

# Attention bias toward noncooperative people. A dot probe classification study in cheating detection

Sven Vanneste<sup>a,\*</sup>, Jan Verplaetse<sup>b</sup>, Alain Van Hiel<sup>a</sup>, Johan Braeckman<sup>c</sup>

<sup>a</sup>Department of Developmental, Personality and Social Psychology, Ghent University, Ghent, Belgium

<sup>b</sup>Department of Jurisprudence and Legal History, Ghent University, Ghent, Belgium

<sup>c</sup>Department of Philosophy and Moral Sciences, Ghent University, Ghent, Belgium

Initial receipt 15 March 2006; final revision received 8 February 2007

## Abstract

Evolution-inspired research assumes the existence of brain mechanisms that scan for information that might signal noncooperative behavior. In this study, we demonstrate an automatic *attention bias* for threatening social interactions involving untrustworthy partners. Using a dot probe classification task, we found that, compared to unknown cooperators, attention was oriented significantly more toward the faces of unknown players who decided not to cooperate during a Prisoner's Dilemma Game. The present results thus suggest that an automatic, preconscious focus of attention underlies our ability to identify noncooperative players in social exchange situations.

© 2007 Elsevier Inc. All rights reserved.

*Keywords:* Attention bias; Cheating detection; Dot probe

## 1. Introduction

Evolution-inspired studies demonstrated that people are proficient when it comes to the detection of noncooperative behavior (Cosmides, 1989; Cosmides & Tooby, 1992). Using the Wason Selection Task, it has been revealed that people are particularly apt at detecting social contract rule violations (Cosmides, 1989; Cosmides & Tooby, 1992). Subsequent research also reported that people are able to memorize the face of a noncooperator more accurately than that of a cooperator (Chiappe & Brown, 2004; Mealy, Daood, & Krage, 1996; Oda, 1997; Yamagishi, Tanida, Mashima, Shimoma, & Kanazawa, 2003). Moreover, humans are, to a certain extent, able to predict crucial social decisions of others without ever having met before and even when information is strictly limited (Brown, Palameta, & Moore, 2003; Frank, Gilovich, & Regan, 1993).

Moreover, a recent study conducted by Verplaetse, Vanneste, and Braeckman (2007) found that snapshots

might be sufficient to predict the cooperativeness of a photographed target at above-chance levels. In this study, participants were asked to rate photographed targets, which were taken during a one-shot Prisoner's Dilemma Game (PDG), while being ignorant to the decisions they then made. Results showed that participants could, in fact, accurately discriminate noncooperative from cooperative players.

In another study (Verplaetse, Vanneste, & Braeckman, unpublished) that used the same photographs, it was further revealed that the pictures of noncooperative targets contained fear-related cues (e.g., fear, threat, anger). By picking up on these emotional cues elicited by noncooperative players, participants might be able to unmask noncooperators they have never met before. This presumption was further substantiated with the help of a group of independent raters ( $N=39$ ), which established a strong correlation ( $r=.74$ ,  $p<.001$ ) between noncooperative behavior and fear-related emotions, such as threat and fear. In agreement with this line of reasoning, our raters judged the pictures originating from noncooperators to be more fearful and more threatening.

Social interactions can be frightening as well. Research has shown abundantly that as soon as fearful features are detected, they immediately capture our attention, even

\* Corresponding author. Department of Developmental, Personality and Social Psychology, Universiteit Ghent, H. Dunantlaan 2, 9000 Ghent, Belgium. Tel.: +32 9 2649134.

E-mail address: sven.vanneste@ugent.be (S. Vanneste).

outside the realms of our conscious awareness (Öhman, 1986, 1993, 1997; Öhman & Mineka, 2001). Until now, this *attention bias* has been extensively investigated in response to explicit fear-related stimuli that are physically menacing or life-threatening, such as snakes or angry faces (Fox, Russo, & Dutton, 2002; Mogg & Bradley, 1999; Robinson, 1998). Studies examining the automatic focus of attention toward threatening social stimuli, such as evil faces, are rare (Stone & Valentine, 2004, 2005) or even nonexistent if one takes more covert fear-related stimuli into consideration, such as social interactions in which untrustworthy, noncooperative partners might be involved.

Here, we investigate whether an automatic attention bias equally applies to threatening social interactions involving untrustworthy partners during a one-shot PDG. A well-known paradigm used to investigate preconscious attention is the *dot probe classification task* (Broadbent & Broadbent, 1988; Mogg & Bradley, 1999). This task measures response latencies to probe stimuli (e.g., small dots), which are shown immediately after the presentation of a stimulus pair. The stimulus pair contains one expressive stimulus (e.g., a fearful face) and one neutral stimulus (e.g., a neutral-expression face). The probe stimuli can emerge in the same visual field as the expressive stimulus (*congruent* presentation) or in the opposite visual field of that stimulus (*incongruent* presentation). Allocation of attention is measured by the time needed to respond to a secondary classification task (e.g., vertical or horizontal dots). The logic of this task entails, first of all, that individuals respond quicker to congruent trials than to incongruent trials because attention was already allocated to the visual field where the probe appears. Second, it assumes that, if expressive stimuli have a preconscious influence, reaction times (RTs) will increase during incongruent trials since these stimuli attract more attention than neutral stimuli and more effort is required when focusing our attention on the opposite visual field.

If our hypothesis holds, an automatic attention bias should occur in response to threatening social stimuli, such as the faces of unknown noncooperators, resulting in larger latencies during incongruent trials of the dot probe task (a) when participants are confronted with pictures originating from noncooperative players and (b) when these pictures are taken at decisive moments during the proper round of the PDG. The latter category is necessary to verify that the automatic attention bias is, in some way, connected to subtle expressive cues that may be particularly salient in proper round pictures.

## 2. Method

### 2.1. Participants and design

Forty-two undergraduate social sciences students (23 female) from Ghent University (Belgium) took part in the experiment. Mean age was 19.24 years (S.D.=2.10; range=17–30). All participants had normal or corrected-

to-normal vision. A 2 (congruence: congruent/incongruent)×2 (picture type: practice/proper round)×2 (player type: cooperating/noncooperating) within-subjects design was used.

### 2.2. Material

A total of 48 pictures (height=50 mm; width=65 mm) originating from 16 individuals (8 women) were selected from Verplaetse et al. (2007). From these 16 individuals, we obtained three pictures, taken at separate moments during the course of a PDG with real money. Players were kept ignorant as to whom they were playing with. Before the game started, we collected neutral-expression pictures of our players, by asking them to relax their facial muscles. During the PDG, pictures were taken at two decision moments, one in the *practice round* without monetary gains and one in the *proper round* in which participants played for real money. With the help of a visible webcam in front of the computer, pictures were taken at the time when the participants made cooperative or noncooperative choices by means of a mouse click. Using Photoshop Elements 2.0, we edited all pictures to obtain a white, neutral, and equal background. Ultimately, our set contained 16 *practice round pictures* and 16 *proper round pictures* from 8 noncooperators and 8 cooperators. Also, to complete the dot probe design, we used 16 *neutral-expression pictures* from the same targets. Three photos of a single target are given as an example (see Fig. 1). In our dot probe design, each practice and proper round picture was paired with the neutral-expression picture of the same person, resulting in a total number of 32 *stimulus pairs*.

### 2.3. Procedure

Participants were tested individually. Before the experiment started, they were informed that they took part in a “visual perception” task and received the standard instructions about the dot probe classification task. The dot probe was either vertical (:) or horizontal (..). Participants were required to classify the dot probe by pressing one of two buttons—a red one on the left (Ctrl-L) and a blue one on the right (Ctrl-R)—as quickly and accurately as possible. Buttons were counterbalanced across participants. Each trial began with a fixation cross (+), presented during 1000 ms in the middle of the screen, followed by the stimulus pair for 500 ms. One picture of the stimulus pair was shown above and the other was shown beneath the central fixation point. The distance between their exterior edges was 50 mm. The

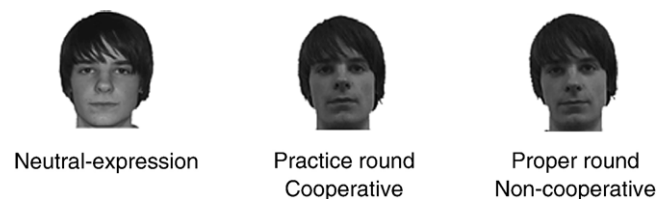


Fig. 1. Illustrative picture of a face with neutral and other expressions for the practice round and the proper round.

probe remained on the screen until a response was made. After an interval of 200 ms, the following trial began.

Participants practiced the dot probe classification task during 20 trials and saw pictures that were different from those during the experimental trials. Subsequently, we randomly presented 256 experimental trials to each participant. Each of the 32 stimulus pairs was presented eight times. Location of stimuli and probe and the type of probe were counterbalanced across trials. Before debriefing, the reliability of the task manipulation (visual perception task) and participants' awareness of the impact of the displayed stimuli were rated on a 5-point Likert scale (1=totally disagree to 5=totally agree). The questions were as follows: "Do you believe this task was a test to verify your visual perception skills?" and "Do you think that certain pictures draw more attention than other pictures?"

### 3. Results

#### 3.1. Manipulation check

Participants believed that they were doing a visual perception task (mean=4.35, S.D.=0.65) and tended to deny that particular pictures attract more attention than others (mean=1.51, S.D.=0.42).

#### 3.2. RT data

RTs diverging more than 3 S.D. from the individual mean were expelled as outliers (1.45%). The percentages of incorrect responses were 1.9% and 1.5% in the case of practice round pictures (cooperative and noncooperative targets, respectively) and 1.0% and 1.3% in the case of the actual round pictures (cooperative and noncooperative targets, respectively). Data from participants who had lost more than 20% of trials due to errors and outliers were not taken into account for further analysis. Based on these criteria, one participant was excluded from further analysis. Moreover, we controlled for possible extreme expressions

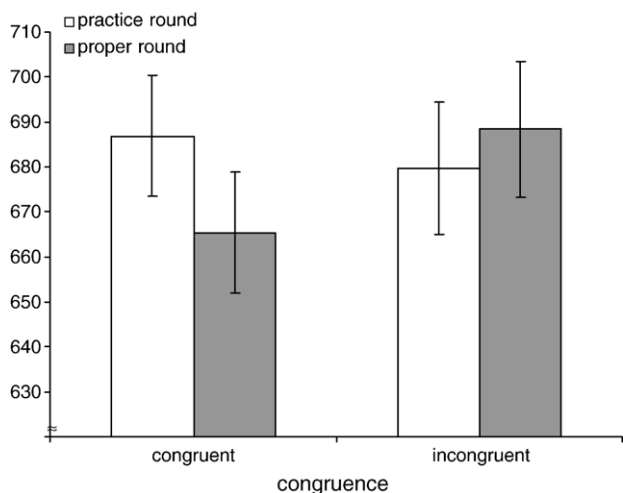


Fig. 2. Mean (and S.E.M.) RT as a function of congruence for the practice round and the proper round.

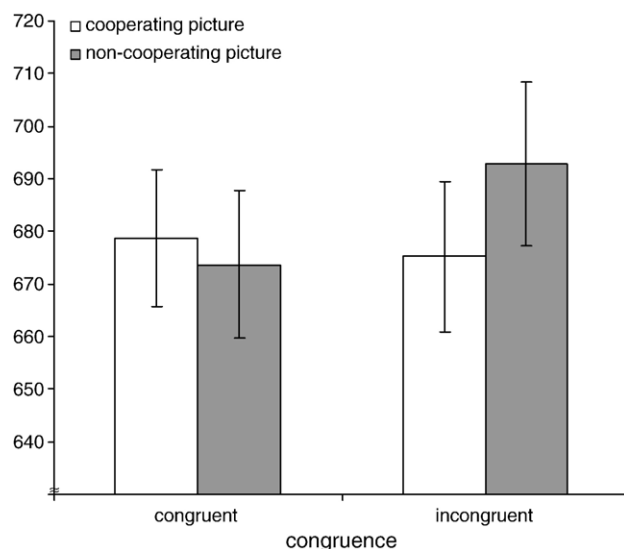


Fig. 3. Mean (and S.E.M.) RT as a function of congruence for the cooperating and cheating behavior.

by verifying whether certain pictures had more extreme latencies for all participants. On this basis, no pictures needed to be excluded.<sup>1</sup>

The overall mean RT was 669 ms (S.D.=87). RTs were analyzed using a 2 (congruence: congruent/incongruent) × 2 (round type: practice/proper round) × 2 (player type: cooperative/ noncooperative targets) repeated measures analysis of variance. This analysis revealed a significant interaction effect for Round Type × Congruence,  $F(1, 40)=19.97, p<.001, \eta^2=.30$  (see Fig. 2). Simple contrasts showed that mean RTs for incongruent trials (mean=688 ms, S.D.=99) were slower than congruent trials (mean=665 ms, S.D.=89) for proper round pictures,  $F(1, 40)=18.75, p<.001, \eta^2=.32$ , but not for the practice round pictures, respectively, mean=687 ms (S.D.=89) and mean=680 ms (S.D.=96),  $F(1, 40)=1.57, p=.21, \eta^2=.02$ .

A significant interaction effect was also found for Congruence × Player Type,  $F(1, 40)=7.66, p<.05, \eta^2=.16$  (see Fig. 3). Simple contrasts revealed that congruent trials (mean=673 ms, S.D.=92) were done faster than incongruent trials (mean=692 ms, S.D.=101) for pictures originating from noncooperative players,  $F(1, 40)=24.37, p<.001, \eta^2=.37$ . For pictures taken from cooperators, no significant differences were found between congruent (mean=678 ms, S.D.=85) and incongruent trials (mean=675 ms, S.D.=94),  $F(1, 40)=0.22, p=.64, \eta^2=.01$ .

### 4. Discussion

Using a dot probe design, the present study has shown that pictures of unfamiliar noncooperative players attract

<sup>1</sup> We further verified whether significant differences were obtained among stimulus persons within the category of cooperators and non-cooperators, but these analyses did not yield significant results.

more automatic attention than pictures of cooperative people. More precisely, significant RT differences were found between congruent and incongruent trials, comparing pictures taken during PDGs with and without monetary gains. More important, the congruency effect was also qualified by type of player, indicating that pictures of noncooperative players elicited more attention.

A plausible reason for the first effect is the notion that stronger emotional cues are present in situations where a monetary gain is at stake. Indeed, in the proper round, the targets anticipated to win or lose an amount of money, and the importance they attach to this may have caused greater emotional reactions (see Hertwig & Ortmann, 2001), which may, in turn, be translated into visible subtle facial expressions. These expressions may have grabbed the attention of our participants, to a greater extent than the less expressive faces of these targets in the practice round. As a consequence, pictures in the proper round yielded greater response latencies compared to the practice round.

These results thus highlight the importance of preconscious processing, especially when important and ecologically valid stimuli are involved. Indeed, dot probe tasks interrupt the conscious processing of attention allocation with the help of a secondary classification task. In fact, these tasks were especially designed to investigate automatic processes of preconscious attention, and the present results therefore suggest that predictive cheating detection is based upon the automatic process of shifting attention rather than on the conscious allocation of attention.

From an evolutionary point of view, it seems very unlikely that noncooperative people display revealing cues at all times. If it would be possible to distinguish noncooperators from cooperators on the basis of permanent facial features, natural selection would drive the noncooperators toward extinction (Brown & Moore, 2002; Frank et al., 1993; Trivers, 1985). In a transparent world, cooperators would always get the highest payoffs. Consequently, noncooperators would only survive when they learn to fake cooperative intentions and mimic altruistic expressions. Since these imitated expressions become unreliable for cooperators, evolution will favor the selection of more subtle expressions and/or more sophisticated detection skills. Taking this ongoing oscillation between signal detection and signal deception in consideration, it is highly unlikely that cues of a less subtle nature, such as permanent facial features or voluntarily produced expressions, are useful to distinguish noncooperators from cooperators. If humans are able to identify noncooperators to a certain extent, this ability is more likely based upon the intake of more delicate signals that capture our attention (Brown & Moore, 2002; Verplaetse et al., 2007). In line with this reasoning, this study confirms that pictures taken at decisive moments contain more reliable emotional cues and are prioritized for attentive processing.

The second finding—the attention-grabbing effect of cheating—corroborates the well-documented attention bias

reported in previous studies using the dot probe task (e.g., Fox et al., 2002; Mogg & Bradley, 1999; Öhman, 1986, 1993, 1997; Öhman & Mineka, 2001; Robinson, 1998). Indeed, using exactly the same methods as in the present study, these previous studies have repeatedly shown that threatening stimuli automatically attract attention, and along similar lines, our study additionally shows that faces of unknown players who decided to defect during a one-shot PDG equally attract attention because of more subtle social features.

The present finding can therefore be attributed to an attention shift toward fear-related emotions (fear, threat, anger) that noncooperative players display. This interpretation is also in line with LeDoux (1996) who asserted that the detection of fear-related stimuli operates at a preattentive stage and automatically triggers appropriate behavioral reactions. In addition, Mogg and Bradley (1999) developed theoretical models that rest upon the assumption that fear-related stimuli are essential to achieve this attention bias, showing that fear-related stimuli are prioritized for attentive processing. Furthermore, Verplaetse et al. (unpublished) were able to show a strong correlation between the cooperativeness of a target and the ratings of fear-related emotions, such as threat and fear. This finding further strengthens the assumption that fear-related cues might be a clue to a better understanding of the precise nature of cheating detection.

The remainder of this discussion will primarily focus on our main result, namely, that pictures of noncooperative players attract more automatic attention than pictures of cooperative people. This outcome is in accordance with various predictions made by evolution-inspired theory. Firstly, since noncooperative deals menace social exchange in a dramatic way, evolution provided humans with cheating detection mechanisms. People are particularly apt at determining when a social contract rule has been violated (Cosmides, 1989; Cosmides & Tooby, 1992), are able to memorize the face of a cheater more accurately than that of a noncheater (Chiape & Brown, 2004; Mealy et al., 1996; Oda, 1997; Yamagishi et al., 2003), and can accurately predict how willing another might be to cooperate (Bond, Berry, & Omar, 1994; Brown et al., 2003; Frank et al., 1993; Verplaetse et al., 2007). A biased orientation to signals of noncooperativeness might be one of the basic mechanisms that underlie the cognitive process of successful cheating detection. By shifting our attention automatically to emotional cues that might reveal people's nonaltruistic intentions, evolutionary adaptation has directed our minds, by default, to crucial social information enhancing our survival chances. To humans, preferential and preconscious attention for subtle noncooperative features might be equally life-rescuing/sustaining as avoiding snakes or angry faces.

Secondly, previous research has revealed that our cheating detection abilities are biased toward those individuals who threaten cooperation for mutual benefit. In human

social life, noncooperative people threaten the viability of social exchange due to the advantages of accepting benefits without paying costs (Axelrod, 1984; Cosmides & Tooby, 1992; Hamilton, 1964). Our finding shows that attention is biased toward noncooperative facial features and, hence, further substantiates the existence of a *cheater bias* in attention allocation (Chiappe & Brown, 2004; Mealy et al., 1996; Oda, 1997).

Finally, Verplaetse et al. (2007) indicated that *predictive cheating detection* did not require overt task motivation and conscious orientation of attention and focus. In the latter study, postexperimental questionnaires revealed that participants had no idea how they detected noncooperators. In addition, they replied that the detection of noncooperators 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%). This striking contrast between estimated competence and task performance additionally suggests that predictive cheating detection might be based upon the automatic process of shifting attention to emotionally relevant stimuli, as found in this study. Due to nonconscious biases toward fear-related facial cues, participants are able to discriminate noncooperators from cooperators, although they are not aware of their own expertise when doing so. Although the relationship between the enhanced attentive orientation and the accuracy and confidence in discriminating noncooperators from cooperators remains far from clear, our main finding supports our presumptive idea that a preconscious attention bias underlies predictive cheater detection.

In summary, using a dot probe classification task, we found that the well-documented attention bias equally applies to threatening social stimuli, more specifically to faces of unknown players who decided to defect during a one-shot PDG. It is suggested that this automatic shift of attention toward faces of noncooperators helps to explain our ability to identify noncooperators in social exchange situations. From an evolutionary point of view, it makes perfect sense to expand the scope of fear-related stimuli to covert social cues such as the possible untrustworthiness of partners during social exchanges.

## Acknowledgments

We thank Margo Wilson, Andy Fiddick, Farah Focquaert, and Anja Demeulenaere for very helpful comments on an earlier version of the manuscript.

## References

- Axelrod, R. (1984). *The evolution of cooperation*. New York: Basic Books.
- 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.
- Broadbent, D. E., & Broadbent, M. (1988). Anxiety and attentional bias: State and trait. *Cognition and Emotion, 2*, 165–183.
- 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: Oxford University Press.
- Fox, E., Russo, R., & Dutton, K. (2002). Attentional bias for threat: Evidence for delayed disengagement from emotional faces. *Cognition and Emotion, 16*, 355–379.
- Frank, R. H., Gilovich, T., & Regan, D. T. (1993). The evolution of one-shot cooperation: An experiment. *Ethology and Sociobiology, 14*, 247–256.
- Hamilton, W. D. (1964). The genetical theory of social behavior. *Journal of Theoretical Biology, 7*, 1–52.
- Hertwig, R., & Ortmann, A. (2001). Experimental practices in economics: A methodological challenge for psychologists. *Behavioral and Brain Sciences, 24*, 383–451.
- LeDoux, J. E. (1996). *The emotional brain*. New York: Simon and Schuster.
- Mealy, L., Daoood, C., & Krage, M. (1996). Enhanced memory for faces of cheaters. *Ethology and Sociobiology, 17*, 119–128.
- Mogg, K., & Bradley, B. P. (1999). Orienting of attention to threatening facial expressions presented under conditions of restricted awareness. *Cognition and Emotion, 13*, 713–740.
- Oda, R. (1997). Biased face recognition in the prisoner’s dilemma games. *Evolution and Human Behavior, 18*, 309–317.
- Öhman, A. (1986). Face the beast and fear the face: Animal and social fears as prototypes for evolutionary analyses of emotion. *Psychophysiology, 23*, 123–145.
- Öhman, A. (1993). Fear and anxiety as emotional phenomena: Clinical phenomenology, evolutionary perspectives and information processing mechanisms. In L. Lewis, & J. M. Haviland (Eds.), *Handbook of emotions* (pp. 511–536). New York: Guilford Press.
- Öhman, A. (1997). As fast as the blink in the eye: Evolutionary preparedness for preattentive processing of threat. In P. J. Lang, R. F. Simons, & M. Balaban (Eds.), *Attention and orienting: Sensory and motivational processes* (pp. 165–184). Hillsdale, NJ: Erlbaum.
- Öhman, A., & Mineka, S. (2001). Fears, phobias, and preparedness: Toward an evolved module of fear and fear learning. *Psychological Review, 108*, 483–522.
- Robinson, M. D. (1998). Running from William James’ Bear: A review of preattentive mechanisms and their contributions to emotional experience. *Cognition and Emotion, 12*, 667–696.
- Stone, A. M., & Valentine, T. (2004). Better the devil you know? Non-conscious processing of identity and affect of famous faces. *Psychonomic Bulletin and Review, 11*, 469–474.
- Stone, A. M., & Valentine, T. (2005). Orientation of attention to nonconsciously recognised famous faces. *Cognition and Emotion, 19*, 537–558.
- Trivers, R. L. (1985). *Social evolution*. Menlo Park: Benjamin/Cummings.
- Verplaetse J., Vanneste S., & Braeckman, J., (2007). You can Judge a Book by its Cover: The Sequel. A Kernel of Evolutionary Truth in Predictive Cheating Detection. *Evolution and Human Behavior, 28*, 260–271.
- Verplaetse J., Vanneste S., & Braeckman, J., (unpublished). The role of fear in detecting cheaters (unpublished).
- 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.