Cognition 117 (2010) 261–275

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/COGNIT

Enhanced old–new recognition and source memory for faces of cooperators and defectors in a social-dilemma game

Raoul Bell *, Axel Buchner, Jochen Musch

Department of Experimental Psychology, Heinrich-Heine-University, Düsseldorf, Germany

ARTICLE INFO

Article history:
Received 26 November 2009
Revised 27 August 2010
Accepted 29 August 2010

Keywords:
Reciprocal altruism
Cheater detection
Cooperation
Source monitoring
Context memory
Cheater recognition

ABSTRACT

A popular assumption in evolutionary psychology is that the human mind comprises specialized cognitive modules for social exchange, including a module that serves to enhance memory for faces of cheaters. In the present study, participants played a trust game with computerized opponents, who either defected or cooperated. In a control condition, no interaction took place. In a surprise memory test, old–new recognition for faces and source memory for the associated cooperative or non-cooperative behavior were assessed. A multinomial model was used to measure old–new discrimination, source memory, and guessing biases separately. Inconsistent with the assumption of a memory mechanism that focuses exclusively on cheating, the present study showed enhanced old–new discrimination and source memory for both cooperators and defectors. Rarity of the behavior strategies within the experiment modulated source memory, but only when the differences in base rates were extreme. The findings can be attributed to a mechanism that focuses on exchange-relevant information and flexibly adapts to take into account the relative significance of this information in the encoding context, which may be more beneficial than focusing exclusively on cheaters.

1. Introduction

Social cooperation among unrelated individuals is an interesting phenomenon from an evolutionary point of view. At first glance, it seems obvious that individuals benefit from mutual cooperation. However, cooperation is costly for the individual providing support for other group members. Hence, natural selection would work against individuals who *unconditionally* provide benefits to others (Axelrod & Hamilton, 1981; Cosmides, 1989; Cosmides & Tooby, 1992, 2005; Trivers, 1971). In a *single-shot Prisoner's Dilemma Game*, defecting is the dominant strategy, because in each interaction that involves a cooperator and a defector, the defector benefits at the expense of the cooperator. However, things change when the same players interact repeatedly with each other. In this situation, cooperation can be very successful if it is reciprocal, that is, if cooperation is made contingent on the opponent's behavior in previous encounters (Axelrod & Hamilton, 1981; Cosmides, 1989; Trivers, 1971). Reciprocal strategies in social exchange require certain cognitive prerequisites such as the ability to detect cheaters, the ability to recognize different individuals, and the ability to "store information about the history of one's past exchanges with other individuals (in order to know when to cooperate, when to defect, and when to punish defection)" (Cosmides & Tooby, 1992, p. 177).

According to social contract theory (Cosmides, 1989; Cosmides & Tooby, 1992, 2005), social exchange is of such crucial importance for the individual's fitness that specialized cognitive modules have evolved that help us to deal with social-exchange situations. The cheater-detection module proposed by this theory allows the individual to

* Corresponding author. Address: Institut für Experimentelle Psychologie, Heinrich-Heine-Universität, D-40225 Düsseldorf, Germany. Tel.: +49 211 811 5643; fax: +49 211 811 5037.
E-mail address: raoul.bell@uni-duesseldorf.de (R. Bell).

0010-0277/$ - see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.cognition.2010.08.020

quickly and easily draw inferences on whether someone has cheated in prior exchanges or is about to cheat in future interactions. Several researchers (Chiappe et al., 2004; Mealey, Daood, & Krage, 1996; Oda, 1997) have argued that—to save the individual from social exploitation—this cheater-detection module has to be complemented by memory mechanisms that are sensitive to violations of social contract laws and that enable the individual to learn from previous negative experiences with cheaters. There are a number of studies that have examined whether there is a specialized module for remembering faces of cheaters. In most of these studies (Barclay & Lalumière, 2006; Bell & Buchner, 2010, in press-a, in press-b; Buchner, Bell, Mehl, & Musch, 2009; Chiappe et al., 2004; Mealey et al., 1996; Mehl & Buchner, 2008), the moral status of the faces was manipulated using short descriptions in which the stimulus characters were associated with cheating, trustworthy, or irrelevant behavior. In their pioneering study, Mealey et al. observed that for faces associated with low-status professions, old–new discrimination was better for faces of cheaters than for faces of trustworthy persons. Unexpectedly, the pattern was descriptively in the opposite direction for faces associated with high-status professions. Subsequent studies failed to replicate the face recognition effect (Barclay & Lalumière, 2006; Mehl & Buchner, 2008). However, simply recognizing a face of a cheater as familiar cannot help to avoid cheaters in social exchange. Source memory for faces of cheaters, that is, better memory for the cheating context in which a face was encountered, in contrast, can be instrumental in avoiding exploitation and is therefore beneficial to cooperating individuals. Consistent with these assumptions, a series of experiments in our lab (Buchner et al., 2009) showed that source memory for faces of cheaters was enhanced compared to source memory for other types of faces. These findings provide support for a functional perspective on human memory (Klein, Cosmides, Tooby, & Chance, 2002; Nairne, 2005; Nairne & Pandeirada, 2008).

Although most of the aforementioned studies were interpreted with reference to models of direct reciprocity (Axelrod & Hamilton, 1981; Trivers, 1971), the moral status of the stimulus characters was manipulated by letting participants read about interactions involving third parties. Cheating and trustworthiness had no negative or positive consequences for participants whatsoever, simply because participants were not directly involved in social exchange with the characters. We henceforth refer to this paradigm as the *description paradigm* (Hammerl, 2000). It is unclear whether the pattern of results obtained in the *description paradigm* may generalize to a situation in which participants are directly involved in social exchange with the stimulus characters. Therefore, it is interesting to examine old–new discrimination and source memory for faces of cheaters in an *involvement paradigm*, that is, using a task in which cheating and trustworthy behavior of the stimulus characters has direct negative or positive consequences for participants. The simplest way to do this is to use a *social-dilemma game*.

To date, memory for faces of cheaters was investigated in three studies using a *social-dilemma game* (Barclay, 2008; Oda, 1997; Singer, Kiebel, Winston, Dolan, & Frith, 2004). Two of these studies manipulated social-exchange status by providing participants with third-party reputational information about the stimulus persons. Oda (1997) required participants to imagine that they were one of two prisoners in the original version of the *Prisoner's Dilemma* and provided information about the opponents' strategies in this game (confessing or keeping silent). The opponents' behavior had only imaginary, but no real negative or positive consequences for the participants. Barclay (2008) informed participants that they would have to play a trust game with computerized players. Before the memory test, participants were given photographs of their prospective opponents and were explicitly informed about who will defect and who will cooperate in the game phase. Given that participants knew that they would benefit directly from encoding and rehearsing the strategies associated with the opponents' faces, the instructions used by Barclay can be equated with explicit learning instructions. Therefore, it is possible that conscious, strategic rehearsal may have overridden the evolved learning biases that may affect memory primarily in situations in which learning is less strategic. This suspicion is substantiated by a study of D'Argembeau and Van der Linden (2004), showing that source memory was enhanced for emotional words when compared to neutral words, but only when learning was incidental and participants had no intention to encode the context of the words anyway. In summary, although the studies of Oda and Barclay manipulated social status using social-dilemma games, participants were provided with third-party reputational information rather than with first-party experience. Consequently, their results are inconclusive regarding the prediction derived from models of direct reciprocity (Axelrod & Hamilton, 1981; Trivers, 1971) that participants *spontaneously* encode and remember their opponents' faces and the associated behavior strategies when they are *directly* involved in social exchange with them.

There is only one study we know of that examined memory for cheaters using an *involvement* paradigm. In this fMRI study (Singer et al., 2004), participants played an iterated trust game with computerized players. Participants could trust their opponents by sending them money. The opponents either reciprocated by sending the money back (which resulted in a moderate profit for both players involved) or defected by keeping the money (whereby they maximized their own profit at the expense of the participants). In a control condition, participants saw the faces of the opponents, but no transaction took place. Each participant interacted repeatedly with the same set of opponents that consisted of five cooperators, three defectors, and three irrelevant control faces. The neuroimaging results revealed enhanced activation in brain regions that are commonly associated with emotional processing and social cognition in response to the cooperator and defector faces in comparison with the irrelevant faces. The study also comprised a source memory test. In this test, participants were asked whether a face belonged to a defector, a cooperator or to an irrelevant control person. The raw number of correct source classifications was more accurate for faces of cooperators than for other faces. The main problem with interpreting ad hoc source memory mea-