How Expression and Context Determine Second-person Judgments of Emotion

Jessie Hoegen
Institute for Creative Technologies
University of Southern California
Los Angeles, CA, USA
jhoegen@ict.usc.edu

Gale Lucas
Institute for Creative Technologies
University of Southern California
Los Angeles, CA, USA
lucas@ict.usc.edu

Danielle Shore
University of Oxford
Oxford, UK
danielle.shore@psy.ox.ac.uk

Brian Parkinson
University of Oxford
Oxford, UK
brian.parkinson@psy.ox.ac.uk

Jonathan Gratch
Institute for Creative Technologies
University of Southern California
Los Angeles, CA, USA
gratch@ict.usc.edu

***Abstract*—Within the field of Affective Computing, facial expressions have traditionally been used as a means of inferring valence and emotions. Most studies have focused on interpreting facial expressions as an isolated signal, typically training algorithms with annotators with a 3rd-person point of view, often without access to the original context. When the context is provided, recent research highlights that the interpretation annotators assign to facial expressions is sometimes more strongly influenced by the context than the facial expression. But even when the context is provided, annotators are psychologically and physically distant from the original setting that produced the emotion. In this paper, we explore how context and facial expressions shape 2nd-person interpretation of facial expressions and compare this to 1st-person self-report. Results show that both expression and context contribute to self and other impressions but in different ways. Expressions and context are independent predictors of 1st-person judgments but interact to determine 2nd-person judgments. In particular, the way players interpret their partner's facial cues changes dramatically based on what just occurred in the game. We discuss the implication of these findings for automatic emotion recognition methods.**

***Index Terms*—context, facial expression, valence, prisoner's dilemma**

I. Introduction

Affective computing has adopted many perspectives on the phenomenon of emotion. Many emotion recognition techniques aim to give insight into a first-person point of view. By examining visible signals like facial expressions, voice or posture, some algorithms seek to reveal insights about an individual's subjective experience such as if they are in pain [1], if they are depressed [2], or if they are happy about a consumer product [3]. More commonly, emotion recognition methods are trained by annotators adopting a third-person point of view. For example, annotators are recruited in the lab or online and asked to annotate previously recorded emotional material. A growing concern with this work is that these annotators are divorced from the original context, yet a growing body of research shows that the interpretation of emotional expressions is context-specific [4]. For example, learning that a person has just won a tennis match can flip the interpretation of a facial expression from anger to joy [5], and learning that someone is sexually aroused can flip the interpretation of an expression from pain to pleasure [6]. Thus, numerous researchers are highlighting the disconnect between using third-person judgments to make first-person inferences [7], or exploring how to incorporate contextual cues into models of emotion inference [8], [9].

In this paper, we examine the role of context in second-person judgments of facially conveyed emotion and contrast this with first-person self-report. Compared to bystanders, people engaged in social interaction hold a much richer understanding of how events, or even their own behavior, might shape the emotions of others engaged with them in a shared activity. Recent psychological research has claimed that people are often quite accurate at inferring their partner's emotions [10]. Some have argued there is a fundamental difference between how people interpret events as social actors compared to being mere observers [11].

While some work has tried to judge affect based on context and facial expressions, as well as body language [12]. Yet this research has rarely examined the mechanism of how facial and contextual cues combine to shape emotion perceptions, nor have such studies used large datasets and automatic methods to study how people form impressions during an ongoing task.

We study emotion perceptions using data from the iterated prisoner's dilemma task. This is a standard social task used to investigate the impact of emotion [17] and expressions [9] on social decisions. In the data we examined, participants engaged in the task online and could only communicate through their actions and facial expressions, making it an ideal setting to examine how facial expressions, in particular, shape the perception of emotion. The game also has simple rules and a small number of possible actions. Prior research on the impact of facial expressions has shown that people are highly expressive when learning about their partner's most recent decision, and these expressions contain information about what just occurred [13]. Research also highlights that the perceived meaning of facial expressions can change dramatically depending on the action that co-occurs with it [14]. For example, seeing one's partner smile after they did something nice for you has a very different emotional meaning than seeing one's partner smile after they betrayed you. Such findings have typically been explored with stylized computer-generated facial expressions [9], [14]. Here we examine how people interpret the expressions of their human partner in the context of the most recent outcome of the task.

Our main goal is to yield insight into how facial expressions and context (meaning the most recent social action) combine to shape the emotion one player attributes to their partner following a given round in the iterated prisoner's dilemma. In particular, we consider two ways contextual and facial cues might combine which each might bring significant implications for the design of automated emotion-recognition technology.

Hypothesis 1: Expression and context are independent sources of information that predict emotion judgments; their additive effects determine interpretation/effect of expression

Hypothesis 2: Expression and context interact to determine emotion judgments; the relationship between a facial cue and perceived emotion changes (or even reverses) depending on the context

If hypothesis 1 holds, this would greatly simplify the development of automated emotion recognition methods. This would mean that if smiles predict positive emotion, this relationship will still hold across contexts, but contextual information might modify the inference when combined with the face. Algorithms would still need to reason about context but wouldn't need to consider all possible combinations of Face-cue cross context-cue combinations. If hypothesis 2 holds the situation becomes far more complex but could be potentially addressed by recent approaches that model multiple cue integration [9].

II. Method

A. Experimental Setup

The analyses use data that was previously analysed in the earlier paper [15]. During the study described in this paper, participants played an iterated prisoner's dilemma with another participant. They specifically chose an iterated prisoner's dilemma to study facial expressions and their influence, as research shows that facial expressions can shape joint decision-making in a prisoner's dilemma [16].

The study was performed in a lab setting, where two participants played at the same time against each other. Participants received instructions on how to play the prisoner's dilemma game and were situated in separate rooms while playing the game on a computer. The game was displayed on the screen, alongside a live video of their opponent recorded by a webcam. These video recordings were stored on a server for future analysis. The participants could see each other while playing the game but were unable to communicate by speech, as audio was not recorded by this system. The prisoner's dilemma was presented as a game where participants were playing for a set of lottery tickets and had to decide whether to *split* (corresponding to the cooperate decision in the prisoner's dilemma) or *steal* (corresponding to the defect decision) the lottery tickets. Figure 1 shows the game interface and Table I shows the payoff matrix for the game. After both participants made a decision, the joint decision would be displayed to both participants (e.g., one participant might have chosen to split, the other to steal) following a short animation. The joint decisions were logged in a database for further analysis. After this the next round of the game would start and participants would make a new decision. This continued for 10 rounds total.

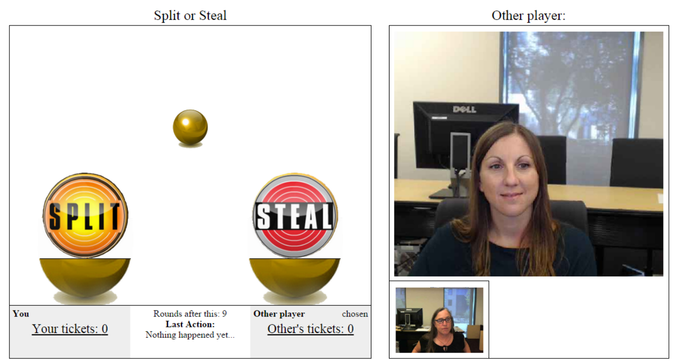

Fig. 1. The game interface. Participants make choices by selecting either the "split" or "steal" option each round. They can also see the number of tickets won by themselves and by the other player as well as the other player's webcam video (top right) and their own webcam video (bottom right).

TABLE I
Payoff matrix the iterated prisoner's dilemma game

		Participant B	
		cooperate (C)	*defect (D)*
Participant A	*cooperate (C)*	A = 5, B = 5	A = 0, B = 10
	defect (D)	A = 10, B = 0	A = 1, B = 1

Participants were randomly matched up with another participant. After the game, participants completed a video-cued recall procedure. During the video-cued recall process participants were presented with 5-second clips of the webcam footage recorded during the game they just played. More specifically, these clips corresponded to the webcam footage in the 5 seconds immediately following the moment that the joint decision was revealed to both participants. Prior research has suggested people are mostly facially expressive during this window [13]. Participants were shown video clips from each round of the game in chronological order of both themselves and their opponent. Per round, participants were first asked to attend to their own reactions and self-report their emotions on a continuous valence scale from -50 (negative valence) to +50 (positive valence). Following this, they would be shown a clip of their partner's reaction to the same result. They were told to attend to the emotions of their partner and rate perceived valence on the same continuous valence scale.

All participants provided informed consent and the study was approved by the university's research ethics committee. Participants were paid £10 for participating after they finished playing. Furthermore, a lottery draw with £100 and £50 prizes was held at the conclusion of the study. Participants who had obtained higher scores (i.e., more lottery tickets) during the game had a higher chance of winning money than those with lower scores.

B. Measures

The analyses use data that was previously described in the earlier paper [15]. It contains participants' individual decisions on each round which we used to determine the joint outcome, as well as the first-person and second-person reports of felt and perceived valence. Finally, we extracted objective information about facial cues from the participant videos in the dataset using facial expression detection software. (See below for further details)

1) Context: For this study we operationalize context as the joint decision that co-occurs with each facial expression being rated. As each participant has the choice to either cooperate or defect during the prisoner's dilemma, there are a total of 4 possible contextual situations associated with facial expressions: The participants both cooperate (joint cooperation or 'CC'), both defect (joint defection or 'DD'), or one of the participants is exploited by the other participant (exploited by other or 'CD' for the participant that was exploited, and exploiting other or 'DC' for the participant who is exploiting).

We will be employing these joint outcomes as *context* in our analysis. We argue that in a social economic game such as the social dilemma where the majority of player actions are the decisions they make, these joint outcomes are a large part of the context within which the facial displays are presented and interpreted. This is especially the case for the events where the outcome is revealed.

2) Video-cued Recall Annotations: By using the video-cued recall procedure, each participant rated videos showing themselves and their partner for all 10 joint reveal outcomes of the game they played. Each of these clips were rated for both the participant self-reported valence and the perceived valence of their partner.

3) Automatic Facial Expression Annotations: Using the timestamps that were stored with each joint decision, 5-second clips of joint outcome events were used during the video-cued recall procedure and the video was stored on a server afterward. Using these videos contained in the dataset, we obtained facial expression ratings automatically using a commercial software system originally based on CERT [17]. This system extracts Action Units (AU) as defined by the Facial Action Coding System (FACS) from each frame of the participant video. The AU values are reported as 'evidence' values, which represents the likelihood of the AU being active in a particular frame.

Action Units correspond to specific facial muscles being active. However, within the iterated prisoner's dilemma video corpus, we observe several sets of AUs that are often active at the same time. Therefore rather than looking at the individual AUs we will be using the facial expression factors [18]. Prior work has found these factors useful for explaining behavior in the prisoner's dilemma [13]. Similar dimension-reduction approaches have been applied to analyze emotional styles [19] or expressions used to deliver good or bad news [20].

Following Stratou and colleagues [18], we computed a set of factors that combine Action Units that commonly co-occur. There are 6 factors total, Factor 1 (F1) through Factor 6 (F6). F1 corresponds to a 'smile' consisting of AU 6, 7 and 12, F2 corresponds to 'eyebrows-up' consisting of AU1 and 2, F3 corresponds to 'open-mouth' consisting of AU 20, 25 and 26, F4 corresponds to 'mouth-tightening' consisting of AU 14, 17 and 23, F5 corresponds to 'eye-tightening' consisting of AU 4, 7 and 9 and F6 corresponds to 'mouth-frown' and consists of AU 10, 15 and 17. During our analysis we will be using this set of 6 factors, rather than individual action units.

C. Analysis

In order to assess H1 (i.e., expression and context are independent sources of information), we looked at the correlation between factors and valence scores. Since we have both valence ratings for displays and interpretations we furthermore looked at whether specific factors are used differently based on the perspective (first-person or second-person) of the annotator. Secondly, we have looked at the overall valence scores per joint outcome, in order to determine the independent impact of context on the ratings.

We used a moderated multiple linear regression approach to investigate our second hypothesis H2 (i.e., expression and context interact). We created regression models to predict the expressed and perceived ratings of valence, in order to find the relation between expression and context. Expressions are represented by the average value for a reveal event of the six previously defined factors that were automatically extracted from the videos following the reveal event.

In order to model context in the linear regression we used dummy coding to encode this state using the variables 'decision self' and 'decision partner.' These decisions will be coded differently based on the joint outcome: For joint cooperation (CC) 'decision self' and 'decision partner' will be coded 0 for cooperating and 1 for defecting. For joint defect (DD) they will be coded reversely, so 0 for defecting and 1 for cooperating. For exploited by other (CD) 'decision self' will be coded as 0 for cooperating and 1 for defecting, while 'decision partner' will be coded as 0 for defecting and 1 for cooperating. Finally for exploiting other (DC) 'decision self' will be coded as 0 for defecting and 1 for cooperating and 'decision partner' as 0 for cooperating and 1 for defecting.

In order to model the relation, the decision variables (i.e., 'decision self' and 'decision partner') will be used as two moderating variables in the regression on the expression, in order to predict the dependent variable of expressed or perceived valence. As such the linear regression will take the form of:

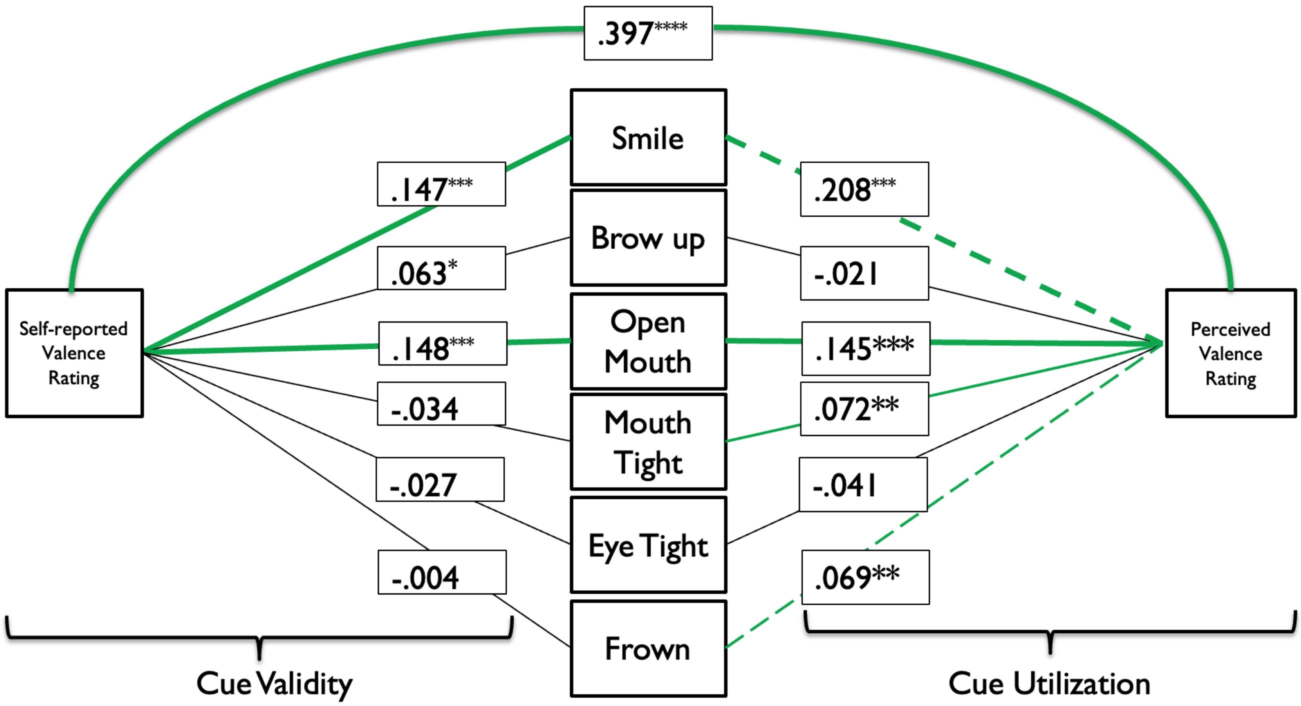

Fig. 2. Lens model of the correlations between the facial factors and participant ratings on valence. Green lines show significant correlations. Dashed lines show factors where we found significant interactions between expression and context.

$$Y = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + b_4(x_1 \times x_2) +$$
$$b_5(x_1 \times x_3) + b_6(x_2 \times x_3) + b_7(x_1 \times x_2 \times x_3) + \varepsilon \quad (1)$$

Where x are the independent values, x_1 is the factor, x_2 is 'decision self' and x_3 is 'decision other'. Y is the dependent value, self-reported or perceived valence and ε is the error term.

III. Results

First, looked at the influence of expression and context on self-reported and perceived valence separately (investigating H1). For expressions, we looked at the correlation between the factors and the valence rating.

We present our results using the lens model in Figure 2. The lens model was developed by Egon Brunswik to assess the accuracy of perceptual judgments [21]. The left-hand side of the diagram (referred to as *cue validity*) shows the relationship between some latent variable (in our case, emotion) and observable perceptual cues. The right-hand side of the diagram (referred to as *cue utilization* illustrates how a perceiver uses cues to reconstruct the latent variable (in our case, the observer's perception of their partner's emotion). Each link in the diagram shows the correlation with the latent construct (i.e., self-reported emotion on the left and perceived emotion on the right). Symmetry between the two sides of the "lens" implies that the perceiver correctly utilizes the valid cues. Asymmetry indicates perceptual errors and helps identify misconceptions.

In our case, we are particularly interested in how context (i.e., joint outcomes of the game) interacts with facial cues: (1) how does context shape the valid cues of self-reported emotion and (2) how does context shape the way perceivers utilize cues to predict how their partner feels? In Figure 2, solid lines indicate that context is an independent predictor. Dashed lines indicate that the cue and context interact. For example, the dashed line between smile and perceived emotion indicates that the way the perceiver uses the smile-cue changes based on the joint-outcome of the game. The top connecting line shows the correlation between self-reported and perceived judgments ($r(907) = 0.397$, $p < .001$).

A. Impact of the face alone

First, we examined cue validity while ignoring context. Looking at the correlation between self-reported valence and each facial factors, we found significant correlations for the smile ($r(907) = .147$, $p < .001$) and for the open mouth ($r(907) = .148$, $p < .001$) factors. This indicates that smiles and open-mouth are valid cues of self-reported emotion and each cue shows a positive relationship (i.e., more smiles mean more positive self-reported valence).

Second, we examined cue utilization while ignoring context. We examined the correlation between the partner's facial expressions and second-person perceptions that the partner felt positive or negative (i.e., perceived valence). We found significant correlations for the smile ($r(907) = 0.208$, $p < .001$), the open-mouth ($r(907) = .145$, $p < .001$), the mouth-tightening ($r(907) = .072$, $p = .30$) and the frown ($r(907) = .069$, $p = .038$) factors. This shows that observers utilize the same two valid cues (smiles and open-mouth), but also attend to two invalid cues (mouth-tightening and frown). Thus, ignoring the role of context, observers pay closest attention to the valid cues of smiling and open mouth, but also attend to irrelevant cues and thus often misinterpret their partner's self-reported feelings.

B. Impact of the context alone

The correlations in Figure 2 highlight how facial cues relate to self-reported and perceived valence but ignore how context might shape these judgments. We can also look at how context shapes judgments while ignoring facial cues. In order to find the impact of context individually we can look at the average valence scores for each context as shown in Figure 3.

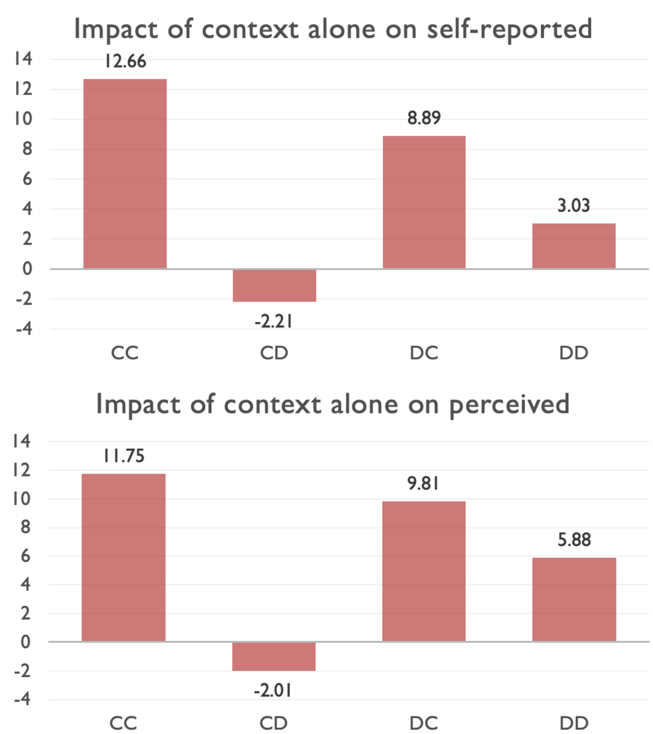

Fig. 3. Difference in valence per condition, showing display on the top and perception below. Valence ratings were collected using a continuous valence scale from -50 (negative valence) to +50 (positive valence) and averaged per round outcome.

What we see here is that context has a similar overall effect on both self-reported and perceived emotion. Regardless of whether participants are reporting their own feelings or estimating their partner's feelings, valence is reported as highest following mutual cooperation (CC). As expected from the structure of the payoff, valence is reported as negative when a player has been exploited (DC) and positive when a player successfully exploits their opponent. Interestingly, valence is reported as somewhat positive following mutual defection, but this might be explained by the fact that some money is better than none.

Just looking at the payoff alone, one might expect the greatest positive emotions when exploiting one's partner, which was not the case. Some research suggests that even in such games, people feel some guilt when they take advantage of their partner which could explain this attenuation [22].

C. How do face and context combine?

Based on the results so far, it is clear that both facial cues and game outcome are sources of information. Participants appear to be attentive to specific facial factors based on their correlation with valence. Participants also judge valence differently based on the context.

In order to contrast H1 (the face and context are independent predictors) with H2 (the face and context interact), we used moderated multiple linear regressions as described in Section II-C. We used moderated regressions for each of the factors based on both display and interpreted valence, looking at the interaction between expression and context, i.e., a total of 6 models for both display and perception.

Although we did not find significant interactions in all regressions, we did in some. Figure 2 displays the moderated linear regressions that were significant with a dashed line. What first stands out is that there appear to only be interactions on the side of the second-person perceivers utilizing facial cues, supporting H2 for second-person judgments. We found that context and the face were independent predictors of self-reported valence, supporting H1 for first-person reports.

We next highlight the multiple linear regression that showed significant interactions for perceiver cues. The first significant interaction that we found was in the regression for smile. This moderated regression showed how own decision (C or D) and partner's decision (C or D; together representing CC CD DC DD) moderated the association between smile (F1) and perceived valence. The interaction term between own decision, partner's decision and factor 1 reached statistical significance, $B = 0.308$, $t = 2.103$, $p = .036$. To break down the interaction we conducted simple slope analysis. This showed that there was a significant effect of factor 1 on the perceived expression in the CC condition ($BB = 0.395$, $t = 9.550$, $p < .001$), as well as in the DC condition ($BB = 0.229$, $t = 2.932$, $p = .003$). However, there were no significant effects of smile on perceived expression in the DD condition ($BB = 0.119$, $t = 1.430$, $p = .153$) or the CD condition ($BB = -0.023$, $t = -0.277$, $p = .781$).

Second, we found a significant interaction in the moderated regression that examined how own decision (C or D) and partners decision (C or D; together representing CC CD DC DD) moderated the association between frown (F6) and perceived valence. The interaction term between own decision, partners decision and frown reached statistical significance, $B = 0.344$, $t = 2.281$, $p = .023$. To break down the interaction we conducted simple slope analysis. This showed that there was a significant effect of frown on the perceived expression in the CD condition ($BB = 0.250$, $t = 2.872$, $p = .004$), but that there were no significant effect of factor 6 on perceived expression in the CC condition (BB = 0.026, t = 0.628, p = .530), the DD condition (BB = -0.026, t = -0.317, p = .751) or the DC condition (BB = 0.095, t = 1.157, p = .248).

Figure 4 gives an overview of the regression coefficients (B) for the moderated regressions on smile and frown.

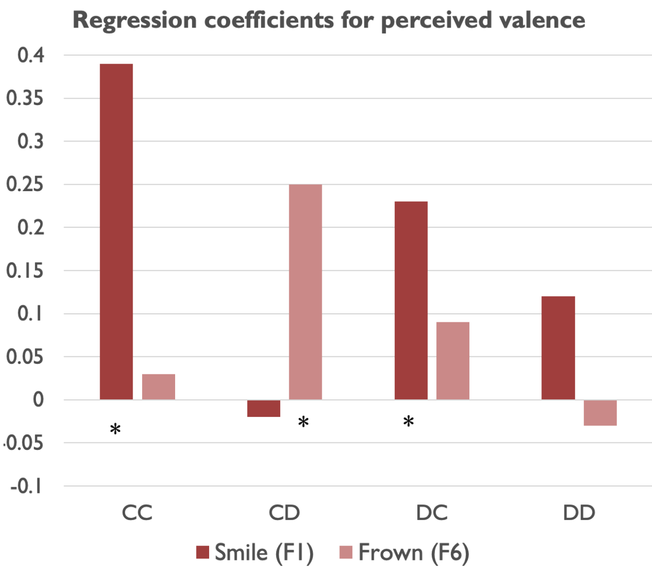

Fig. 4. The regression coefficients for smile and frown based on gamestate for perceived valence.

From this graph we can see that participants appear to use smiles and frowns differently across conditions. The significant effects for smile in the CC and DC-state in the simple slope analysis show that participants rate the associated expressions more positively if their partner displays a smile. The significant effect for frown in the CD-state means that, perhaps surprisingly, participants rate a partner showing a frown after being exploited as feeling more positive than one who does not.

IV. Conclusion

In this paper we have looked at the information contained within expression and context and their effect on the display and interpretation of emotion.

We looked at two hypotheses. Our first hypothesis stated that expression and context are independent sources of information and we found some evidence for this. Facial cues related to both self-reported and perceived emotion, as shown by the significant correlations of the smile and the open mouth factors with self-reported valence, and the significant correlation of the smile, open-mouth, mouth-tightening and frown factors with perceived valence (see Figure 2). Although smile and open-mouth appears to be used for both display and interpretation, there are some difference as well between these two perspectives. When people rated their own displays they would rate displays with only smiles and open mouth factor higher, whereas when interpreting other's displays there was a significant correlation with the the mouth-tightening and frown factors as well. These results are interesting as generally mouth-tightening and frown displays are not necessarily thought of as positive.

Whereas there were some differences in use of facial factors by self-reporters and perceivers, there appear to be no differences in their use of context when rating valence (ignoring facial expression), as shown in Figure 3. Both self-reported and perceived valence judgments were shaped by context in almost identical ways. Participants rated emotion as most positive in the joint cooperation outcome, and as more negative in the exploited by other outcome. Feelings were also reported and perceived to be more positive following successful exploitation and, interestingly, somewhat positive following mutual defection.

Finally, we examined the relationship between facial and contextual sources of information about valence. We found that facial cues and context were independent predictors of self-reported emotion (supporting H1 for first-person judgments) but that facial cues and context interacted to determine perceived emotion (supporting H2 for second-person judgments). In particular, perceivers used smiles as a cue to predict positive feelings following mutual cooperation but completely ignored smiles when inferring how someone felt after being exploited. Surprisingly, people used frowns to predict their partner was happy following exploitation (after being exploited. if they frowned they were seen as feeling more positive). In other contexts, the frown was not utilized.

The results relating to smiling might seem intuitive. As people smiling in CC might be expected since they are in a mutually beneficial state. While in the DC-state they won a lot of points themselves, perhaps making the opponent perceive these smiles as more genuine and, as such, correlated with positive valence. However, the results relating to frowning might appear counter-intuitive initially. One possible explanation based on anecdotal observations is that frowns are often accompanied by smiles as well. In many of these cases the participant appears to be signaling something along the lines of "You got me," perhaps showing that they are willing to forgive their partner for exploiting them. However further work will need to be done in order to confirm this.

Taken together, our results reinforce prior findings that context and facial displays provide important cues about emotion. Prior work has suggested that people are sometimes poor judges of emotion. Our work supports this in that people sometimes utilize invalid cues when estimating their partners feelings. But our more novel finding is that the way that context and facial cues interact also differs between self-reported and perceived emotion. This may prove challenging for approaches that seek to improve the accuracy of emotion recognition by simply providing context to emotion annotators.

Ethical Impact Statement

This work highlights the importance of considering context when interpreting facial expressions, as this can significantly affect the emotional interpretation of a second party. All study participants provided informed consent and the study was approved by the university's research ethics committee. Our findings highlight the importance of context and suggest that automatic emotion recognition methods that use facial expression may need to be adjusted to account for contextual cues. Since our work is focused on a very specific setting, that of a social dilemma, such methods may require further research to be generalizable. As such, any methods that are developed based on this work, should be deployed with care. Finally, we hope that our research contributes to a better understanding of human emotions and how they are perceived in social interactions, and that this may lead to informing strategies to improve interpersonal communication and social functioning.

Acknowledgment

This research was sponsored by the Army Research Office and was accomplished under Cooperative Agreement Number W911NF-20-2-0053. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Office or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation herein.

References

[1] T. Hassan, D. Seuß, J. Wollenberg, K. Weitz, M. Kunz, S. Lautenbacher, J.-U. Garbas, and U. Schmid, "Automatic detection of pain from facial expressions: a survey," *IEEE transactions on pattern analysis and machine intelligence*, vol. 43, no. 6, pp. 1815–1831, 2019.

[2] A. Jan, H. Meng, Y. F. A. Gaus, F. Zhang, and S. Turabzadeh, "Automatic depression scale prediction using facial expression dynamics and regression," *Avec '14*, p. 73–80, 2014.

[3] D. McDuff, R. E. Kaliouby, J. F. Cohn, and R. W. Picard, "Predicting ad liking and purchase intent: Large-scale analysis of facial responses to ads," *IEEE Transactions on Affective Computing*, vol. 6, no. 3, pp. 223–235, 2015.

[4] M. J. Wieser and T. Brosch, "Faces in context: A review and systematization of contextual influences on affective face processing," *Frontiers in psychology*, vol. 3, p. 471, 2012.

[5] L. F. Barrett, B. Mesquita, and M. Gendron, "Context in emotion perception," *Current Directions in Psychological Science*, vol. 20, no. 5, pp. 286–290, 2011.

[6] J.-M. Fernández-Dols, P. Carrera, and C. Crivelli, "Facial behavior while experiencing sexual excitement," *Journal of Nonverbal Behavior*, vol. 35, no. 1, pp. 63–71, 2011.

[7] L. F. Barrett, R. Adolphs, S. Marsella, A. M. Martinez, and S. D. Pollak, "Emotional expressions reconsidered: Challenges to inferring emotion from human facial movements," *Psychological Science in the Public Interest*, vol. 20, no. 1, pp. 1–68, 2019.

[8] R. Kosti, J. M. Alvarez, A. Recasens, and A. Lapedriza, "Context based emotion recognition using emotic dataset," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, no. 11, pp. 2755–2766, 2020.

[9] D. C. Ong, J. Zaki, and N. D. Goodman, "Affective cognition: Exploring lay theories of emotion," *Cognition*, vol. 143, pp. 141–162, 2015.

[10] J. Lange, M. W. Heerdink, and G. A. van Kleef, "Reading emotions, reading people: Emotion perception and inferences drawn from perceived emotions," *Current Opinion in Psychology*, vol. 43, pp. 85–90, 2022.

[11] L. Schilbach, B. Timmermans, V. Reddy, A. Costall, G. Bente, T. Schlicht, and K. Vogeley, "Toward a second-person neuroscience," *Behavioral and Brain Sciences*, vol. 36, no. 4, pp. 393–414, 2013.

[12] S. K. D'mello and A. Graesser, "Multimodal semi-automated affect detection from conversational cues, gross body language, and facial features," *User Modeling and User-Adapted Interaction*, vol. 20, pp. 147–187, 2010.

[13] S. Lei and J. Gratch, "Emotional expressivity is a reliable signal of surprise," *IEEE Transactions on Affective Computing*, pp. 1–12, 2023.

[14] C. de Melo, P. J. Carnevale, S. J. Read, and J. Gratch, "Reading people's minds from emotion expressions in interdependent decision making," *Journal of Personality and Social Psychology*, vol. 106, no. 1, pp. 73–88, 2014.

[15] R. Hoegen, J. Gratch, B. Parkinson, and D. Shore, "Signals of emotion regulation in a social dilemma: Detection from face and context," in *2019 8th International Conference on Affective Computing and Intelligent Interaction (ACII)*. IEEE, 2019, pp. 1–7.

[16] R. H. Frank, *Passions within reason: The strategic role of the emotions.* WW Norton & Co, 1988.

[17] G. Littlewort, J. Whitehill, T. Wu, I. Fasel, M. Frank, J. Movellan, and M. Bartlett, "The computer expression recognition toolbox (cert)," in *2011 IEEE International Conference on Automatic Face & Gesture Recognition (FG)*. IEEE, 2011, pp. 298–305.

[18] G. Stratou, J. Van Der Schalk, R. Hoegen, and J. Gratch, "Refactoring facial expressions: An automatic analysis of natural occurring facial expressions in iterative social dilemma," in *2017 Seventh international conference on affective computing and intelligent interaction (ACII)*. IEEE, 2017, pp. 427–433.

[19] H. Ilgen, J. Israelashvili, and A. Fischer, "Personal nonverbal repertoires in facial displays and their relation to individual differences in social and emotional styles," *Cognition and Emotion*, vol. 35, no. 5, pp. 999–1008, 2021.

[20] D. M. Watson, B. B. Brown, and A. Johnston, "A data-driven characterisation of natural facial expressions when giving good and bad news," *PLOS Computational Biology*, vol. 16, no. 10, p. e1008335, 2020.

[21] E. Brunswik, *Perception and the representative design of psychological experiments*. Univ of California Press, 1956.

[22] E. Fehr and K. M. Schmidt, "A theory of fairness, competition, and cooperation," *The quarterly journal of economics*, vol. 114, no. 3, pp. 817–868, 1999.