ELSEVIER

Evolution and Human Behavior 28 (2007) 260 – 271

Evolution and Human Behavior

You can judge a book by its cover: the sequel. A kernel of truth in predictive cheating detection

Jan Verplaetse[a,*], Sven Vanneste[b], Johan Braeckman[c]

[a]Department of Jurisprudence and Legal History, Ghent University, 9000 Ghent, Belgium
[b]Department of Developmental, Personality, and Social Psychology, Ghent University, 9000 Ghent, Belgium
[c]Department of Philosophy and Moral Science, Ghent University, 9000 Ghent, Belgium

Initial receipt 9 January 2006; final revision received 24 April 2007

Abstract

In accordance with evolutionary models of social exchange, we suggest the possible existence of a limited predictive cheater detection module. This module enables humans, to a certain extent, to predict how willing another might be to cooperate or not. Using unknown target subjects who had played a one-shot prisoner's dilemma game earlier, we asked participants in two experiments to rate how cooperative these target subjects were. Pictures were taken of the target subjects at three different moments: a neutral-expression picture taken prior to the game, an event-related picture taken during the decision-making moment of a practice round, and an event-related picture taken during the decision-making moment of a proper round. We found that participants in the experiments could accurately discriminate noncooperative pictures from cooperative ones, but only in response to those taken during the proper round. In both neutral-expression pictures and practice-round pictures, identification rates did not exceed chance level. These findings leave room for the existence of a predictive cheater detection module that deduces someone's decision to cooperate from event-related facial expressions.

Keywords: Cheating detection; Altruism; One-shot prisoner's dilemma game; Facial expression; Cooperation

1. Introduction

It is widely believed that the human face reveals information about the self—that even without prior acquaintance, one can deduce personality traits, moral virtues, or social characteristics from the face of another (Hassin & Trope, 2000; Liggett, 1974). In Europe, from the time of ancient Greece, cultivated men practiced physiognomy, the art of reading traits from faces. In the second half of the 18th century, Lavater's physiognomy came into fashion and played a significant role among intellectual circles of the time (Zebrowitz, 1997). Despite continual criticism that physiognomy is analogous with charlatanism and despite the political dangers that its stereotypes tend to foster, the belief and practice of physiognomy continue to the present day.

Experimental studies in social cognition suggest that people can, in fact, infer personality traits from another's face. Studies indicate that face-based judgments are often more accurate than skeptics would like to think. Bond, Berry, and Omar (1994) found a reliable relation between face-based impressions of honesty and the willingness to deceive others, and Borkenau, Mauer, Riemann, Spinath, and Angleitner (2004) found that inferences of intelligence and personality from thin slices of behavior strongly predicted intelligence and personality test scores. These studies in social psychology document the so-called kernel-of-truth hypothesis, which maintains that there is some validity to social judgments based upon facial appearances (Bond et al., 1994). More precisely, this hypothesis suggests that people are able to distinguish personality traits from faces with a greater-than-chance accuracy rate. This rate is likely to

* Corresponding author. Department of Jurisprudence and Legal History, Ghent University, Universiteitstraat 4, 9000 Ghent, Belgium. Tel.: +32 9 264 67 93.
E-mail address: jan.verplaetse@ugent.be (J. Verplaetse).
URL: http://www.themoralbrain.be (J. Verplaetse).

1090-5138/$ – see front matter © 2007 Elsevier Inc. All rights reserved.
doi:10.1016/j.evolhumbehav.2007.04.006

increase if the target information contains more dynamic features (a 90-s videotape), if the target information is event related (telling a lie), or if people actually meet one another (a 30-min interaction). In settings of a less contrived nature, where more relevant information is available, social impressions concerning an individual's personality are likely to converge.

Face-based impressions do prove more complicated, however. Again and again, lie detection research has revealed our poor capacity to identify liars (DePaulo & Rosenthal, 1979; Ekman, O'Sullivan, & Frank, 1999). Even in real-life settings, unskilled perceivers fail to pick up relevant signals that are likely to display untruthfulness. Considerable evidence demonstrates that our social cognition system is not fine-tuned enough to receive and process deception cues that make it easier to identify a lie. Moreover, while evolutionary psychologists attach great significance to the human capacity of cheating detection (Cosmides, 1989; Cosmides & Tooby, 1992), individuals fail to discriminate defectors from cooperators when confronted with neutral-expression pictures taken from subjects who played a prisoner's dilemma game (PDG). Perceivers fail to identify the type of photographed targets with an accuracy any better than that of chance (Yamagishi et al., unpublished data). Until now, no study in the field of cheater detection research has surpassed the greater-than-chance accuracy threshold via the use of still photographs. The successful detection of cheaters apparently requires a lot more information. Brown, Palameta, and Moore (2003) found a significant altruist detection effect in response to videotaped storytelling. However, this study did not examine whether information of a more limited nature (nonverbal images or stills, for instance) yields the same detection effect. In contrast, Frank, Gilovich, and Regan (1993) found that participants needed a 30-min "get-acquainted" meeting before they were able to predict individual behavior in a one-shot PDG with any degree of accuracy. Although it seems reasonable "to suppose that cooperators can identify one another, at least in a statistical sense" (p. 256), prediction accuracy did not significantly exceed chance when subjects had access to information of only a limited nature. The same conclusion was reached with regards to cheater detection.

Several arguments, however, cast doubt on this conclusion. Evolutionary studies in psychology document that individuals are not indifferent to cooperative behavior. Cooperation for mutual benefit is a pervasive feature of social living and has been of crucial importance to the evolution of hominids (Trivers, 1971). Evolutionarily inspired reasoning assumes the existence of adaptive skills for discovering when someone has cheated and for remembering the person who has done it. Studies have abundantly demonstrated that people are proficient when it comes to these abilities. The Wason Selection Task research shows that people are particularly apt at determining when a social contract rule has been violated (Cosmides, 1989; Cosmides & Tooby, 1992). Further research indicates that people are also able to memorize the face of a cheater more accurately than that of a noncheater (Chiappe & Brown, 2004; Mealy, Daood, & Krage, 1996; Oda, 1997). We think, however, that evolutionary models of social cooperation offer the possibility of a cheater detection skill that is predictive as well. Humans may be equipped not only to discover and remember noncooperative individuals but also to predict whether an individual will be cooperative. Since noncooperative deals could have fatal consequences for deceived parties, the possession of a limited predictive cheater detection mechanism may be of vital importance in order to pass over or break off potentially defective deals. Consequently, it is not unreasonable to assume that evolution designed the human mind to scan others for information that might signal intentions to defect (for instance, by scrutinizing the presence or the absence of emotional cues that cannot be faked).

In this study, we examine the possibility of a predictive cheater detection module that functions within certain realistic limits. Our main objective is to precisely delineate these limits. On theoretical grounds, it is unlikely that predictive cheater detection mechanisms could rely on permanent facial features revealed by neutral facial expressions (Brown & Moore, 2002; Frank et al., 1993; Trivers, 1985). If differences between cooperators and cheaters were obvious, natural selection would move cheaters towards extinction. In a transparent world, cooperators who assortatively interact only with cooperators always get the highest payoffs. Consequently, cheaters only survive when they learn to fake cooperative intentions or to mimic expressions that might distinguish cooperators from defectors (or vice versa). Since these imitated expressions become unreliable for cooperators, evolution favors the selection of more subtle expressions and/or more sophisticated detection skills. Taking this ongoing oscillation between signal detection and signal deception into consideration, we do not believe that cues of a less subtle nature (such as permanent facial features or voluntarily produced expressions) demarcate defectors from cooperators. If humans are able to identify cheaters to a certain extent, this ability must be based upon the reception of more delicate signals.

On the other hand, we believe that a cheater detection mechanism can prove more powerful than previous results have suggested. We do not believe that 30 min of verbal interaction is necessary to predict whether someone will be cooperative. In Paleolithic times, it is quite plausible that quick nonverbal impressions mattered more than lengthy chats—that it was not conversation length or verbal coding but facial expression that was the decisive factor in the detection of cheaters during hominid interactions. Although a multitude of facial expressions can be gathered during a conversation (videotaped or real), we speculate that a single event-related picture may do the same job. If the event strongly resembles one that may invite defective behavior,

accuracy rates ought to increase. If someone decides to cheat, facial cues related to expressions of anxiety, guilt, or greed are likely to reveal one's intentions and future behavior. One quick impression, laid down in one photograph, should suffice to stop the transaction or to break off the deal.

We suggest that subjects ought to be able to predict the behavior of unknown individuals in response to a single event-related picture. A similar study (Yamagishi, Tanida, Mashima, Shimoma, & Kanazawa, 2003) found that humans were better able to remember the faces of cheaters than the faces of cooperators, even though no information (no labels attached to the photographs) as to whether the subjects had cheated was given when the faces were first seen (or memorized). Since Yamagishi and Tanida (unpublished data) also found that participants were unable to identify either type of player above chance when gauging neutral-expression photographs, this result came across as somewhat puzzling. In fact, there seems to be more than accurate remembrance but less than accurate identification. Unfortunately, Yamagishi and Tanida never asked their participants to discriminate between event-related pictures of cheaters and cooperators. They used these pictures in memory tasks alone, not in identification tasks. Here we explore the possibility that this omission is the key to a more precise delineation of a predictive cheater detection module. By incorporating the experimental design of Yamagishi and Tanida, we were able to demonstrate that people can indeed identify unknown cheaters when confronted with a single event-related picture.

2. Stimuli

We asked participants to rate stimulus pictures that were obtained from target subjects at three distinct moments. After an initial neutral-expression picture was taken in front of a white wall, target subjects played a computer-mediated one-shot PDG for real money (€0>€1>€3>€5). Players were kept ignorant as to whom they were playing with. Recognition and acquaintance were prevented by using separate entrances for each player and by hanging sheets between computers. From their computer screens, target subjects could read self-paced instructions that explained the PDG. It was explicitly communicated that, although defection was the most rational choice, cooperation was the most socially minded choice. All players read the following lines (originally in Dutch):

> "Some people maintain that you should always choose the Big Prize (€5). This seems the most rational choice. At worst, you always get something; at best, you get the large amount. You never go home empty handed. Choosing the Modest Reward (€3) seems a little foolish. You risk gaining nothing from the experiment. Although the other player might not even have wanted to share, he gets everything while you get nothing. Other people claim that you should always choose the Modest Reward (€3). This seems the most socially minded choice. Only in this way do you avoid the other player going home empty handed. Choosing the Big Prize (€5) hardly even seems moral. The other player risks gaining nothing from this experiment. Although the other player might have wanted to share, you receive everything. These are just opinions."

They were then asked to complete a two-question multiple-choice quiz designed to assess their comprehension of the game. If both players completed the quiz, a signal sounded, followed by a countdown indicating that decisions were to be expected within 10 s. At the end of the countdown, two buttons appeared: red (defection) and green (cooperation). Decisions were made with a click of the mouse. Our software mutually communicated the decisions. With the help of a visible webcam in front of the monitor (Logitech Quickcam 8.0), pictures were taken at the very moment of the mouse click. Target subjects then received the money they earned and were asked to complete a postexperimental questionnaire.

In order to assess the impact of event-related expressions on cheater detection, we introduced an intermediary recording moment. A practice round, without payoffs, preceded the proper round. The practice round was categorically announced and served to familiarize subjects to this rather short game. Again, webcams took pictures at the moment of decision; decisions were mutually communicated. From each player, we thus acquired three pictures: one non-event-related neutral picture and two event-related pictures (one practice-round picture and one proper-round picture). Due to the absence of real payoffs in the practice round, we predicted that the accuracy rate for correctly identifying the practice-round pictures would stand about midway between the rate of the proper-round pictures and the rate of the neutral-expression pictures.

Pictures of 112 target subjects were taken. The average age was 21.18 years (S.D.=5.09). Five players were excluded because they believed that they were recognized by their opponents. Two other players were removed because postexperimental questionnaires revealed that they did not understand the PDG. In the practice round, 67 subjects (60%) decided to cooperate and 45 subjects (40%) opted to defect. In the proper round, 58 subjects (52%) ultimately cooperated and 54 subjects (48%) decided to defect.

3. Experiment 1

3.1. Method

3.1.1. Identification

3.1.1.1. Identification set. From the pool of 105 one-shot PDG players, we selected 64 players whose pictures were without photographic deficiencies; faces that were not entirely framed or faces that were partly covered by hair, garments, eyeglasses, or hands were excluded. From the remaining players, pictures were chosen at random, provided that they met the following requirements: all pictures were

equated for background, brilliance, and luminance; and all pictures had to be more or less equal in size. Using Photoshop Elements 2.0, we edited all selected pictures to obtain a white, neutral, and equal background. Our final identification set contained 13 noncooperative and 13 cooperative proper-round pictures, 13 noncooperative (8 of which also defected during the proper round) and 12 cooperative practice-round pictures (9 of which also cooperated during the proper round), and 12 neutral-expression pictures (6 of which defected during the proper round). We selected an equal number of males and females.

3.1.1.2. Identification task. In Experiment 1, a group of perceivers was asked collectively to guess who cooperated and who did not cooperate via a paper-and-pencil test. The participants were 106 social science students (55 females) at Ghent University (Ghent, Belgium). The average age was 20.24 years (S.D.=3.02). Stimulus participants and perceivers had approximately the same age. Before the identification task began, an introduction on the PDG was given (the same as the one given to our players), and the students were given the opportunity to ask questions should anything remain unclear. Participants then viewed 64 pictures from our identification set in a PowerPoint presentation for 5 s each, with a 2-s interval between pictures. A questionnaire asked for the target's cooperativeness (Had X cooperated or not?), confidence in judgment (How certain are you?), and whether the subject knew the target (Do you know X?). If subjects recognized a target, answers were removed. Once the presentation was over, subjects had to rate the task effort (How difficult was this task?) on a 5-point Likert scale. In order to facilitate rating and to avoid mistakes, all pictures were numbered. During the 2-s interval, the next picture's number was briefly presented.

3.1.2. Memory

3.1.2.1. Memory set. Although our research primarily addresses cheater detection rather than recall of cheaters, we nonetheless composed a memory set. We randomly mixed the faces of our identification set with new faces chosen from the reserve set. Our reserve collection consisted mainly of players for whom we could only obtain one or two successful photographs. The memory set used in Experiment 1 ultimately contained 30 pictures: 15 randomly selected faces from our reserve set and 15 faces from our identification set.

3.1.2.2. Memory task. Approximately 1 h after the identification task, roughly the same group (106 participants, minus 27 who quit the experiment) was asked to participate in a previously unannounced memory task. Seventy-nine participants viewed 30 pictures from our memory set in a fixed randomized sequence. A PowerPoint presentation showed these pictures, again for 5 s each, again with a 2-s interval between the pictures. A paper-and-pencil questionnaire asked for the target recall (Have you seen X during the first task?), confidence in judgment (How certain are you?), and again whether the subject was familiar with the target (Do you know X?). Answers were removed if subjects were familiar with a target. Once the presentation was over, subjects rated task effort (How difficult was this task?) on a 5-point Likert scale. The same precautions were taken as in Experiment 1.

3.2. Results

3.2.1. Identification

3.2.1.1. Frequency rate. Two pictures were eliminated because some participants recognized the subjects (one noncooperative practice-round picture and one cooperative proper-round picture). Our participants rated 48% of our set as cooperators and 52% of our set as defectors. Nothing was mentioned beforehand about our identification set's equally balanced composition. For each separate picture category, we controlled for outliers. All respondents who exceeded the 95% confidence interval for mean were to be excluded from the analysis, but none fell into this category.

3.2.1.2. Identification rate. A one-sample t test was conducted to determine whether various true-positive identification rates exceeded chance level (0.50), the results of which are shown in Table 1. A significant effect was found for cooperative proper-round pictures (mean=0.59) and noncooperative proper-round pictures (mean=0.66) [$t(105)=6.52$, $p<.001$ and $t(105)=10.91$, $p<.001$, respectively]. In contrast, identification rates for neutral-expression pictures and for both cooperative and noncooperative practice-round pictures did not exceed chance level. Varying accuracy rates seem to indicate that proper-round pictures must differ in some way from both neutral-expression and practice-round pictures.

We also conducted a supplementary analysis to calculate the possible impact of expected accuracy rates versus the overall actual accuracy rate. The difference between both rates is informative inasmuch as perceivers who expect more cooperators than cheaters will identify more cooperators by random guessing alone (and vice versa). If expected

Table 1
One-sample t test of the true-positive identification rate (Experiment 1)

Pictures	True-positive identification rate (S.D.)	p
Neutral-expression cooperators	0.48 (0.12)	ns
Neutral-expression defectors	0.48 (0.08)	ns
Practice round: cooperators	0.52 (0.13)	ns
Practice round: defectors	0.53 (0.10)	ns
Proper round: cooperators	0.59 (0.15)	<.001
Proper round: defectors	0.66 (0.16)	<.001

ns=not significant.

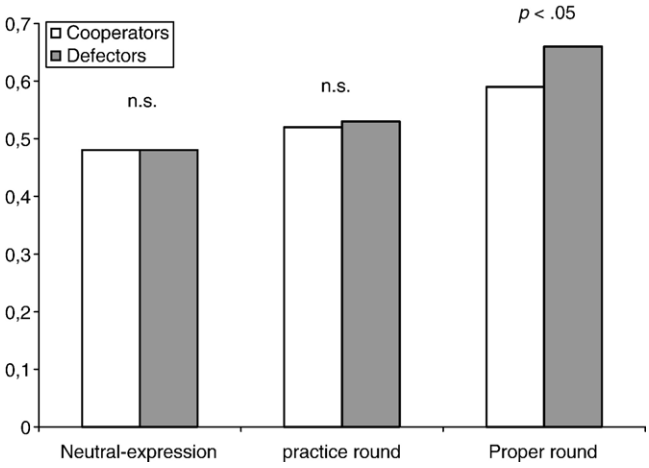

Fig. 1. Identification rates for different conditions in Experiment 1.

accuracy rates approach overall actual accuracy rates, even a significant high true-positive identification rate will, of course, be less robust. To calculate the expected accuracy rate, Frank et al. (1993) formulated a mathematical standard.[1] Because our stimulus set was equally balanced, perceivers had a 50% chance of randomly identifying targets as cooperators or defectors. Since the actual accuracy rates of cooperative and noncooperative pictures are 0.47 and 0.53, respectively, the expected accuracy rate is 50%=(0.47)(50) +(0.53)(50). This is 13 percentage points lower than the 63% overall accuracy rate achieved (Appendix A).

3.2.1.3. Cheater versus cooperator detection. Similar to Yamagishi et al. (2003), a paired *t* test was conducted on the true-positive identification rate of the proper-round condition. Confronted with proper-round pictures, we found that participants identified noncooperative pictures (mean=0.66) more accurately than cooperative pictures (mean=0.59) [$t(105)=3.44$, $p<.05$] (Fig. 1).[2] Moreover, we found that classifying faces as cooperative or noncooperative was done with unequal confidence. A paired *t* test showed that participants were more confident when classifying faces as noncooperative (mean=2.91) than when classifying pictures as cooperative (mean=2.66) [$t(105)=4.02$, $p<.001$]. Again, this cheater bias only held true when subjects were confronted with proper-round pictures. Neutral-expression and practice-round pictures did not show this bias.

Again we conducted a supplementary analysis to calculate the impact of unbalanced random guessing, which might affect our documented bias towards noncooperative proper-round pictures. This time we compared the accuracy rates of cooperative and noncooperative proper-round pictures with those of neutral pictures. These latter accuracy rates can be considered a priori expected accuracy rates since neutral pictures (cooperative or not) are not identified above chance level. A chi-square test revealed a significant effect, revealing differential accuracy rates in detecting cooperators versus defectors (Appendix B).

3.2.2. Memory

3.2.2.1. Recognition accuracy. The true-positive recall rate (participants correctly recalling previously presented faces) was 0.74, while the false-positive rate (participants incorrectly recognizing faces not previously shown) was 0.22. The discrimination measure d' was 1.42, which means that, in general, participants were indeed better able to recognize faces previously shown than faces that were not shown. Recall of previously shown faces (72%) was more accurate than the recognition of pictures that had not been shown before (65%). On the other hand, none of the different picture categories delivered recognition effects that significantly stood out from the others. Thus, proper-round pictures (cooperative or not) were no better recognized than practice-round or neutral-expression pictures.

3.2.2.2. Cheater or cooperator recall. Defectors were easier to remember than cooperators (0.72 vs. 0.66, respectively) [$t(78)=-2.59$, $p<.05$]. This effect was only achieved with proper-round pictures, however. Recall accuracy rates with practice-round pictures and neutral-expression pictures revealed no biases. This finding reinforces the well-documented cheater bias characteristic of memory tasks. We summarize these findings in Fig. 2.

3.3. Discussion

The findings in Experiment 1 substantiate the existence of a cheater detection effect. To a certain extent, participants were able to accurately differentiate defectors from cooperators even though they had never met them before. This identification only occurred, however, with

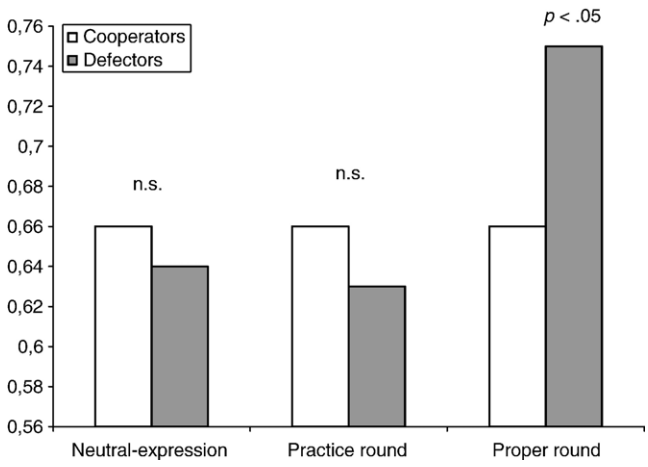

Fig. 2. Recalling rates for different conditions in Experiment 1.

[1] This standard rested on research conducted by Dawes, McTavish, and Shaklee (1977).

[2] The statistical power was .66 (α=.001).

event-related pictures (i.e., pictures that were taken at the moment of decision in a one-shot PDG). In response to both practice-round and neutral-expression pictures, participants were unable to discriminate cheaters from cooperators. Proper-round pictures must differ in some way from neutral-expression and practice-round pictures. Our participants were not informed about the different recording moments. It seems reasonable to assume that proper-round pictures capture subtle visual cues, whereas practice-round pictures lack such cues because nothing was at stake. To conclude, our perceivers were able to successfully pick up event-related expressions, which only proper-round pictures captured. These results suggest that cheating detection depends on event-related facial signals and not on permanent facial features. In addition, Experiment 1 suggests a bias to cheaters with regards to identification and memory. People identify cheaters better than cooperators, and they also seem more confident in doing so. As documented in previous studies, cheaters also proved easier to recall. Although we replicated this bias, in contrast to previous studies, we found it solely in relation to event-related proper-round photographs.

It must be noted, however, that these results are not entirely conclusive. Several weak points may have confounded the findings we obtained in Experiment 1. To begin with, it is possible that our stimulus sets were biased. Perhaps we unintentionally selected noncooperative proper-round pictures from cheaters with more striking permanent features. In this case, our claim that accuracy rates depend on event-related cues would obviously be unfounded. To anticipate this criticism, we conducted a second experiment in which we counterbalanced our identification and memory set by switching practice-round pictures for proper-round ones. If identification hinges on permanent features, this alteration should influence accuracy rates (i.e., target subjects who would now be presented in proper-round pictures should be identified and recalled more accurately, and vice versa).

In addition, classroom presentations and paper-and-pencil tests potentially entail several confounding effects. Because all participants answered simultaneously, a degree of conformism (group effect) could not be avoided. It is possible that a few participants, particularly gifted at discerning cheaters, were imitated by less gifted participants, giving the false impression that people, in general, are quite good at detecting cheaters. To overcome this confound, in the second experiment, participation was private. Because all participants in Experiment 1 perceived the stimuli in a fixed sequence, pictures were possibly compared with previous ones, which could act as prototypical examples of cheaters or cooperators. Presentation order could, therefore, have affected accuracy rates. Moreover, the pictures shown at the beginning or at the end of the presentation were perhaps remembered with greater ease, according to well-documented primacy and recency effects. In Experiment 2, identification and memory tasks were therefore self paced, computer mediated, and administered individually. A counterbalanced stimulus set was presented in randomized order.

4. Experiment 2

4.1. Method

4.1.1. Identification

4.1.1.1. Identification set. Modifying the identification set used in Experiment 1, we exchanged practice-round pictures and proper-round pictures: If we previously selected pictures of players during their practice round, we now took their proper-round pictures and vice versa. The composition of the neutral-expression picture set was kept similar to that used in Experiment 1. Because the decisions to cooperate or to not cooperate could differ from one round to the next (from the practice round to the proper round), this alteration changed the composition of our identification set. Our identification set now contained 10 noncooperative and 17 cooperative practice-round pictures, 11 noncooperative and 14 cooperative proper-round pictures, and 12 neutral-expression pictures (6 defected and 6 cooperated in the proper round).

4.1.1.2. Identification task. We now asked 28 participants to rate who may or who may not have cooperated. This time our identification task was self paced, computer mediated, and administered on an individual basis. The order of presentation was randomized. Participants were social science students (18 females, 10 males) at Ghent University. The average age was 22.24 years (S.D.=1.45). Stimulus participants and perceivers had approximately the same age. We gave the same introduction to the PDG and administered a two-question multiple-choice quiz. Subjects were then shown 64 pictures from our identification set. One half of the participants gave ratings by pressing Ctrl-L key (cooperation) and Ctrl-R key (noncooperation); the other half pressed conversely. Confidence in judgment was gauged by measuring the reaction times of participants. If subjects recognized a target, they were asked to skip the picture. A postexperimental questionnaire asked for task effort (How difficult was this task?).

4.1.2. Memory

4.1.2.1. Memory set. In order to obtain more conclusive results, we expanded our memory set from 30 to 50 pictures: 25 from our reserve collection and 25 from our newly formed identification set, with each category equally represented.

4.1.2.2. Memory task. Approximately 20 min after the identification task, following a distracting calculation exercise, participants were asked to complete an unanticipated memory task. The memory task was also self paced, computer mediated, and administered on an individual basis. Twenty-eight participants viewed the 50 pictures in the

Table 2
One-sample *t* test of the true-positive identification rate (Experiment 2)

Pictures	True-positive identification rate (S.D.)	*p*
Neutral-expression cooperators	0.49 (0.14)	ns
Neutral-expression defectors	0.52 (0.11)	ns
Practice round: cooperators	0.48 (0.12)	ns
Practice round: defectors	0.50 (0.17)	ns
Proper round: cooperators	0.52 (0.09)	ns
Proper round: defectors	0.66 (0.12)	<.001

memory set. One half of the participants gave ratings by pressing Ctrl-L key for seen and Ctrl-R key for not seen; the other half pressed conversely. Confidence in judgment was gauged by measuring the participants' reaction times. If subjects recognized a target, they were asked to skip the picture. A postexperimental questionnaire asked for task effort (How difficult was this task?).

4.2. Results

4.2.1. Identification

4.2.1.1. Frequency rate. Two noncooperative proper-round pictures were eliminated because some participants recognized the target subjects. Participants rated 47% of our set as cooperators and 53% of our set as defectors. As in the previous experiment, the composition of our identification set was not mentioned to the participants beforehand. As in Experiment 1, we controlled for outliers, but no participants were excluded thereby.

4.2.1.2. Identification accuracy. We conducted a one-sample *t* test to determine whether participants scored above the 50% chance level (Table 2). As in Experiment 1, analyses revealed a significant effect for proper-round pictures [$t(27)=8.20$, $p<.001$]; however, in contrast to Experiment 1, this held only for noncooperative pictures. When participants were confronted with cooperative proper-round pictures, the identification rate dropped to chance level.

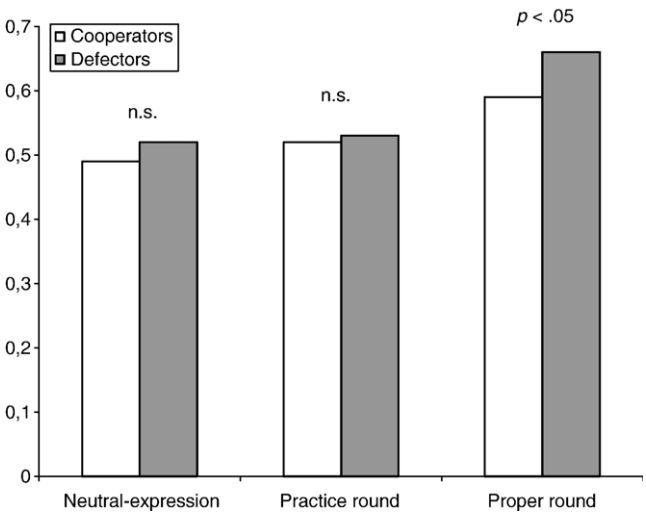

Fig. 3. Identification rates for different conditions in Experiment 2.

Table 3
Time (in ms) needed to make a decision (Experiment 2)

Pictures	Mean (S.D.)
Neutral: cooperators	2310 (763)
Neutral: defectors	2240 (721)
Practice round: cooperators	2137 (805)
Practice round: defectors	1935 (758)
Proper round: cooperators	1958 (436)
Proper round: defectors	1634 (343)

Again we tested for possible deviations between expected and overall accuracy rates of noncooperative and cooperative proper-round pictures to assess the impact of disproportionate random guessing. In contrast with Experiment 1, our stimuli set was not equally balanced, as it now contained more cooperative (62%) than noncooperative pictures. Since the actual accuracy rates of cooperative and noncooperative pictures were 0.44 and 0.56, the expected accuracy rate now was 49%=(0.44)(56)+(0.56)(44). Following the mathematical standard used in Frank et al. (1993), this is 9 percentage points lower than the 58% overall accuracy rate achieved. The likelihood of such a high overall accuracy rate occurring by chance is <1 in 1000 (Appendix A).

4.2.1.3. Cheater versus cooperator detection. In addition to the above results, a significant effect was found between cooperative and noncooperative proper-round pictures [$t(27)=-6.27$, $p<.001$]. Cheaters were more easily detected than cooperators, but only when a proper-round picture was presented (Fig. 3). In Experiment 2, confidence in judgment was measured via reaction times. A compared *t* test analyzed the average time each participant needed in order to make one's decisions (Table 3). A significant effect was found between cooperative and noncooperative pictures [$t(27)=4.61$, $p<.001$], again only for proper-round pictures.[3] Proper-round pictures of cheaters were classified significantly faster than proper-round pictures of cooperators. In addition, reaction times decreased as pictures became more expressive.

As in Experiment 1, we compared the accuracy rate for proper-round pictures with the a priori expected accuracy rate (e.g., accuracy rate with neutral pictures) to assess the impact of unbalanced random guessing on the here documented bias towards noncooperative proper-round pictures. A chi-square test yielded a significant effect, revealing a differential accuracy rate of detecting defectors versus cooperators (Appendix B).

[3] The statistical power was .96 (α=.001).

4.2.2. Memory

4.2.2.1. Recognition accuracy. A measure of discrimination d' was 1.26, whereby the true-positive recognition rate was 0.68. The false-positive rate was 0.23. These figures indicate that, in general, participants were able to discriminate between faces that had been shown and faces that had not been shown. The participants' abilities to recall pictures previously shown (pressing "seen"; 71%) did not differ from their ability to recognize pictures that were not shown (pressing "unseen"; 69%). No significant effects were discovered between the different categories. Proper-round pictures (cooperative or not) were no better recognized than practice-round pictures or neutral-expression pictures.

4.2.2.2. Cheater or cooperator recall. Defectors were recalled more readily than cooperators (0.79 vs. 0.60, respectively) [$t(27)=-4.93$, $p<.001$] (Fig. 4). Again, this effect only occurred with proper-round pictures. Recall accuracy rates with practice-round pictures and neutral-expression pictures revealed no biases.

4.3. Discussion

In Experiment 2, the existence of a cheater detection effect was replicated; perceivers were able to discriminate between cheaters and cooperators without ever having met them. As in Experiment 1, only event-related proper-round pictures caused this effect. In less expressive images, both cheaters and cooperators could not be identified. These findings demonstrate that predictive cheating detection requires event-related facial information. A propensity to defect cannot be predicted from permanent physiognomic features. Contrary to Experiment 1, however, identification of cooperators, even in the case of proper-round pictures, failed to cross the threshold of chance. Although this finding indicates a limit to the possible scope of cheater detection skill, it is somewhat expected. Cheater detection is likely reliant on cues that reveal emotions (anxiety, guilt, or greed) associated with intention to defect rather than on cues suggesting cooperation.

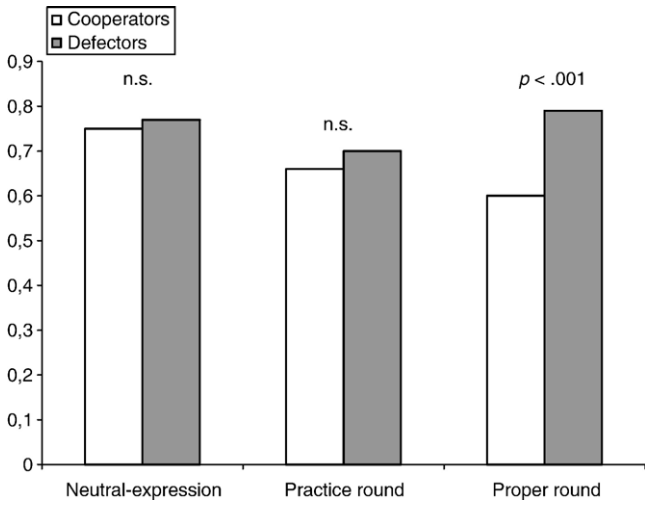

Fig. 4. Recalling rates for different conditions in Experiment 2.

5. General discussion

Evolutionary models of social exchange leave room for a predictive cheater detection mechanism that allows individuals to predict, to a certain extent, the social propensities of another. This mechanism processes facial information during the process of making a decision, before any potential cheating might begin. It differs substantially, therefore, from cheater detection skills documented to date (i.e., discovering and remembering defectors; Chiappe & Brown, 2004; Cosmides & Tooby, 1992). Although it may seem to parallel the dubious ancient art of physiognomy, we argue that face-reading skill makes sense from an evolutionary point of view. When noncooperative deals threaten social exchange in a dramatic manner, a predictive cheating detection mechanism offers the opportunity to pass over or break off potentially costly deals. In accord with Frank (1988) and Hirshleifer (1987), one can assume that the ability to scrutinize emotion-based facial cues might contribute to solving commitment problems. Commitment problems arise when information about potential partners' future cooperative behavior is lacking. Since the likelihood that potential partners will defect during future interactions cannot be dismissed, partners are reluctant to engage in risky transactions without sufficient guarantee of another's commitment to a strategy of cooperation. A predictive cheating detection mechanism that is sufficiently sensitive to signals of noncooperative intentions may help to overcome this information deficit.

In this study, we presumed the existence of such a predictive cheater detection mechanism. We asked participants in two experiments to rate the cooperativeness of unknown target subjects who had played a one-shot PDG earlier. We took photographs of these target subjects at three different moments: a neutral-expression picture prior to the game, and event-related pictures taken during decision-making moments of both a practice round and a proper round. We found that participants in both experiments were able to discriminate noncooperative from cooperative pictures with a 66% accuracy rate in response to pictures taken during the proper round. For the two remaining categories (neutral-expression pictures and practice-round pictures), identification rates did not exceed chance. In the case of identifying cooperators in proper-round pictures, the accuracy rate only proved significant in Experiment 1.

These findings demonstrate that predictive cheating detection contains a kernel of truth. To a certain degree, people reveal their noncooperative intentions. A single picture, taken at the moment of decision, can apparently give enough visual information to expose these intentions. This

conclusion contradicts that of Frank et al. (1993). Cheater detection does not require a lengthy videotaped talk. One event-related picture taken at the moment of a defective decision can produce the same results. To our knowledge, this is the first time in the field of cheater detection research that the greater-than-chance accuracy rate has been successfully obtained from (nonmoving) photographic stimuli. We did not, however, confirm the kernel-of-truth hypothesis with regard to the less expressive pictures. Our study indicates that defective behavior cannot be predicted from facial appearances in which only permanent facial features are exhibited. In this respect, evolutionary psychology does not support physiognomy. In our view, faces are not untrustworthy in themselves; yet, when in provocative social settings, they can radiate cues of noncooperativeness. This view also explains the puzzling results in Yamagishi et al. (2003). Even unknown cheaters are better remembered because observers can separate them from cooperative players, but only when the pictures are taken in provocative social settings. The more expressive the picture, the more pronounced the cheater bias will be in recall and identification. To restate, aside from accurate remembrance, there is accurate identification as well, but only in response to event-related pictures.

From a theoretical perspective, these results are in line with Brown and Moore's (2002) conceptual analysis of the evolution of reliable altruism indicators (reliable signaling theory), which presumes a continuous oscillation between signal detection and signal deception in evolutionary time. Following Frank (1988) and Trivers (1985), the generation of reliable altruism signals (which allow cooperators to avoid cheaters) and their faked counterparts (which seduce altruists to cooperate with defectors) is subject to a predator–prey arms race. If it becomes too easy to mimic altruism signals, evolution favors cooperators who give off more subtle cues of cooperativeness that are more difficult to fake. The cheater bias replicated here suggests that our attention is focused towards noncooperative people and their absent or failed faking of reliable signals. However, alternative scenarios are conceivable. Instead of producing more sophisticated honesty cues, evolution might stimulate the selection of more fine-tuned detection skills. According to this alternative model, the prey–predator arms race between detection and deception improves the quality of receivers' cheater detection skills, whereas the range of expressions emitted by senders remains unchanged. While evolution does not affect the complexity of the signaling, it does generate detection mechanisms able to pick up more subtle expressive differences between cheaters and cooperators. Despite dissimilarities, both models nonetheless assume a trend towards more sophistication. This corresponds to our main finding that predictive cheater detection requires event-related cues rather than permanent facial features.

It should be noted that there is plenty of room for future studies to complement our findings. Firstly, we are aware that defection in a one-shot PDG differs from cheating in a repeated PDG. Game theorists characterize cooperation in a one-shot PDG as a strongly altruistic behavior (hardcore altruism), whereas cooperation in a repeated PDG requires only reciprocal altruism (soft core altruism). Accordingly, we may distinguish noncooperative behavior in a one-shot PDG (soft cheating) from noncooperative behavior in a repeated PDG (hard defection). For several reasons, we preferred to limit our experiments to a one-shot PDG design. Undoubtedly, since evidence for a soft cheating detection module a fortiori substantiates the case for a hard cheating detection module, this option does not undermine our conclusions. We, nevertheless, admit that a comparison with pictures taken during a repeated PDG would be welcome. Secondly, it might be objected that in our experimental setting, in contrast to real life, the photographed players had no incentive to conceal their facial expressions. This might facilitate the detection of noncooperative players. To increase the percentage of noncooperative players, we opted for an anonymous setting in which visual communication between the players was prevented. It might be assumed that open settings, in contrast, benefit defectors who are adept at hiding their intentions during direct interactions. It should be noted, however, that the players' gestures were recorded by two visible webcams, that a third video camera stood near the participants and overlooked the whole scene, and that an experimenter observed both players. Recent evolutionary psychological research has shown that even subtle cues suggesting the presence of others (e.g., eye spots) change our social choices, arguably because reputation might be at stake (Haley & Fessler, 2005). Although anonymity was assured with the players, our setting offered sufficient "eyes" to encourage players to conceal their defective intentions from possibly influential witnesses. Nevertheless, we again welcome experiments in which more incentives to hide facial expressions are introduced.

Three further issues can be highlighted from this study. A first concerns the possible differences between cooperators and cheaters. Which precise facial features are involved in the self-betrayal of (unsuccessful) defectors? Our study did not investigate the matter but, nevertheless, offers background for future research. Brown and Moore (2002), for instance, used smiling icons to show that posed smiles are observed differently from involuntary or spontaneous smiles (e.g., asymmetrical) and negatively influence reputations in cooperative contexts. It would be interesting to investigate whether manipulations of the subject's pose, causing expressions such as smile symmetry/asymmetry, could affect identification rates. Our identification set offers the chance to investigate a picture sample more realistic than that of the Brown and Moore experiment.

A second intriguing issue concerns the difference between unsuccessful and successful defectors. Why were some cheaters discovered while others were left unexposed? This may be due, of course, to extraneous elements. Our photographs were limited to decision-making moments alone and did not frame all of the subjects' expressions.

Cheaters could have displayed facial cues at different points in time and thus remained unnoticed. Presumably, if more slices of visual information were presented, accuracy rates would increase. Since we also videotaped the games, an opportunity to investigate this hypothesis exists. We did not expect, however, that participants would be able to identify all of the cheaters. As already mentioned, predictive cheating detection is only possible within certain limits. When cues of noncooperativeness would prove too obvious, the logic of natural selection predicts that cheaters would disappear.

A last issue concerns the (neuro)psychological mechanism that underlies the predictive cheating detection mechanism. Postexperimental questionnaires revealed that participants did not really know how they were able to detect cheaters. In addition, participants answered that the detection of defectors was ("very") difficult (Experiment 1: 86%; Experiment 2: 90%), and, although they actually performed quite well ("very"), bad performance was to be expected (Experiment 1: 89%; Experiment 2: 87%). These poor estimations suggest that predictive cheating detection is an implicit automatic processing skill. In a recent study, we confirmed this presumption using a dot-probe task, in which conscious shifts of attention were obviated. Results showed that participants unconsciously pay more attention to pictures of cheaters who decide to defect during the proper round (as opposed to players who decide to cooperate) (Vanneste, Verplaetse, & Braeckman, in press). In an explicit task, social impression assessment is more accurate, participants are more confident, and they are quicker in responding to noncooperative proper-round pictures. In an implicit task, more and longer attention is allocated to these stimuli.

These findings may help us to disentangle the neural underpinnings of a predictive cheater detection module. Noncooperative proper-round pictures appear to automatically trigger neural regions that are involved in assessing the trustworthiness of a given face. Recent functional imaging studies show that certain brain regions become more active when responding to faces judged untrustworthy (Adolphs, 1999, 2003). Until now, however, no fMRI study has made use of event-related pictures. Winston, Strange, O'Doherty, and Dolan (2002) found increased activation in the bilateral amygdalar and right insular cortices, yet used previously selected trustworthy/untrustworthy faces that originated from individuals whose social adjustment was of no importance. Singer, Kiebel, Winston, Dolan, and Frith (2004) found that faces of (known) cheaters and defectors who played a PDG activated distinct brain circuits, but event-related pictures were not administrated in this case, either. It may thus prove productive to extend this line of neuroimaging research to different stimulus sets and to investigate, for instance, whether the neural activity located in distinct brain circuits (including the amygdalar and/or insular cortices) gradually increases in response to noncooperative event-related pictures. If so, this knowledge might shed light on neural matrices that allow us to glean social information from another's face.

In conclusion, we have established here that humans can predict noncooperativeness from facial photographs within a limited degree of accuracy. We presume that people subconsciously pick up cues of noncooperativeness—cues that do not involve permanent facial features, but rather consist of facial expressions elicited by significant social decisions. These findings are in accord with the possible existence of a distinct cheater detection mechanism that deduces someone's willingness to cooperate from event-related visual materials.

Appendix A

Study 1

To eliminate the influence of disproportionate random guessing (if one judges all pictures as defectors, the true-positive identification rate is 100% for defectors), which could affect our finding that proper-round pictures are more accurately identified, we conducted a supplementary analysis suggested by Frank et al. (1993). According to this, Study 1 might calculate this impact by comparing expected accuracy rates with the overall accuracy rates of proper-round pictures. Let us start with a recapitulation of our main results before we calculate both indicative rates below. From the 13 cooperative and 13 noncooperative proper-round pictures, our participants correctly classified 59% and 66%, respectively (Table 4). Consequently, from the 13 cooperative pictures, 7.67 pictures (59%) were correctly classified and 5.33 pictures (13; 7.67) were wrongly classified. From the 13 noncooperative pictures, 8.58 pictures (66%) were correctly classified while 4.42 pictures (8.58) were classified wrongly (Table 4).

We now calculated the expected accuracy rate of subjects who randomly predicted cooperative pictures 50% of the time and noncooperative pictures 50% of the time (due to the equal balance of our stimulus set in Study 1). Since the actual rates of cooperative and noncooperative pictures were 47% and 53%, respectively, the expected accuracy rate for these subjects shall be:

$$(0.47)(50) + (0.53)(50) = 50\%.$$

According to Frank et al. (1993), the overall accuracy rate can be calculated as follows. Take the sum of the correctly classified cooperative (7.67) and noncooperative (8.58) pictures and divide this by the total sum of the proper-

Table 4
Predicted versus actual accuracy of cooperative and noncooperative proper-round pictures (Experiment 1)

	Actual [*n* (%)]		Total predicted [*n* (%)]
	Cooperative	Noncooperative	
Predicted			
Cooperative	7.67	5.33	13 (50)
Noncooperative	4.42	8.58	13 (50)
Total actual	12.09 (47)	13.91 (53)	26

Table 5
Predicted versus actual accuracy of cooperative and noncooperative proper-round pictures (Experiment 2)

	Actual [*n* (%)]		Total predicted [*n* (%)]
	Cooperative	Noncooperative	
Predicted			
Cooperative	7.28	6.72	14 (56)
Noncooperative	3.74	7.26	11 (44)
Total actual	11.02 (44)	13.98 (56)	25

round pictures (26). The overall accuracy rate is 63%. If we compare this 63% overall accuracy rate with the expected accuracy rate (50%), this is 13 percentage points higher. A binomial analysis of proportions revealed that these properties differ significantly. A Two Dependent Proportions Testing revealed a significant effect [$t(25)=11.50$, $p<.001$; Arsham, H.;1994. *A Two Dependent Proportions Testing*. Available: http://home.ubalt.edu/ntsbarsh/Business-stat/otherapplets/PairedProp.htm].

Study 2

A similar analysis was conducted for Study 2. From the 14 cooperative and 11 noncooperative proper-round pictures, our participants correctly classified 52% and 66%, respectively (Table 2). Consequently, from the 14 cooperative pictures, 7.28 pictures (52%) were correctly classified and 6.72 pictures (14; 2.28) were wrongly classified. From the 11 noncooperative pictures, 7.26 pictures (66%) were correctly classified while 3.74 pictures (7.26) were classified wrongly (Table 5).

Since the actual rates of cooperative and noncooperative pictures were 44% and 56% for subjects who randomly predicted cooperative pictures 56% of the time and noncooperative pictures 44% of the time, respectively (Table 6), the expected accuracy rate is:

$$(0.44)(56) + (0.56)(44) = 49\%.$$

This is 9% points lower than the than the 58% accuracy rate achieved [=(7.28+7.26)/25]. A Two Dependent Proportions Testing revealed a significant effect [$t(24)=7.23$, $p<.001$].

Table 6
Expected versus actual accuracy of cooperative and noncooperative proper-round and neutral-expression pictures (Experiment 1)

	Actual	
	Cooperative	Noncooperative
Cooperative	0.59	0.41
Noncooperative	0.34	0.66
	Expected (a priori)	
	Cooperative	Cooperative
Neutral round		
Cooperative	0.48	0.52
Noncooperative	0.52	0.48

Table 7
Expected versus actual accuracy of cooperative and noncooperative proper-round and neutral-expression pictures (Experiment 2)

	Actual	
	Cooperative	Noncooperative
Proper round		
Cooperative	0.52	0.48
Noncooperative	0.34	0.66
	Expected (a priori)	
	Cooperative	Cooperative
Neutral round		
Cooperative	0.49	0.51
Noncooperative	0.48	0.52

Appendix B. Study 2

Study 1

A sound method, suggested by one of our reviewers, to assess the impact of unbalanced random guessing on the bias towards proper-round noncooperative pictures documented here entails a comparison with cooperative and noncooperative pictures originating from neutral expressions. Since judgment of neutral pictures is random, the actual accuracy rates in the judgment of neutral pictures can be used as a priori bases (or expected accuracy rates) to which actual accuracy rates in the judgment of proper pictures are compared. These rates are presented in Table 6. By calculating for χ^2, this analysis again yields a significant effect [$\chi^2(1)=13.92$, $p<.001$].

Study 2

Similar to Study 1, we calculated for χ^2 and used the accuracy rate of cooperative and noncooperative neutral pictures as the (a priori) expected accuracy, as indicated in Table 7. This analysis yielded a significant effect [$\chi^2(1)=8.21$, $p<.001$].

References

Adolphs, R. (1999). Social cognition and the human brain. *Trends in Cognitive Science*, *3*, 469–479.

Adolphs, R. (2003). Cognitive neuroscience of human social behavior. *Nature Reviews Neuroscience*, *4*, 165–178.

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

Borkenau, P., Mauer, N., Riemann, R., Spinath, F. M., & Angleitner, A. (2004). Thin slices of behavior as cues to personality and intelligence. *Journal of Personality and Social Psychology*, *86*, 599–614.

Brown, W. M., & Moore, C. (2002). Smile asymmetries and reputation as reliable indicators of likelihood to cooperate: An evolutionary approach. *Advances in Psychological Research*, *11*, 59–78.

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.

Chiappe, D., & Brown, A. (2004). Cheaters are looked at longer and remembered better than cooperators in social exchange situations. *Evolutionary Psychology*, *2*, 108−120.

Cosmides, L. (1989). The logic of social exchange: Has natural selection shaped how humans reason? Studies in the Wason Selection Task. *Cognition*, *31*, 187−276.

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

Dawes, R., McTavish, J., & Shaklee, H. (1977). Behavior, communication, and assumptions about other people's behavior in a commons dilemma. *Journal of Personality and Social Psychology*, *35*, 1−11.

DePaulo, B. M., & Rosenthal, R. (1979). Telling lies. *Journal of Personality and Social Psychology*, *37*, 1713−1722.

Ekman, P., O'Sullivan, M., & Frank, M. G. (1999). A few can catch a liar. *Psychological Review*, *10*, 263−266.

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

Frank, R. H., Gilovich, T., & Regan, D. T. (1993). The evolution of one-shot cooperation: An experiment. *Ethology and Sociobiology*, *14*, 247−256.

Haley, K. J., & Fessler, D. M. T. (2005). Nobody's watching? Subtle cues affect generosity in an anonymous economic game. *Evolution and Human Behavior*, *26*, 245−256.

Hassin, R., & Trope, Y. (2000). Facing faces: Studies on the cognitive aspects of physiognomy. *Journal of Personality and Social Psychology*, *78*, 837−852.

Hirshleifer, J. (1987). On the emotions as guarantors of threats and promises. In J. Dupré, (Ed.), *The latest on the best: Essays on evolution and optimality* (pp. 307–326) Cambridge, MA: MIT.

Liggett, J. C. (1974). *The human face*. New York: Stein & Day; 1974.

Mealy, L., Daood, C., & Krage, M. (1996). Enhanced memory for faces of cheaters. *Ethology and Sociobiology*, *17*, 119−128.

Oda, R. (1997). Biased face recognition in the prisoner's dilemma games. *Evolution and Human Behavior*, *18*, 309−317.

Singer, T., Kiebel, S. J., Winston, J. S., Dolan, R. J., & Frith, C. D. (2004). Brain responses to the acquired moral status of faces. *Neuron*, *41*, 653−662.

Trivers, R. L. (1971). The evolution of reciprocal altruism. *Quarterly Review of Biology*, *46*, 35−55.

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

Vanneste, S., Verplaetse, J., Van Hiel, A., & Braeckman, J. (2007). Attention bias toward non-cooperative people. A dot probe classification study in cheating detection. *Evolution and Human Behavior*, *28*, 272−276.

Winston, J. S., Strange, B. A., O'Doherty, J., & Dolan, R. J. (2002). Automatic and intentional brain responses during evaluation of trustworthiness of faces. *Nature Reviews Neuroscience*, *5*, 277−283.

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*, 290−301.

Zebrowitz, L. A. (1997). *Reading faces: Windows in the soul?* Boulder, CO: Westview Press; 1997.