ELSEVIER

Evolution and Human Behavior 24 (2003) 290–301

Evolution and Human Behavior

You can judge a book by its cover Evidence that cheaters may look different from cooperators

Toshio Yamagishi[a,b,*], Shigehito Tanida[b], Rie Mashima[b], Eri Shimoma[b], Satoshi Kanazawa[c,1]

[a] *Center for Advanced Study in the Behavioral Sciences, Stanford,CA*
[b] *Hokkaido University, N10 W7 Kita-ku, Sapporo 060-0810, Japan*
[c] *Interdisciplinary Institute of Management, London School of Economics and Political Science, Houghton Street, London WC2A 2AE, UK*

Received 18 November 2002; received in revised form 24 May 2003

Abstract

Cosmides and Tooby argue that humans possess a domain-specific cheater detection module, which allows them to keep track of who has honored and who has violated social contracts. Consistent with this logic, others demonstrate that humans better recognize faces of known cheaters than those of known cooperators. We show, in Experiments 1–3, that humans better recognize faces of cheaters than those of cooperators when they do not know who are cheaters and cooperators. Experiment 4 demonstrates, however, that humans think they recognize cheaters' faces even when they have not seen them before. The results of these experiments suggest that cheaters might look different from cooperators, possibly due to beliefs and personality traits that make them less ideal exchange partners, and the human mind might be capable of picking up on subtle visual cues that cheaters' faces give off.

Keywords: Face recognition; Cheater detection; Prisoner's dilemma; Cooperation

* Corresponding author. Graduate School of Letters, Hokkaido University, N10 W7 Kita-ku, Sapporo 060-0810, Japan.
E-mail addresses: toshio@let.hokudai.ac.jp (T. Yamagishi), s.kanazawa@lse.ac.uk (S. Kanazawa).
[1] Also corresponding author.

1090-5138/03/$ – see front matter
doi:10.1016/S1090-5138(03)00035-7

1. Introduction

Cosmides (1989) and Cosmides and Tooby (1992) have argued that humans possess a cognitive mechanism for social exchange in the form of a domain-specific cheater detection module, which allows them to keep track of who has honored and who has violated social contracts in the past. Consistent with this logic, Mealey, Daood, and Krage (1996) and Oda (1997) found that people remember faces of known cheaters better than those of known cooperators. Mealey et al.'s (1996) subjects were shown pictures of men fictitiously labeled either as trustworthy or untrustworthy, whereas Oda's (1997) subjects saw pictures of individuals with fictitious information indicating whether they had cooperated or defected in a Prisoner's Dilemma Game (PDG). In both experiments, the subjects were then shown the original photographs together with a set of filler pictures and asked which faces they had seen before. Recognition rates for faces of putatively untrustworthy men or defectors were higher than for faces of putatively trustworthy men or cooperators.

In this paper, we address a related but separate question of whether people remember real defectors' faces better than those of cooperators without being told who are defectors or cooperators. In the experiments by Mealey et al. (1996) and Oda (1997), subjects were provided with explicit, fictitious information about the prior behavior of each photographed person, with stimuli counterbalanced in such a way that better recall of the faces of alleged cheaters could not reflect any difference in the memorability of the faces themselves. Here, we ask whether people recognize faces of real cheaters better than those of real cooperators without explicit designation of the faces as cheaters or cooperators. In Experiments 1–3, we show that humans indeed remember faces of real cheaters better than those of cooperators. Experiment 4 demonstrates, however, that people think they recognize the faces of cheaters even when they have not seen them before. The results of these experiments suggest that cheaters actually look different from cooperators, possibly due to certain beliefs and personality traits that make them less ideal exchange partners, and that the human mind is capable of picking up on subtle visual cues that cheaters' faces give off.

2. Stimulus pictures

The stimulus pictures used in the following experiments are those of the participants in Kikuchi, Yamagishi, and Yamamoto's (1998) one-shot PDG experiment and in Matsuda and Yamagishi's (2001) repeated PDG experiment. Participants in Kikuchi et al.'s study played a one-shot PDG with real money against one unidentified member of a six-person group. The cooperation rate in the PDG was 41%. After they made the cooperation–defection decision, they completed a postexperimental questionnaire. Then, they were paid the money they had earned in the experiment and were asked if they would agree to have a facial photograph taken for later use. Eighty-seven of 90 participants agreed, and they were photographed in front of a white wall.

For Experiment 1, 36 pictures (18 males and 18 females) were selected as stimulus pictures from the pool of 87 pictures, based on their clarity and a lack of obtrusive features such as

eyeglasses; 18 were used as target pictures and 18 as filler pictures (see Section 3.1 for details). In Experiment 2, 56 pictures (28 males and 28 females) were selected as target pictures. In Experiment 4, the same 56 pictures were used as target pictures or filler pictures depending on the experimental condition.

In contrast to the pictures used in Experiments 1, 2, and 4, which were taken after participants of a one-shot PDG had made a decision and filled out a rather lengthy postexperimental questionnaire, the pictures used in Experiment 3 were taken at the exact moment when participants in a repeated PDG (Matsuda & Yamagishi, 2001) made their decisions between cooperation and defection. These participants played a one-shot PDG repeatedly with a different anonymous partner each time. They made the decision between cooperation and defection based on the information provided to them on their computer monitor and made their decision with a click of a mouse. A digital camera placed on top of the monitor took the picture of the participant's face at the moment of the click of the mouse. One hundred and fifty participants made decisions 24 times each in this repeated one-shot PDG condition. Because the participant's body and face were not fixed to a particular position, and the camera did not follow the participant's face, many of the pictures failed to capture the participant's face clearly and thus were unusable. We selected 48 participants with at least three usable pictures of the cooperation decision and three of the defection decision.

We then divided these 48 participants into those with cooperation rate in the upper 40% of the entire 150 participants (cooperators) and those in the lower 40% (defectors). We then showed 45 pictures of the 15 cooperators (three pictures each) and 45 pictures of the 15 defectors in random order to 30 judges and asked them on each picture if they thought the photographed person made a cooperation or defection decision. We selected one picture out of three pictures for each cooperator that was judged to be most cooperative, and one for each defector that was judged to be least cooperative, and used these 15 cooperators' pictures and 15 defectors' pictures as the target pictures for Experiment 3. Thirty pictures in the middle 15% of the overall cooperation rates were used as filler pictures.

Although the target persons in the stimulus pictures come from the same general population as most (though not all) of the experimental subjects, several factors make it very unlikely that the subjects recognized the targets from interactions with fellow students on campus. First, Hokkaido University is one of the largest public universities in Japan, with more than 10,000 undergraduate students. Second, Sapporo, where Hokkaido University is located, is a metropolitan area with a population of about 2 million. Third, and most important, none of the stimulus pictures were ever used until 4 years after they were taken, giving the target persons a high probability of having graduated and left the campus.

3. Experiment 1

3.1. Procedure

Thirty-seven subjects (18 male, 19 female) participated. Each participant was led to a private room with a networked computer, on which all instructions were displayed. For

distraction, "emblem sessions" were inserted between the picture presentation session and the picture recognition session. The experiment started with the first picture presentation session in which 18 faces (either all males or all females) were presented on the computer at once in a 3×6 grid for 30 s. Each picture was 4.8×4.4 cm (portrait). The first face presentation session ended after 30 s, and the first emblem session began a few seconds later. In the emblem session, 18 emblems of European aristocratic families were first presented in a 3×6 grid for 30 s. After the emblem presentation session was over, an emblem recognition session began in which one emblem was presented at a time and the subject was asked to judge if he or she had seen the emblem in the preceding presentation session. The subject judged a total of 18 emblems, nine of which were presented in the emblem presentation session and nine of which were new, filler emblems. The entire emblem session (both presentation and recognition) took about 3 min.

After this distraction task, the first face recognition session began. One picture (5.6×5.0 cm, portrait) was displayed at a time, and the subject was asked to judge if he or she had seen the face before. The subject judged a total of 18 pictures, nine of which were from the first face presentation session, while the other nine were new, filler pictures. Subjects received feedback of "correct" or "incorrect" for each judgment. Another cycle of a second face presentation session, a second emblem session, and a second face recognition session followed the end of the first face recognition session. For half the subjects, the first face session involved male faces and the second involved female faces, with this order reversed for the other half.

3.2. Results

Disregarding the distinction between cooperators and defectors, the rate of true positive identification (where subjects correctly identified pictures they had earlier seen as such) was 0.66, and the rate of false positives (where subjects erroneously believed they had seen pictures in the earlier set that they had not) was 0.21. A measure of discrimination, d', is 1.28. Thus, subjects could generally discriminate faces that they had seen before from those they had not seen. As shown in Fig. 1, the true positive ID rate for defectors was greater than that for cooperators (0.67 vs. 0.64), but in the analysis of variance (ANOVA), there was no significant main effect of the target type (cooperator vs. defector), $F(1,36) = 0.78$. Subjects did, however, recognize faces of male targets better than those of female targets (0.69 vs. 0.63), $F(1,36) = 5.85$, $P < .05$. The interaction between target type and target sex was not significant, $F(1,36) = 3.08$, $P < .09$, but separate analyses of male and female target pictures indicated that subjects recognized defectors better than cooperators among the female targets (0.67 vs. 0.55, $t(36) = 2.08$, $P < .05$), but not among the male targets (0.71 vs. 0.67).

The rate of false positives was again slightly higher for the defectors' faces (0.23) than for the cooperators' faces (0.19), but the difference is not statistically significant. The Target Type × Target Sex ANOVA revealed no significant effects on the false positive rate, nor did an ANOVA on the d' index.

The results of Experiment 1 suggest that cheaters may look different from cooperators, at least among female faces, and that humans may be able to pick up on the subtle difference to better remember cheaters' faces than cooperator's faces, but the effect was weak at best.

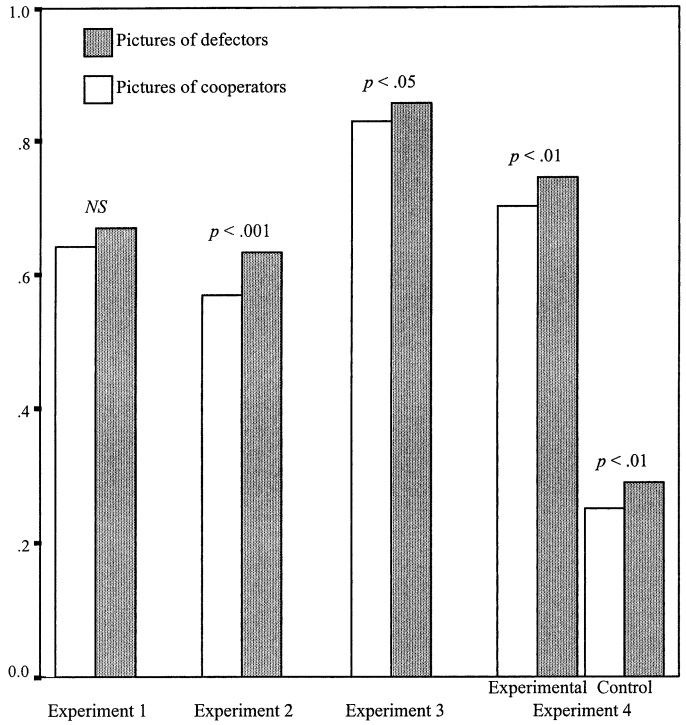

Fig. 1. Recognition rates for pictures of cooperators and defectors.

Because the weakness of the evidence may have been due to the small number of subjects and target pictures used in this initial experiment, we replicated the same experiment with a larger number of subjects and target pictures.

4. Experiment 2

4.1. Procedure

Fifty-five subjects (6 males, 49 females) participated in the experiment in two groups. The number of target pictures was increased from 18 to 56. These included the 18 target pictures used in Experiment 1, and 38 others, also drawn from among participants in Kikuchi, Yamagishi, and Yamamoto (1998) who had either cooperated or defected on a one-shot PDG. The same number of cooperators and defectors (14 pictures each) were included for each sex of the target. The experiment took place in a classroom rather than in private rooms in a laboratory.

Subjects viewed each of the 56 pictures one at a time projected on a screen in a random order for 3 s each. After seeing all 56 pictures, they received a 14-page brochure containing 112 pictures. Each page contained eight pictures of one sex, with order of pictures within the page

randomized between subjects. Subjects were told (truthfully) that some of the pictures in the brochure were from the set that they had seen on the screen, and others not, and were instructed to mark which ones they had seen before, by encircling either "YES" or "NO" by each picture.

In order to maximize the reliability of our recognition rate measures, we used all "usable" pictures (those in which the target did not have distinctive features like eyeglasses) of individuals whose choices in a previous PDG were known to us as the target pictures in this experiment. This forced us to use as fillers newly taken photographs of individuals who had not participated in a PDG and thus were not known to us as either cooperators or defectors. We are therefore unable to calculate the false positive rates or d' separately for cooperators and defectors in this experiment. Potential problems arising from this feature of the experimental design are addressed in Experiment 4.

4.2. Results

The correct recognition (true positive) rate was 0.60, while the false positive rate was 0.22 ($d' = 1.15$), again demonstrating the subjects' ability to recognize faces correctly. The correct recognition rate was significantly higher for defectors (0.63) than for cooperators (0.57), $F(1,54) = 14.97$, $P < .001$. The main effect of target's sex and the interaction effect between the type and sex of the target were not statistically significant. Thus, unlike Experiment 1, the results of Experiment 2 demonstrate that the subjects can better recognize faces of defectors of both sexes.

Some research suggests that the differential recognition rates for defectors and cooperators found in this experiment could be spurious due to physical attractiveness of the target person, although the relevant results have been inconsistent. According to Mulford, Orbell, Shatto, and Stockard (1998), physically attractive women are more likely to defect than less attractive women, although Kikuchi, Yamagishi, Shinagawa, and Sasaki (1998) failed to replicate this finding using Japanese subjects. Moreover, Shepherd and Ellis (1973) found that faces that were either high or low in attractiveness were easier to recognize than those of moderate attractiveness, although Little, Hollander, and Kayra-Stuart (1981) found that faces of attractive men were more difficult to remember because their features were less distinct and closer to the population average. To examine the possibility that better recognition of defectors than cooperators reflected differential attractiveness, we analyzed physical attractiveness of the pictures used in this experiment. Note that the unit of analysis for the following analyses is the picture, not the subject.

The participants in Kikuchi, Yamagishi, and Yamamoto's (1998) experiment evaluated the other members of their group on physical attractiveness. In addition, physical attractiveness of their pictures has been evaluated by three groups of judges (38 college students, 12 students of a lecture, and 54 subjects of the current experiment). Each picture thus has four mean attractiveness scores provided by independent groups of judges. We use the unweighted means of the standardized scores of those four average scores as the physical attractiveness measures for the target pictures used in this experiment. The average attractiveness score of the 28 defectors (0.30, S.D.=.82) is higher than that of the 28 cooperators (-0.08, S.D.=.70), but the difference is not statistically significant, $F(1,52) = 3.54$, $P < .07$. In a multiple

regression analysis, with the true recognition rate as the dependent variable and target type, target sex, and the attractiveness score of the picture as independent variables, the effect of target type remains significant, $b = -0.080$, $t(52) = 3.06$, $P < .01$. The independent effect of attractiveness is also significant, $b = -0.036$, $t(52) = 2.25$, $P < .05$. These results indicate that differential recognition of defectors vs. cooperators was not simply an artifact of differential attractiveness, although this factor may have played some role.

We also examined the relationship between the targets' responses to the postexperimental questionnaire in Kikuchi, Yamagishi, and Yamamoto (1998) and their correct recognition rates in the present study. As mentioned before, Kikuchi et al.'s participants answered a set of questions before their pictures were taken. Furthermore, their pictures were judged by 38 college students in a separate study with respect to whether they thought the targets had cooperated or defected in a PDG, if they wanted to play a PDG with them, and if they wanted to cooperate with them given a chance to play a PDG with them. We correlated the target-specific correct recognition rates with the targets' responses to the postexperimental questionnaire and with the average evaluations of the 38 judges, with interesting results: the targets who were recognized at a higher rate were those who (1) expected less cooperation from the other five members of their group ($r = -.35$, $P < .01$), (2) regarded the partner's welfare as less important ($r = -.30$, $P < .05$), (3) considered cooperators less clever than defectors ($r = -.30$, $P < .05$), and (4) the judges did not desire to cooperate with ($r = -.28$, $P < .05$). Controlling for these variables in a multiple regression eliminates the effect of target type on the recognition rate, $b = -.001$, $t(43) = .01$, ns.

We then examined the possibility that differential recognition rate was produced by a small number of outliers. Because there were two pictures of defectors for which recognition rates were very high, we repeated the analyses without these two outliers. The correct recognition rate for the remaining defectors (0.62) was still significantly higher than that for the cooperators (0.57), $F(1,54) = 9.45$, $P < .01$. We also examined potential effects of outlier subjects (as opposed to outlier targets) by correlating picture-specific recognition rates with subjects' binary responses to the picture (recognize or not), and we found four subjects who had negative correlations with the average recognition rate. If we eliminate these subjects and recalculate the recognition rate, the main effect of target type on this recalculated recognition rate is still significant, $F(1,50) = 16.16$, $P < .001$.

Finally, to reduce the influence of extreme recognition rates in either direction, we also categorize the recognition rates into three groups: high, medium, and low. The numbers of the high, medium, and low recognition pictures are 14, 8, and 6, respectively, among the defectors, and 5, 12, and 11, respectively, among the cooperators. These ratios are significantly different in the two target type conditions, $\chi^2(2) = 6.53$, $P < .05$, indicating that our results are not attributable to either extremely high or extremely low recognition rates of a few pictures.

5. Experiment 3

This experiment used an entirely different set of target pictures from Matsuda and Yamagishi (2001). In that experiment, participants played repeated one-shot PDGs on

computer terminals and were photographed by a small camera placed on top of the monitor at the moment they made cooperative or defecting choices with a click of the mouse. We used pictures of 15 high cooperators (eight males, seven females) and 15 low cooperators (eight males, seven females) as target pictures.

5.1. Procedure

In the presentation session, 75 subjects (48 males, 27 females) viewed pictures of the 30 targets on a computer, one at a time in a randomized order, for 3 s each, with a 1-s interval between pictures. The subjects were told that the amount of their reward for participation depended on how many pictures they correctly remembered, and were thus motivated to remember as many faces as possible. In the recognition phase, the subjects viewed 60 pictures (30 pictures they had seen and 30 filler pictures) in four sets of 15 pictures in 3×5 grids and indicated whether they had seen each of them in the previous session. The pictures were 5.6×5.0 cm (portrait) in the presentation session, and 4.8×4.4 cm (portrait) in the recognition session. As in Experiment 2, the experimental design allowed us to calculate only the true positive recognition rate, and not the false positive rate or d' separately for cooperators and defectors. Again, this design feature is a consequence of using the maximum number of pictures for detecting differential recognition rates (HIT rates) of cooperators and defectors.

5.2. Results

The correct recognition rate was 0.84 and the false positive rate 0.13 ($d' = 2.25$), again indicating genuine recognition. In a 2×2 ANOVA, target type had a significant main effect on correct recognition (0.86 vs. 0.83, $F(1,71) = 4.32$, $P < .05$); the main effect of target's sex and the interaction effect were not significant. The results of Experiment 3 thus replicate those of Experiment 2, by indicating that human subjects can recognize the faces of defectors more accurately than those of cooperators.

In order to examine possible effect of targets' attractiveness on recognition rates, we asked a group of 18 judges (not the subjects of this experiment) to evaluate physical attractiveness of the 30 pictures used in the experiment. A Target Type × Target Sex ANOVA of the attractiveness ratings, using the picture as the unit of analysis, produced no significant effect, and in this case, the defectors' average attractiveness (2.05, S.D.=.52) was almost identical to that of cooperators (2.06, S.D.=.61), thus eliminating the possibility that the differential recognition rates of defectors and cooperators resulted from differential attractiveness.

Analysis of the targets' responses to questionnaire items in Matsuda and Yamagishi's (2001) experiment revealed that individuals who were recognized at higher rates were those who (1) did not trust the partners in the PDG ($r = -.36$, $P < .06$), (2) did not think that the partners trusted them ($r = -.39$, $P < .05$), (3) expressed less desire to play against the same partners again ($r = -.38$, $P < .05$), (4) regarded mutual cooperation as less important ($r = -.37$, $P < .05$), (5) regarded joint gains in the PDG as less important ($r = -.44$, $P < .05$), (6) considered cooperators less clever than defectors ($r = -.31$, $P < .10$), (7) regarded mutual help as less of the reality in the social life ($r = -.40$, $P < .05$), and (8) were

less willing to help others in situations where there is no one around to witness it ($r = -.37$, $P < .05$). As in Experiment 2, controlling for these variables in multiple regression reduces the effect of the target type on the true positive recognition rate to a nonsignificant level, $b = 1.09$, $t(20) = .18$, ns.

There were no outliers among the target pictures in terms of recognition rates, so we did not conduct the same picture outlier analysis as in Experiment 2. However, we did conduct a similar subject outlier analysis and found four subjects whose recognition patterns correlated negatively with the average recognition rates. Eliminating those four did not affect the result much, although the effect of the target type was weakened, $F(1,67) = 3.89$, $P < .06$.

6. Experiment 4

The results of Experiments 1–3 showed that defectors' faces were remembered better than those of cooperators: Although differential recognition was only marginal in Experiment 1, it was highly significant in Experiments 2 and 3, using different stimulus materials and different presentation methods. However, one question remains unanswered in the three experiments. Filler pictures included in the recognition phases of the Experiment 2 (pictures not originally shown to the subjects, but later added to see if they could select only the ones they had seen before) belonged to individuals who had not played a PDG and who were therefore not identifiable as either cooperators or defectors. Filler pictures in Experiment 3 were of the medium cooperators. As a result, although subjects in these experiments were significantly more likely to recognize pictures of defectors than of cooperators, we cannot tell whether they truly recognize defectors at a higher rate (genuine HIT) or are instead more likely to think they recognize a defector's face whether they have seen it before or not. If the latter is the case, then there should also be more false positives (which we call "False Alarm" or FA) with unfamiliar defectors than with unfamiliar cooperators. We address this question in the next experiment.

6.1. Procedure

One hundred and fifty-four subjects (115 males and 39 females) participated in this experiment, which had two between-subjects conditions. In the experimental condition, we replicated the procedures of Experiment 2: 79 subjects viewed 56 pictures (28 male, 28 female) of cooperators and defectors, and later judged whether they had seen them before in a set of 112, 56 of which were in the original set and 56 of which were filler pictures. In the control condition, 75 subjects viewed a different set of 56 pictures (which were the same as the filler pictures in the experimental condition) of people who had not participated in a PDG and were thus not known to us as either cooperators or defectors. Then, during the recognition phase, they viewed the same set of 112 pictures used in the experimental condition. Thus, all pictures of known cooperators and defectors were fillers in the control condition, and any pictures that the control subjects could truly recognize were of individuals who had not participated in PDGs. If the control subjects claim to recognize more defectors than cooperators, it can only be because of differential rates of false positives.

This between-subjects design allows as to maximize the number of pictures of cooperators and defectors used in the calculation of reliable true positive and false positive recognition rates on the one hand, and the separate measurements of the false positive recognition rates for cooperators and defectors that are not possible in the Experiments 2 and 3 on the other. The only drawback of this design is that we cannot calculate d' for each individual subject, but because we were not interested in individual differences in the differential recognition of cooperators and defectors, this did not pose a serious concern. This between-subjects design was chosen because it gave us a more reliable measure of the true positive and the false positive recognition rates than a within-subject design in which all 154 subjects view 28 target pictures and 28 filler pictures in the recognition phase.

6.2. Results

The overall correct identification (HIT) rate was 0.68, and the false positive (FA) rate 0.25 ($d' = 1.27$). In the experimental condition, HIT rate = 0.72, FA rate = 0.22 ($d' = 1.49$). In the control condition, HIT rate = 0.64, FA rate = 0.27 ($d' = 1.03$). These results indicate that, as before, subjects were able to discriminate fairly well between faces that they had and had not seen before.

The analysis of the HIT rate in the experimental condition replicates the differential HIT rates found in Experiments 2 and 3. The HIT rate for defectors (0.74) was significantly higher than that for cooperators (0.70) in the Target Type × Target Sex ANOVA, $F(1,78) = 8.99$, $P < .01$, sex of target, $F(1,78) = 5.47$, $P < .05$, and the Target Type × Target Sex interaction, $F(1,78) = 5.63$, $P < .05$, were also significant. As in Experiment 1, the higher HIT rate for defectors was more pronounced for female faces (0.74 vs. 0.67) than for male faces (0.75 vs. 0.74).

The important novel result comes from the control condition, in which subjects were more likely to think they recalled the faces of defectors (0.29) than those of cooperators (0.25), having actually seen neither. This difference in FA rates is significant, $F(1,74) = 7.85$, $P < .01$. Again, the main effect of target sex, $F(1,74) = 10.60$, $P < .01$, and the Target Type × Target Sex interaction, $F(1,74) = 33.70$, $P < .0001$, were also significant. Thus, control subjects incorrectly "recognize" defectors' faces more often than cooperators' faces, and, as in the experimental condition, this tendency was stronger for female faces (0.31 vs. 0.19); among male faces, the FA rate was actually slightly higher for cooperators (0.31) than for defectors (0.27).

7. General discussion

The results of the four experiments presented in this paper indicate that subjects recognize the faces of defectors more often than those of cooperators when they do not know who are cooperators and defectors. At the same time, Experiment 4 shows that not only could subjects recognize faces of defectors better than those of cooperators, but they also (erroneously) believed that they had seen the defectors before when they had not. These findings raise more questions than we can answer. The fact that the subjects "recognize" defectors' faces more

often than cooperators' faces either correctly or incorrectly strongly suggests that some facial features distinguish defectors from cooperators, whether we can consciously identify these features or not. In an unpublished study, Yamagishi et al. analyzed whether people can identify cooperators and defectors from the 56 pictures used in the second and the fourth experiments and found that they could not consciously discriminate the two types of targets better than chance. Nevertheless, the subjects of the current experiments recognized, or thought they recognized, defectors' faces more frequently than cooperators' faces. Some distinct facial features of the defectors must feel familiar to the subjects. Thus, the next questions are (1) what facial features distinguish defectors from cooperators, and (2) why these features make the subjects feel familiar to them. Physical attractiveness immediately comes to mind as a candidate for such a feature, but our analysis indicates that it is not the main source of the differential recognition rates.

There are two other possible answers for the first question. The first is emotional expressions on the faces of targets as they made cooperative vs. defecting decisions in the PDG. The expressions of guilt, anxiety, or greed associated with defection may produce arousal in the targets, and their heightened level of arousal may lead to the (correct or incorrect) conclusion the subjects have seen the targets' faces before. Another possibility is that the distinguishing cues are more permanent than the ephemeral expressions of emotions. The correlations between the recognition rate and the targets' responses to postexperimental questionnaires suggest this possibility, because their responses probably reflect not only their immediate emotional states but also more stable attitudes and beliefs. The correlations reported for Experiments 2 and 3 tend to suggest that subjects recognized the faces of untrusting and untrustworthy people, i.e., faces of those who are not desirable as exchange partners. If this is the case, we must face a difficult challenge of identifying facial features of chronic cheaters.

To identify which of the two possibilities — emotional expressions vs. permanent features — is behind the differential rates of "recognition" observed in the experiments presented in this paper requires a new experimental design in which pictures of the targets are taken before they play the PDG. If we still observe the differential "recognition" rates even with the pre-PDG pictures of cooperators and defectors, then we can rule out the expressions of ephemeral emotions arising from their choices in the PDG as the explanation for the differential "recognition" rates. If cheaters genuinely look different from cooperators because of certain beliefs and personality traits that make them less ideal exchange partners, and if humans can pick up on these differences, by recognizing their faces more accurately (or by thinking that they do), then our findings support the claim of Cosmides (1989) and Cosmides and Tooby (1992) that humans have an evolved adaptation for detecting cheaters. Human tendency to pay greater attention to cheaters than to cooperators is likely a part of such a domain-specific cheater detection module.

Acknowledgments

We thank the editors and two anonymous reviewers for criticism of a prior draft, as well as colleagues at Hokkaido University who helped us recruit potential participants from their

classes. The research reported in this paper has been supported by a grant from Japan Society for the Promotion of Science. Preparation of the paper has been facilitated by a fellowship to the first author from Center for Advanced Study in the Behavioral Sciences.

References

Cosmides, L. (1989). The logic of social exchange: has natural selection shaped how humans reason? Studies with 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 Univ. Press.

Kikuchi, M., Yamagishi, T., Shinagawa, M., & Sasaki, A. (1998). Accuracy of judgment of trustworthiness of photographed faces and physical attractiveness of the faces. *Proceedings of the 46th Annual Meetings of the Japanese Group Dynamics Association*, 76–77.

Kikuchi, M., Yamagishi, T., & Yamamoto, S. (1998). Effects of attractive appearance in social exchange. *Proceedings of the Annual Meetings of the Japanese Society of Social Psychology*, *39*, 80–81.

Little, L. L., Hollander, S., & Kayra-Stuart, F. (1981). Why attractive people are harder to remember. *Personality and Social Psychology Bulletin*, *7*, 269–276.

Matsuda, M., & Yamagishi, T. (2001). Trust and cooperation: an experimental study of PD with choice of dependence. *Japanese Journal of Psychology*, *72*, 413–421.

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

Mulford, M., Orbell, J., Shatto, C., & Stockard, J. (1998). Physical attractiveness, opportunity, and success in everyday exchange. *American Journal of Sociology*, *103*, 1565–1592.

Oda, R. (1997). Biased face recognition in the prisoner's dilemma games. *Evolution and Human Behavior*, *18*, 309–315.

Shepherd, J. W., & Ellis, H. D. (1973). The effect of attractiveness on recognition memory for faces. *American Journal of Psychology*, *86*, 627–633.