dorsomedial PFC – and affective ToM – focused on inferences about others' emotions, typically measured using Reading the Mind in the Eyes or faux pas tasks, and involving the ventromedial PFC (Baron-Cohen, Wheelwright, Hill, Raste, & Plumb, 2001; Shamay-Tsoory & Aharon-Peretz, 2007; Shamay-Tsoory, Aharon-Peretz & Perry, 2009). Reverse appraisal presumably involves both of these ToM components, as the individual must interpret emotion displays (affective ToM) for the purpose of inferring the other's intentions (cognitive ToM). Recent research has, in fact, begun finding important relationships between the affective and cognitive ToM regions when people engage in social cognition tasks (Poletti, Enrici, & Adenzato, 2012). Future studies are needed to clarify brain regions involved in decision making tasks with emotionally expressive others to identify the neural substrate for the reverse appraisal process.

Finally, in de Melo et al.'s studies, the use of virtual confederates allowed for a high degree of experimental control, but they could be criticized on the grounds that the findings would not translate to real face-to-face interactions. Thus, another line of our research sought evidence of reverse appraisal in more natural interactions. Stratou, Hoegen, Lucas, and Gratch (2015) allowed pairs of participants to play the iterated prisoner's dilemma game over a videoconferencing interface that allowed them to see each other's facial expressions but prevented them from communicating verbally (Figure 4). Using this setup, we were able to capture facial expressions and relate these expressions to in-game events using automated expression recognition software (e.g., one player exploiting the other) and subsequent decisions to cooperate or compete. First, the results suggest that expressions do communicate reliable information about the signaler's future intentions (e.g., a player's smiles predicted cooperative choices in the game). Second, the results suggest that players can recover these intentions by observing their opponent's expression (e.g., players were more likely to defect against an opponent that signals they will cooperate on the next round). Together, these findings suggest that reverse appraisal operates in face-to-face interactions.

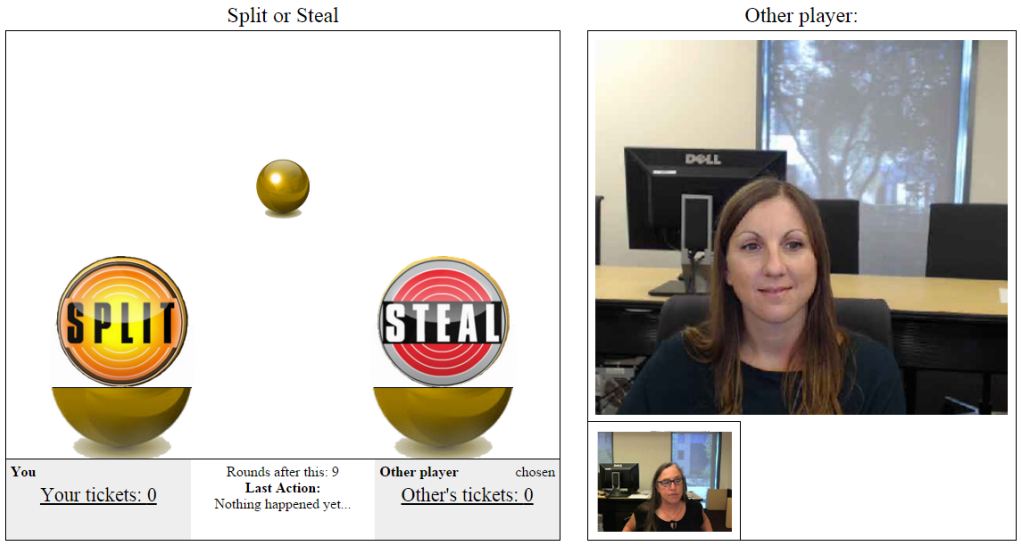

Figure 4. Software used to study the relationship between (real) emotion expressions and decisions in the split-steal social dilemma (Giota et al., 2015).

4. Practical Implications

4.1 Computational Modeling

A basic application of a computational model of reverse appraisal is to test psychological theory by exposing hidden assumptions and identifying gaps (Scherer, 2010). A complementary application is to drive the behavior of computational agents that need to understand the emotions of their human counterparts (e.g., social robots; Breazeal, 2003). To accomplish this, a theory-driven, data-driven, or hybrid approach could be followed. Theory-driven models rely on various computational techniques to simulate the psychological mechanism specified by theory. Several such models have been proposed for emotion synthesis (Picard, 1997), most of which based on appraisal theories (Petta, Marsella, & Gratch, 2010). However, a purely theory-driven model for reverse appraisal is still missing. Data-driven approaches, instead, rely on data (about emotion expressions and their meaning) and use machine learning or model fitting techniques to automatically identify the patterns in the data. These models can then generalize to new unseen situations. An example of this approach is the work by Wu, Baker, Tenenbaum, and Schulz (2018), which developed a Bayesian model of the beliefs, desires, and intentions people retrieve from emotion expressions. The prior probabilities in the model were calculated from data collected with one set of participants, and the predictions were tested with another set of participants in emotion-eliciting scenarios – e.g., a person learns about a plane crash on a route flown by her coworker, she learns that the coworker died and she is surprisingly happy about this; did she initially believed that the coworker was on the plane and did she desire harm upon her coworker? Though demonstrating the plausibility of using a Bayesian approach to model human behavior, the proposed model is, however, mostly atheoretical, in the sense that it does not follow any (reverse) appraisal theory to specify which kinds of beliefs, desires, and intentions can be retrieved from emotion expressions. In contrast, de Melo, Carnevale, Read, Antos, & Gratch (2012) proposed a Bayesian model that explicitly represents appraisals and can be used to make predictions about others' intentions to cooperate from their contextualized emotion expressions. This hybrid approach relied on data from previous experimental studies (Experiments 4 and 5 in de Melo et al., 2014) to specify the priors in the model. They demonstrated that this model outperformed variations that did not account for appraisal perceptions, and an even simpler model that ignored emotion displays.

4.2 Emotion Recognition and Understanding

The last two decades witnessed the emergence of the field of automatic *affect recognition* due to consistent improvements in computational power, advances in computational techniques (e.g., machine learning algorithms), the emergence of ubiquitous technology (e.g., smartphones and sensors), and availability of more data (for recent reviews of the field see: D'Mello & Kory, 2015; Sariyanidi, Gunes, & Cavallaro, 2015; Zeng, Pantic, Roisman, & Huang, 2009). Automatic affect recognition is of considerable practical interest across various fields: in the medical domain, it could help monitor patient's (emotional) state of mind during treatment administration and in the post-treatment stage; in marketing research, it could be used to perceive, in real-time, the target audience's (emotional) reaction to the product and to ad campaigns; in transportation, it could continuously gauge the driver's emotions (e.g., for fatigue or anger), and make appropriate recommendations that increase safety; and so on. Much progress has been made identifying and extracting relevant facial, acoustic, and physiological features

(Calvo & D'Mello, 2010), developing new multimodal fusion techniques (Baltrušaitis, Ahuja, & Morency, 2018), and tackling affect detection in non-posed expressions (Krumhuber, Kappas, & Manstead, 2013). However, considerably less effort has been spent on the harder, but arguably more interesting, challenge of *understanding* affect. In this case, the goal is not only to acquire a probability distribution over the possible emotions being expressed, but to infer what those expressions mean in terms of the expresser's mental states and possible behavioral consequences. We argue that reverse appraisal can serve as the mechanism tying low-level perceptions to higher-level inferences about the expresser's beliefs, desires, and intentions (for a similar argument using a Bayesian approach see: Alfonso, Pynadath, Lhommet, & Marsella, 2015). Reverse appraisal can also provide insight on relevant features. For instance, reverse appraisal emphasizes the importance of situational context for the interpretation of emotion. Accordingly, and in line with the current trend to fuse multiple modalities to improve recognition (Baltrušaitis et al., 2018), Rens, Stratou, and Gratch (2017), using information about both prior actions and emotion expressions led to higher counterpart action prediction than just looking at the expressions.

4.3 Autonomous Socially Intelligent Machines

A relatively recent trend is the advent of autonomous technology. These intelligent systems – such as self-driving cars (Dally, Medasani, Behringer, & Trivedl, 2017; de Melo, Marsella, & Gratch, 2019; Waldrop, 2015), drones (Floreano & Wood, 2015), personal assistants (López, Quesada, & Guerrero, 2017), and social anthropomorphic robots (Breazeal, 2003; Kachouie, Sedighadeli, Khosla, & Chu, 2014; Stone & Lavine, 2014) – are being designed to act autonomously on people's behalf. In the near future, it is expected that people will increasingly often interact and cooperate with autonomous machines in their personal, professional, and social lives (Al-Fuqaha, Guizani, Mohammadi, Aledhari, & Ayyash, 2015; Kott, Swami, & West, 2016). Consequently, these machines will have to be designed to not only understand the functional aspects of the tasks, but the broader social context. Our research illustrates that machines can promote cooperation with humans through expression of simulated emotion, and perception and understanding of human emotion. On the one hand, appraisal theory serves as a concrete mathematical framework for translating the beliefs, desires and intentions of machines into intuitive human-readable signals. On the other hand, reverse appraisal provides a mechanism for decoding human emotion signals in the specific situational context. More broadly, these findings highlight that emotions play a functional role in social interactions – i.e., communicating mental state – and that autonomous systems can draw on this function to enhance their own efficacy.

5. Conclusion

The last fifteen years have seen increasing interest, across scientific disciplines, on the interpersonal effects of emotion. This cross-disciplinary research has shed light on mechanisms driving these effects, and their implications to the decisions people make in mixed-motives social settings. Reverse appraisal (de Melo et al., 2014) is one of those mechanisms, which is based in theory of mind reasoning, whereby people retrieve information about other appraisals, from emotion expressions, and use this information to make inference about others' mental states. Reverse appraisal emphasizes the importance of context to understanding the behavioral consequences of emotion expressions in others (i.e., the same display can lead to different consequences in different contexts) and clarifies that the same

effects can be achieved through different expression modalities, provided that these channels can truthfully communicate appraisals to others.

However, reverse appraisal is only one mechanism for the interpersonal effect of emotion expression, and it is important to further understand how it relates to other mechanisms that account for effects through experienced emotion (Hatfield et al., 1994; Lanzetta & Englis, 1989; Parkinson & Simons, 2009) and social appraisals (Manstead & Fischer, 2001; Parkinson, 2019). A cross-disciplinary approach is necessary to tackle this challenge. More experiments, in the lab (e.g., de Melo et al., 2014) and in the field (e.g., Parkinson & Simons, 2009), focusing on establishing the causal models for the identified effects are needed, preferably using complementary techniques (Preacher & Hayes, 2008; Spencer et al., 2005). Recent work on the neural substrates of decision making emphasizes the influence of experienced emotion and theory of mind reasoning on social decision making (Amodio & Frith, 2006; Rilling & Sanfey, 2011; Shamay-Tsoory, 2011). Understanding how the proposed mechanisms for the interpersonal effects of emotion reflect in the brain would also be useful in clarifying where there are overlaps and what is distinct about each process. Finally, computational modeling, by either forcing the programmer to explicitly write the rules (in a theory-driven approach) or automatically learning patterns using large amounts of data (in a data-driven or hybrid approach), can shed light on limitations and gaps in current theorizing (de Melo et al., 2012; Petta et al., 2010; Wu et al., 2018).

Despite the progress seen in the last years, several important topics still need further attention. People can use emotion expressions strategically to influence others' decision making (Andrade & Ho, 2009; Fitness, 2000; Sutton, 1991; Tamir, Mitchell, & Gross, 2008), but this can backfire and lead to negative consequences (Côté et al., 2013). Further research, thus, is needed to understand whether and how people perceive misleading communication of emotions and appraisals, across different expression modalities, and what are the corresponding behavioral consequences. Cultural rules can also shape the emotions people express (Ekman & Friesen, 1971) and, accordingly, it is necessary to further understand the implications of expressing inappropriate emotions (van Kleef, Homan, & Cheshin, 2012). Finally, complementing recent work identifying important moderators for the social effects of emotion (van Kleef, 2016; van Kleef et al., 2010), it is necessary to understand individual (Lewis, 2001) and cultural (Mesquita & Ellsworth, 2001) factors shaping the way we appraise events and situations and, thus, possibly the way we perceive appraisals in others.

Recent times have seen increasing interest in socially intelligent machines - such as digital home assistants (López et al., 2017), self-driving cars (de Melo et al., 2019), social robots (Breazeal, 2003), and automatic negotiators (Lin & Kraus, 2010). These autonomous machines are meant to interact with humans in social settings and make decisions with or on people's behalf. Research on the social functions of emotion is, therefore, invaluable to the design of machines that will succeed in social interaction with humans. This research could inform the machine's ability to recognize human emotional displays (e.g., Calvo & D'Mello, 2010; Zeng et al., 2009) and even endow them with anthropomorphic characteristics such as the ability to signal emotion (DeSteno et al., 2012; Gratch et al., 2002). Researchers should, thus, strive to study how theories of the interpersonal effects of emotion expression apply in these emerging domains.

References

Al-Fuqaha, A., Guizani, M., Mohammadi, M., Aledhari, M., & Ayyash, M. (2015). Internet of things: A survey on enabling technologies, protocols, and applications. *IEEE Communication Surveys & Tutorials*, *17*, 2347-2376. https://doi.org/10.1109/COMST.2015.2444095

Alfonso, B., Pynadath, D., Lhommet, M., & Marsella, S. (2013). Emotional perception for updating agents' beliefs. In *Proceedings of the International Conference on Affective Computing and Intelligent Interaction (ACII)*. https://dx.doi.org/10.1016%2Fj.tics.2012.07.009

Algoe, S. B., Buswell, B. N., & DeLamater, J. D. (2000). Gender and job status as contextual cues for the interpretation of facial expression of emotion. *Sex Roles*, *42*, 183-208.

Amodio, D., & Frith, C. (2006). Meeting of minds: The medial frontal cortex and social cognition. *Nature Reviews Neuroscience*, *7*, 268-277. https://doi.org/10.1038/nrn1884

Andrade, E., & Ho, T.-H. (2009). Gaming emotions in social interactions. *Journal of Consumer Research*, *36*, 539-551. https://psycnet.apa.org/doi/10.1086/599221

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

Baltrušaitis, T., Ahuja, C., & Morency, L.P. (2018). Multimodal machine learning: A survey and taxonomy. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. https://doi.org/10.1109/TPAMI.2018.2798607

Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y., Plumb, I., 2001. The "Reading the Mind in the Eyes" Test revised version: a study with normal adults, and adults with Asperger syndrome or high-functioning autism. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *42*, 241-251.

Boone, R., & Buck, R. (2003). Emotional expressivity and trustworthiness: The role of nonverbal behavior in the evolution of cooperation. *Journal of Nonverbal Behavior*, *27*, 163-182.

Breazeal, C. (2003). Toward sociable robots. *Robotics and Autonomous Systems*, *42*, 167-175. https://doi.org/10.1016/S0921-8890(02)00373-1

Brown, W., Palameta, B., & Moore, C. (2003). Are there nonverbal cues to commitment? An exploratory study using the zero-acquaintance video presentation paradigm. *Evolutionary Psychology*, *1*, 42-69. https://doi.org/10.1177%2F147470490300100104

Calvo, R., & D'Mello, S. (2010). Affect detection: An interdisciplinary review of models, methods, and their applications. *IEEE Transactions on Affective Computing*, *1*, 16-37. https://doi.org/10.1109/T-AFFC.2010.1

Camerer, C. (2003). Psychology and economics. Strategizing in the brain. *Science*, *300*, 1673-1675. https://doi.org/10.1126/science.1086215

Côté, S., Hideg, I., & van Kleef, G. (2013). The consequences of faking anger in negotiations. *Journal of Experimental Social Psychology*, *49*, 453-463. https://doi.org/10.1016/j.jesp.2012.12.015

Dally, M., Medasani, S., Behringer, R., & Trivedl, M. (2017). Self-driving cars. *IEEE Computer*, *50*, 18-23. https://doi.org/10.1109/MC.2017.4451204

Damasio, A. (1994). *Descartes' error: Emotion, reason and the human brain*. New York, NY: Putnam.

Dawes, R. (1980). Social dilemmas. *Annual Review of Psychology*, *31*, 169-93. https://doi.org/10.1146/annurev.ps.31.020180.001125

de Melo, C., Carnevale, P., & Gratch, J. (2014). Using virtual confederates to research intergroup bias and conflict. In: *Academy of Management Proceedings*, Vol. 2014, No. 1.

de Melo, C., Carnevale, P., Read, S., Antos, D., & Gratch, J. (2012). Bayesian model of the social effects of emotion in decision-making in multiagent systems. In: *Proceedings of Autonomous Agents and Multiagent Systems (AAMAS 12)*.

de Melo, C., Carnevale, P., Read, S., & Gratch, J. (2014). Reading people's minds from emotion expressions in interdependent decision making. *Journal of Personality and Social Psychology*, *106*, 73-88. https://doi.org/10.1037/a0034251

de Melo, C., Marsella, S., & Gratch, J. (2019). Human cooperation when acting through autonomous machines. *Proceedings of the National Academy of Sciences U.S.A.,* 116, 3482-3487. https://doi.org/10.1073/pnas.1817656116

DeSteno, D., Breazeal, C., Frank, R., Pizarro, D., Baumann, J., Dickens, L., et al. (2012). Detecting the trustworthiness of novel partners in economic exchange. *Psychological Science*, *23*, 1549-1556. https://doi.org/10.1177%2F0956797612448793

D'Mello, S., & K., Jacqueline. (2015). A review and meta-analysis of multimodal affect detection systems. *ACM Computing Surveys*, *47*. https://doi.org/10.1145/2682899

Ekman, P. (1993). Facial expression and emotion. *American Psychologist*, *48*, 384-392. https://psycnet.apa.org/doi/10.1037/0003-066X.48.4.384

Ekman, P., & Friesen, W. V. (1971). Constants across cultures in the face and emotion. *Journal of Personality and Social Psychology*, *17*, 124-129. https://psycnet.apa.org/doi/10.1037/h0030377

Ellsworth, P., & Scherer, K. (2003). Appraisal processes in emotion. In R. Davidson, K. Scherer, H. Goldsmith (Eds.), *Handbook of Affective Sciences* (pp. 572-595). New York, NY: Oxford University Press.

Farrell, J., & Rabin, M. (1996). Cheap talk. *Journal of Economics Perspectives*, *10*, 103-118.

Floreano, D., & Wood, R. (2015). Science, technology and the future of small autonomous drones. *Nature*, *521*, 460-466.

Fischer, A., Rotteveel, M., Evers, C., & Manstead, A. (2004). Emotional assimilation: How we are influenced by others' emotions. *Cahiers de Psychologie Cognitive*, *22*, 223-245.

Fitness, J. (2000). Anger in the workplace: An emotion script approach to anger episodes between workers and their superiors, co-workers and subordinates. *Journal of Organizational Behavior*, *21*, 147-162. https://doi.org/10.1002/(SICI)1099-1379(200003)21:2%3C147::AID-JOB35%3E3.0.CO;2-T

Frank, R. (1988). *Passions within reason: The strategic role of the emotions*. New York: Norton.

Frank, R. (2004). Introducing moral emotions into models of rational choice. In A. Manstead, N. Frijda & A. Fischer (Eds.), *Feelings and emotions* (pp. 422-440). New York, NY: Cambridge University Press. https://psycnet.apa.org/doi/10.1017/CBO9780511806582.024

Frijda, N. (1986). *The emotions*. Cambridge, UK: Cambridge University Press.

Frijda, N., & Mesquita, B. (1994). The social roles and functions of emotions. In S. Kitayama & H. Markus (Eds.), *Emotion and culture: Empirical studies of mutual influence* (pp. 51-87). Washington, DC: American Psychological Association. https://psycnet.apa.org/doi/10.1037/10152-002

Frijda, N., Kuipers, P., & ter Schure, E. (1989). Relations among emotion, appraisal, and emotional action readiness. *Journal of Personality and Social Psychology*, *57*, 212-228. https://psycnet.apa.org/doi/10.1037/0022-3514.57.2.212

Gallese V. (2007). Before and below 'theory of mind': embodied simulation and the neural correlates of social cognition. *Philosophical Transactions Royal Social London B Biological Science*, *362*, 659-669. https://dx.doi.org/10.1098%2Frstb.2006.2002

Giota, S., Hoegen, R., Lucas, G., & Gratch, J. (2015). Emotional signaling in a social dilemma: An automatic analysis. In: *Proceedings of the International Conference on Affective Computing and Intelligent Interaction (ACII)*.

Gratch, J., Rickel, J., Andre, E., Badler, N., Cassell, J., & Petajan, E. (2002). Creating interactive virtual humans: Some assembly required. *IEEE Intelligent Systems*, 17, 54-63. https://doi.org/10.1109/MIS.2002.1024753

Gunnery, S., Hall, J., & Ruben, M. (2013). The deliberate Duchenne smile: Individual differences in expressive control. *Journal of Nonverbal Behavior*, *37*, 29-41.

Haidt, J. (2003). Autonomic specificity and emotion. In R. J. Davidson, K. R. Scherer & H. H. Goldsmith (Eds.), *Handbook of Affective Sciences*, 852-870. New York, NY: Oxford University Press.

Hareli, S. (2014). Making sense of the social world and influencing it by using a naïve attribution theory of emotions. *Emotion Review*, *6*, 336–343. https://doi.org/10.1177/1754073914534501

Hareli, S., & Hess, U. (2010). What emotional reactions can tell us about the nature of others: An appraisal perspective on person perception. *Cognition and Emotion*, *24*, 128-140. https://doi.org/10.1080/02699930802613828

Hareli, S., Elkabetz, S., & Hess, U. (2019). Drawing inferences from emotion expressions: The role of situative informativeness and context. *Emotion*. Advance online publication. http://dx.doi.org/10.1037/emo0000368

Hatfield, E., Cacioppo, J., & Rapson, R. (1994). *Emotional contagion*. New York, NY: Cambridge University Press. https://doi.org/10.1080/00029157.1997.10403399

Hess, U., Blaison, C., & Kafetsios, K. (2016). Judging facial emotion expressions in context: The influence of culture and self-construal orientation. *Journal of Nonverbal Behavior*, *40*, 50-64.

Hess, U., & Hareli, S. (2016). The impact of context on the perception of emotions. In C. Abell & J. Smith (Eds.), *The expression of emotion: Philosophical, psychological, and legal perspectives* (pp. 199–218). Cambridge: Cambridge University Press. https://doi.org/10.2466/pms.1986.62.1.79

Hess, U., & Hareli, S. (2018). On the malleability of the meaning of contexts: the influence of another person's emotion expressions on situation perception. *Cognition and Emotion*, *32*, 185-191. https://doi.org/10.1080/02699931.2016.1269725

Ickes, W., & Simpson, J. A. (2004). Motivational Aspects of Empathic Accuracy. In M. B. Brewer & M. Hewstone (Eds.), *Emotion and motivation. Perspectives on social psychology* (pp. 225-246). Malden, MA: Blackwell Publishing.

Kachouie, R., Sedighadeli, S., Khosla, R., & Chu, M.-T. (2014). Socially assistive robots in elderly care: A mixed-method systematic literature review. *International Journal of Human-Computer Interaction*, *30*, 369-393. https://doi.org/10.1080/10447318.2013.873278

Keltner, D., & Ekman, P. (2000). Facial expression of emotion. In M. Lewis & J. Haviland-Jones (Eds.), *Handbook of Emotion* (pp. 236-249). New York, NY: Guilford Press.

Keltner, D., & Kring, A. M. (1998). Emotion, social function, and psychopathology. *Review of General Psychology*, *2*, 320-342. https://psycnet.apa.org/doi/10.1037/1089-2680.2.3.320

Keltner, D., & Haidt, J. (1999). Social functions of emotions at four levels of analysis. *Cognition and Emotion*, *13*, 505-521. https://doi.org/10.1080/026999399379168

Kott, A., Swami, A., & West, B. (2016). The internet of battle things. *IEEE Computer*, *49*, 70-75. https://doi.org/10.1109/MC.2016.355

Kreibig, S. (2010). Autonomic nervous system activity in emotion: A review. *Biological Psychology*, *84*, 394-421. https://doi.org/10.1016/j.biopsycho.2010.03.010

Krumhuber, E., Manstead, A., & Kappas, A. (2007). Facial dynamics as indicators of trustworthiness and cooperative behavior. *Emotion*, *7*, 730-735. https://doi.org/10.1037/1528-3542.7.4.730

Krumhuber, E., Kappas, A., & Manstead, A. (2013). Effects of dynamic aspects of facial expressions: A review. *Emotion Review*, *5*, 41-46. https://doi.org/10.1177%2F1754073912451349

Lanzetta, J., & Englis, B. (1989). Expectations of cooperation and competition and their effects on observer's vicarious emotional responses. *Journal of Personality and Social Psychology*, *36*, 543-554. https://psycnet.apa.org/doi/10.1037/0022-3514.56.4.543

Lerner, J., Li, Y., Valdesolo, P., & Kassam, K. (2015). Emotion and decision making. *Annual Review of Psychology*, *66*, 799-823. https://doi.org/10.1146/annurev-psych-010213-115043

Levenson, R. (2003). Autonomic specificity and emotion. In R. J. Davidson, K. R. Scherer & H. H. Goldsmith (Eds.), *Handbook of Affective Sciences*, 212-224. New York, NY: Oxford University Press.

Leventhal, H., & Scherer, K. (1987). The relationship of emotion to cognition: A functional approach to a semantic controversy. *Cognition and Emotion*, *1*, 3-28. https://doi.org/10.1080/02699938708408361

Lewis, M. (2001). Personal pathways in the development of appraisal: A complex systems/stage theory perspective. In K. Scherer, A. Schorr & T. Johnstone (Eds.), *Appraisal processes in emotion: Theory, methods, research* (pp. 205-220). New York, NY: Oxford University Press.

Lin, R., & Kraus, S. (2010). Can automated agents proficiently negotiate with humans? *Communications of the ACM*, *53*, 78-88. https://doi.org/10.1145/1629175.1629199

Loewenstein, G., & Lerner, J. (2003). The role of affect in decision making. In R. Davidson, K. Scherer & H. Goldsmith (Eds.), *Handbook of Affective Sciences* (pp. 619-642). New York, NY: Oxford University Press.

López, G., Quesada, L., & Guerrero, L. (2017). Alexa vs. Siri vs. Cortana vs. Google Assistant: A comparison of speech-based natural user interfaces. In *Proceedings of International Conference on Applied Human Factors and Ergonomics (AHFE 2017)*.

Manstead, A., & Fischer, A. (2001). Social appraisal: The social world as object of and influence on appraisal processes. In K. Scherer, A. Schorr, & T. Johnstone (Eds.), *Appraisal processes in emotion: Theory, methods, research* (pp. 221–232). Oxford, England: Oxford University Press.

Mehu, M., Grammer, K., & Dunbar, R. (2007). Smiles when sharing. *Evolution and Human Behavior*, *28*, 415-422. https://doi.org/10.1016/j.evolhumbehav.2007.05.010

Mesquita, B., & Ellsworth, P. (2001). The role of culture in appraisal. In K. Scherer, A. Schorr & T. Johnstone (Eds.), *Appraisal processes in emotion: Theory, methods, research* (pp. 233-248). New York, NY: Oxford University Press.

Morris, M., & Keltner, D. (2000). How emotions work: An analysis of the social functions of emotional expression in negotiations. *Research in Organizational Behavior*, *22*, 1-50. http://dx.doi.org/10.1016/S0191-3085(00)22002-9

Niedenthal, P., Mermillod, M., Maringer, M., & Hess, U. (2010). The Simulation of Smiles (SIMS) model: Embodied simulation and the meaning of facial expression. *Behavioral and Brain Sciences*, *33*, 417-480. https://doi.org/10.1017/S0140525X10000865

Ortony, A., Clore, G., & Collins, A. (1988). *The cognitive structure of emotions*. New York, NY: Cambridge University Press.

Parkinson, B. (2019). Calibrating emotional orientations: Social appraisal and other kinds of relation alignment. In D. Dukes & F. Clément (Eds.), *Foundations of Affective Social Learning: Conceptualising the transmission of social value*. Cambridge: Cambridge University Press.

Parkinson, B., & Simons, G. (2009). Affecting others: Social appraisal and emotion contagion in everyday decision making. *Personality and Social Psychology Bulletin, 35*, 1071-1084. https://doi.org/10.1177%2F0146167209336611

Petta, P., Marsella, S., & Gratch, J. (2010). Computational models of emotion. In K. Scherer, T. Bänziger, & E. Roesch (Eds.), *A Blueprint for Affective Computing: A Sourcebook and Manual* (pp. 21-40), Oxford University Press.

Picard, R. (1997). *Affective computing*. Cambridge, MA: The MIT Press.

Poletti, M., Enrici, I., & Adenzato, M. (2012). Cognitive and affective Theory of Mind in neurodegenerative diseases: Neuropsychological, neuroanatomical and neurochemical levels. *Neuroscience and Biobehavioral Reviews*, *26*, 2147-2164. https://doi.org/10.1016/j.neubiorev.2012.07.004

Preacher, K., & Hayes, A. (2008). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behavior Research Methods*, *40*, 879-891.

Reisenzein, R. (2009). Emotions as metarepresentational states of mind: Naturalizing the belief–desire theory of emotion. *Cognitive Systems Research*, *10*, 6-20. https://doi.org/10.1016/j.cogsys.2008.03.001

Rens, H., Stratou, G., & Gratch, J. (2017). Incorporating emotion perception into opponent modeling for social dilemmas. In *Proceedings of the International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*.

Rilling, J., & Sanfey, A. (2011). The neuroscience of social decision-making. *Annual Review of Psychology*, *62*, 23-48. https://doi.org/10.1146/annurev.psych.121208.131647

Roseman, I. (2001). A model of appraisal in the emotion system: Integrating theory, research, and applications. In K. Scherer, A. Schorr & T. Johnstone (Eds.), *Appraisal processes in emotion: Theory, methods, research* (pp. 68-91). New York, NY: Oxford University Press.

Sariyanidi, E., Gunes, H., & Cavallaro, A. (2015). Automatic analysis of facial affect: A survey of registration, representation, and recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *37*, 1113-1133. https://doi.org/10.1109/TPAMI.2014.2366127

Scharlemann, J., Eckel, C., Kacelnik, A., & Wilson, R. (2001). The value of a smile: Game theory with a human face. *Journal of Economic Psychology*, *22*, 617-640. https://doi.org/10.1016/S0167-4870(01)00059-9

Scherer, K. (2001). Appraisal considered as a process of multi-level sequential checking. In K. Scherer, A. Schorr & T. Johnstone (Eds.), *Appraisal processes in emotion: Theory, methods, research* (pp. 92-120). New York, NY: Oxford University Press.

Scherer, K. (2010). The component process model: Architecture for a comprehensive computational model of emergent emotion. In K. Scherer, T. Bänziger, & E. Roesch (Eds.), *A Blueprint for Affective Computing: A Sourcebook and Manual* (pp. 47-70), Oxford University Press.

Scherer, K., & Grandjean, D. (2008). Facial expressions allow inference of both emotions and their components. *Cognition and Emotion*, *22*, 789-801. https://doi.org/10.1080/02699930701516791

Scherer, K., & Moors, A. (2019). The emotion process: Event appraisal and component differentiation. *Annual Review of Psychology*, *70*. https://doi.org/10.1146/annurev-psych-122216-011854

Scherer, K., Johnstone, T., & Klasmeyer, G. (2003). Vocal expression of emotion. In R. J. Davidson, K. R. Scherer & H. H. Goldsmith (Eds.), *Handbook of Affective Sciences*, 433-456. New York, NY: Oxford University Press.

Scherer, K., Mortillaro, M., Rotondi, I., Sergi, I., & Trznadel, S. (2018). Appraisal-driven facial actions as building blocks for emotion inference. *Journal of Personality and Social Psychology*, *114*, 358-379. https://doi.org/10.1037/pspa0000107

Schug, J., Matsumoto, D., Horita, Y., Yamagishi, T., & Bonnet, K. (2010). Emotional expressivity as a signal of cooperation. *Evolution and Human Behavior*, *31*, 87-94. https://doi.org/10.1016/j.evolhumbehav.2009.09.006

Shamay-Tsoory, S. (2011). The neural bases for empathy. *The Neuroscientist*, *17*, 18-24. https://doi.org/10.1177/1073858410379268

Shamay-Tsoory, S., & Aharon-Peretz, J. (2007). Dissociable prefrontal networks for cognitive and affective theory of mind: a lesion study. *Neuropsychologia*, *45*, 3054-3067. https://doi.org/10.1016/j.neuropsychologia.2007.05.021

Shamay-Tsoory, S., Aharon-Peretz, J., & Perry, D. (2009). Two systems for empathy: a double dissociation between emotional and cognitive empathy in inferior frontal gyrus versus ventromedial prefrontal lesions. *Brain*, *132*, 617-627. https://doi.org/10.1093/brain/awn279

Sorce, J. F., Emde, R. N., Campos, J., & Klinnert, M. D. (1985). Maternal emotional signaling: Its effect on the visual cliff behavior of 1 year olds. *Developmental Psychology*, *21*, 195-200. https://psycnet.apa.org/doi/10.1037/0012-1649.21.1.195

Spencer, S., Zanna, M., & Fong, G. (2005). Establishing a causal chain: Why experiments are often more effective than mediational analyses in examining psychological processes. *Journal of Personality and Social Psychology*, *89*, 845-851. https://doi.org/10.1037/0022-3514.89.6.845

Stone, R., & Lavine, M. (2014). The social life of robots. *Science*, *346*, 178-179.

Sutton, R. (1991). Maintaining norms about expressed emotions: The case of bill collectors. *Administrative Science Quarterly*, *36*, 245-268. https://psycnet.apa.org/doi/10.2307/2393355

Tamir, M., Mitchell, C., & Gross, J. (2008). Hedonic and instrumental motives in anger regulation. *Psychological Science*, *19*, 324-328. https://doi.org/10.1111/j.1467-9280.2008.02088.x

Terada, K., & Takeuchi, C. (2017). Emotional expression in simple line drawings of a robot's face leads to higher offers in the ultimatum game. *Frontiers of Psychology*, *8*, Article 724. https://dx.doi.org/10.3389%2Ffpsyg.2017.00724

van Kleef, G., De Dreu, C., & Manstead, A. (2004). The interpersonal effects of anger and happiness in negotiations. *Journal of Personality and Social Psychology*, *86*, 57-76. https://dx.doi.org/10.2139/ssrn.305070

van Kleef, G., De Dreu, C., & Manstead, A. (2006). Supplication and appeasement in negotiation: The interpersonal effects of disappointment, worry, guilt, and regret. *Journal of Personality and Social Psychology*, *91*, 124-142. https://psycnet.apa.org/doi/10.1037/0022-3514.91.1.124

van Dijk, E., van Kleef, G. A., Steinel, W., & van Beest, I. (2008). A social functional approach to emotions in bargaining: When communicating anger pays and when it backfires. *Journal of Personality and Social Psychology*, *94*, 600–614. https://doi.org/10.1037/0022-3514.94.4.600

van Kleef, G. (2016). *The interpersonal dynamics of emotion: Toward an integrative theory of emotions as social information*. Cambridge University Press.

van Kleef, G., De Dreu, C., & Manstead, A. (2010). An interpersonal approach to emotion in social decision making: The emotions as social information model. *Advances in Experimental Social Psychology*, *42*, 45-96. https://doi.org/10.1016/S0065-2601(10)42002-X

van Kleef, G., Homan, A., & Cheshin, A. (2012). Emotional influence at work: Take it EASI. *Organizational Psychology Review*, *2*, 311-339. https://doi.org/10.1177%2F2041386612454911

Waldrop, M. (2015). No drivers required. *Nature*, *518*, 20-23.

Weiner, B. (1987). The social psychology of emotion: Applications of a naive psychology. *Journal of Social and Clinical Psychology*, 5, 405-419. doi:10.1521/jscp.1987.5.4.405

Wu, Y., Baker, C., Tenenbaum, J., & Schulz, L. (2018). Rational inference of beliefs and desires from emotional expressions. *Cognitive Science*, *42*, 850-884. https://doi.org/10.1521/jscp.1987.5.4.405

Zeng, Z., Pantic, M., Roisman, G., & Huang, T. (2009). A survey of affect recognition methods: audio, visual, and spontaneous expressions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *31*, 39-58. https://doi.org/10.1109/TPAMI.2008.52