Identifying Meaningful Facial Configurations during Iterative Prisoner's Dilemma Games

Olly Robertson (ORCID 0000-0002-7333-0903)[1, 2], Brian Parkinson (ORCID 0000-0001-8065-5725)[1] and Danielle Shore (ORCID 0000-0001-7290-3696)[1]

[1] Department of Experimental Psychology, University of Oxford, UK

[2] Department of Psychiatry, University of Oxford, UK

Author Note

This research was funded by the European Office of Aerospace Research and Development grant FA9550-18-1-0060 awarded to Brian Parkinson and Danielle Shore.

Correspondence concerning this article should be addressed to Olly Robertson, Department of Psychiatry, University of Oxford, Oxford, UK. Email: olly.robertson@psych.ox.ac.uk

Abstract

The contraction and relaxation of facial muscles in humans is widely assumed to fulfil communicative and adaptive functions. However, to date most work has focussed either on individual muscle movements (action units) in isolation or on a small set of configurations commonly assumed to express “basic emotions”. As such, it is as yet unclear what information is communicated between individuals during naturalistic social interactions and how contextual cues influence facial activity occurring in these exchanges. The present study investigated whether consistent patterns of facial action units occur during dyadic iterative prisoners’ dilemma games, and what these patterns of facial activity might mean. Using exploratory and confirmatory factor analyses, we identified three distinct and consistent configurations of facial musculature change across three different datasets. These configurations were associated with specific gameplay outcomes, suggesting that they perform psychologically meaningful context-related functions. The first configuration communicated enjoyment and the second communicated affiliation and appeasement, both indicating cooperative intentions after cooperation or defection respectively. The third configuration communicated disapproval and encouraged social partners not to defect again. Future work should validate the occurrence and functionality of these facial configurations across other kinds of social interaction.

Keywords: Facial expression, trust, cooperation, FACS, emotion, communication

Identifying Meaningful Facial Configurations during Iterative Prisoner's Dilemma Games

"There's no art to find the mind's construction in the face"

(Shakespeare, 1710)

In Shakespeare's Macbeth (1710), King Duncan tries, and later fails, to make sense of the defection of a trusted companion. He finally contends that it is impossible to truly know or predict the behaviours of another simply by looking at their face. The scientific literature is seemingly at odds with these musings. Psychological research consistently informs us that humans extrapolate a wide range of information from facial movements (Barrett et al., 2019; Calvo & Nummenmaa, 2016). The contraction of facial muscles creates folds, wrinkles, and areas of tightness which are assumed to fulfil communicative and adaptive functions (Rinn, 1984). While thousands of potential configurations are possible, most research focuses on deriving communicative value from either individual muscle movements (e.g., Tian et al., 2000) or configurations of movements based on theoretical assumptions relating to categories of emotion (e.g., Ekman, 1992; Morais, 2022). Evidence suggests that individuals navigate their everyday social world by making inferences about an interaction partner's state of mind from these clusters of facial muscle movements (de Melo et al., 2012).

However, there is little consensus about the precise nature of the information that facial muscle movements provide (Barrett & Satpute, 2019; Scarantino, 2017; Siegel et al., 2018). In addition, expressions in everyday interactions often do not resemble the patterns of muscle movements supposedly associated with basic emotions (e.g., Ekman, 1992), and perceivers' interpretation of these non-prototypical facial configurations is correspondingly less consistent (Hess & Hareli, 2017; Hoegen et al., 2019). These findings have led to a shift in research focus towards assessing the facial movements that are actually produced during realistic interpersonal interactions without making restrictive assumptions about the patterns that faces should adopt during emotional experience (e.g., Hyniewska et al., 2019; Stratou et al., 2017; Zhou et al., 2020). Such a shift facilitates exploration of how facial configurations vary based on intrapersonal, interpersonal, and contextual cues as opposed to predefined emotion categories. Recent research highlights that context is important for recognition of facial expressions in interpersonal interactions (Hoegen et al., 2019). The study reported here therefore aimed to examine the configurations of facial muscle movements occurring during a specific interpersonal context (iterated prisoner's dilemma game) and to evaluate what information these expressions may convey within that context.

As noted above, researchers disagree about the kinds of information that can be extracted from facial activity. Some categorical conceptualisations suggest that faces display information about a person's emotional state. In particular, Basic Emotion Theory (e.g., (Ekman, 1992; Ekman & Keltner, 1997) holds that a limited number of canonical facial expressions clearly and reliably signal specific emotion categories, providing a means through which interaction partners can express and interpret otherwise unknowable emotional qualia (for discussion, see Barrett & Satpute, 2019). According to some recent exponents of this view, upwards of 20 specific emotions can be reliably diagnosed from dynamic expressions (Keltner et al., 2019). Each expression allows interaction partners to draw systematic inferences about the expresser's inclinations, guided by the specific emotion expressed. A wealth of literature in this area focuses on the recognition and perception of prototypical, static basic emotion expressions (Scherer et al., 2011), and suggests that humans consistently associate specific emotions with posed prototypical facial expressions when relevant categories are provided as response options (Lopes et al., 2017).

Despite the high levels of consistency in categorising static images of prototypical facial expressions, correlations between basic emotions and predicted facial expressions are generally low in both naturalistic (Fernández-Dols & Crivelli, 2013) and laboratory settings (Durán & Fernández-Dols, 2021; Reisenzein et al., 2013). The fact that prototypical basic emotion expressions occur relatively infrequently during interpersonal interactions (Gaspar et al., 2014) may suggest that facial muscle movements serve a range of purposes in addition to any role they might play in emotion expression. Indeed, research suggests that patterns of muscle activity vary across eliciting contexts, such that diverse patterns of facial movements convey specific information relevant to the current social interaction (Aviezer et al., 2008; Crivelli & Fridlund, 2019). According to this functional perspective, facial movements are flexible, context dependent, and contingent upon the dynamics of the interpersonal interactions in which they occur (Crivelli & Fridlund, 2018).

To determine different dimensions of naturalistic facial activity during ongoing social interactions a previous study conducted an exploratory factor analysis on a large dataset of naturalistic expressions produced during dyadic experimental tasks (Stratou et al., 2017). Frames from these videos were processed, using commercial software based on the computer expression recognition toolbox (CERT; Littlewort et al., 2011). The resulting output provided likelihood scores for 16 specific facial muscle movements, with higher scores indicating a higher likelihood that the specific muscle had been activated in that frame. These *Action Units* (AUs; (Ekman & Friesen, 1978) derived from the Facial Action unit Coding System (FACS) were then analysed to elucidate meaningful groupings of facial muscles. These likelihood scores were subsequently subjected to exploratory factor analysis.

From this analysis, six consistent configurations of AU activation were identified, described as Enjoyment Smile, Eyebrows Up, Open Mouth, Mouth Tightening, Eye Tightening, and Mouth Frown (Stratou et al., 2017). These configurations mainly failed to map directly onto basic emotion categories. For data derived from Prisoner's Dilemma gameplay, the authors later correlated each participant's factor scores with specific outcomes (e.g., number of rounds resulting in mutual cooperation), and concluded that the documented 6 configurations were psychologically meaningful, as they were likely related to contextual cues in social interactions rather than to theoretical emotion categories.

The present study builds upon this prior research by focusing on a circumscribed interpersonal situation during an experimental game (Prisoner's Dilemma, PD) in order to clarify the specific interpersonal contingencies that give rise to particular patterns of changes to facial musculature. While Stratou and colleagues (2017) assessed facial muscle configurations across entire interactions (i.e., IPD gameplay, negotiations and diagnostic interviews), the current research focuses on a specific period in IPD gameplay (when the round outcome was revealed to both players simultaneously). By limiting the analysis of facial activity to this period, we were able to assess the impact of different gameplay outcomes such as mutual cooperation or mutual defection.

As operationalized in our study, PD is a two-player task, where outcomes are predicated on the simultaneous choices of both players (Poundstone, 1993). Cooperation yields the largest mutual reward, but successful defection yields a greater individual reward, thus creating a dilemma of trust. We used an Iterated prisoner's dilemma (IPD) task in which this dilemma is repeated across multiple rounds, weaving a dynamic narrative of trust and deceit distinct to each dyad. To aid decision-making, players typically use a range of cues to draw inferences about their opponent to predict their most likely action. There is evidence that inferences made from facial expressions significantly impact decisions in IPD (Hoegen et al., 2019; Lei et al., 2020).

As facial muscle movements are highly context-dependent (e.g., (Parkinson, 2013), it is important to delineate how they are influenced by variations in gameplay outcome. By identifying clusters of facial muscle movements which convey important information about intentions and orientations in situations where trust can be violated, research may shed light on how those facial configurations can predict or promote mutually beneficial outcomes.

Using a similar approach to Stratou and colleagues (2017) we processed our videos of participants engaged in dyadic IPD gameplay using the automated Facial Action Coding System AFFDEX (McDuff et al., 2016). AFFDEX automatically provides intensity scores for specific facial actions across 34 AUS which were then analysed to elucidate meaningful groupings of facial activity associated with specific outcome contexts. Following the same procedure as Stratou et al (2017), we then conducted EFA on the data; we extended this approach by undertaking multiple EFAs and a CFA to assess the extent to which our findings were a robust reflection of patterns in facial muscle movements, rather than random noise. When assessing latent structures, multiple EFAs on separate datasets are recommended as a form of cross-validation (Thompson, 2004). Subsequent CFA of unrelated data provides a further test of model consistency. We therefore divided our video data into three sets and explored the first two using EFA before applying CFA to the final set. To establish the meaning of any discovered facial action factors, the relationship between the factors, game states and decisions were explored subsequent to factor analysis. Through assessing the relationships between patterns of facial muscle movements to specific game states, the current study aims to assess whether there are reliable patterns of facial activity that respond to these contexts and that provide meaningful context-related information to interaction partners.

General Methods

Participants

Three different datasets were used in the present analyses (see table 1 for demographic information), containing a total of 334 participants who each engaged in 10 rounds of IPD gameplay video-recorded using a webcam. In order to create datasets of comparable size, the second dataset pooled observations from two studies (Study One *n* = 62; Study Two *n* = 46). All participants were recruited through local advertisements or from local community panels and entered into a lottery to win monetary prizes (1 x £100 and 2 x £50 for dataset 1 and 3; 8 x £25 for dataset 2). Participants whose data were included in dataset 1 and 3 were also paid £10 for participation. Pairs of participants did not know each other prior to taking part in the study. All participants provided written informed consent and the University of Oxford's Ethics Committee approved all studies.

Table 1. Demographic information for samples across the 3 datasets

Dataset	*N*	Gender	M Age (SD)	Age Range	Ethnicity	*N* Factor Analysis	*N Rounds* Correlation
1	100	67% Female, 33% Male	26.35 (7.44)	18-59	23.4% BAME	91	840
2	108	53.70% Female, 46.30% Male	20.74 (1.38)	19-24	49.64% BAME	103	980
3	126	65.08% Female, 34.92% Male	25.54 (7.78)	18-66	38.1% BAME	114	110

Procedure

Across the three studies, all participants played 10 rounds of a computer-mediated PD in pairs. Participants were paired randomly and seated in separate cubicles. Both participants could see each other throughout the game via a live webcam feed that did not include audio. Participants were instructed not to talk or use hand gestures to communicate. The IPD task was developed by Hoegen and colleagues (2015) and modelled on the British TV show Golden Balls (see Figure 2 for game interface). In each round, participants played for a set of lottery tickets and selected whether to "Split" (cooperate) or "Steal" (defect) by clicking the relevant button on the screen. Once both participants had made their choices, the outcome of the round was displayed to both participants. A ticket counter allowed participants to track their current scores. The outcome for both players depended on their joint decisions, with maximum joint profit for mutual co-operation (CC) and minimum joint profit for mutual defection (DD). In mixed outcomes (CD and DC), the individual profit for the defector exceeded the individual profit in mutual cooperation (see Table 2 for the payoff matrix). Participants were informed that each ticket that they earned would be entered into a lottery draw for a monetary prize, and as such, by obtaining more tickets they could increase their chances of winning a prize. After the game, participants answered questions assessing their general game experience and impressions of their game partner (see supplementary materials for item detail[1]). Following completion of the questionnaires, participants were debriefed and thanked for their participation. The procedure in the second and third dataset involved additional manipulations which varied slightly from the above and whose effects are not assessed here (please see supplementary materials for details).

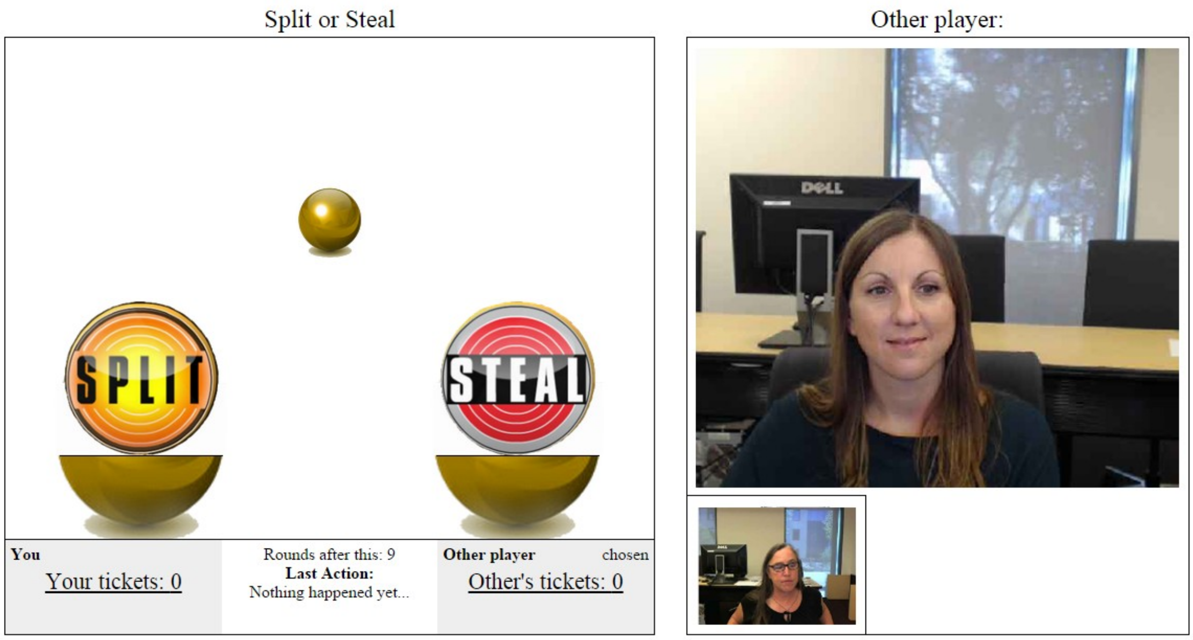

Figure 1. The IPD game interface. Participants selected either the "split" or "steal" ball on each round. Participants could see the number of tickets won by both players as well as the other player's live video feed (big window) and their own video feed (small window) throughout the game. (Image taken from Hoegen et al., 2015)

Table 2. Payoff Matrix for Levels of Cooperation and Defection between Dyad Members

		Participant B	
		Cooperate (Split)	Defect (Steal)
Participant A	Cooperate (Split)	A = 5, B = 5	A = 0, B = 10
	Defect (Steal)	A = 10, B = 0	A = 1, B = 1

[1] Responses to these questionnaires are not used in the present paper's analyses, for further details please see supplementary materials. The procedures for the second and third dataset involved additional manipulations, either a short interaction or instructions to regulate/suspect regulation prior to the IPD game. The effects of these manipulations are not assessed here, see supplementary materials for further details.

Data Preparation and Analysis

Facial configurations from the webcam videos were automatically analysed using the AFFDEX by Affectiva module of iMotions software (McDuff et al., 2016). AFFDEX outputs frame-by-frame intensity values for 20 observable muscle movements (AUs), as defined by the Facial Action Coding System (Ekman & Friesen, 1978). AFFDEX has been found to reliably detect and report AU activation, with Receiver Operator Characteristic scores ranging from .75 to .96 (McDuff, 2016). AU activation values range from 0-100, where 0 is no activity and 100 indicates full activation. Video footage was synchronised with actions in the IPD game, and the 5-second (150 frames) "outcome" clips were analysed for each of the 10 trials for each participant. In line with prior research (Kulke et al., 2020), any participant with greater than 10% of data missing (because of undue head movement or computer error) was excluded from analyses (dataset 1 *N* = 9; dataset 2 *N* = 5; dataset 3 *N* = 12).

The first two datasets were analysed using Exploratory Factor Analysis (EFA) to examine the underlying factor structure of AU activation during IPD game play. By analysing two separate datasets the degree of agreement in factors could be evaluated. The third dataset was analysed using Confirmatory Factor Analysis to test the model derived from the two EFAs. We than used correlation and moderation analyses to assess the relationship between factor scores and IPD gameplay outcomes.

Factor analyses and correlations were conducted using R version 1.1.456, and moderation analyses were conducted using the PROCESS macro (Hayes, 2022) in SPSS version 26. Because each participant not only made a decision but also responded to their partner's decision during each trial, their data is represented twice in the analyses. Further each pair provides data across repeated trials. For both of these reasons, it should be noted that the observations used in our analyses (like those used in Stratou et al's, 2017, study) are not all statistically independent.

Analysis Phase 1: Assessing the Structure of Facial Activity in IPD

The structure of facial activity in IPD was evaluated using factor analyses. Factor analysis is a statistical technique which assesses patterns of multidimensional constructs available in the measured variables. Factor analyses are dichotomised as being either exploratory or confirmatory in nature, with the differing types meeting different but complementary analytical requirements. In the current work, exploratory factor analysis on data sets 1 and 2 was used to delineate a structure of distinct patterns of facial movement (AUs which co-occurred with one another activity). We then used confirmatory factor analysis on data set 3 was used to investigate test factor structure validity and whether the structure of facial activity specified in the exploratory analyses represents AU activation available in a further dataset thereby providing an indication for the construct validity of the factor structure. We discuss the analytical assumptions, decisions (i.e. rotation), and results are discussed in turn below (see supplementary materials for all statistical output).

Exploratory Factor Analyses

To test the underlying factor structure, we subjected the first and second datasets to exploratory factor analysis (EFA) using varimax rotation, with the observed variables being the 20 AUs available from AFFDEX analysis. Varimax rotation was used to ensure that multiple observed variables (AUs) did not load onto multiple factors. This provided pure cluster solutions, where variables do not cross-load, for further analysis. As values in both datasets did not follow a gaussian distribution, were skewed, and kurtosed, all data were log-transformed prior to analysis. The KMO test indicated sampling adequacy at .78 and .75 for the first and second dataset respectively (Kaiser, 1974). Bartlett's test of sphericity was significant for both the first ($\chi 2$ (171) = 12033.38, $p < .001$) and second ($\chi 2$ (171) = 12834.07, $p < .001$) dataset, indicating suitable factorability of the correlation matrix (Field, 2009). We used Stevens' (1992) criterion such that items needed to load ≥ .4 in order to be retained. For the first dataset, the scree plot indicated a three-factor solution; with two factors having eigenvalues > 1 and the point of inflexion falling at the third factor. For the second dataset, analysis of the scree plot indicated three eigenvalues > 1. While the point of inflexion indicated a four-factor model, only one variable loaded at a value of ≥. 4 on the fourth factor (see supplemental materials for further details and scree plots). A three-factor model was therefore preferred, which accounted for 43% and 44.16% of the variance in the first and second datasets, respectively. The patterns of loading for each of the three factors were similar across data sets (see Table 3).

Factor 1's highest loading items were lip corner puller (AU 12), mouth open (AU 27), and cheek raiser (AU 6). This factor accounted for 15.8% of the model variance in the first dataset and 15.41% in the second. Factor 2's highest-loading items were dimpler (AU 14), chin raiser (AU 17), lip pressor (AU 24), lip stretch (AU 20), and lip suck (AU 28). This factor accounted for 15.7% of the model variance in the first dataset and 13.32% in the second. Factor 3's highest-loading items were brow furrow (AU 4), nose wrinkle (AU 9), and upper lip raiser (AU 10). This factor accounted for 11.47% of model variance in the first dataset and 15.42% in the second. It is worth noting that some AUs did not load highly on any of these three factors. This does not mean that participants did not produce these AUs, but that they were not strongly associated with other AUs across each dataset. The presence of such low-loading items is not unusual in EFA and is not considered to be problematic.

Table 3. Exploratory Factor Structure from IPD Data

	First Dataset			Second Dataset		
	Factor 1	Factor 2	Factor 3	Factor 1	Factor 2	Factor 3
Cheek Raise (AU 6)	.94			.95		
Lip Corner Puller (AU 12)	.92			.92		
Mouth Open (AU 27)	.56			.58		
Jaw Drop (AU 26)						
Lid Tighten (AU 7)				.40		
Lip Pressor (AU 24)		.79			.85	
Dimpler (AU 14)		.84			.89	
Lip Stretcher (AU 20)		.65			.66	
Lip Suck (AU 28)		.57			.61	
Chin Raiser (AU 17)		.65			.61	
Brow Furrow (AU 4)			.73			.82
Nose Wrinkler (AU 9)			.88			.88
Upper Lip Raiser (AU 10)			.86			.90
Eye Closure (AU 43)						
Eye Widen (AU 5)						
Inner Brow Raiser (AU 1)						
Lip Corner Depressor (AU 15)						
Lip Pucker (AU 18)						
Outer Brow Raiser (AU 2)						

The EFAs produced highly similar factor solutions across the two datasets. Activation of the same AUs loaded highly on the three factors in each case, with the exception of Lid Tighten (AU 7) which had a marginal loading (0.40) on Factor 1 in the first dataset but did not meet the retention threshold in the second dataset. The reliability of the underlying factor solution speaks to the robustness of the proposed model and makes it an appropriate selection for further assessment using confirmatory factor analyses.

Confirmatory Factor Analysis

We conducted a Confirmatory Factor Analysis (CFA) on dataset 3 to test the three-factor model found in datasets 1 and 2. To evaluate model fit, both the above three-factor model and Stratou and colleagues' (2017) six-factor model were assessed using CFA. The goodness-of-fit for each model was also subsequently compared with a one-factor model. The comparisons were undertaken as Jackson and colleagues (2009) recommend assessing competing model fit to avoid modifying models post-hoc to improve fit.

Prior to CFA using maximum likelihood, we log-transformed data to alleviate skew and kurtosis. Some variables (lip corner puller [AU14]; cheek raiser [AU6]; and eye closure [AU43]) remained non-normally distributed, and all variables were kurtosed > 2.0. CFA undertaken on skewed or kurtosed variables is likely to yield inflated chi-square and RMSEA fit index statistics (Beauducel & Herzberg, 2006). As such RMSEA and the ratio of chi-square to degrees of freedom are not reported in the current paper (see supplementary materials for these fit statistics). Based on Hu and Bentler's (1999) recommendations, we assessed the model using the following fit index thresholds: CFI ≈ .95 and SRMR < .09. A measure of absolute and incremental fit is theorised to reflect the whole model (Beauducel & Herzberg, 2006).

*Three-Factor Model***.** The three-factor model based on our earlier EFA met the recommended fit indices. Fit indices were CFI = .917 and SRMR = .088. All items loaded well onto the factors, with loadings ranging from .49 to 1.09 (see Figure 4). Because data were not normally distributed, we conducted non-parametric bootstrapping in 2000 simulated samples with normal Maximum Likelihood (Bollen & Stine, 1992). The bootstrap procedure indicated that the proposed three-factor model was consistent with the data, $p = .163$ (Walker & Smith, 2017). The three-factor model also fit the data significantly better than a single latent factor model for action unit activation, $\chi^2(44) = 6115.20$, $p < .001$. The internal reliability of scales based on each the three factors was satisfactory at Cronbach's $\alpha = .81$ for Factor 1, $\alpha = .78$ for Factor 2, and $\alpha = .64$ for Factor 3.

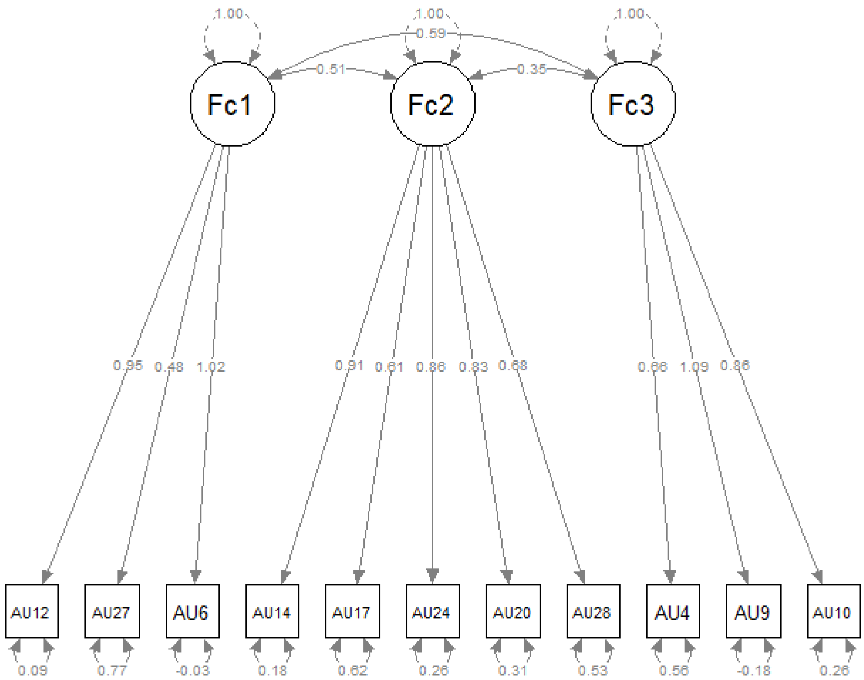

Figure 4. CFA model with 3 latent factors; Fc1 = Factor 1, Fc2 = Factor 2, Fc3 = Factor 3

Six-Factor Model. Stratou and colleagues' (2017) six-factor model with (1) Enjoyment Smile (AUs 6, 7, and 12), (2) Eyebrows Up (AUs 1 and 2), (3) Open Mouth (AUs 20, 25 and 26), (4) Mouth Tightening (AUs 14, 17 and 23), (5) Eye Tightening (AUs 5, 7 and 9), and (6) Mouth Frown (AUs 10, 15 and 17), did not meet the recommended fit indices, CFI = .872 and SRMR = .113. Not all items loaded onto Stratou et al's (2017) previously stipulated latent factors. Other than for the factor of (2) 'Eyebrows Up' – comprised of outer brow raiser (AU 1) and inner brow raiser (AU 2) – which resulted in Ultra Heywood cases, item loadings ranged from .009 to 1.056, with (1) Enjoyment Smile and (5) Eye Tightening being the only factors where ≥ 2 variables loaded above .4 (see figure 5). Due to non-normal data, non-parametric bootstrapping was conducted in 2000 simulated samples with normal Maximum Likelihood (Bollen & Stine, 1992). Results indicated that the six-factor model was also consistent with the data, p = .114 (Walker & Smith, 2017).

Compared to a single latent factor model for action unit activation, the six-factor model provided a significantly better fit, $\chi^2(74) = 1612.4$, $p < .001$. However, evidence of ultra-Heywood cases, as found in the 6-factor model, likely render a factor solution invalid (Cooperman & Waller, 2022). In any case, our 3-factor model was a superior fit for the data to the 6-factor model and met the recommended fit indices.

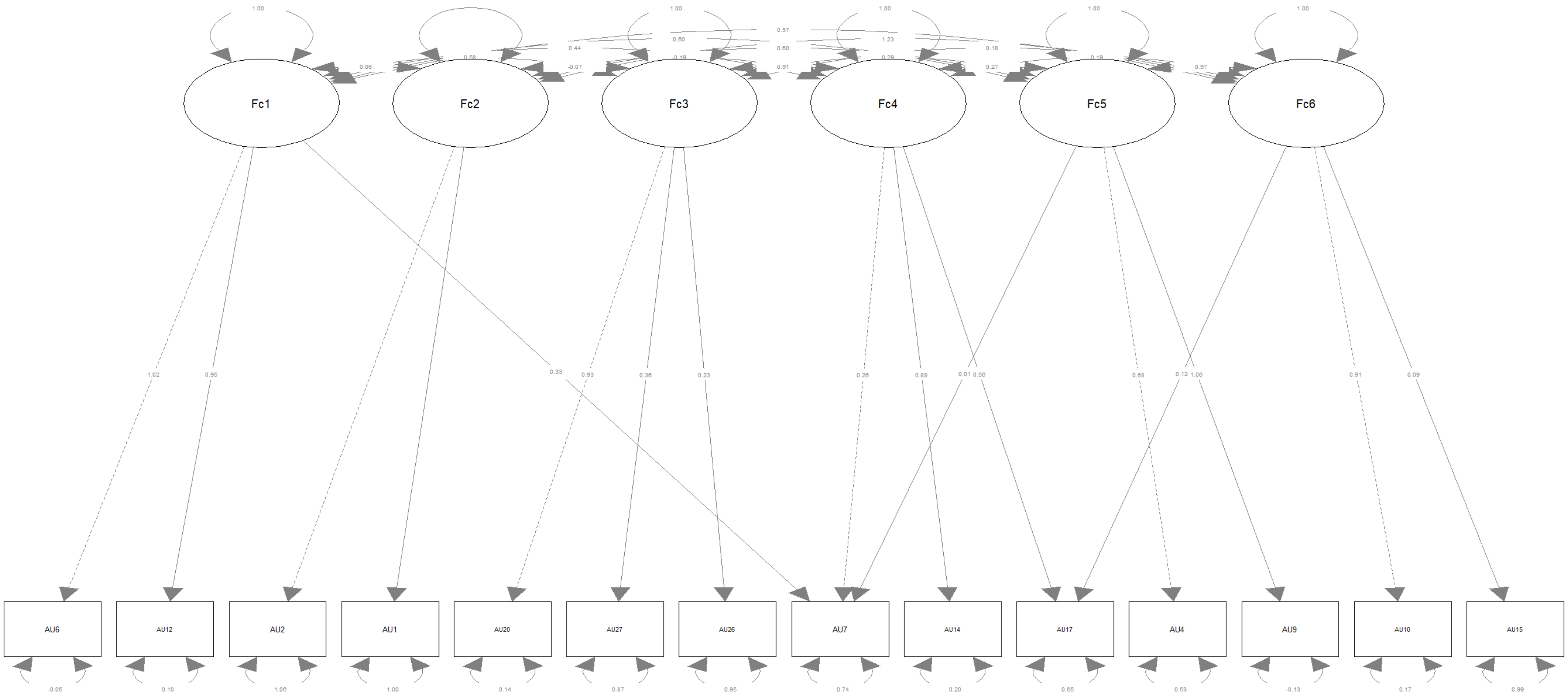

Figure 5. CFA model with 6 latent factors; Fc1 = (1) Enjoyment Smile, Fc2 = (2) Eyebrows Up, Fc3 = (3) Open Mouth, Fc4 = (4) Mouth Tightening, Fc5 = (5) Eye Tightening , Fc6 = (6) Mouth Frown.

Discussion

The aim of the first analysis phase was to examine patterns of facial muscle activity in IPD contexts and to elucidate meaningful groupings of facial activity that likely communicated specific interaction relevant information. Two EFAs and a CFA yielded a consistent 3-factor model of AU activation. The CFA showed that this three-factor model had superior fit scores to both a single factor model and the six-factor model previously obtained by Stratou and colleagues (2017).

Factor Interpretation

Three AUs loaded high on our first factor: Lip Corner Puller (AU12), Mouth Open (AU27), and Cheek Raiser (AU6). In previous research AU12 and AU6 are central to smile taxonomies and combined with Lips Parted (AU25) indicate Reward Smiles (Martin et al., 2017; Rychlowska et al., 2017). As iMotions does not provide data relating to AU25, we suggest AU27 may indicate a proxy measure for this AU. Smiles which include AU6 and AU12 activation are assumed to represent a genuine smile denoting enjoyment or amusement (Gunnery et al., 2012). As such, Factor 1 likely reflects a common expression inherent in social interactions, rather than a context or IPD gameplay specific pattern of musculature change. Thus, we name our first factor 'Reward Smile' (RS).

It is noted that Reward Smiles resemble AU activation associated with prototypical displays of the basic emotion 'Joy' (Ekman, 1992), which are typified by AU6 and AU12. There is also overlap between the present RS factor and Stratou and colleagues' (2017) 'Enjoyment Smile' (AU6, AU7, and AU12) factor. These smiles are associated with receiving or giving rewards or positively valenced emotion experiences. It is, therefore, reasonable to suggest that RSs may be associated with mutually positive or rewarding experiences (e.g., rounds in which both players cooperate).

Five AUs had high loadings on our second factor: Dimpler (AU 14), Chin Raiser (AU 17), Lip Pressor (AU 24), Lip Stretch (AU 20) and Lip Suck (AU 28). The morphology is similar to that of affiliative smiles (AU 12, AU 17) as characterized by Martin et al. (2017) and Rychlowska et al (2017). Affiliative smiles are theorised to reflect positive social intentions and outcomes; as well as indicating low levels of aggression or threat, thereby promoting approach motivations, social bonds, and intimacy (Niedenthal et al., 2010). The patterns of AU activation for Factor 2 also resemble expressions of regret, which are theoretically associated with Dimpler (AU 14), Lip Tightening (AU 23), and Lip Pressor (AU 24) activation (Rychlowska et al., 2017). Therefore, Factor 2 arguably adds further nuance in communicating context-dependent information. In particular, Factor 2 may not specifically fulfil affiliative functions, but rather may suggest some form of appeasement (Keltner, 1995; Rychlowska et al., 2017). Appeasement is theorised to be an inhibited, submissive behaviour associated with the recognition and commitment to social norms (Keltner et al., 1997). Through the theoretical association between affiliation and appeasement, factor 2 is therefore named Mouth-Based Appeasement (MBA).

This factor shares some similarities to Stratou et al's (2017) 'Mouth Tightening' (AU14, AU17 and AU23) factor. This factor was suggested to indicate appeasement or affiliation, and display of this pattern of muscle movements was associated with greater pro-social behaviours within dyads.Three AUs had high loadings on our third factor: Brow Furrow (AU4), Nose Wrinkler (AU9), and Upper Lip Raiser (AU10). Activation of AU4, AU9, and AU10 are characteristically associated with goal-obstruction and anger (e.g., Pope & Smith, 1994). Within IPD gameplay, goal obstruction occurs where the partner defects, thereby removing the opportunity for a mutually beneficial outcome. When moving beyond the scope of anger, AU4 and AU9 activation are associated with incidences of sincere remorse (Baker, 2018; ten Brinke et al., 2012). Remorse is the negative emotional experience which results from violating one's moral standards, such as in lying or deceiving, and remorse displays are theorised to indicate a willingness for social reintegration into the dyad or group (Jack et al., 2016). Moreover, activation of AU9 and AU10 are associated with deception in real life criminal trial video footage (Şen et al., 2022). It is reasonable to suggest that the Factor 3 configuration conveys information relating to the individual's perceptions of their behaviour and the implicit desire to remediate the negative consequences which in turn leads to fewer deceptions. As such the third factor is named Brow and Upper Lip Disapproval (BULD).

This factor is similar to Stratou et al's (2017) 'Eye Tightening' (AU4, AU7 and AU9) factor, which was suggested to reflect participant concentration or, potential, anger. However, Stratou and colleagues (2017) found no evidence that this factor was related to any specific behaviour occurring within their data, and so concentration during the experimental task was proposed as the most likely motivator.

Relation to Prior Research

There was some correspondence between our three-factor solution and the factors identified previously by Stratou and colleagues (2017). For example, our factor 1 (RS) shared AUs with an "enjoyment smile", our Factor 2 (MBA) had some shared AUs with Mouth Tightening and Factor 3 (BULD) had some shared AUs with Eye tightening. It is important to note, however, that the 6-factor model had poor fit in the present analyses. This suggests that the 3-factor solution extracted from our data likely associate with relational processes operating between participants in IPD gameplay. The disparity between the two factor structures is likely due to the greater variation in context available in Stratou and colleagues' (2017) dataset, as our 3-factor structure relates specifically to social dilemma contexts. The reliable patterns of AU activity found in the present study likely reflect meaningful communication of information relevant to the interaction context.

Interestingly, only one of the three configurations we found had any similarity to a prototypical emotion expression. This may reflect the fact that participants deployed context-dependent communicative displays (e.g., Crivelli & Fridlund, 2018) in our IPD games rather than expressing basic emotions. However, it is also possible that participants experienced emotion blends that produced mixtures of the facial patterns emphasised by Ekman and colleagues (e.g., Ekman & Keltner, 1997).

Limitations and Future Directions

A limitation of this work is that factor analyses are sensitive to researcher degrees of freedom which can yield different factor solutions. In the present paper, the factor solution was replicated across three studies, suggesting that it is a robust reflection of patterns of facial muscle movements in IPD gameplay. Thus, while the evidence suggests that the three-factor structure is an appropriate model, conclusions should be reasonably tempered to reflect that the model proposed is only a plausible – although well supported – model. Models derived from alternative statistical pathways (e.g., different rotation methods) may yield other plausible models.

The present analyses provide evidence for consistent patterns of facial expression occurring in an IPD context, some of which relate to meaningful configurations found in previous research. Given the context-dependent nature of these muscle movements during IPD gameplay and the suggestions presented above about how each factor may relate to specific relational contexts, it is important to assess how gameplay actions are associated with the identified patterns of facial activity. In the next section, we report within-pair correlations between factor scores and relations of factor scores with frequencies of the different IPD gameplay outcomes sampled in our dataset in order to check our interpretations of factor meaning.

Analysis Phase 2: Relating Factor Scores to Game Outcomes

In order to establish the role of each of the three identified facial configurations during IPD gameplay, we correlated factor scores for each player with game outcomes experienced by the dyad. We also used regression analyses to assess whether game outcome moderated the association between the two dyad members' factor scores. This approach mirrors that taken by Stratou and colleagues (2017). These analyses serve as a means of checking whether the factors derived from our data have meaningful associations with decisions and behaviours made during IPD gameplay. See Table 4 for descriptive statistics for the contextual variables.

Table 4. Means and standard deviations for gameplay outcomes across all rounds of IPD.

	Dataset 1		Dataset 2		Dataset 3	
	M	SD	M	SD	M	SD
Total Cooperation Choices	7.56	2.72	7.33	2.42	8.01	2.51
Total Defection Choice	2.44	2.72	2.67	2.42	1.99	2.51
Total CC	6.26	3.50	5.88	3.13	6.93	3.39
Total CD	1.30	1.36	1.45	1.27	1.08	1.47
Total DC	1.30	1.36	1.45	1.27	1.08	1.47
Total DD	1.14	2.09	1.22	1.78	0.91	1.57

Note. CC = Mutual cooperation; CD = Actor cooperated, partner defected; DC = Actor defected, partner cooperated; DD = Mutual defection.

A series of Pearson correlations, with alpha set to $p < .005$ (Bonferroni correction), assessed the relationship between expression of the three factors between Actor and Partner, and overall gameplay outcomes (total CC, CD, DC, DD; see Figures 2 and 3 for correlations). We also conducted moderation analyses to assess the influence of round outcome on these associations, with p set at $< .05$. Consistent results across all datasets are discussed below and all significant results are available in both figures and tables, any results not discussed were not significant (see supplementary materials).

Associations between Actor and Partner Factor Scores

Table 5. Correlations between Actor factor scores and Partner factor scores across all IPD rounds.

	Actor RS			Actor MBA			Actor BULD		
	DS1	DS2	DS3	DS1	DS2	DS3	DS1	DS2	DS3
Partner RS	.29	.41	.18	.28	.26				
Partner MBA	.28	.26		.28	.26				
Partner BULD								-.11	-.06

Note. DS1 = Dataset 1; DS2 = Dataset 2; DS3 = Dataset 3; RS = Reward Smile; MBA = Mouth-Based Appeasement; BULD = Brow and Upper Lip Disapproval. Unless otherwise indicated, all r statistics reported were $p < .005$.

When considering the relationship between dyad partners' scores on the same factor, results indicated significant positive associations between actor and partner RS scores across all three datasets, r = .18 to .41, ps < .005. There was no evidence that game outcome (from the actor's perspective, i.e., CD, DC, etc) significantly moderated this relationship in either the first or second data set. In the third dataset, moderation analyses showed this positive relationship was statistically significant only for CC, b = .22, t = 6.06, 95% CI = .15 to .29, p < .001, and CD, b = .27, t = 2.52, 95% CI = .06 to .47, p = .012, outcomes. Similarly, there was a significant positive association between actor and partner MBA scores, r = 26 to .28, in the first and second datasets, but not in the third (p > .05). Finally, the correlation between actor and partner BULD scores was not significant (p > .05) in any of the three datasets.

There were also significant associations between dyad members' scores on different factors. For example, there was a positive association between one dyad member's RS score and the other member's MBA score in both the first and second dataset, r = .26 to .28, p < .005. Moderation analyses indicated that this relationship was statistically significant and positive in these two datasets for the CC outcome, and for the DC outcome where the actor displayed RS (p < .05), and the CD outcome where the actor displayed MBA (p < .05). There was no evidence of a significant association or of significant moderation in the third dataset (p > .05). No other correlations were significant (p > .05).

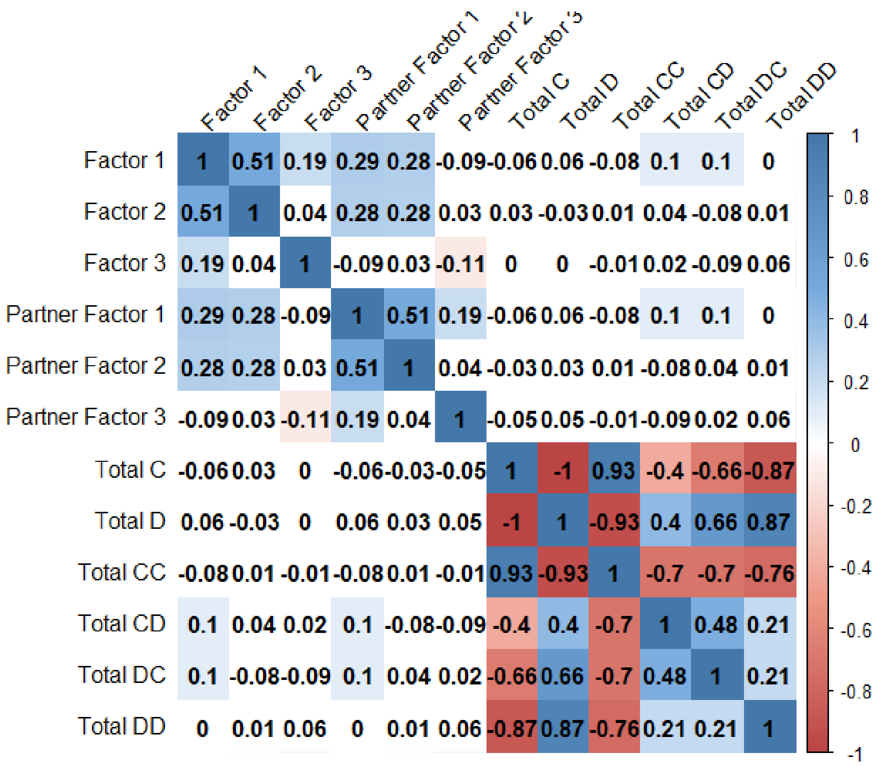

Figure 2. Pearson Correlations between Factors and IPD Outcomes in first dataset with alpha set at $p < .005$; blank regions signify a non-significant result; Factor 1 = Reward Smile; Factor 2 = Mouth Based Appeasement; Factor 3 = Brown and Upper Lip Disapproval.

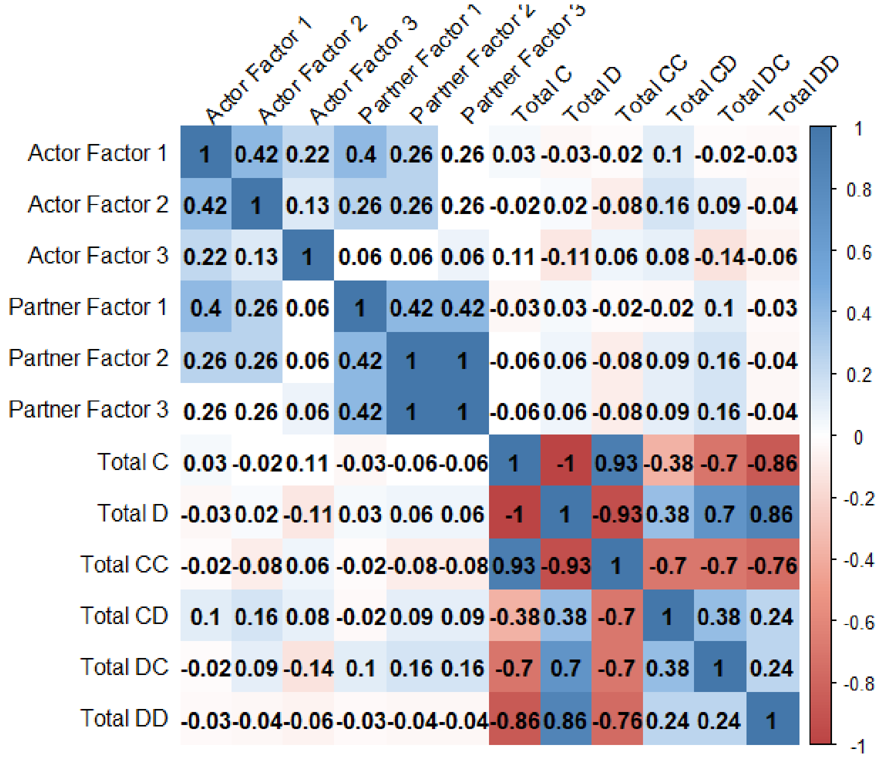

Figure 3. Pearson Correlations between Factors and IPD Outcomes in second dataset with alpha set at $p < .005$; blank regions signify a non-significant result; ; Factor 1 = Reward Smile; Factor 2 = Mouth Based Appeasement; Factor 3 = Brown and Upper Lip Disapproval.

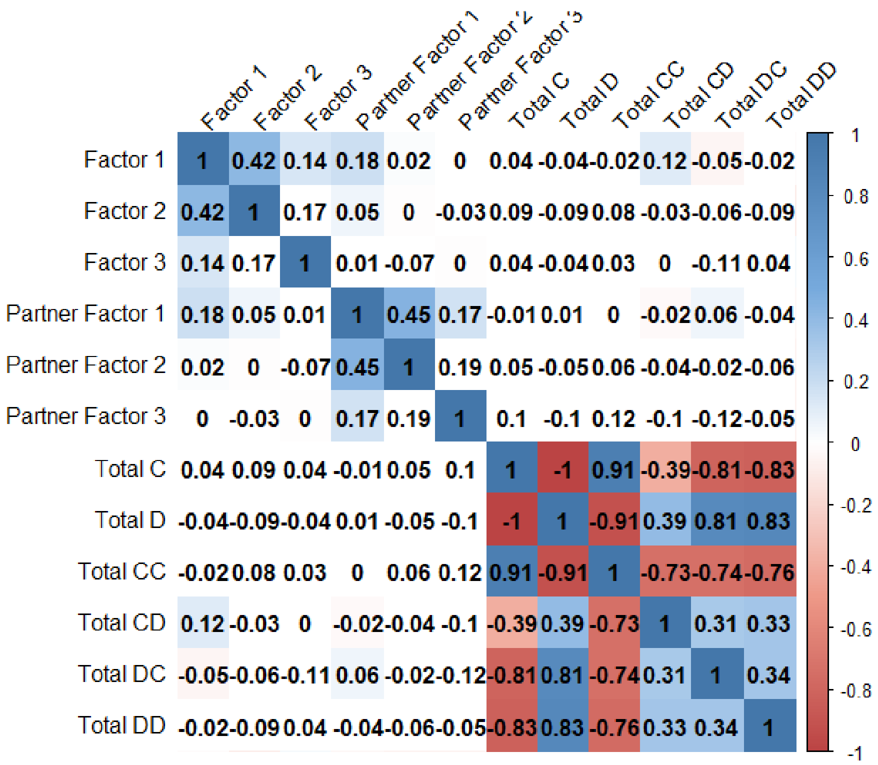

Figure 6. Pearson Correlations between Factors and IPD Outcomes in third dataset with alpha set at p < .005; blank regions signify a non-significant result.

Associations Between Factor Scores and Gameplay Outcome Frequencies

A series of Pearson correlations, with alpha set to p < .005 (Bonferroni correction), assessed the relationships between each participant's scores on the three facial factors and measures of the frequency of each of the four possible game-play outcomes (total CC, CD, DC, DD; see Table 5 for overall gameplay correlations). Each participant was included in the correlational analyses once and only significant results are discussed here (please see supplementary materials for all results).

Table 6. Correlations between Gameplay Outcome Frequencies and Factor Expression Scores across Datasets One to Three

	Total CC			Total CD			Total DC			Total DD		
	RS	MBA	BULD	RS	MBA	BULD	RS	MBA	BULD	RS	MBA	BULD
Dataset 1				.10*			.10*		-.09			
Dataset 2				.10*	.16*			.09*	-.14*			
Dataset 3		.10*		.12*					-.10*		-.13*	.10*

Note. RS = Reward Smile; MBA = Mouth-Based Appeasement; BULD = Brow and Upper Lip Disapproval; * = $p \leq .005$. Unless otherwise indicated, reported correlations are $p \leq .01$.

Across all three datasets, there was a significant positive association between RS scores and total CD outcomes (rs = .10 to .12, ps < .005). There were no significant associations between MBA scores and total gameplay outcomes in any of the three datasets (p > .005). BULD scores were significantly negatively associated with total DC outcomes (rs = -.09 to .14) across all three datasets, but the relationship did not meet the pre-determined threshold of p = .005 in the third dataset (with p only being less than .01).

Discussion

The aim of the second analysis phase was to assess whether the patterns of facial activation we identified are associated across partners and with IPD gameplay outcomes. The correlational analyses showed that, across all three datasets, each factor was associated with different gameplay outcomes and behavioural actions; thus, suggesting that each factor differentially influenced, or was differentially influenced by, variations in gameplay.

Factor Scores and Gameplay Events

As previously noted, Reward Smiles resemble the AU activation thought to be associated with Joy. However, RS was displayed significantly more during CD gameplay outcomes, an outcome which is unlikely to be consistently pleasant. Rather, this and the other identified factors appear to represent facial muscle movements which fulfil social functions during naturalistic dyadic interactions. Thus, the results suggest that RSs are tools that fulfil particular social functions and convey pertinent information related to the specific context (Crivelli & Fridlund, 2018). Within the Simulation of Smiles model (Niedenthal et al., 2010), RSs are defined as reflecting a happy response and/or a method of rewarding other people (Rychlowska, et al., 2017). As the AFFDEX module used in the present analysis does not measure AU25 activation, the AU27 activation may be analogous to movement of AU25. Thus, IPD contexts appear to elicit RSs consistent with empirical evidence and they are associated with game states that include successful defection. Previous evidence also suggests that RSs are unlikely to signal deceptive motives consequently leaving the individual vulnerable to exploitation (Centorrino et al., 2010). That is, RS displays communicate that the actor is open to exploitation. Thus, they are not necessarily specific responses to CD outcomes, but instead may makes CD outcome more likely when they occur more commonly across trials. The reported association between RS and CD is not only not consistent with theoretical accounts, but also replicates positive correlations between morphologically similar Joyful Smiles and betrayal obtained in Stratou et al.'s (2017) study.

When considering Mouth Based Appeasement, a factor assumed to denote affiliation, evidence suggests that affiliative smiles within social dilemmas provide a mechanism through which interpersonal cooperation can be facilitated (Senior et al., 2019). Indeed, the communication of affiliation through facial muscle movements is theorised to convey an individual's positive social motives and to create/maintain social bonds (Orlowska et al., 2018). The present finding that MBA correlated with opponent RS in the CC and CD conditions in the first two datasets may reflect the expresser's desire to promote pro-social outcomes in future rounds (i.e., mutual cooperation).

AUs loading high on the Brow and Upper Lip Disapproval factor, specifically AU 4 and AU 9, are associated with goal obstruction motivated anger. Expressions of anger in response to goal obstruction are suggested to motivate a restoration of equity between individuals (Van Doorn et al., 2014). Indeed, an actor's expression of anger often leads to greater prosocial actions and compensatory behaviours in the partner (Hareli et al., 2009). Within the current study BULD was negatively correlated with total DC outcomes, a finding which may evidence the direct pro-social consequences of displaying this facial configuration. Indeed, BULD displays are assumed to be signals of appeasement or sadness, and often lead to future compliance or pro-sociality (Fehr & Schurtenberger, 2018). As a factor, BULD appears less frequently for outcomes where the actor has betrayed the cooperating partner. Thus, rather than being detrimental to trust and future outcomes, BULD displays may contribute to the restoration of trust and the fostering of positive social interactions in IPD gameplay.

The lack of any significant relationships between MBA scores and frequency of any of the gameplay outcomes may suggest that this facial configuration is used functionally in response to specific interpersonal cues from the other dyad member (as considered below) rather than being directly associated with overall reward payoffs.

Associations between Actor and Partner Factor Scores

Our analyses found evidence consistent with actor-partner reciprocity in both RS and MBA. Reciprocity is a social phenomenon whereby interaction partners trade equivalent behaviours in order to achieve a mutual benefit (Cialdini & Goldstein, 2004). Indeed, both genuine and polite smiles are likely to be reciprocated in face-to-face social interactions (Heerey & Crossley, 2013). Within the context of economic choices, a lack of smile reciprocity is associated with perceptions of reduced interaction and relationship quality; a factor which is highly salient in IPD where both individual and mutual goals are predicated on fostering perceptions of trust in the other player. It is, therefore, unsurprising that the results indicated a relationship between participants' smiles.

We also found some evidence of actor-partner associations across different factors. In particular, there were significant positive correlations between one dyad member's RS displays and the other dyad member's MBA presentation. One possible interpretation is that appeasing displays solicit positive responses or encouragement from the other dyad member. However, further clarification of the association depends on considering the game outcomes that moderated it.

Game Outcomes as Moderators of Actor-Partner Factor Score Associations

Our regression-based analyses found evidence that the significant positive association between actor RS and partner MBA displays in the first and second dataset was found in CC outcomes and CD and DC outcomes when the MBA score applied to the cooperating dyad member. MBA from a cooperative player in response to the other player's defection may clearly be interpreted as a display that is oriented to appeasing that other player in order to reduce the chances of future defection. This in turn may recruit RS from the defector, who may be reassured that there is unlikely to be retaliatory defection in subsequent trials. Similarly, in the context of mutual cooperation RS-MBA associations may reflect a dialogue of appeasement oriented to past or possible future defection by the other dyad member and reassurance and reward from the other dyad member. Understanding this pattern of interpersonal exchange requires more fine-grained dynamic investigation of sequences of dyadic facial behaviour.

Alternatively, the defector's presentation of RS may be seen as a means of soliciting MBA from the cooperative dyad member rather than as a response to it. Such an interpretation first requires us to unpack the reasons why a defector might deliver an RS when their partner has cooperated. One possibility is that the defector's RS serves to minimise any potential negative emotional or social repercussions of defection within the dyad, analogous to 'smiling away' deception (Stratou et al., 2017). Another possibility is that RS following defection reflects what Ekman (1981) refers to as "duping delight", namely the expression of exhilaration, pleasure, or satisfaction experienced during the process of deception. In the present context, this delight may relate to the defector's success in achieving the most successful points outcome at the expense of the other dyad member rather than specifically misleading them. It has largely been assumed that examples of duping delight are unlikely to occur within a laboratory setting (Hess & Kleck, 1990), due to social behaviours expected of participants within these contexts. However, the present research paradigm may provide an appropriate social context in which duping delight may be displayed as the interactions between dyad members are naturalistic and short-lived. Smiling away the negative result by the partner and the actor expressing duping delight are not necessarily mutually exclusive functions of the same facial movements. However, it is also reasonable to suggest that players may be conveying delight related to other gameplay related contexts (e.g., the reward of winning a round), and further research should be conducted to better understand the rationale for this expression.

If defectors' RS smiles reflect duping delight then cooperators' responses are more likely to be appeasing because a satisfied defector needs more persuasion to stop defecting. By contrast, if defectors' RS smiles are oriented to smiling away the interpersonal offence then cooperators may be more inclined to appease than display disapproval or threaten retaliation.

Limitations and Future Directions

Some of the effects discussed above were only present in datasets one and two, and not in the third dataset. As such the inferences should be approached with caution. However, it is possible that the experimental manipulation used on some of the participants contributing to dataset three may have introduced noise which masked or reduced the above effects, thus explaining the lack of significant results. Regardless, further replication work should assess the robustness of the moderation effects found in datasets one and two.

General Discussion

The present research used two EFAs and a CFA to derive a three-factor model of facial configurations produced following the revealed outcomes in IPD games. Across three independent datasets, we found that Factor 1 corresponded to a Reward Smile (Martin, et al., 2017), Factor 2 reflected Mouth-Based Appeasement, and Factor 3 indicated Brow and Upper Lip Disapproval. The combined results from correlational and factor analyses indicated that there are reliable patterns of AU activation in IPD interactions. These patterns were not consistently associated with AU activation thought to be indicative of prototypical emotional expressions. Rather, they seemed to reflect context-specific facial displays which likely serve social and communicative functions (Crivelli & Fridlund, 2018). Indeed, there is evidence to support the functional argument that facial muscle configurations serve to facilitate interpersonal relationships and interactions.

The reported associations between specific factors and gameplay outcomes suggests that the extracted patterns of activity are psychological meaningful within the IPD context. The functions fulfilled by each factor are – like the patterns of AU activation in the factors themselves – distinctly different from one another, and replicate across multiple datasets. Thus, these factors may be used as units of analysis in future research as a means of exploring and understanding facial expressions.

There are limitations to the present analysis, however. While the context-specific nature of the present study allows for analysis of the impact of specific gameplay outcomes on AU activation, it is also a limitation. While the data was collected as part of a naturalistic interaction, the extent to which the 3-factor solution will replicate in other contexts outside of social dilemma decision making remains unclear. Future research could assess whether the 3-factor model accurately captures patterns of AU activation in other social contexts, such as negotiation tasks (e.g., Stratou, et al., 2015), to assess the validity of the model. Similarly, video-mediated interactions may yield slight variations in non-verbal behaviour when compared to face-to-face interactions (Shahid et al., 2012). As such, future research should assess to what extent patterns of facial musculature change differ between video-mediated and face-to-face interactions. Future work should compare the usefulness of the two models in analysis of facial expressions in social interaction.

Finally, the correlational nature of our data means that we are unable to draw strong conclusions about causal effects of presenting each of the identified facial configurations. For example, there remains ambiguity about whether MBA solicits RS in certain game contexts, whether RS encourages MBA, or whether the relationship between these configurations operates bidirectionally. In order to clarify the functionality of these facial displays, it is necessary either to manipulate them experimentally or to analyse the temporal dynamics of the established interpersonal interdependencies using thinner slices of behaviour.

Conclusion

Taken together, our results suggest that there are three distinct patterns of facial muscle activity that relate to context-specific events and behaviours in IPD games. This indicates facial expressions perform psychologically meaningful context related functions. Specifically, Reward Smiles provide reward information; Mouth-Based Appeasement promotes affiliation following actual or anticipated interpersonal conflict; and Brow and Upper Lip Disapproval indicates the actor's disapproval with the status quo and serves to encourage a return to a more equitable outcome.

Further work could validate the functionality of these three expressions, examining expressions elicited by a wider range of social events.

References

Aviezer, H., Hassin, R. R., Bentin, S., & Trope, Y. (2008). Putting facial expressions back in context. In *First impressions*. (pp. 255–286). Guilford Publications. https://doi.org/10.1037/10559-014

Baker, A. T. (2018). *The elusive moral construct : an examination of remorse from a functional perspective*. https://open.library.ubc.ca/collections/24/items/1.0372338

Barrett, L. F., Adolphs, R., Marsella, S., Martinez, A. M., & Pollak, S. D. (2019). Emotional Expressions Reconsidered: Challenges to Inferring Emotion From Human Facial Movements. *Psychological Science in the Public Interest*, *20*(1), 1–68. https://doi.org/10.1177/1529100619832930

Barrett, L. F., & Satpute, A. B. (2019). Historical pitfalls and new directions in the neuroscience of emotion. *Neuroscience Letters*, *693*, 9–18. https://doi.org/10.1016/j.neulet.2017.07.045

Beauducel, A., & Herzberg, P. Y. (2006). On the Performance of Maximum Likelihood Versus Means and Variance Adjusted Weighted Least Squares Estimation in CFA. *Structural Equation Modeling: A Multidisciplinary Journal*, *13*(2), 186–203. https://doi.org/10.1207/s15328007sem1302_2

Bollen, K. A., & Stine, R. A. (1992). Bootstrapping Goodness-of-Fit Measures in Structural Equation Models. *Sociological Methods & Research*, *21*(2), 205–229. https://doi.org/10.1177/0049124192021002004

Calvo, M. G., & Nummenmaa, L. (2016). Perceptual and affective mechanisms in facial expression recognition: An integrative review. *Cognition and Emotion*, *30*(6), 1081–1106. https://doi.org/10.1080/02699931.2015.1049124

Centorrino, S., Djemai, E., Hopfensitz, A., Milinski, M., & Seabright, P. (2010). Honest smiles as a costly signal in social exchange. *Behavioral and Brain Sciences*, *33*(6), 439–439. https://doi.org/DOI: 10.1017/S0140525X10001287

Cialdini, R. B., & Goldstein, N. J. (2004). Social Influence: Compliance and Conformity. *Annual Review of Psychology*, *55*(1), 591–621. https://doi.org/10.1146/annurev.psych.55.090902.142015

Cooperman, A. W., & Waller, N. G. (2022). Heywood you go away! Examining causes, effects, and treatments for Heywood cases in exploratory factor analysis. *Psychological Methods*, *27*(2), 156–176. https://doi.org/10.1037/met0000384

Crivelli, C., & Fridlund, A. J. (2018). Facial Displays Are Tools for Social Influence. *Trends in Cognitive Sciences*, *22*(5), 388–399. https://doi.org/10.1016/j.tics.2018.02.006

Crivelli, C., & Fridlund, A. J. (2019). Inside-Out: From Basic Emotions Theory to the Behavioral Ecology View. *Journal of Nonverbal Behavior*, *43*(2), 161–194. https://doi.org/10.1007/s10919-019-00294-2

de Melo, C., Gratch, J., Carnevale, P. J., & Read, S. J. (2012). Reverse appraisal: The importance of appraisals for the effect of emotion displays on people's decision making in a social dilemma. *Cognitive Science*, *34*.

Durán, J. I., & Fernández-Dols, J.-M. (2021). Do emotions result in their predicted facial expressions? A meta-analysis of studies on the co-occurrence of expression and emotion. *Emotion*, *21*(7), 1550–1569. https://doi.org/10.1037/emo0001015

Ekman, P. (1981). Mistakes When Deceiving. *Annals of the New York Academy of Sciences*, *364*, 269–278. https://doi.org/10.1111/j.1749-6632.1981.tb34479.x

Ekman, P. (1992). Are there basic emotions? *Psychological Review*, *99*(3), 550–553. https://doi.org/10.1037/0033-295X.99.3.550

Ekman, P., & Friesen, W. V. (1978). Facial Action Coding System (FACS). *Environmental Psychology & Nonverbal Behavior*.

Ekman, P., & Keltner, D. (1997). Universal facial expressions of emotion. In U. P. Segerstrale & P. Molnar (Eds.), *Nonverbal communication: Where nature meets culture* (pp. 27–46).

Fehr, E., & Schurtenberger, I. (2018). Normative foundations of human cooperation. *Nature Human Behaviour*, *2*(7), 458–468. https://doi.org/10.1038/s41562-018-0385-5

Fernández-Dols, J.-M., & Crivelli, C. (2013). Emotion and Expression: Naturalistic Studies. *Emotion Review*, *5*(1), 24–29. https://doi.org/10.1177/1754073912457229

Gaspar, A., Esteves, F., & Arriaga, P. (2014). *On Prototypical Facial Expressions Versus Variation in Facial Behavior: What Have We Learned on the "Visibility" of Emotions from Measuring Facial Actions in Humans and Apes* (pp. 101–126). https://doi.org/10.1007/978-3-319-02669-5_6

Hareli, S., Harush, R., Suleiman, R., Cossette, M., Bergeron, S., Lavoie, V., Dugay, G., & Hess, U. (2009). When scowling may be a good thing: The influence of anger expressions on credibility. *European Journal of Social Psychology*, *39*(4), 631–638. https://doi.org/https://doi.org/10.1002/ejsp.573

Hayes, A. (2022). *Introduction to Mediation, Moderation, and Conditional Process Analysis* (3rd ed.). Guilford Press.

Heerey, E. A., & Crossley, H. M. (2013). Predictive and Reactive Mechanisms in Smile Reciprocity. *Psychological Science*, *24*(8), 1446–1455. https://doi.org/10.1177/0956797612472203

Hess, U., & Hareli, S. (2017). The social signal value of emotions: The role of contextual factors in social inferences drawn from emotion display. In J. M. Fernández-Dols & J. A. Russell (Eds.), *The science of facial expression* (pp. 375–393). Oxford University Press.

Hess, U., & Kleck, R. E. (1990). Differentiating emotion elicited and deliberate emotional facial expressions. *European Journal of Social Psychology*, *20*(5), 369–385. https://doi.org/https://doi.org/10.1002/ejsp.2420200502

Hoegen, R., Gratch, J., Parkinson, B., & Shore, D. (2019). Signals of Emotion Regulation in a Social Dilemma: Detection from Face and Context. *2019 8th International Conference on Affective Computing and Intelligent Interaction (ACII)*, 1–7. https://doi.org/10.1109/ACII.2019.8925478

Hoegen, R., Stratou, G., Lucas, G. M., & Gratch, J. (2015). *Comparing Behavior Towards Humans and Virtual Humans in a Social Dilemma* (pp. 452–460). https://doi.org/10.1007/978-3-319-21996-7_48

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1–55. https://doi.org/10.1080/10705519909540118

Hyniewska, S., Sato, W., Kaiser, S., & Pelachaud, C. (2019). Naturalistic Emotion Decoding From Facial Action Sets. *Frontiers in Psychology*, *9*. https://www.frontiersin.org/articles/10.3389/fpsyg.2018.02678

Jack, R. E., Sun, W., Delis, I., Garrod, O. G. B., & Schyns, P. G. (2016). Four not six: Revealing culturally common facial expressions of emotion. *Journal of Experimental Psychology: General*, *145*(6), 708–730. https://doi.org/10.1037/xge0000162

Kaiser, H. F. (1974). An index of factorial simplicity. *Psychometrika*, *39*(1), 31–36. https://doi.org/10.1007/BF02291575

Keltner, D., Sauter, D., Tracy, J., & Cowen, A. (2019). Emotional Expression: Advances in Basic Emotion Theory. *Journal of Nonverbal Behavior*, *43*(2), 133–160. https://doi.org/10.1007/s10919-019-00293-3

Keltner, D., Young, R. C., & Buswell, B. N. (1997). Appeasement in human emotion, social practice, and personality. *Aggressive Behavior*, *23*(5), 359–374. https://doi.org/https://doi.org/10.1002/(SICI)1098-2337(1997)23:5<359::AID-AB5>3.0.CO;2-D

Kulke, L., Feyerabend, D., & Schacht, A. (2020). A Comparison of the Affectiva iMotions Facial Expression Analysis Software With EMG for Identifying Facial Expressions of Emotion. *Frontiers in Psychology*, *11*. https://doi.org/10.3389/fpsyg.2020.00329

Lei, S., Stefanov, K., & Gratch, J. (2020). Emotion or expressivity? An automated analysis of nonverbal perception in a social dilemma. *2020 15th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2020)*, 544–551. https://doi.org/10.1109/FG47880.2020.00123

Littlewort, G., Whitehill, J., Wu, T., Fasel, I., Frank, M., Movellan, J., & Bartlett, M. (2011). The computer expression recognition toolbox (CERT). *2011 IEEE International Conference on Automatic Face & Gesture Recognition (FG)*, 298–305. https://doi.org/10.1109/FG.2011.5771414

Lopes, A. T., de Aguiar, E., De Souza, A. F., & Oliveira-Santos, T. (2017). Facial expression recognition with Convolutional Neural Networks: Coping with few data and the training sample order. *Pattern Recognition*, *61*, 610–628. https://doi.org/https://doi.org/10.1016/j.patcog.2016.07.026

Martin, J., Rychlowska, M., Wood, A., & Niedenthal, P. (2017). Smiles as Multipurpose Social Signals. *Trends in Cognitive Sciences*, *21*(11), 864–877. https://doi.org/10.1016/j.tics.2017.08.007

Morais, R. (2022). *Facial Expression Recognition* (https://github.com/rondinellimorais/facial-expression-recognition/tree/master#readme).

Niedenthal, P. M., 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*(6), 417–433. https://doi.org/DOI: 10.1017/S0140525X10000865

Orlowska, A. B., Krumhuber, E. G., Rychlowska, M., & Szarota, P. (2018). Dynamics Matter: Recognition of Reward, Affiliative, and Dominance Smiles From Dynamic vs. Static Displays. *Frontiers in Psychology*, *9*. https://www.frontiersin.org/articles/10.3389/fpsyg.2018.00938

Parkinson, B. (2013). Contextualizing Facial Activity. *Emotion Review*, *5*(1), 97–103. https://doi.org/10.1177/1754073912457230

Pope, L. K., & Smith, C. A. (1994). On the distinct meanings of smiles and frowns. *Cognition and Emotion*, *8*(1), 65–72. https://doi.org/10.1080/02699939408408929

Poundstone, W. (1993). *Prisoner's dilemma: John von Neumann, game theory, and the puzzle of the bomb*. Anchor.

Reisenzein, R., Studtmann, M., & Horstmann, G. (2013). Coherence between Emotion and Facial Expression: Evidence from Laboratory Experiments. *Emotion Review*, *5*(1), 16–23. https://doi.org/10.1177/1754073912457228

Rinn, W. E. (1984). The neuropsychology of facial expression: A review of the neurological and psychological mechanisms for producing facial expressions. *Psychological Bulletin*, *95*(1), 52–77. https://doi.org/10.1037/0033-2909.95.1.52

Rychlowska, M., Jack, R. E., Garrod, O. G. B., Schyns, P. G., Martin, J. D., & Niedenthal, P. M. (2017). Functional Smiles: Tools for Love, Sympathy, and War. *Psychological Science*, *28*(9), 1259–1270. https://doi.org/10.1177/0956797617706082

Scarantino, A. (2017). How to Do Things with Emotional Expressions: The Theory of Affective Pragmatics. *Psychological Inquiry*, *28*(2–3), 165–185. https://doi.org/10.1080/1047840X.2017.1328951

Scherer, K. R., Clark-Polner, E., & Mortillaro, M. (2011). In the eye of the beholder? Universality and cultural specificity in the expression and perception of emotion. *International Journal of Psychology*, *46*(6), 401–435. https://doi.org/10.1080/00207594.2011.626049

Senior, C., Ridout, N., & Stewart, P. A. (2019). Smiles of Affiliation and Dominance when competing for Leadership: The effect of gender in the 2016 US Presidential Debates. *PsyArXiv*.

Şen, M. U., Pérez-Rosas, V., Yanikoglu, B., Abouelenien, M., Burzo, M., & Mihalcea, R. (2022). Multimodal Deception Detection Using Real-Life Trial Data. *IEEE Transactions on Affective Computing*, *13*(1), 306–319. https://doi.org/10.1109/TAFFC.2020.3015684

Shahid, S., Krahmer, E., & Swerts, M. (2012). Video-mediated and co-present gameplay: Effects of mutual gaze on game experience, expressiveness and perceived social presence. *Interacting with Computers*, *24*(4), 292–305. https://doi.org/10.1016/j.intcom.2012.04.006

Siegel, E. H., Sands, M. K., den Noortgate, W., Condon, P., Chang, Y., Dy, J., Quigley, K. S., & Barrett, L. F. (2018). Emotion fingerprints or emotion populations? A meta-analytic investigation of autonomic features of emotion categories. *Psychological Bulletin*, *144*(4), 343.

Stevens, J. (1992). Applied multivariate statistics for the social sciences, 2nd ed. In *Applied multivariate statistics for the social sciences, 2nd ed.* Lawrence Erlbaum Associates, Inc.

Stratou, G., Van Der Schalk, J., Hoegen, R., & Gratch, J. (2017). Refactoring facial expressions: An automatic analysis of natural occurring facial expressions in iterative social dilemma. *2017 Seventh International Conference on Affective Computing and Intelligent Interaction (ACII)*, 427–433. https://doi.org/10.1109/ACII.2017.8273635

ten Brinke, L., Porter, S., & Baker, A. (2012). Darwin the detective: Observable facial muscle contractions reveal emotional high-stakes lies. *Evolution and Human Behavior*, *33*(4), 411–416. https://doi.org/10.1016/j.evolhumbehav.2011.12.003

Thompson, B. (2004). Exploratory and confirmatory factor analysis: Understanding concepts and applications. In *Exploratory and confirmatory factor analysis: Understanding concepts and applications.* American Psychological Association. https://doi.org/10.1037/10694-000

Tian, Y., Kanade, T., & Cohn, J. F. (2000). Recognizing lower face action units for facial expression analysis. *Proceedings Fourth IEEE International Conference on Automatic Face and Gesture Recognition (Cat. No. PR00580)*, 484–490. https://doi.org/10.1109/AFGR.2000.840678

Walker, D. A., & Smith, T. J. (2017). Computing Robust, Bootstrap-Adjusted Fit Indices for Use With Nonnormal Data. *Measurement and Evaluation in Counseling and Development*, *50*(1–2), 131–137. https://doi.org/10.1080/07481756.2017.1326748

Zhou, F., Kong, S., Fowlkes, C. C., Chen, T., & Lei, B. (2020). Fine-grained facial expression analysis using dimensional emotion model. *Neurocomputing*, *392*, 38–49. https://doi.org/https://doi.org/10.1016/j.neucom.2020.01.067