Journal of Theoretical Biology 317 (2013) 338–347

Contents lists available at SciVerse ScienceDirect

Journal of Theoretical Biology

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

Indirect reciprocity with trinary reputations

Shoma Tanabe [a], Hideyuki Suzuki [b], Naoki Masuda [a,*]

[a] Department of Mathematical Informatics, The University of Tokyo, 7-3-1 Hongo, Bunkyo, Tokyo 113-8656, Japan
[b] Institute of Industrial Science, The University of Tokyo, 4-6-1 Komaba, Meguro, Tokyo 153-8505, Japan

HIGHLIGHTS

- We examined the possibility of cooperation when a player owns a trinary reputation.
- We identified cooperative and stable populations in a game of indirect reciprocity.
- Cooperation occurs in our model under a simple reputation assignment rule.

ARTICLE INFO

Article history:
Received 4 July 2012
Received in revised form
20 October 2012
Accepted 25 October 2012
Available online 2 November 2012

Keywords:
Cooperation
Altruism
Indirect reciprocity
Social norm

ABSTRACT

Indirect reciprocity is a reputation-based mechanism for cooperation in social dilemma situations when individuals do not repeatedly meet. The conditions under which cooperation based on indirect reciprocity occurs have been examined in great details. Most previous theoretical analysis assumed for mathematical tractability that an individual possesses a binary reputation value, i.e., good or bad, which depends on their past actions and other factors. However, in real situations, reputations of individuals may be multiple valued. Another puzzling discrepancy between the theory and experiments is the status of the so-called image scoring, in which cooperation and defection are judged to be good and bad, respectively, independent of other factors. Such an assessment rule is found in behavioral experiments, whereas it is known to be unstable in theory. In the present study, we fill both gaps by analyzing a trinary reputation model. By an exhaustive search, we identify all the cooperative and stable equilibria composed of a homogeneous population or a heterogeneous population containing two types of players. Some results derived for the trinary reputation model are direct extensions of those for the binary model. However, we find that the trinary model allows cooperation under image scoring under some mild conditions.

1. Introduction

Humans and other animals often cooperate even when cooperation is more costly than defection. In such social dilemma situations, direct reciprocity is among main reasons for cooperation between pairs of individuals that repeatedly meet each other (Trivers, 1971; Axelrod, 1984). However, individuals, in particular humans, cooperate with others even when they seldom meet the same partners more than once, as is the case for large populations. Reputation-based indirect reciprocity (also called downstream reciprocity; we simply call it indirect reciprocity in this paper) seems to be a dominant mechanism that enables cooperation in this situation. In indirect reciprocity, individuals cooperate with others with good reputations and they in turn gain good reputations if they appositely behave (e.g., cooperate) toward somebody else. The conditions under which indirect reciprocity realizes cooperation have been theoretically and numerically clarified in great details (Nowak and Sigmund, 1998a,b; Leimar and Hammerstein, 2001; Panchanathan and Boyd, 2003; Mohtashemi and Mui, 2003; Fishman, 2003; Ohtsuki, 2004; Ohtsuki and Iwasa, 2004, 2006, 2007; Nowak and Sigmund, 2005; Brandt and Sigmund, 2004, 2005, 2006; Pacheco et al., 2006; Roberts, 2008; Uchida, 2010; Nakamura and Masuda, 2011; Berger, 2011; Sigmnud, 2012).

Under the so-called image scoring, cooperation and defection are regarded to be good and bad behavior, respectively (Nowak and Sigmund, 1998a,b). Laboratory experiments suggest that humans use image scoring to evaluate others' behavior (Wedekind and Milinski, 2000; Milinski et al., 2001; Seinen and Schram, 2006). However, main theories attain that image scoring does not stabilize cooperation (Leimar and Hammerstein, 2001; Panchanathan and Boyd, 2003; Ohtsuki, 2004; Ohtsuki and Iwasa, 2004, 2007; Roberts, 2008). Although some studies have shown the viability of cooperation under image scoring (Nowak and

* Corresponding author. Tel.: +81 3 3812 2111.
E-mail address: masuda@mist.i.u-tokyo.ac.jp (N. Masuda).

0022-5193/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.jtbi.2012.10.031

Sigmund, 1998a; Fishman, 2003; Brandt and Sigmund, 2004, 2005, 2006; Uchida, 2010), the situations in which cooperation occurs are, in our view, quite restricted (but see Berger, 2011; we discuss this reference in Section 5). Under image scoring, cooperation occurs only when individuals always cooperate in the first round (Nowak and Sigmund, 1998a), unconditional defectors sometimes cooperate (Fishman, 2003), the number of interaction obeys the binomial distribution (Brandt and Sigmund, 2004) or Poisson distribution (Brandt and Sigmund, 2006), the probability that individuals recognize others' reputations increases in time (Brandt and Sigmund, 2005), or the reputations of individuals are revealed to others with a small probability (Uchida, 2010). Therefore, the reason for the discrepancy between the experiments and theory remains obscure.

For mathematical tractability and possible influences of the first seminal theoretical papers on this subject (Nowak and Sigmund, 1998a,b), most theoretical results of indirect reciprocity are derived from the analysis of binary reputation models. In other words, individuals are endowed with the binary reputation, i.e., good (+) or bad (−), depending on the last action toward others and other factors. However, the binary reputation may not be realistic in that only the last behavior of an individual in the social dilemma situation determines the reputation of the individual. In fact, experimental (Wedekind and Milinski, 2000; Milinski et al., 2001, 2002; Seinen and Schram, 2006; Wedekind and Braithwaite, 2002; Keser, 2003; Bolton and Katok, 2004; Bolton et al., 2005; Engelmann and Fischbacher, 2009) and numerical (Diekmann and Przepiorka, 2005) studies of indirect reciprocity in the context of online marketplaces often assume that the reputations are many valued, which complies with the reality of online marketplaces (Resnick and Zeckhauser, 2002; Resnick et al., 2006). More than binary valued reputations have also been employed in numerical studies of indirect reciprocity in theoretical biology literature (Nowak and Sigmund, 1998b; Leimar and Hammerstein, 2001; Mohtashemi and Mui, 2003; Roberts, 2008). Nevertheless, these studies are not concerned with relationships between the degree of cooperation and the number of the possible reputation values.

In this paper we analyze a trinary reputation model to identify stable populations that realize cooperation. The difference between the present results and those derived from the binary reputation models is remarkable. In particular, we find that image scoring can stabilize cooperation in the trinary reputation model.

2. Model

2.1. Donation game with reputations

We consider an infinitely large population. In each generation, the so-called donation game is repeated for sufficiently many rounds. Fig. 1(A) illustrates the interaction in each round. Two players are randomly selected from the population, one as donor and the other as recipient, with the equal probability. The donor intends to cooperate (C) or defect (D) toward the recipient according to the action rule σ, which we define below. We assume that the donor misimplements intended C such that the donor actually defects with probability $\epsilon_i > 0$ and that the intended D is always correctly implemented. We seek the possibility of cooperation in the population under this kind of implementation error, which is adverse to cooperation (Panchanathan and Boyd, 2003; Fishman, 2003; Ohtsuki, 2004; Ohtsuki and Iwasa, 2004, 2006, 2007; Nowak and Sigmund, 2005; Brandt and Sigmund, 2004, 2005, 2006; Uchida, 2010; Berger, 2011). If the donor implements C, the donor pays cost c, and the recipient obtains benefit b. If the donor implements D, the payoffs to the donor and recipient do not change. We assume that $0 < c < b$ such that the donation game is essentially the prisoner's dilemma.

We assume that each player possesses a reputation that takes one of the three values, i.e., G (Good), N (Neutral), or B (Bad). The action rule σ is a function from the recipient's reputation to the donor's intended action (i.e., C or D). Therefore, there are $2^3 = 8$ action rules, as shown in Fig. 1(B). For example, the AllC and AllD intend C and D regardless of the recipient's reputation, respectively. The so-called generous discriminator (gDisc) intends C when the recipient's reputation is either G or N and D otherwise.

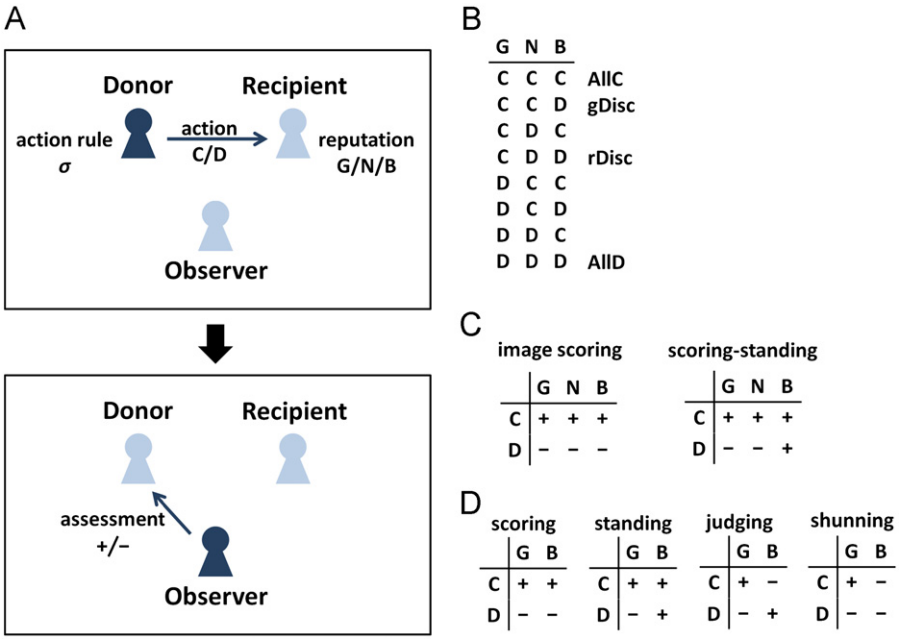

Fig. 1. Rule of the donation game with trinary reputations. (A) Illustration of the interaction in a single game. (B) Eight action rules. (C) Representative social norms. The rows represent the donor's actions (i.e., C and D), the columns represent the recipient's reputations (G, N, and B), and + and − represent the assessments that observer assigns to the donor. (D) Representative social norms in the binary reputation model.