ELSEVIER

Evolution and Human Behavior 31 (2010) 109–117

Evolution and Human Behavior

Altruism toward in-group members as a reputation mechanism☆

Nobuhiro Mifune, Hirofumi Hashimoto, Toshio Yamagishi*

Hokkaido University, Sapporo, Hokkaido, Japan

Initial receipt 18 March 2009; final revision received 4 September 2009

Abstract

To test the hypothesis that sensitivity to monitoring drives people to act altruistically toward members of their own community, two experiments investigated whether an eye-like painting promotes altruism toward in-group members, but not toward out-group members. Participants played the role of dictator in a dictator game with another participant (a recipient) who was from the minimal in-group or out-group. Participants knew whether their recipient was an in-group member or an out-group member, but were informed that their recipient did not know the group membership of the dictator. In-group favoritism occurred only when participants were facing a computer desktop which displayed a painting of eyes, but did not occur in the absence of eyes. These findings demonstrate that the eye painting displayed on the participant's computer screen worked as a cue for monitoring and thus enhanced the participant's altruistic behavior.

Keywords: Altruism; Cooperation; Reputation; Indirect reciprocity; Dictator game; In-group favoritism

1. Introduction

Altruism toward genetically unrelated individuals whom one is not likely to meet again is an evolutionary puzzle. This is because the standard explanations of altruism among animal species, including kin selection (Hamilton, 1964) and reciprocal altruism (Trivers, 1971), are incapable of explaining altruism toward genetically unrelated individuals whom one is not likely to interact with again. From an evolutionary perspective, such altruistic behavior seems to be, at least on the surface, only detrimental to one's inclusive fitness. And yet, such behavior is not unusual; many people anonymously donate money to various charities, and participants in both laboratory and field experiments who are strangers to each other behave in altruistic and cooperative manners even in one-shot encounters (Berg, Dickhaut & McCabe, 1995; Fehr & Fischbacher, 2003; Forsythe, Horowitz, Savin & Sefton, 1994; Gintis, Bowles, Boyd & Fehr, 2003; Henrich et al., 2006; Hoffman, McCabe & Smith, 1996; Ledyard, 1995; Messick & McClintock, 1968). In this study, we present experimental evidence in support of the proposal that such altruism can be understood as an implicit strategy to adapt to a system of indirect reciprocity.

The growing literature on indirect reciprocity in theoretical biology aims to provide an answer to this evolutionary puzzle. Altruistic behavior toward someone whom one never sees again will not be rewarded by the recipient of such behavior. This effectively eliminates the possibility for altruists to acquire gains through mutually altruistic behavior, that is, what Trivers (1971) calls reciprocal altruism. This, however, does not eliminate the possibility that altruists are indirectly reciprocated by other altruists. When other altruists are conditional altruists, or those who behave altruistically only toward other altruists but not toward egoists, it is possible for altruists to acquire more benefits than egoists. Thus, altruism can evolve when there are a sufficiently large number of conditional altruists in a community, such that the extra benefits an altruist receives from conditional altruists exceed the cost of behaving altruistically toward others.

It should be noted, however, that conditional altruism can evolve only under specific conditions. Several models have been proposed as candidates for specifying the exact nature of these conditions (e.g., Leimar & Hammerstein, 2001; Nowak & Sigmund, 1998; Panchanathan & Boyd, 2004;

☆ The studies reported in this article were supported by Grant-in-aid 19330139 and 08J05455 by the Japan Society for the Promotion of Science.

* Corresponding author.

E-mail addresses: toshio@let.hokudai.ac.jp, TYamagish@aol.com (T. Yamagishi).

1090-5138/$ – see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.evolhumbehav.2009.09.004

Takahashi & Mashima, 2006). In this article, we do not discuss the similarities and differences among those models. Instead, we would like to point out that, despite some differences in the specification of the conditions for being a recipient of conditional altruism, information about one's past behavior toward others plays a crucial role in the success of all indirect reciprocity models. Unless one's altruistic or egoistic behavior toward others is known to the other conditional altruists, conditional altruism cannot function. In other words, reputation about one's behavior plays an indispensable role for the success of conditional altruism as a fitness-enhancing strategy.

In all of the models of indirect reciprocity cited above, transparency of information about all community members' past behavior is assumed, although the models differ from each other in the nature of the information available to other members. For example, the "image scoring" model proposed by Nowak and Sigmund (1998) requires the availability of information about others' immediate past — i.e., whether each of the other members of the community acted altruistically or not in the immediately preceding trial. The other models, such as the "standing" model proposed by Leimar and Hammerstein (2001), the "extra standing" model and the "strict discriminator" model proposed by Takahashi and Mashima (2006), require further information; the players in these models are assumed to know, for example, whether someone behaved altruistically toward someone else in the previous trial, and, furthermore, whether that someone else behaved altruistically toward yet another person two trials ago. Despite these differences, a fundamental assumption shared by all of these models is that all players' immediate past behavior is known to all of the other players. In this study, we seek to examine the psychological implications of this fundamental assumption used in various models of indirect reciprocity.

One immediate psychological implication of the above assumption is that humans are sensitive to how others behave toward members of their own community. In their cheater detection studies, Cosmides and Tooby (1989) have demonstrated that humans are sensitive to others' behavior, especially if others' behavior does not meet the moral standard of the community. Wang (1996) further demonstrated that such sensitivity to other members' behavior is stronger in smaller groups than in larger groups.

What we want to pursue in this study is another psychological implication of the above assumption, which is as obvious as the first one, but has not been fully explored yet. That is, humans are sensitive to the fact that their behavior is being observed by other community members. People must behave in a way to enhance their reputation of behaving altruistically toward their community members in order to qualify as someone who deserves similar treatment from other members of the community. The fact that people generally behave in a more cooperative and altruistic manner when their behavior is observed as opposed to when under complete anonymity has been repeatedly demonstrated in experimental studies. For example, face-to-face contact between game players (a situation where players can observe each other's behavior) has been repeatedly demonstrated to enhance cooperation level in prisoner's dilemma and other related games (e.g., Bixenstine, Levitt & Wilson, 1966; Brechner, 1977; Dawes, McTavish & Shaklee, 1977; Kurzban, 2001). In other studies, participants were found to behave in a more altruistic manner when they expected that their reputation would spread to other participants (Barclay, 2004; Barclay & Willer, 2007; Milinski, Semmann & Krambeck, 2002).

Accumulating evidence suggests that people are sensitive not only to the explicit information that their future rewards depend on how others perceive and judge their behavior, but also to subtle cues suggesting that they may be monitored by others. A good example is found in a study by Bateson, Nettle and Roberts (2006), which demonstrated that users of a common coffee machine in an office voluntarily donated more money to maintain the supply of coffee when a poster of eyes was posted on the wall than when a poster of flowers was posted. The presence of the eyes is assumed to function as a cue to the presence of monitoring in this situation, thus enhancing the altruistic behavior. Burnham (2003) demonstrated that presenting a picture of the recipient to the dictator in a dictator game prompted the dictator to allocate more money to the recipient. In another experiment of a dictator game, Haley and Fessler (2005) found that participants who played the role of a dictator voluntarily gave more share of an endowment of $10 when their computer screen displayed a painting of eyes. Burnham and Hare (2007) added further evidence to the effect of eyes on a computer screen by presenting a robot figure which appeared to be looking forward. They found that the presence of the robot on the computer screen promoted the cooperation level in a public goods game. Findings in these studies suggest that the presence of eye figures functions as a cue to the operation of monitoring in the situation, prompting participants to act in a more altruistic manner. Furthermore, Rigdon, Ishii, Watabe and Kitayama (2009) found that even an abstract figure remotely resembling a human face can have the same effect in a dictator game experiment in which participants were presented with three black dots in either the shape of a triangle, "∴", or an upside-down triangle. In their study, male participants allocated more money to the recipient when they saw an upside-down triangle of three dots, which remotely resembles human face, than when they saw a straight triangle. These findings provide support to the view that at least some of prosocial behavior frequently observed in one-shot games is a product of intuitive decision making driven by subtle but salient cues. Such a cue-driven psychological mechanism operates independently of the process of explicit decision making in which relevant pieces of information are deliberately evaluated and compared.

As shown above, we already have substantial evidence that humans are sensitive to cues suggesting the presence of monitoring and respond to such cues in the form of