LEBS

Vol. 1 No. 1 (2010) 19-22.

Letters on Evolutionary Behavioral Science

Detecting Defectors When They Have Incentives to Manipulate Their Impressions

Toko Kiyonari*

Aoyama Gakuin University, 5-10-1 Fuchinobe, Chuo-ku, Sagamihara-city, Kanagawa, 229-8558, Japan

*Author for correspondence (kiyonari@si.aoyama.ac.jp)

We examined if naive observers can distinguish defectors from cooperators even when defectors may be motivated to present themselves positively. In Study 1, 150 participants played a "semi-sequential" Prisoner's Dilemma Game (PDG) with real monetary incentives, half as first players and half as second players. First players decided to cooperate or defect, and second players made the same decision without knowing the first player's choice. The first player was given a chance to present a video message to the second player before the latter made their decision. After the PDG, players played a separate one-shot, semi-sequential Stag Hunt Game (SHG), a coordination game where cooperation is the best choice insofar as the other also cooperates. In this game, the first player was not given a chance to send a video message. When the players had incentives to hide intentions or manipulate impressions of themselves, even motivated judges (whose monetary gain depended on the accuracy of cheater/cooperator detection) could not distinguish defectors from cooperators in either the PDG or SHG. However, they were able to discriminate "hard-core defectors" who defected in both games. In Study 2, however, in which judges had no monetary incentives to detect targets' choices, participants were unable to discern even hard-core defectors. The contents of the messages did not provide help discerning defectors.

Keywords

cheater detection, cooperation, prisoner's dilemma.

Introduction

Driven by the theoretical possibility that human cooperation can evolve when humans are capable of discriminating cheaters or non-cooperators from cooperators (e.g., Cosmides & Tooby, 1992), many researchers have conducted experimental studies to examine if humans can discriminate non-cooperators from cooperators using facial cues. Brown, Palameta, and Moore (2003) found a significant altruist detection effect in response to video-taped storytelling. Verplaetse, Vanneste, and Braeckman (2007) showed that participants could discriminate defectors from cooperators when they saw pictures taken at the moment participants decided to defect or cooperate. Frank, Gilovich, and Regan (1993) found that participants were able to predict the player's choice in a one-shot PD when they had a 30-min "get-acquainted" meeting before the game. Except for Frank et al.'s (1993) study in which participants could make promises regarding their game behavior, however, these results were obtained only in situations where defectors had no incentives to deceive others. The paucity of evidence that humans can discriminate non-cooperators from cooperators even when non-cooperators have incentives to mimic cooperators call into question theoretical account of human cooperation based on signal detection ability. If we assume that cooperators exclusively select each other for mutual benefit, non-cooperators should be motivated to mimic cooperators. The ability to discriminate cooperators from non-cooperators should thus be viable even when non-cooperators have incentives to mimic cooperators. In this study, we examine if participants can successfully distinguish cooperators from defectors when defectors have incentives to conceal or disguise their facial or verbal expressions.

Study 1

Summary of the Experiment

Seventy-nine undergraduate students (41 males and 38 females) played the role of the first player in a "semi-sequential" Prisoner's Dilemma Game and Stag Hunt Game, and 79 students (41 males and 38 females) played the role of the second player in those games. After making a decision between cooperation and defection, the first player in the PDG was given an opportunity to send a video-taped message to the second player. The first player then played the SHG. The second player was shown the messages video without audio from first players (but not the first players' decisions), judged if each of them cooperated or defected, and made his or her own decision against each of the first players. Then, the second player played the one-shot SHG.

Procedure

Participants played a PDG with a payoff-matrix shown in Figure 1. First players were told that their partner would play the game a few days later, and thus they and their partner would receive the

doi: 10.5178/lebs.2010.5
Published online 1 August 2010.

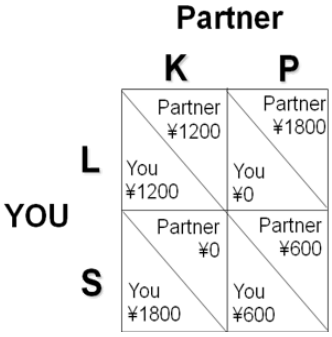

Figure 1 The payoff matrix of the Prisoner's Dilemma Game used in the study

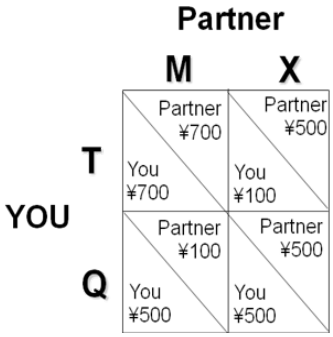

Figure 2 The payoff matrix of the Stag Hunt Game used in the study

payoffs of the game after their partner decided his/her choice. However, first players were told that their partner would not be informed of the first player's decision before he/she made his/her decision. We call this a "semi-sequential game" because the game is played sequentially, but the second player makes a decision without knowledge of the first player's decision. After the first player decided whether to cooperate or defect (L or S in Figure 1), they were provided with an unexpected opportunity to send a 30 sec video message to their partner. The first player was told that the videotape would be shown to their partner either with or without audio. Since first players were told about the video messages only after they had made the decision, they played the PDG without knowledge of it. The video message was taken in a sound-proof room. The participant was alone during the recording.

After recoding the message, first players were told that they would play another game with another partner, who would also play the second game later in the week. The first player was explicitly told that he/she would not have the opportunity to send a message to the partner of this game. The second game was a Stag Hunt Game (Skyrms, 2004) shown in Figure 2. The SHG differs from the PDG in several aspects. The SHG is a coordination game in which cooperation is an individually more profitable choice than defection insofar as the partner also chooses cooperation. Self-regarding players should thus choose cooperation when they expect the second player to cooperate. When first players expect that the second player would choose to defect, defection is an individually more profitable choice.

During weekdays of each week, 10 to 14 first players participated in the study. During the weekend, the same number of second players participated. Second players were shown the video-clips of all of the first players who participated in the immediately preceding week, without sound, and decided whether to cooperate or defect with each of them, on the assumption that he/she would be paid for the outcome of a randomly matched game with one of the first players. Before they made the decision, second players were asked to estimate whether each of the first players cooperated or defected. Next, second players played a SHG and did some other tasks. After all tasks were completed, they were paid based on the results of the PDG and SHG matched with randomly chosen partners.

Results

On average, 35.6% of the first players cooperated in the PDG, 49.3% in the SHG, and 43.8% defected in both games. The second players, on average, cooperated with 33.4% of the 10-14 potential partners in the PDG, and 53.2% of them cooperated in the SHG. No sex differences were observed in both games. The 26 first players who actually cooperated in PDG were judged to have cooperated, on average, by 54.6% of the second players who saw their videos. In contrast, the 47 first players who actually defected in the PDG were judged to have cooperated, on average, by 49.9% of the second players who saw their videos. The difference was in the predicted direction, but was not statistically significant, $t(71) = .99$, $p = .32$. The first players who cooperated in the SHG were judged to have cooperated, on average, by 54.8% of the second players, and those who defected in the SHG were judged to have cooperated by 48.5% of the second players, and the difference was not significant, $t(71) = 1.39$, $p = .17$. The 32 "hard-core defectors" who defected in both games were judged by 46.7% of the second players, and the remaining 41 first players were judged by 55.4% of the second players, and the difference was marginally significant, $t(71) = 1.96$, $p = .054$. This comparison, however, is misleading since the proportion of the cooperative first players judged by second players differed from week to week. If second players made relative judgments rather than judging each first player independently, the differential composition of cooperators might have introduced biases in the second player's judgment. In order to eliminate the effects of the composition of judges for the week, we calculated the adjusted judgment score for each first player, which is a deviation score from the mean proportion of cooperator judgments for that week. The use of the adjusted judgment score did not affect the above conclusion regarding the lack of significant differences in the judgments of cooperators and defectors, in either game. However, the difference

between the hard-core defectors (M = -0.05, SD = 0.16) and the rest (M = 0.04, SD = 0.18) was significant, t(71) = 2.37, p = .02.

Study 2

The first study provide evidence that interaction partners who have incentives to detect cooperators and defectors were able to discriminate actual cooperators from defectors even when the defectors had incentives to disguise as cooperators. To examine if the third-party who have no personal stake in discriminating cooperators from non-cooperators are also able to tell cooperators from non-cooperators, we showed the video-clips of the first players used in Study 1 to another group of judges. Thirty judges from another university judged all of the 73 first players' video-clips.

Procedure

The thirty judges participated in a series of experiments as part of an undergraduate class exercise. They were shown video-clips of 41 male first players from the first study, and judged whether each of the video-taped players cooperated or defected in the PDG. In the next week, they were shown 32 female first players' video-clips and judged if each of them cooperated or defected in the PDG. In yet another week, the same participants participated in a message judgment study, in which the verbal messages by the first players in the first study had been transcribed and were presented to them without visual or audio stimuli. The judges read each player's transcribed message and evaluated the player who sent the message on the following criteria: 1) Whether or not the message implied that the message sender cooperated; 2) whether the sender was equivocating which decision the sender chose; 3) whether the sender tried to communicate that he/she had cooperated; and 4) whether the message was a lie if the message sender had actually defected. Finally, judges estimated whether the player who wrote the message had cooperated or defected in the PDG.

Results

Judgment of cooperators and defectors from the video-clips

The first players who had actually cooperated in the PDG were judged to have cooperated in the PDG, on average, by 16.85 of the 30 judges, and first players who had actually defected in the PDG were judged to have cooperated, on average, by 16.68 judges. Those who had actually cooperated in the SHG were judged to have cooperated in the PDG, on average, by 17.17 of the 30 judges, and those who actually defected in the SHG were judged to have cooperated, on average, by 16.32 judges. Obviously, judges were unable to tell cooperators from defectors in either game, t(71) = .12, ns., and t(71) = .05, ns., respectively. Furthermore, the judges were unable to discriminate hard-core defectors from the rest, either. The hard-core defectors were judged by 16.28 judges (54.3%) to have cooperated in the PD, and the rest were judged by 17.10 judges (57.0%) to have cooperated, t(17) = .62, ns.

Analysis of message transcripts

Whether the sender of the message actually cooperated or defected did not affect the judges' evaluations of the messages on the four criteria above (all ps > .25). While more judges correctly identified actual cooperators as cooperators (M = 9.85 or 32.8%) than mistakenly rated as defectors (M = 6.54 or 21.8%) from the messages, the difference was not significant, t(25) = 1.34, p = .19. The remaining raters (or 44.6%) selected "unsure." Similarly, more judges correctly rated the actual defectors as defectors (M = 9.74 or 32.5%) than as cooperators (M = 7.30 or 24.3%), though the difference was not significant, t(46) = 1.17, p = .25. The remaining raters (42.9%) selected "unsure." The difference between the differences was marginally significant, t(71) = 1.71, p = .09.

Discussion

When defectors had incentives to disguise as cooperators, even motivated judges (whose monetary gain depended on the accuracy of cheater/cooperator detection) could not distinguish defectors from cooperators either in the PDG or SHG. However, they were able to discriminate hard-core defectors from the rest. When judges had no monetary incentives to detect targets' choice, they were unable to detect even hard-core defectors. The contents of the messages did not provide a sufficient help to distinguish defectors from cooperators.

Who Are "Hard-Core Defectors"?

The hard-core defectors were those who defected in both games. In the PDG, there are three reasons for defection: greed, competitiveness, and fear. In the SHG, greed cannot be a reason for defection. Thus, the hard-core defectors shall be the ones who are motivated by competitiveness and/or fear. However, none of hard-core defectors indicated in the post-experimental questionnaire the QM cell in the SHG, in which the difference between their own payoff and the partner's payoff was the largest, as personally most desirable. Twenty-nine of the 32 hard-core defectors (90.6%) indicated that mutual cooperation was most desirable. The competitive social motivation thus cannot explain the defection choice by the hard-core defector, leaving fear the defining characteristics of the hard-core defector. Regardless of their willingness, they couldn't choose cooperation in the SHG due to low expectation toward their potential partner's decision. Actually, 27 of 32 of the hard-core defectors (84.4%), compared to 8 of 41 remaining participants expected that their partner would choose defection in the SHG. The difference was strongly significant, $X^2(1) = 30.30$, $p < .0001$. If this interpretation of the hard-core defector is correct, fear of exploitation

by the interaction partners is the factor that motivated judges were discerning. More detailed study identifying what motivated judges actually discerning is definitely needed.

Acknowledgment

I gratefully thank all those who helped conducting this study especially Joanna Schug, Mizuho Shinada, Yang Li, Taiki Takahashi and Toshio Yamagishi.

References

Brown, W. M., Palameta, B., & Moore, C. (2003). Are there nonverbal cues to commitment? An exploratory study using the zeroacquaintance video presentation paradigm. Evolutionary Psychology, 1, 42-69.

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.

Frank, R. H., Gilovich, T., & Regan, D. T. (1993). The evolution of one-shot cooperation: An experiment. Ethology and Sociobiology, 14, 247-256. (doi:10.1016/0162-3095(93)90020-I)

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