Journal of Theoretical Biology 346 (2014) 34–46

Contents lists available at ScienceDirect

Journal of Theoretical Biology

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

The conflict of social norms may cause the collapse of cooperation: Indirect reciprocity with opposing attitudes towards in-group favoritism

Tadasu Matsuo [a,*], Marko Jusup [b], Yoh Iwasa [b]

[a] Faculty of Economics, Ritsumeikan University, 1-1-1 Noji-Higashi, Kusatsu 525-8677, Japan
[b] Department of Biology, Kyushu University, 6-10-1 Hakozaki, Higashi-ku, Fukuoka 812-8581, Japan

HIGHLIGHTS

- We study the replicator dynamics of two social norms engaged in reciprocal altruism.
- The norms differ only in their attitude towards cooperation with outsiders.
- Each social norm can maintain full cooperation if dominant.
- Both norms evaluate each other negatively, impeding cooperation between them.
- Defectors can successfully invade if both social norms are equally effective.

ARTICLE INFO

Article history:
Received 2 July 2013
Received in revised form
11 December 2013
Accepted 16 December 2013
Available online 28 December 2013

Keywords:
Social norm
In-group favoritism
Indirect reciprocity
Replicator dynamics
Reputation dynamics

ABSTRACT

Indirect reciprocity is a cooperation maintaining mechanism based on the social evaluation of players. Here, we consider the case of a group in which two social norms with opposing attitudes towards in-group favoritism are mixed. One norm, called Bushido (the way of warriors), regards cooperation with outsiders as betrayal, whereas the second norm, called Shonindo (the way of merchants), regards cooperation with outsiders as desirable. Each member of the group, irrespective of being a Bushido or a Shonindo player, is evaluated in two different ways and assigned two different labels: "ally" or "enemy" according to the Bushido evaluation; "good" or "bad" according to the Shonindo evaluation. These labels change in response to the action taken (cooperation or defection) when acting as a donor, as well as the label attached to the recipient. In addition to Bushido players, who cooperate with an ally and defect from an enemy, and Shonindo players, who cooperate with a good recipient and defect from a bad recipient, the group contains a third kind of players – unconditional defectors. The fractions of the three types of players follow the replicator dynamics. If the probability of interacting with outsiders is small, and if the cost-to-benefit ratio of cooperation is low, we observe several important patterns. Each social norm is able to maintain a high level of cooperation when dominant. Bushido and Shonindo players evaluate each other unfavorably and engage in a severe conflict. In the end, only one norm permeates the whole group driving the other to the extinction. When both social norms are equally effective, a rare occurrence of unconditional defectors may lead to a successful invasion.

1. Introduction

Many social conflicts in a human society are caused by the pursuit of material interests, pleasure, safety, power, or honor. Social norms, oftentimes backed by a formal punishment system, are established to cope with a wide variety of potential conflict areas. However, occasional severe conflicts, ranging from personal quarrels to national wars, are either caused or worsened by social norms themselves. Especially the hatred over national, ethnical, regional, cultural, or physical differences may be spurred by the sense of loyalty towards a group to which one belongs. In this context, it is of little help that loyalty-driven norms are often identified with the phenomenon of in-group favoritism – a human tendency to discriminately favor members of one's group over outsiders (for a social experiment, see Yamagishi et al., 1998; for anthropological observations, see Bernhard et al., 2006 and Diamond, 2012).

Not all social norms in a human society are loyalty-driven. Jacobs (1992), for instance, classifies social norms into two types,

* Corresponding author. Tel.: +81 942 33 8487.
E-mail address: matsuo-t@ec.ritsumei.ac.jp (T. Matsuo).

0022-5193/$ - see front matter © 2014 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.jtbi.2013.12.018

termed the “Guardian Syndrome” and the “Commerce Syndrome”. The former type (i) has “loyalty” as its main value, (ii) is guided by such moral principles as “be exclusive”, and (iii) is considered necessary for establishing and defending territories. The latter type, in contrast, (i) has “honesty” as its main value, (ii) is embodied by the moral principles like “collaborate with strangers and aliens”, and (iii) is regarded as a prerequisite for trading activities. In addition, the two types of norms differ in the level of compatibility with in-group favoritism (the former being largely compatible, while the latter being mostly incompatible).

In pre-modern Japan, two dominant types of social norms closely resembled Jacobs′ division. One type was called “Bushido”, after the warrior class (jap. bushi) which governed Japan at the time. The other type was called “Shonindo”, after the class of merchants (jap. shonin), which commanded little political power, but amassed a considerable wealth. Matsuo (2009) identified the universality-oriented attitude of Shonindo, contrary to the Bushido′s in-group favoritism. The author also pointed out that Bushido and Shonindo have been engaged in a continuous conflict throughout the history of Japan. Readers less familiar with the Japanese history may draw some interesting parallels by considering the case of the Maghribis and the Genoese – two pre-modern societies with origins around the Mediterranean Sea. Greif (1994), for example, traces the important differences between the two societies to collectivist cultural beliefs of the Maghribis (similar to Bushido) and individualist cultural beliefs of the Genoese (similar to Shonindo).

Motivated by the above ethical considerations and the historical context, we consider the conditions that lead to the domination or the collapse of a social norm supportive of strong in-group favoritism. Conversely, we explore the conditions that promote a social norm with strong distaste for in-group favoritism. Jacobs (1992) argued that the two types of norms could not be mixed, otherwise they would lead to corruption. Here, we construct a mathematical model in an attempt to assess the validity of such an argument and, more generally, expand the theoretical basis for similar considerations.

Indirect reciprocity is a mechanism to realize stable cooperation in a society based on social information on each of its members (Nowak and Sigmund, 1998a, 2005; Leimer and Hammerstein, 2001; Ohtsuki and Iwasa, 2004, 2006, 2007; Ohtsuki et al., 2009; Pacheco et al., 2006; Brandt and Sigmund, 2004). The society may, for example, assign each member a simple binary evaluation, such as "good" or "bad", based on particular actions and situations in which these actions were adopted. Subsequent interactions between members of the society may then spontaneously exhibit a high level of cooperation and maintain it stably. The same formalism can be extended to discuss the competition between different social norms (Pacheco et al., 2006; Uchida and Sigmund, 2010), including the effect of the group structure on interactions (Nakamura and Masuda, 2012; Masuda, 2012) and the role of the costly punishment (Ohtsuki et al., 2009).

We study the dynamics of a group in which two evaluation-action strategies are in existence. Both strategies adopt (i) the same method of evaluating the reputation (called "the reputation dynamics" by Ohtsuki and Iwasa, 2004; "the assessment rule" by Brandt and Sigmund, 2004), (ii) the same rule for choosing an action in response to other member′s reputations (called "the behavioral strategy" by Ohtsuki and Iwasa, 2004; "the action rule" by Brandt and Sigmund, 2004), but (iii) diverging attitudes towards in-group favoritism (or equivalently towards cooperation with the outside world). One strategy, named Bushido, refuses to have anything to do with members of the society outside of a given group. The other strategy, named Shonindo, accepts and encourages the exchange with anyone. Each strategy uses two reputation levels (favorable and unfavorable) that are applied to all members of the society. Group members choose to cooperate with someone who has a favorable reputation, but defect from anyone with an unfavorable reputation; namely, they are discriminators (Nowak and Sigmund, 1998b). Because the two evaluation-action strategies are distinct, a member of the society may have a high reputation according to one strategy, but a low reputation according to the other. To allow for the possibility that some members of the group refuse to accept either evaluation-action strategy, we introduce unconditional defectors labeled AllD afterwards (Ohtsuki and Iwasa, 2004).

In what follows, we provide a detailed overview of the methodology, especially with respect to the conflict of social norms (Section 2). We then formulate the model in mathematical terms and examine the reputation dynamics with its equilibrium states (Section 3). The evolutionary dynamics of norms is explored next (Section 4) and illustrated in great detail with numerical examples (Section 5). A discussion towards the end (Section 6) elaborates the main results. In particular, we discuss why strong conflict arises when two social norms are mixed and why one norm tends to disappear. We also observe that each social norm is able to establish cooperation upon becoming dominant within the group, refuting the invasion of unconditional defectors in most cases. If both norms are equally effective, however, the conflict between the two may result in a successful invasion of unconditional defectors.

2. Indirect reciprocity games

In its simplest form, indirect reciprocity can be formalized in the following manner. Members of a society, or players, encounter each other randomly. Upon an encounter, one player takes the role of a donor, while the other acts as a recipient. The donor may choose to cooperate with the recipient, and consequently bear a cost c, or to defect and avoid any cost. If players manage to cooperate, the recipient benefits from a payoff b, resulting in the net gain of $b-c>0$ for the society. To a donor, however, defection may appear more advantageous than cooperation because only the latter choice incurs a cost. The decision on whether to cooperate or defect is further complicated by reputation considerations. Based on the recipient′s reputation and the donor′s action, the donor is assigned a favorable or unfavorable reputation by the members of the society. Because maintaining a good reputation raises the prospect of a high payoff in the future, every time the decision has to be made, the donor faces a trade-off between avoiding an immediate cost or improving the chances of a long term benefit. If the future payoff is sufficiently high, the society can maintain cooperation in a stable manner. Players participate in many rounds of the same game within a generation, and hence their reputation changes over time. Those players that are able to maintain a higher average payoff per round are considered to be more successful. More precisely, we assume that the fitness of each player is an increasing function of the average payoff. The fractions of different player types in a group then follow the replicator dynamics.

Next, we proceed to specify reputation assessment rules, i.e. the stylized, theoretical representations of social norms. The simplest reputation assessment rule is called scoring or image-score: a player who cooperates is regarded favorably, whereas a player who defects is regarded unfavorably (Nowak and Sigmund, 1998b). Such a simple rule, however, cannot maintain a cooperative society because the slow initial increase of unconditional cooperators (hereafter AllC) is followed by the spread of AllD players and the eventual collapse of cooperation (Leimer and Hammerstin, 2001; Ohtsuki, 2004). Reputation assessment rules that can foster cooperation in the society must allow for two donor′s actions (to cooperate or to defect) and must consider the reputation of the recipient (favorable or unfavorable). An example is provided by the Kandori rule, also called stern-judging (Kandori, 1992; Ohtsuki and Iwasa, 2007; Pacheco et al., 2006).