British Journal of Psychology (2012)
© 2012 The British Psychological Society

The British Psychological Society

www.wileyonlinelibrary.com

The telltale face: Possible mechanisms behind defector and cooperator recognition revealed by emotional facial expression metrics

Zsófia Kovács-Bálint[1], Tamás Bereczkei[2] and István Hernádi[1]*

[1]Institute of Biology, University of Pécs, Hungary
[2]Institute of Psychology, University of Pécs, Hungary

In this study, we investigated the role of facial cues in cooperator and defector recognition. First, a face image database was constructed from pairs of full face portraits of target subjects taken at the moment of decision-making in a prisoner's dilemma game (PDG) and in a preceding neutral task. Image pairs with no deficiencies ($n = 67$) were standardized for orientation and luminance. Then, confidence in defector and cooperator recognition was tested with image rating in a different group of lay judges ($n = 62$). Results indicate that (1) defectors were better recognized (58% vs. 47%), (2) they looked different from cooperators ($p < .01$), (3) males but not females evaluated the images with a relative bias towards the cooperator category ($p < .01$), and (4) females were more confident in detecting defectors ($p < .05$). According to facial microexpression analysis, defection was strongly linked with depressed lower lips and less opened eyes. Significant correlation was found between the intensity of micromimics and the rating of images in the cooperator-defector dimension. In summary, facial expressions can be considered as reliable indicators of momentary social dispositions in the PDG. Females may exhibit an evolutionary-based overestimation bias to detecting social visual cues of the defector face.

Social cooperation among strangers with limited information about reliability is a crucial question in psychology. Studies in evolutionary psychology suggest that specific cognitive mechanisms have been selected for enabling cooperators to avoid exploitation from defectors (Cosmides, Barrett, & Tooby, 2010; Cosmides & Tooby, 1992). Humans may possess a cognitive adaptation to recognize defectors using various psychological mechanisms.

Yamagishi, Tanida, Mashima, Shimoma, and Kanazawa (2003) used a one-shot prisoner's dilemma game (PDG) to demonstrate that humans remember faces of defectors better than those of cooperators, without explicit designation of the faces as defectors or cooperators. However, subjects also (erroneously) believed that they had seen the defectors before when they actually had not. Some of these results were confirmed by

*Correspondence should be addressed to István Hernádi, Department of Experimental Neurobiology, Institute of Biology, University of Pécs, 6 ifjusag str. Pecs H 7624, Hungary (e-mail: hernadi@ttk.pte.hu).

DOI:10.1111/bjop.12007

Verplaetse, Vanneste, and Braeckman (2007), who added that participants could accurately discriminate non-cooperative pictures from cooperative ones even during the decision-making phase of the PDG. A related experiment suggested that an automatic attention bias for threatening social interactions involving untrustworthy partners may help us to identify non-cooperative partners (Vanneste, Verplaetse, Van Hiel, & Braeckman, 2007). The authors argued that these findings do not support old-fashioned physiognomy; faces may not be untrustworthy themselves, but may express strong emotions (fear, guilt, anger) at the moment of cheating that would grab the attention of the observers.

Other studies, using different methods, suggest that people can infer personality traits, including trustworthiness, from another's face or behaviour. Borkenau, Mauer, Riemann, Spinath, and Angleitner (2004) found that we can access personality traits and intelligence on the basis of thin slices of behaviour. Oosterhof and Todorov (2008) developed a 2D model for face evaluation over two identified dimensions: valence (trustworthiness) and dominance. They found that perception of trustworthiness is related to emotional facial expressions, such as anger and happiness. These findings support that trustworthiness of the face modulates the intensity of the perceived emotions – the more trustworthy a face is, the happier it seems (Oosterhof & Todorov, 2009). Several authors (Todorov, Pakrashi, & Oosterhof, 2009; Willis & Todorov, 2006) demonstrated that we need no more than approximately 100 ms stimulus exposure duration to make a final judgment about the trustworthiness of a face. Others found that cooperators displayed a lot more emotional expressions when responding to unfair offers in the ultimatum game compared with non-cooperators, and these facial expressions included both positive and negative emotional reactions (Schug, Matsumoto, Horita, Yamagishi, & Bonnet, 2010). Furthermore, individuals were found to be able to estimate the level of altruism of unfamiliar target persons on the basis of watching a 20-s long silent video clip that was completely unrelated to altruistic behaviour, suggesting that individuals can identify permanent altruistic traits in others (Fetchenhauer, Groothuis, & Pradel, 2010).

The above experiments imply that there may be certain emotional features of the face that we can utilize to rapidly identify somebody as 'cooperator' or 'defector' in social exchange situations. However, these studies have not identified the specific facial traits that are involved in the self-betrayal of defectors. Other studies demonstrated the role of emotional facial expressions in various social relationships focusing on cooperative situations rather than on those associated with deception. In a different experiment, perceivers made assessments of video clips depicting self-reported altruists and self-reported non-altruists (Brown, Palameta, & Moore, 2003). During a self-presentation, altruists produced significantly more felt, shorter, and symmetrical smiles, more concern furrows, and more head nods than non-altruists. Several of these non-verbal behaviours may be difficult to fake because they are mostly automatically controlled (e.g., the activation of facial muscles, especially the m. orbicularis oculi). In a subsequent study, in which authors repeated and improved the method of Brown's previous experiment, felt smile was also found as a basic cue for detecting altruists (Oda, Yamagata, Yabiku, & Matsumoto-Oda, 2009). Mehu, Grammer, and Dunbar (2007) also found that Duchenne (spontaneous, felt) smiles were displayed at higher rates in the sharing situation as opposed to the control situation, and were positively correlated with inclination to altruism measured by questionnaires. Another study showed that smiling photographs, by conveying information about positive emotions, can elicit cooperation from strangers in a trust game (Scharlemann, Eckel, Kacelnik, & Wilson, 2001).

In this study, therefore, we aimed to identify particular changes in such facial expressions during the decision-making process in a PDG. We also aimed to describe those facial areas ('action units', see later) which may signal cooperation or defection. In accordance with the former results, we assume that people are more likely to identify the faces of unknown defectors than those of unknown altruists. Based on other studies (Vanneste *et al.*, 2007; Verplaetse *et al.*, 2007), we assume that the anticipation of winning or losing associated with cheating in the social dilemma task cause specific emotional reactions in the defector which may, in turn, be translated into facial expressions. The small, transient but identifiable changes in the facial expressions are expected, primarily, to serve as basic features for the detection of defectors. In the light of previous studies focusing on facial expressions of negative emotions (Ekman & O'Sullivan, 2006; Schug *et al.*, 2010), we expect that observing certain facial areas, especially around the mouth and the eyes, will enhance the recognition of defectors.

However, the production and detection of facial expressions related to deceptive behaviour should be considered as separate processes. Why would defectors reveal their emotions if these emotions would betray them to their partners and group members? The question will rise particularly sharply when altruism and defection are compared. Altruists reliably signal their altruistic dispositions that may enable them to choose each other for mutual cooperation. Defectors are not expected to send honest signals about their deceptive inclinations since such signals are likely to increase the chance of being recognized and avoided by others. Many studies have found that cooperative and altruistic individuals display a high level of positive emotions, such as Duchenne smiles involving the movements of the m. orbicularis oculi (a facial muscle surrounding the eyes) that is difficult to intentionally control (Ekman & Friesen, 1982; Mehu *et al.*, 2007). In contrast, deception may be rather linked to the expression of negative emotions (fear, sadness, anger, guilt) involving the operation of other muscles.

We assume, however, that under certain circumstances, the emotions underlying the act of deception may be openly signalled. Evidence suggests that many facial expressions of negative emotions are difficult to fake because they are involuntarily controlled (Schug *et al.*, 2010). Furthermore, in anonymous conditions when deception cannot be observed by the partner – for example computerized PDGs provide such conditions – defectors are not necessarily forced to conceal their feelings when making decisions on social relationships.

We hypothesize then, that individuals playing an experimental game may openly express their emotions that accompany their deceptive act because (1) these expressions are difficult to intentionally suppress (and may be identifiable as microexpressions or micromimics) and (2) because they make decisions in an anonymous situation.

We also hypothesize that women can recognize defectors' faces more accurately than men. In the light of previous empirical data (Geary, 2006), the reason for this difference may be that females have an advantage for processing and interpreting non-verbal behaviour and facial expression, and for displaying emotions. On the other hand, they are more likely than men to show sensitivity to potential defectors because of the higher cost they are imposed on in exploitative situations (Geary, 1998).

FACE IMAGE DATABASE DEVELOPMENT

Method

A custom-built computer program was developed and used for recording the decision of participants in the PDG and to take their full face photographs, similarly to procedures of previous investigations (Verplaetse *et al.*, 2007). The most important features of the software were (1) to display the PDG; (2) to enable taking photos at the moment of decision with high temporal resolution (in the first 100 ms after the subjects indicate their decision by pressing a mouse button).

The protocol of the whole study was approved by the local University Ethical Committee and conformed to international standards. We asked all participants to give a written approval after the subject matter of the given study phase had been fully explained. Participants received partial credits towards their class grades for participation in the study. For the construction of the face image database, target subjects played the computer-mediated PDG. They were seated in front of a computer screen which displayed all the instructions. They played two games: (1) decision in a control (neutral) dilemma (whether to buy a new pet) and (2) decision in a one-shot classical PDG (whether to cooperate with a fictional partner in a hypothetical game where 'cooperating' means holding out on the police after a bank robbery and 'non-cooperating' means confessing to the police). After the PDG session, participants were asked a series of questions regarding how they understood the story. They concordantly considered non-cooperation as a kind of defection (cheating) because the defector receives the benefit but he/she fails to reciprocate.

One hundred and sixteen subjects participated in this phase of the study; 90 of them were females (34 defectors, 56 cooperators), and 26 of them were males (11 defectors, 15 cooperators). At the very moment of their decision (within a 100 ms time frame), both in the neutral and in the PDG game, a web camera photo of their full face (a neutral photo and an 'action' photo, respectively) was saved on the computer for further use in this study. Subjects were informed about the exact moment of photographing after the PDG was completed. They all agreed to let their photos be used in further phases of the study. In conjunction with the PDG, subjects were not otherwise rewarded for their instantaneous decisions (cooperate vs. defect).

Results

We selected 67 from the initial 116 pairs of photos that were free from photographic deficiencies (faces that were partly covered by hair, garments, eyeglasses, or hands were excluded). In addition, all subjects had closed lips on the selected photos. The mean (± *SE*) age of the cooperators ($n = 36$, 27 females) was 21.27 (± 0.36) years, whereas the mean age of the defectors ($n = 31$, 21 females) was 21.51 (± 0.44) years.

Selected photos were cropped within an oval frame and the background was masked with a black ellipse [the procedure was based on the study of Gronenschild, Smeets, Vuurman, van Boxtel, and Jolles (2009)]. Luminance and RGB values were adjusted to the average to standardize the photos for use as images. Figure 1 depicts the procedure of image standardization.

Then, a pixel-based image analysis was performed to ensure that images were the same in size, colour and luminance. The result of statistical comparisons between defector and cooperator images (independent-sample *t*-test for each corresponding pixel) did not reveal any difference between these two decision categories. After the above procedure of standardization, we used the images as face stimuli in further phases of the study.

IMAGE RATING

Method

In this phase, our aim was to evaluate the previously developed face database of standardized cooperator and defector images which had been taken at the moment of

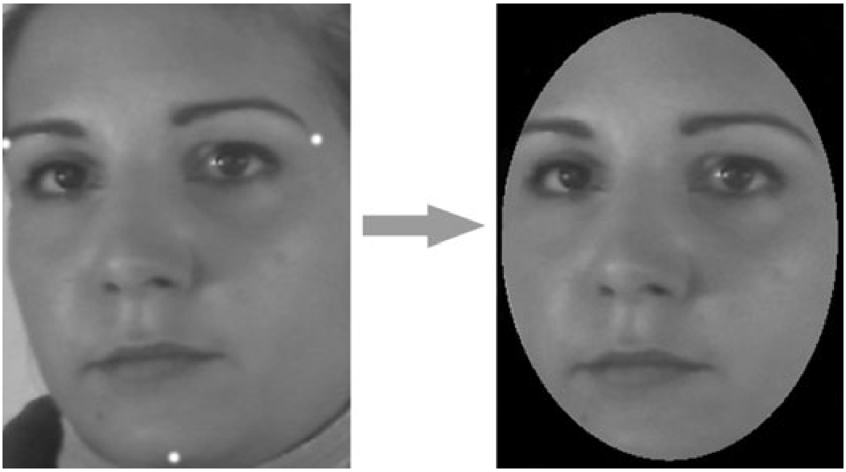

Figure 1. The procedure of stimulus standardization. Left: Example of original picture with indication of corner positions for eyes and cheek. Through these marker corner positions (indicated by small white spots), an ellipse was constructed. The ellipse was then adjusted to a common final axis size (240 × 320 pixels). Right: The final image was masked with a black background outside the ellipse.

decision-making in a computer-mediated PDG (see paragraph 'Face image database development'). The evaluation comprised of testing whether defector faces were identifiable as defectors, and cooperator faces as cooperators. Sixty-two healthy university students (33 females) were aged 22.05 ± 0.45 years volunteered to participate in this phase of the study as lay judges. They evaluated face stimuli ('action photos') along the intensity dimension on an 11-point Likert scale. Their task was to decide whether the person shown in the image could have been a *possible cooperator* (1–5), *a defector* (−1 to −5), or *none of those* (0). The higher the detected saliency of an image was along the cooperative-defective 'facial expression' dimension, the more positively or negatively it had to be scored. Each image was presented until a decision has been recorded. Then, in a 2-s inter-trial interval (ITI), non-figurative pictures (e.g., fractals) were presented to mask the previous facial image. An equal ratio of cooperator/defector images (36 and 31 images, respectively) was spontaneously scored by the lay judges, previously not knowing how many images they would have to score in each category. At the end of the procedure, judges were routinely asked to report familiar facial images – these images were excluded from further analysis.

Average score ± *SE* for each facial image was computed. From the pool of 67 images, true positively identified images were selected for further analysis (i.e., those cooperator images which were recognized as cooperators and defector images which were recognized as defectors). The criteria for true-positive identification were defined so that the average score of a given image had to fall in the category of the image (i.e., positive average score for cooperator images, and negative average score for defector images).

First, data were subjected to repeated measures analyses of variances (rANOVAs) to compare group effects for TARGET TYPE (the category of the image: defector vs. cooperator), TARGET GENDER (the gender of the target subject shown in the image: male vs. female), and JUDGE GENDER (the gender of the lay judge: male vs. female).

For the analysis of gender differences within the intensity dimension of image rating, absolute values of data were computed for true positively identified images. Then, rANOVA was computed to test main effects for TARGET TYPE (the category of the image: defector vs. cooperator) and JUDGE GENDER (the gender of the lay judge: male vs. female).

In all statistical analyses, where main effects or the interactions between variables were statistically significant, *post-hoc* Fisher LSD tests were performed to determine pairwise group differences.

Results

Frequency rate

No images were eliminated from the database because of familiarity: none of the lay judges could reportedly identify any of the presented facial images. We tested the image database for recognition bias and found that lay judges rated 30/67 images (44.8%) as cooperators and 33/67 images (49.3%) as defectors, which means that the image database was judged as equally balanced.

Identification rate

We found that 17/36 (47.2%) of the cooperator images and 18/31 (58.1%) of the defector images were correctly (true positively) identified.

True positively identified images (n = 35) were further analysed by rANOVAs where group effects revealed that cooperator images were scored significantly higher than defector images indicating the clear difference between the two categories (average ± SE for cooperator images: 1.26 ± 0.3; defector images: −0.91 ± 0.14; main effect of TARGET TYPE: $F_{1,\ 31} = 43.82, p < .01$; Fig. 2), with no effect of the gender of the target images (average ± SE for male target images: 0.25 ± 0.3, females: 0.1 ± 0.12; TARGET GENDER main effect: $F_{1,\ 31} = 0.2, p =$ N.S.).

However, lay judges showed gender-related difference in evaluating target images: males evaluated the images with a relative bias towards the cooperator category compared with females who did not show such bias (grand average score of male lay judges: 0.52 ± 0.19; females: 0.18 ± 0.22; main effect of JUDGE GENDER: $F_{1,\ 31} = 7.58, p < .01$).

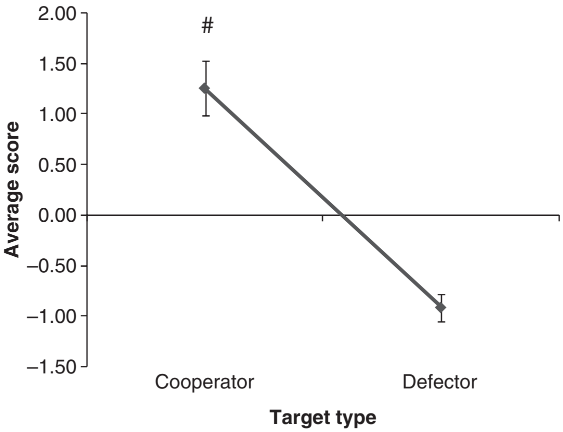

Figure 2. Grand average scores (± SE) for true positively identified PDG-related face images. Significant main effect of TARGET TYPE revealed that cooperators were rated with higher scores compared with defectors. (#:$p < .01$).

Furthermore, analysis of intensity scores revealed a significant interaction between JUDGE GENDER × TARGET TYPE ($F_{1,\ 31} = 10.79, p < .01$). Results of *post-hoc* individual comparison (Fisher LSD test) indicated that females gave more intensive scores for defector faces compared with males (average intensity scores for defector faces: females: 1.2 ± 0.15; males: 0.84 ± 0.14; $p < .05$; Fig. 3).

FACIAL EXPRESSION CODING

Method

Based on earlier studies (Verplaetse *et al.*, 2007; Yamagishi *et al.*, 2003), we hypothesized that we can recognize defectors due to their facial expressions. To test our hypothesis, three independent coders compared 'action' facial images (defector or cooperator) with their neutral counterparts (N.B., one neutral and one 'action' photo was taken of each player as described in the section 'Face image database development') using the 'Facial action coding system' (FACS) developed by Ekman and Friesen (1976). Coders were university academics, who were trained by FACS experts. They evaluated the entire PDG face database (67 image pairs of which 35 pairs were true positively identified) for facial expression metrics along 27 facial areas, termed as action units (AUs), all of which represented contraction or relaxation of one or more facial or neck muscles. Coders compared the individual AUs on each pair of 'neutral' and 'action' photos on a −5 to +5 scale using positive values for increased muscle tension, negative values for decreased muscle tension and zero for no change. Coders were only instructed to compare two photos of each individual and were not aware of any background information related to the purpose of the study.

First, to test for inter-rater reliability we computed the rate of total agreement between coders on a given AU across the images, that is, the number of images rated with three

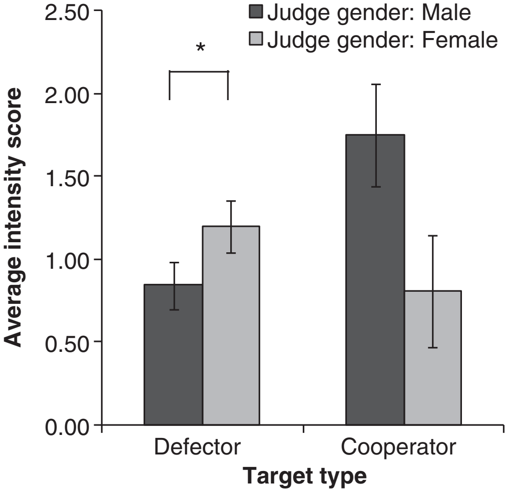

Figure 3. Average (± *SE*) intensity scores for true positively identified cooperator and defector images. *Post-hoc* comparison revealed that female lay judges gave higher intensity scores for defector images compared with males. (*:*p* < .05).

identical scores (*increased* [+] or *decreased* [−] or *no change* [0] of muscle tension) on a given AU divided by the number of all rated images on that AU.

Then, mean difference scores ± *SE* were computed for each AU, then AUs (n = 27) were divided into four major groups as regions of interests namely: lower face up-down movement AUs or LFUD (n = 8; AU4, AU9, AU10, AU11, AU12, AU13, AU14, AU17), lower face lip rounder AUs or LFLR (n = 6; AU18, AU22, AU25, AU26, AU27, AU28), lower face lip stretcher AUs or LFLS (n = 5; AU15, AU16, AU20, AU23, AU24), and upper face and eye movement AUs or UFEM (n = 8; AU1, AU2, AU5, AU6, AU7, AU43, AU45, AU46). Data of each AU group (LFUD, LFLR, LFLS or UFEM) versus image type (TARGET TYPE [cooperators, defectors]) were analysed separately by rANOVAs. Finally, to reveal possible relation between the intensity of micromimics and the confidence of defector/cooperator recognition of the faces, we analysed the correlation between the scores of FACS evaluation (for each relevant AU) and the cooperator or defector intensity ratings separately for true positively or incorrectly identified images. After normal distribution of data was verified (Shapiro–Wilk test), average scores versus individual AUs of the LFLS and UFEM regions were subjected to Pearson correlations. Only those AU groups (namely LFLS and UFEM) were analysed which were found to be relevant according to the analysis of FACS data. Images which were judged as ‘unchanged’ by the coders (i.e., received zero scores for all 13 relevant AUs) were eliminated from further analysis (n = 6, 4 cooperators).

Results

Inter-rater reliability analysis

Inter-rater reliability analysis revealed a good agreement between coders across the individual AUs (0.85 ± 0.02, with a range between 0.59 and 1.00) indicating a high overall reliability of the results.

Analysis of micromimics by FACS

Analysis of defector and cooperator micromimics revealed that the defectors’ lips became more stretched and tightened compared with cooperators at the moment of decision-making in the PDG (rANOVA, significant main effect of the TARGET TYPE for the LFLS group of AUs, $F_{1,64} = 6.015, p < .05$; Fig. 5A). *Post-hoc* individual comparisons (Fisher LSD test) revealed significant differences between cooperator and defector facial movements along AU 20 (average scores for cooperators: 0.000 ± 0.02, defectors: 0.102 ± 0.03; $p < .05$), and a tendency along AU23 (average scores for cooperators: −0.028 ± 0.01, defectors: 0.054 ± 0.02; $p = .07$; Figs. 4 and 5A). In addition, for the upper face region, rANOVA showed significant UFEM × TARGET TYPE interaction ($F_{6,384} = 7.15, p < .01$). *Post-hoc* individual comparisons revealed strong difference between the average scores (± *SE*) of cooperators and defectors along AU5 (cooperators: 0.343 ± 0.05, defectors: 0.032 ± 0.02, $p < .01$), AU7 (cooperators: 0.056 ± 0.03, defectors: 0.193 ± 0.03, $p < .05$), and AU43 (cooperators: −0.074 ± 0.03, defectors: 0.086 ± 0.03, $p < .01$), indicating that cooperators opened their eyes more widely (Figs. 4 and 5B).

Correlation analysis

For true positively identified images, a weak negative correlation was found between image rating and FACS scores in the LFLS region (AU16: $r = -.33, p = .09$), suggesting that the more the target subjects' lips were depressed (according to the FACS analysis), the more defectors they seemed (according to the image rating). Similarly, for the UFEM region, correlation coefficients showed a significant positive relationship between image rating and FACS scores in AU5 ($r = .49, p < .01$), and a tendency of negative correlation in AU7 ($r = -.33, p = .08$), both suggesting that the more the target subjects opened their eyes, the more positively they were rated along the cooperator/defector dimension (i.e., they were more confidently judged as cooperators).

However, similar analysis of incorrectly identified cooperator images revealed strong negative correlation between image rating and FACS scores for the UFEM region (AU43: $r = -.46, p < .05$), suggesting that there was a tendency to make a mistake in identifying those cooperators who happened to lower their eyelids when the photo was taken: the more the eyelids were lowered, the more defector they seemed.

Discussion

In this study, we developed and standardized an action-based image battery based on cooperator and defector full face photos that were taken at the moment of decision in an experimental situation using computer-mediated PDG for constructing a PDG face database. By evaluating the database in an image-rating task, we found that (1) scores for true positively identified cooperator images differed from the scores for true positively identified defector images; and that (2) males but not females were biased in identifying images towards the cooperator dimension. By coding these cooperator and defector images using FACS, we demonstrated that, at the moment of their decision-making, the defectors' facial expression differed from that of cooperators. Defectors but not cooperators closed their upper eyelids as if they were blinking and also depressed and

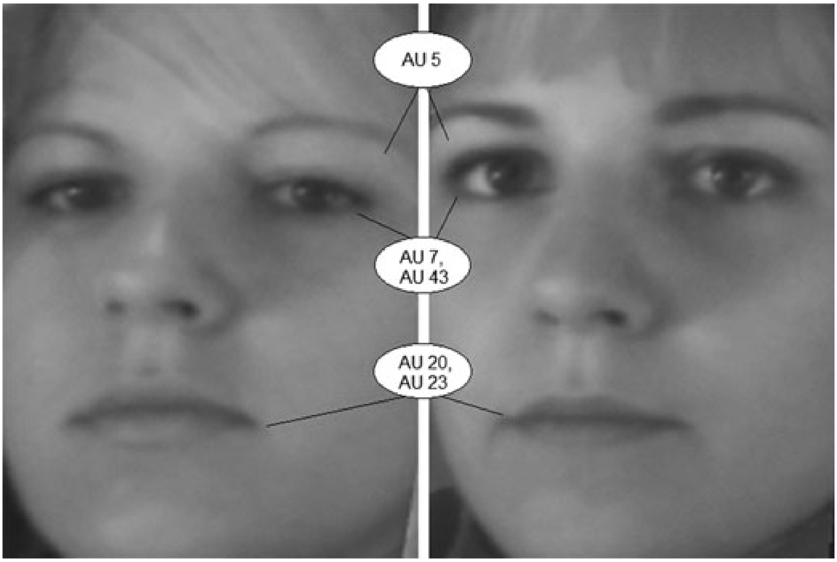

Figure 4. Typical cooperator (right) and defector (left) facial expression. For anonymity reasons, images were averaged from individual photos by morphing technique. Bubbles indicate the approximate positions of the characteristic action units (AUs) corresponding to the facial action coding system (FACS) by Ekman and Friesen (1976).

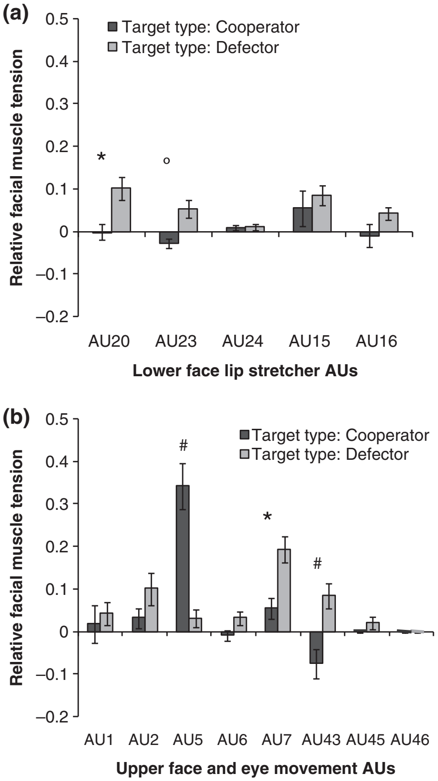

Figure 5. Results of the FACS analysis. Bar charts represent relative differences (± *SE*) of defector and cooperator facial movements. (A) Lower face lip stretcher action units (LFLS AUs). Significant group difference in AU20 (lip stretcher) and tendency in AU23 (lip tightener) represent that defectors tighten their lips and pull their lip corners laterally as compared with cooperators. (B) Upper face and eye movement action units (UFEM AUs). Significant difference in AU5 (upper eyelid raiser), AU7 (lid tightener) and AU43 (eye closure) revealed that cooperators open their eyes more widely than defectors. (#:*p* < .01, *:*p* < .05, °:*p* < .1).

tightened their (lower) lips. In addition, significant correlations between micromimic changes and recognition rates indicated that the more intense the microexpression changes of the faces were, the more confidently the faces were identified by the lay judges. In addition, those cooperators, who tended to close their eyes (as real defectors did), were mistakenly identified as defectors.

As the image-rating method and the FACS are both commonly used in studies concerning emotional facial expression (Melfsen, Osterlow, & Florin, 2000; Tracy, Robins, & Schriber, 2009), we believe that the present evaluation provides reliable combination of subjective rating with objective facial metrics data concerning event related, transient emotional expression changes at the moment of the subjects' decision on cooperation or deception.

In general, the function of cheater detection is to defend the cooperator against exploitation (Cosmides *et al.*, 2010). A serious question may arise about the possible evolutionary reason for the defectors' signalling their intention. Why do defectors disclose their intentions by recognizable facial expressions? A related question would concern the possible costs associated with the development and production of such facial patterns. Are they under automatic control and, as a consequence, function as a reliable indicator of deceptive dispositions?

As we described in the Introduction, cooperators may produce facial expressions that enable observers to differentiate them from defectors and they would have been selected for emitting the right facial signals whereas defectors might not have been selected against withholding their facial expressions. In general, the evolution of reliable signalling has been subjected to an arms race between the signaller and the receiver (Brown *et al.*, 2003; Trivers, 1985). This arms race must be asymmetrical, given that the cost of being exploited is greater to an altruist than the cost to a defector who fails to exploit another individual. It is possible that this asymmetry has selected for an improvement in the receiver's cheater detection skills, whereas the quality of the signals emitted by the senders remained basically unchanged (Verplaetse *et al.*, 2007). Evolution retains the subtlety and complexity of the signalling system but selects for a more sophisticated recognition of defectors. They may not control the expressions of their emotions that accompany their deceptive act, but the facial signals they send may alert the receivers to avoid being defected.

Former studies have shown that pictures of non-cooperative players attract more automatic attention than pictures of cooperative individuals (Vanneste *et al.*, 2007). At the moment of their decisions, defectors might feel strong emotional reactions (fear, anger, guilt) which may, in turn, be translated into subtle facial expressions. Because of these automatically developing strong emotions, the resulting facial expressions may not be under voluntary control. Ekman (1985) reasoned that under high-stake deception situation people are likely to feel strong emotions when lying. It was found that the ability to accurately detect lying is related to the ability to accurately recognize micromomentary facial expressions of emotion (Frank & Ekman, 1997). Based on the present data, we can speculate that a social dilemma situation can evoke automatic emotional expressions that can often disclose the action of deception. The strength of those facial expressions may depend on the risk that is involved in the deceptive decision (Ekman, 1985).

At the moment, we do not know whether these facial cues are costly signals or not, and to what extent they are easy to fake. However, several studies have revealed that facial expressions of negative emotions such as sadness, outrage, or disgust are difficult to fake (Schug *et al.*, 2010). One study found higher levels of activation of the m. levatorlabii, a facial muscle which elevates the upper lip in facial expression of disgust, as a response to unfair offer in the ultimatum game (Chapman, Kim, Susskind, & Anderson, 2009).

The present results may also confirm the possible relationship between the facial correlates of deception in a social dilemma game and the facial expressions of negative emotions. Our result – that defection is linked to closed upper eyelids – can be related to the previous finding that fear is expressed by a combination of muscular activities that are involved in the movements of the eyebrow and eyes (m. frontalis, m. corrugator; Ekman & Friesen, 1976). Similarly, an angry face can be created by muscles that operate both around the eyes (m. depressor, m. procerus) and around the mouth (m. depressor labii inf., m. mentalis). In accordance, we found that defection is strongly linked to tightly depressed lower lips.

To sum up, we suggest that the action of closing the upper eyelids and depressing the lower lips serve as reliable indicators of deceptive dispositions because it mirrors negative social feelings that may be informative for the receiver. But why do defectors not try to hide these emotions? The relative strength and quality of the signal emitted by the defectors reportedly depend on the circumstances (Bell, Buchner, & Musch, 2010). This conditional rule may refer to the regulation of the emotions that the defectors feel in a social dilemma encounter. They may express or hide their automatically triggered emotions, depending on the social environment and their skills. When, for instance, defectors cannot be observed by the other party, they do not necessarily need to control these emotions. This is what may happen in the anonymous PDG situation: players who do not see each other can openly express emotions associated with their decisions without a risk of being detected. Furthermore, in the anonymous conditions related to the computer-mediated experimental game, the emotions being aroused on the participants' face were probably weak, and we expect that the intensity of their motivation to conceal their hints to defection was also limited. Further studies are needed to explore how defectors regulate their expressive behaviour when the others can observe their decisions.

We also found that females more confidently evaluated facial expressions. Former studies have revealed a natural superior ability of women to identify facial emotions; they exceeded men in their ability to recognize happiness, sadness, and fear (Rotter & Rotter, 1988; Timmers, Fischer, & Manstead, 1998). A more recent study showed that women were faster than men at recognizing both positive and negative emotions from facial cues, and this female advantage was not due to a gender difference in perceptual speed (Hampson, van Anders & Mullin, 2006). Furthermore, the gender difference was accentuated for negative emotions; the female superiority in RT was even higher for negative (fear, anger, disgust) than for positive (happiness) emotions. The explanation of the gender differences in recognizing emotional facial expressions may result from a possible evolutionary scenario according to which men (as a philopatric sex in humans) tend to stay in the birth group and women tend to emigrate to a different group (Geary, 1998, 2006). Consequently, female ancestors lived in a community with non-kin, and under these circumstances they were more likely to be socially isolated and at a greater risk for exploitation than men. The greater attentiveness of women to social cues such as facial expressions might reflect an adaptation to these social conditions. However, this explanation may not be entirely convincing because hundreds of matrilocal cultures might have existed in the past (Ember & Ember, 1971). Since in this marriage arrangement females stayed among their relatives after their marriage, they were less likely to be exploited by males in these cultures.

An alternative explanation may suggest that, compared to males, females in threatening social situations are more cautious to evaluate possible partners on the basis of their trustworthiness. Their ability to detect transiently evoked facial expressions may be a bias towards making less costly errors. In light of error management theory (Haselton & Buss, 2000), whenever the costs of different types of errors are asymmetrical, people favour biased decision rules that produce more beneficial or less costly outcomes relative to the alternative decision rules. As females have carried most of the burden of raising children, it may be beneficial for them to have an advantage in the accuracy of recognizing untrustworthy expressions of their potential partners. During evolution, they might have been selected for overestimating the importance of facial features associated with cheating to avoid deception (the partner's desertion), whereas an underestimation of these stimuli might have been harmful for their own and their children's survival chance. This adaptive bias may explain the present tendency demonstrating that females can recognize defectors more appropriately compared to males. Further studies should be carried out to provide more details of sex-specific cognitive bias in this field.

Acknowledgements

This work was supported by the PTE IKT research grant (University of Pecs, Research Team on Innovation, SROP-4.2.2/08/1/2008-0011), the Hungarian National Science Fund (OTKA, K101762 to T.B.) and the Janos Szentagothai Research Center. The authors thank Dr. Árpád Csathó for valuable discussion on the statistical analysis of the data, and Ferenc Kocsor for his kind help in image morphing.

References

Bell, R., Buchner, A., & Musch, J. (2010). Enhanced old-new recognition and source memory for faces of cooperators and defectors in a social-dilemma game. *Cognition*, *117*(3), 261–275. doi:10.1016/j.cognition.2010.08.020

Borkenau, P., Mauer, N., Riemann, R., Spinath, F. M., & Angleitner, A. (2004). Thin slices of behavior as cues of personality and intelligence. *Journal of Personality and Social Psychology*, *86*(4), 599–614. doi:10.1037/0022-3514.86.4.599 2004-12052-006

Brown, W. M., Palameta, B., & Moore, C. (2003). Are there nonverbal cues to commitment? An exploratory study using the zero-acquaintance video presentation paradigm. *Evolutionary Psychology*, *1*, 42–69. Retrieved from http://www.epjournal.net/wp-content/uploads/2011/08/ep014269.pdf

Chapman, H. A., Kim, D. A., Susskind, J. M., & Anderson, A. K. (2009). In bad taste: Evidence for the oral origins of moral disgust. *Science*, *323*(5918), 1222–1226. doi:10.1126/science.1165565

Cosmides, L., Barrett, H. C., & Tooby, J. (2010). Adaptive specializations, social exchange, and the evolution of human intelligence. *Proceedings of the National Academy of Sciences of the United States of America*, *107*, 9007–9014. doi:10.1073/pnas.0914623107

Cosmides, L., & Tooby, J. (1992). Cognitive adaptations for social exchange. In J. E. Barkow, L. Cosmides, & J. Tooby (Eds.), *The adapted mind: Evolutionary psychology and the generation of culture* (pp. 163–228). New York: Oxford University Press.

Ekman, P. (1985). *Telling lies: Clues to deceit in the marketplace, politics, and marriage*. New York: Norton.

Ekman, P., & Friesen, W. V. (1976). Measuring facial movement. *Environmental Psychology and Nonverbal Behavior*, *1*(1), 56–75. doi:10.1007/BF01115465

Ekman, P., & Friesen, W. V. (1982). Felt, false, and miserable smiles. *Journal of Nonverbal Behavior*, *6*(4), 238–252.

Ekman, P., & O'Sullivan, M. (2006). From flawed self-assessment to blatant whoppers: The utility of voluntary and involuntary behavior in detecting deception. *Behavioral Sciences & the Law*, *24* (5), 673–686. doi:10.1002/Bsl.729

Ember, M., & Ember, C. R. (1971). The conditions favoring matrilocal versus patrilocal residence. *American Anthropologist*, *73*, 571–594. doi:10.1525/aa.1971.73.3.02a00040

Fetchenhauer, D., Groothuis, T., & Pradel, J. (2010). Not only states but traits - humans can identify permanent altruistic dispositions in 20 s. *Evolution and Human Behavior*, *31*(2), 80–86. doi:10.1016/j.evolhumbehav.2009.06.009

Frank, M. G., & Ekman, P. (1997). The ability to detect deceit generalizes across different types of high-stake lies. *Journal of Personality and Social Psychology*, *72*(6), 1429–1439. doi:10.1037/0022-3514.72.6.1429

Geary, D. C. (1998). *Male, female. The evolution of human sex differences*. Washington, DC: American Psychological Association.

Geary, D. C. (2006). Sex differences in social behavior and cognition: Utility of sexual selection for hypothesis generation. *Hormones and Behavior*, *49*(3), 273–275. doi:10.1016/j.yhbeh.2005.07.014

Gronenschild, E. H., Smeets, F., Vuurman, E. F., van Boxtel, M. P., & Jolles, J. (2009). The use of faces as stimuli in neuroimaging and psychological experiments: A procedure to standardize stimulus features. *Behaviour Research Methods*, *41*(4), 1053–1060. doi:10.3758/BRM.41.4.1053

Hampson, E., van Anders, S. M., & Mullin, L. I. (2006). A female advantage in the recognition of emotional facial expressions: Test of an evolutionary hypothesis. *Evolution and Human Behavior*, *27*(6), 401–416. doi:10.1016/j.evolhumbehav.2006.05.002

Haselton, M. G., & Buss, D. M. (2000). Error management theory: A new perspective on biases in cross-sex mind reading. *Journal of Personality and Social Psychology*, *78*(1), 81–91. doi:10.1037/0022-3514.78.1.81

Mehu, M., Grammer, K., & Dunbar, R. I. M. (2007). Smiles when sharing. *Evolution and Human Behavior*, *28*(6), 415–422. doi:10.1016/j.evolhumbehav.2007.05.010

Melfsen, S., Osterlow, J., & Florin, I. (2000). Deliberate emotional expressions of socially anxious children and their mothers. *Journal of Anxiety Disorders*, *14*(3), 249–261. doi:10.1016/S0887-6185(99)00037-7

Oda, R., Yamagata, N., Yabiku, Y., & Matsumoto-Oda, A. (2009). Altruism can be assessed correctly based on impression. *Human Nature-an Interdisciplinary Biosocial Perspective*, *20* (3), 331–341. doi:10.1007/s12110-009-9070-8

Oosterhof, N. N., & Todorov, A. (2008). The functional basis of face evaluation. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(32), 11087–11092. doi:10.1073/pnas.0805664105

Oosterhof, N. N., & Todorov, A. (2009). Shared perceptual basis of emotional expressions and trustworthiness impressions from faces. *Emotion*, *9*(1), 128–133. doi:10.1037/a0014520

Rotter, N. G., & Rotter, G. S. (1988). Sex-differences in the encoding and decoding of negative facial emotions. *Journal of Nonverbal Behavior*, *12*(2), 139–148. doi:10.1007/BF00986931

Scharlemann, J. P. W., Eckel, C. C., Kacelnik, A., & Wilson, R. K. (2001). The value of a smile: Game theory with a human face. *Journal of Economic Psychology*, *22*(5), 617–640. doi:10.1016/S0167-4870(01)00059-9

Schug, J., Matsumoto, D., Horita, Y., Yamagishi, T., & Bonnet, K. (2010). Emotional expressivity as a signal of cooperation. *Evolution and Human Behavior*, *31*(2), 87–94. doi:10.1016/j.evolhumbehav.2009.09.006

Timmers, M., Fischer, A. H., & Manstead, A. S. R. (1998). Gender differences in motives for regulating emotions. *Personality and Social Psychology Bulletin*, *24*(9), 974–985. doi:10.1177/0146167298249005

Todorov, A., Pakrashi, M., & Oosterhof, N. N. (2009). Evaluating faces on trustworthiness after minimal time exposure. *Social Cognition*, *27*(6), 813–833. doi:10.1521/soco.2009.27.6.963

Tracy, J. L., Robins, R. W., & Schriber, R. A. (2009). Development of a FACS-verified set of basic and self-conscious emotion expressions. *Emotion*, *9*(4), 554–559. doi:10.1037/A0015766

Trivers, R. L. (1985). *Social evolution*. Menlo Park, CA: Benjamin/Cummings.

Vanneste, S., Verplaetse, J., Van Hiel, A., & Braeckman, J. (2007). Attention bias toward noncooperative people. A dot probe classification study in cheating detection. *Evolution and Human Behavior*, *28*(4), 272–276. doi:10.1016/j.evolhumbehav.2007.02.005

Verplaetse, J., Vanneste, S., & Braeckman, J. (2007). You can judge a book by its cover: The sequel. A kernel of truth in predictive cheating detection. *Evolution and Human Behavior*, *28*(4), 260–271. doi:10.1016/j.evolhumbehav.2007.04.006

Willis, J., & Todorov, A. (2006). First impressions: Making up your mind after a 100-ms exposure to a face. *Psychological Science*, *17*(7), 592–598. doi:10.1111/j.1467-9280.2006.01750.x

Yamagishi, T., Tanida, S., Mashima, R., Shimoma, E., & Kanazawa, S. (2003). You can judge a book by its cover - Evidence that cheaters may look different from cooperators. *Evolution and Human Behavior*, *24*(4), 290–301. doi:10.1016/S1090-5138(03)00035-7

Received 11 January 2012; revised version received 24 July 2012