MATHEMATICAL ENGINEERING
TECHNICAL REPORTS

Groupwise Information Sharing Promotes Ingroup Favoritism in Indirect Reciprocity

Mitsuhiro NAKAMURA and Naoki MASUDA

METR 2012–19 October 2012

DEPARTMENT OF MATHEMATICAL INFORMATICS
GRADUATE SCHOOL OF INFORMATION SCIENCE AND TECHNOLOGY
THE UNIVERSITY OF TOKYO
BUNKYO-KU, TOKYO 113-8656, JAPAN

WWW page: http://www.keisu.t.u-tokyo.ac.jp/research/techrep/index.html

Groupwise Information Sharing Promotes Ingroup Favoritism in Indirect Reciprocity

Mitsuhiro NAKAMURA and Naoki MASUDA*

Department of Mathematical Informatics
Graduate School of Information Science and Technology
The University of Tokyo

*Corresponding author (masuda@mist.i.u-tokyo.ac.jp)

October 9th, 2012

Abstract

Indirect reciprocity is a mechanism for cooperation in social dilemma situations. In indirect reciprocity, an individual is motivated to help another to acquire a good reputation and receive help from others afterwards. Another aspect of human cooperation is ingroup favoritism, whereby individuals help members in their own group more often than those in other groups. Ingroup favoritism is a puzzle for the theory of cooperation because it is not easily evolutionarily stable. In the context of indirect reciprocity, ingroup favoritism has been shown to be a consequence of employing a double standard when assigning reputations to ingroup and outgroup members. An example of such a double standard is the situation in which helping an ingroup member is regarded as good, whereas the same action toward an outgroup member is regarded as bad. We analyze a computational model of indirect reciprocity in which information sharing is conducted groupwise. In our model, individuals play social dilemma games within and across groups, and the information about their reputations is shared within each group. We show that evolutionarily stable ingroup favoritism emerges even if all the players use the same reputation assignment rule regardless of group (i.e., a single standard). Two reputation assignment rules called simple standing and stern judging yield ingroup favoritism; under these rules, cooperation with (defection against) good individuals is regarded as good (bad) and defection against bad individuals is regarded as good. Stern judging induces much stronger ingroup favoritism than does simple standing. Simple standing and stern judging are evolutionarily stable against each other when groups employing different assignment rules compete and the number of groups is sufficiently large. In addition, we analytically show as a limiting case that homogeneous populations of reciprocators that use reputations are unstable when individuals independently infer reputations of individuals, which is consistent with previously reported numerical results. Our results suggest that ingroup favoritism can be promoted in indirect reciprocity by the groupwise information sharing, in particular under the stern judging assignment rule.

Background

Behavioral nature of humans depends on the economy of reputations, where praise and blame often lead to gain and loss of material benefits [1, 2]. Humans, among other animals, cooperate via indirect reciprocity, which involves cooperation beyond pairwise relationships [3, 4, 5, 6]. In indirect reciprocity based on reputations, an individual acquires a good reputation by behaving cooperatively in apposite situations. The cost of maintaining a good reputation is compensated for by other individuals' future cooperation toward the individual possessing the good reputation. Indirect reciprocity has been extensively studied in both theories [5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19] and experiments [20, 21, 22, 23, 2].

Another facet of human cooperation is that an individual often cooperates with members in the same group and not with others, a phenomenon called ingroup favoritism [24, 25, 26, 27, 28, 29, 30, 31, 32, 33]. Ingroup favoritism poses a puzzle for the theory of cooperation because it is usually not Pareto efficient; i.e., the payoff to an individual in the case of ingroup favoritism is smaller than that in the case of group-independent all-out cooperation. In addition, an individual implementing ingroup favoritism is worse off than an individual defecting against both ingroup and outgroup members unless a specific assumption is imposed. In fact, known mechanisms for stable ingroup favoritism (e.g., correlation between altruistic traits and phenotypic tags [34, 35, 36], incomplete observability of tags [37], combination of mutation of tags and limited dispersal [38]) are, in our view, complicated. Otherwise, stable ingroup favoritism requires an additional mechanism (e.g., intergroup conflict [39, 40]) that is capable of stabilizing cooperation on its own.

If maintaining a good reputation is a concern, why do individuals want to discriminate between ingroup and outgroup fellows? One of the present authors has shown that ingroup favoritism is evolutionarily stable in various situations when only group-level reputations are available in regard to outgroup members [41]. In the model, an individual's action changes the individual's reputation in the eyes of the ingroup members, and the action also changes the reputation of the group to which the individual belongs. It was revealed that the action rule of individuals (i.e., the strategy depending on the reputation of the coplayer) toward ingroup and outgroup members and the reputation assignment rule (also called the social norm) used for evaluating ingroup and outgroup interactions, or at least the latter, must discriminate between ingroup and outgroup members for stabilizing ingroup favoritism. An example is a rule whereby cooperation toward outgroup members is frowned upon, whereas the same behavior toward ingroup members leads to a good reputation. Consistent with this theoretical example, Yamagishi and colleagues had conducted behavioral experiments suggesting that ingroup favoritism occurs because subjects anticipate that the reputation mechanism is functional only inside the group [25, 26, 27, 29, 33]. These theoretical and experimental results suggest that double standards, in terms of the action rule or the reputation assignment rule, may underpin ingroup favoritism.

In the context of indirect reciprocity, group structure may play a crucial role in spreading reputations of individuals via rumor and gossip. In general, individuals interact more frequently with ingroup members than with outgroup members [42]. Therefore, rumor and gossip may enable sharing of reputations of individuals more smoothly within a group than between different groups. Most theoretical studies of indirect reciprocity have assumed that information sharing and interactions occur randomly in a well-mixed population. Otherwise, individuals are assumed to not exchange information about reputations [7, 10, 13, 19].

In the present study, we explore a scenario of ingroup favoritism without resorting to rules that apply double standards. In practice, humans may not differentiate between ingroup and outgroup coplayers with regard to their action rules or reputation assignment rules. We analyze a group-structured model of indirect reciprocity, in which an individual's reputation is shared by each group but not between groups. We study the case in which all the players use the same reputation assignment rule and the case in which players in different groups use different reputation assignment rules. We show that ingroup favoritism can emerge when players simply implement reputation-based decision making and do not favor ingroup members. Because of the assumed groupwise information sharing and some reputation assignment error, ingroup and outgroup members tend to possess good and bad reputations, respectively, without further assumptions. In particular, ingroup favoritism is strong when individuals adopt a reputation assignment rule called stern judging, under which helping bad individuals is regarded as bad.

Methods

Model

We consider an infinitely large population of players divided into M (≥ 2) groups. Each group is assumed to contain the equal fraction, $1/M$, of players. In the population, players are involved in sufficiently many rounds of the so-called donation game. In a one-shot donation game, two players are randomly selected from the population, one as donor and the other as recipient. We assume that the donor and recipient belong to the same group with probability θ. The donor cooperates (C), i.e., provides help, or defects (D), i.e., refrains from helping, depending on the donor's action rule and the recipient's reputation (good (G) or bad (B)). Action C imposes cost c (> 0) on the donor and results in benefit b $(> c)$ imparted to the recipient. Action D does not change the payoff to either the donor or recipient. A donor adopting action rule ALLC cooperates with any recipient. A donor adopting action rule ALLD defects against any recipient. A donor adopting action rule DISC cooperates with G recipients and defects against B recipients.

To know a recipient's reputation, the donor consults the unique information source, called the observer, that is shared by the group to which the donor belongs. Therefore, players in different groups may perceive different reputations (i.e., G or B) of the same player. The observer in each group independently assigns a reputation to the donor and shares it with the other players in the observer's group. Observers intend the predefined reputation assignment toward a donor's action but may assign a reputation opposite to the intention. The M observers independently commit such assignment error with probability μ $(\ll 1)$. In the example of intragroup interaction shown in Fig. 1, all the three observers intended to assign G to the donor, and one observer erroneously assigned B to the donor. If the assignment error occurs, the "wrong" reputation is shared by all the players in the group to which the observer belongs.

Observers assign reputations according to a common reputation assignment rule unless otherwise stated. We principally compare three rules: image scoring (IM), simple standing (ST), and stern judging (JG) [5, 6, 10, 11, 12, 14, 18], symbolically shown in Fig. 2. Among the three rules, IM is the simplest rule under which observers assign G and B to a donor that has selected C and D, respectively. ST and JG are simplest among the so-called "leading eight" reputation assignment rules that stabilize cooperation in well-mixed populations [11, 12]. Under ST and JG, the new reputation of the donor depends on the action of the donor (i.e., C or D) and the reputation of the recipient (i.e., G or B). When a recipient has a G reputation, observers assign G and B to a C and D donor, respectively, under both ST and JG. When a recipient has a B reputation, observers assign G to a D donor under both rules. The two rules are different in that helping bad individuals (i.e., donor's C) is appreciated (i.e., G imparted by the observer) under ST, whereas the same action of the donor is punished (i.e., B imparted by the observer) under JG; JG is sterner than ST [18, 1].

After sufficiently many rounds of the donation game involving reputation updates, the reputation distribution in the eyes of each group-specific observer reaches a unique equilibrium. In the equilibrium, we measure the quantities of interest such as the fractions of G players, the probability of cooperation, and their dependence on groups.

Analysis Methods

Equilibria of the reputation dynamics

Table 1 summarizes the definitions of the symbols used in this section.

We examine the stability of a homogeneous population of DISC players. Each player bears a reputation vector, $\boldsymbol{r} = (r_1, r_2, \dots, r_M) \in \{\mathrm{G}, \mathrm{B}\}^{\mathrm{M}}$, in the eyes of M observers, each representing a group. We denote by $p_k(\boldsymbol{r})$ the probability that a player in group k has reputation vector $\boldsymbol{r}$. By adopting the formalism developed by Ohtsuki & Iwasa [11], we obtain the following reputation dynamics:

$$\frac{\mathrm{d}}{\mathrm{d}t} p_k(\boldsymbol{r}) = -p_k(\boldsymbol{r}) + \sum_{\boldsymbol{r}' \in \{\mathrm{G},\mathrm{B}\}^{\mathrm{M}}} \left\{ \theta p_k(\boldsymbol{r}') + (1-\theta) p_{-k}(\boldsymbol{r}') \right\} \prod_{k'=1}^{M} \Phi_{r_{k'}}(\sigma(r'_k), r'_{k'}). \tag{1}$$

The summation on the right-hand side of Eq. (1) represents the average over the recipient's reputation vector $\boldsymbol{r}'$. With probability θ, a game involves a donor and a recipient in group k, and the recipient has reputation vector $\boldsymbol{r}'$ with probability $p_k(\boldsymbol{r}')$. With probability $1-\theta$, a donor and a recipient belong to group k and another group, respectively, and the recipient has reputation vector $\boldsymbol{r}'$ with probability $p_{-k}(\boldsymbol{r}') \equiv \sum_{k'=1, k' \neq k}^{M} p_{k'}(\boldsymbol{r}')/(M-1)$. $\sigma(r'_k)$ represents a donor's action toward a recipient having reputation r'_k. Because we assume DISC donors, $\sigma(\mathrm{G}) = \mathrm{C}$ and $\sigma(\mathrm{B}) = \mathrm{D}$. The reputation assignment rule is essentially given by $\Phi_{r_{k'}}(\sigma(r'_k), r'_{k'}) \in \{1-\mu, \mu\}$, which is the probability that an observer in group k' assigns reputation $r_{k'}$ to a donor in group k. This probability depends on the donor's action $\sigma(r'_k)$ toward a recipient having reputation $r'_{k'}$ in the eyes of the observer in each group k'. In Tab. 2, we list the Φ values under different assignment rules. It should be noted that all the observers use a unique assignment rule unless otherwise stated; we do not basically assume that observers employ different assignment rules as in previous studies [10, 15, 16, 18, 19].

We reduce Eq. (1) to mean field dynamics of two reputation distributions. First, we apply summation $\sum_{\boldsymbol{r}_{-k}} \equiv \sum_{r_1} \sum_{r_2} \cdots \sum_{r_{k-1}} \sum_{r_{k+1}} \cdots \sum_{r_M}$ to both sides of Eq. (1) to obtain the reputation dynamics in the eyes of ingroup observers as follows:

$$\frac{\mathrm{d}}{\mathrm{d}t} p_{\mathrm{in}}(r) = -p_{\mathrm{in}}(r) + \sum_{r' \in \{\mathrm{G},\mathrm{B}\}} \left\{ \theta p_{\mathrm{in}}(r') + (1-\theta) p_{\mathrm{out}}(r') \right\} \Phi_r(\sigma(r'), r'), \tag{2}$$

where $p_{\mathrm{in}}(r) \equiv \sum_{\boldsymbol{r}_{-k}} p_k(\boldsymbol{r})$ and $p_{\mathrm{out}}(r) \equiv \sum_{\boldsymbol{r}_{-k}} p_{-k}(\boldsymbol{r})$ are the probabilities that a player has reputation $r \in \{\mathrm{G}, \mathrm{B}\}$ in the eyes of ingroup and outgroup observers, respectively. The two terms inside the curly brackets on the right-hand side of Eq. (2) correspond to the two situations shown in Fig. 3 **(a)** and **(b)**. With probability $\theta p_{\mathrm{in}}(r')$, the recipient belongs to the donor and observer's group, and has reputation r' (Fig. 3 **(a)**). With probability $(1-\theta)p_{\mathrm{out}}(r')$, the recipient does not belong to the donor and observer's group, and has reputation r' (Fig. 3 **(b)**).

Second, by applying summation $\sum_{\boldsymbol{r}_{-\ell}} \equiv \sum_{r_1}\sum_{r_2}\cdots\sum_{r_{\ell-1}}\sum_{r_{\ell+1}}\cdots\sum_{r_M}$, where $\ell \neq k$, to both sides of Eq. (1), we obtain

$$\frac{\mathrm{d}}{\mathrm{d}t}p_k(r_\ell) = -p_k(r_\ell) + \sum_{r'_k\in\{\mathrm{G,B}\}}\sum_{r'_\ell\in\{\mathrm{G,B}\}} \left\{\theta p_k(r'_k, r'_\ell) + (1-\theta)\left[\frac{1}{M-1}\,p_\ell(r'_k, r'_\ell) + \left(1-\frac{1}{M-1}\right)\,p_{-k\ell}(r'_k, r'_\ell)\right]\right\}\Phi_{r_\ell}(\sigma(r'_k), r'_\ell), \tag{3}$$

where $p_k(r'_k, r'_\ell)$, $p_\ell(r'_k, r'_\ell)$, and $p_{-k\ell}(r'_k, r'_\ell)$ are the probabilities that a player in group k, group ℓ, and a group other than k and ℓ, respectively, has reputation r'_k and r'_ℓ in the eyes of observers in groups k and ℓ. By approximating the three probabilities by $p_{\mathrm{in}}(r'_k)p_{\mathrm{out}}(r'_\ell)$, $p_{\mathrm{out}}(r'_k)p_{\mathrm{in}}(r'_\ell)$, and $p_{\mathrm{out}}(r'_k)p_{\mathrm{out}}(r'_\ell)$, respectively, we obtain the reputation dynamics in the eyes of outgroup observers as follows:

$$\frac{\mathrm{d}}{\mathrm{d}t}p_{\mathrm{out}}(r) = -p_{\mathrm{out}}(r) + \sum_{r'\in\{\mathrm{G,B}\}}\sum_{r''\in\{\mathrm{G,B}\}} \left\{\theta p_{\mathrm{in}}(r')p_{\mathrm{out}}(r'') + (1-\theta)\left[\frac{1}{M-1}\,p_{\mathrm{out}}(r')p_{\mathrm{in}}(r'') + \left(1-\frac{1}{M-1}\right)\,p_{\mathrm{out}}(r')p_{\mathrm{out}}(r'')\right]\right\}\times \Phi_r(\sigma(r'), r''). \tag{4}$$

The three terms inside the curly brackets on the right-hand side of Eq. (4) correspond to the three situations shown in Fig. 3 **(c)**, **(d)**, and **(e)**. With probability $\theta p_{\mathrm{in}}(r')p_{\mathrm{out}}(r'')$, the recipient belongs to the donor's group, which differs from the observer's group, and has reputation r' and r'' in the eyes of the donor and observer, respectively (Fig. 3 **(c)**). With probability $(1-\theta)[1/(M-1)]p_{\mathrm{out}}(r')p_{\mathrm{in}}(r'')$, the recipient belongs to the observer's group, which differs from the donor's group, and has reputation r' and r'' in the eyes of the donor and observer, respectively (Fig. 3 **(d)**). With probability $(1-\theta)[1-1/(M-1)]p_{\mathrm{out}}(r')p_{\mathrm{out}}(r'')$, the recipient belongs to a group different from the donor's and observer's groups, and has reputation r' and r'' in the eyes of the donor and observer, respectively (Fig. 3 **(e)**).

By setting $\mathrm{d}p_{\mathrm{in}}(r)/\mathrm{d}t = \mathrm{d}p_{\mathrm{out}}(r)/\mathrm{d}t = 0$ in Eqs. (2) and (4), we identify stationary points that are candidates of stable equilibria of the reputation dynamics. We examine the conditions $\det \boldsymbol{J} > 0$ and $\mathrm{Tr}\ \boldsymbol{J} < 0$, where $\boldsymbol{J}$ is the Jacobian matrix, at each stationary point to identify all the stable equilibria. We confirmed that the stable equilibrium denoted by $p^*_{\mathrm{in}}(r)$ and $p^*_{\mathrm{out}}(r)$ is unique under each reputation assignment rule.

Stability against invasion by ALLC and ALLD mutants

We check the evolutionary stability of a homogeneous population composed of DISC players against invasion by an infinitesimal fraction of mutants adopting ALLC or ALLD. The payoff to a DISC resident player is given by

$$\pi_{\mathrm{DISC}} = (b-c)\left[\theta p^*_{\mathrm{in}}(\mathrm{G}) + (1-\theta)p^*_{\mathrm{out}}(\mathrm{G})\right], \tag{5}$$

and those to ALLC and ALLD mutants are given by

$$\pi_{\mathrm{ALLC}} = -c + b\left[\theta p^{\mathrm{C}}_{\mathrm{in}}(\mathrm{G}) + (1-\theta)p^{\mathrm{C}}_{\mathrm{out}}(\mathrm{G})\right] \tag{6}$$

and

$$\pi_{\mathrm{ALLD}} = b\left[\theta p^{\mathrm{D}}_{\mathrm{in}}(\mathrm{G}) + (1-\theta)p^{\mathrm{D}}_{\mathrm{out}}(\mathrm{G})\right], \tag{7}$$

respectively. In Eqs. (6) and (7), $p^{\mathrm{C}}_{\mathrm{in}}(\mathrm{G})$, $p^{\mathrm{C}}_{\mathrm{out}}(\mathrm{G})$, $p^{\mathrm{D}}_{\mathrm{in}}(\mathrm{G})$, and $p^{\mathrm{D}}_{\mathrm{out}}(\mathrm{G})$ represent the probabilities that the mutants selecting C and D gain G reputations in the eyes of ingroup and outgroup observers, and are given by

$$p^a_{\mathrm{in}}(\mathrm{G}) = \sum_{r'\in\{\mathrm{G},\mathrm{B}\}} \left\{\theta p^*_{\mathrm{in}}(r') + (1-\theta)p^*_{\mathrm{out}}(r')\right\}\Phi_{\mathrm{G}}(a,r')$$

and

$$p^a_{\mathrm{out}}(\mathrm{G}) = \sum_{r'\in\{\mathrm{G},\mathrm{B}\}}\sum_{r''\in\{\mathrm{G},\mathrm{B}\}} \left\{\theta p^*_{\mathrm{in}}(r')p^*_{\mathrm{out}}(r'') + (1-\theta)\left[\frac{1}{M-1}p^*_{\mathrm{out}}(r')p^*_{\mathrm{in}}(r'') + (1-\frac{1}{M-1})p^*_{\mathrm{out}}(r')p^*_{\mathrm{out}}(r'')\right]\right\}\times \Phi_{\mathrm{G}}(a,r''),$$

where $a = \mathrm{C}$ or D. The population of DISC players is stable against invasion by ALLC and ALLD mutants if

$$\pi_{\mathrm{DISC}} > \max\left\{\pi_{\mathrm{ALLC}}, \pi_{\mathrm{ALLD}}\right\}. \tag{8}$$

Cooperativeness

DISC donors cooperate exclusively with G recipients. Therefore, in each stable equilibrium, the probability of cooperation, which we call the cooperativeness, toward ingroup and outgroup recipients is given by $p^*_{\mathrm{in}}(\mathrm{G})$ and $p^*_{\mathrm{out}}(\mathrm{G})$, respectively. The cooperativeness for the entire population is given by

$$\psi \equiv \theta p^*_{\mathrm{in}}(\mathrm{G}) + (1-\theta)p^*_{\mathrm{out}}(\mathrm{G}). \tag{9}$$

Measurement of ingroup bias

To quantify the degree of ingroup bias, we measure the difference between ingroup and outgroup cooperativeness, defined by

$$\rho \equiv p^*_{\mathrm{in}}(\mathrm{G}) - p^*_{\mathrm{out}}(\mathrm{G}). \tag{10}$$

When $\rho \approx -1$, players basically cooperate with outgroup recipients and defect against ingroup recipients, implying outgroup favoritism. When $\rho \approx 0$, players equally likely cooperate with ingroup and outgroup recipients. When $\rho \approx 1$, players cooperate with ingroup recipients and defect against outgroup recipients, implying ingroup favoritism.

Results

Table 3 summarizes the results obtained under the three reputation assignment rules. It shows the stable fractions of G players in the eyes of ingroup and outgroup observers (i.e., $p^*_{\text{in}}(\text{G})$ and $p^*_{\text{out}}(\text{G})$), the stability conditions, cooperativeness (i.e., ψ), and the degree of ingroup bias (i.e., ρ).

IM

Under IM, the equilibrium fractions of G players in the eyes of ingroup and outgroup observers are both equal to $\psi = 1/2$. Therefore, ingroup favoritism does not occur, i.e., $\rho = 0$. Furthermore, the population of DISC players is invaded by ALLC mutants such that it is unstable. This result is consistent with the established result that cooperation is usually unstable under IM because observers do not distinguish between selfish defection (i.e., D against G recipients) and justified defection (i.e., D against B recipients) [8, 10, 43, 1].

ST

Under ST, DISC players almost always cooperate with ingroup recipients, i.e., $p^*_{\text{in}}(\text{G}) = 1 - \mu$. This result is consistent with the previous results in which ST enables perfect cooperation when a population does not possess group structure (corresponding to $M = 1$) [11, 12].

The fraction of G players in the eyes of outgroup observers is given by

$$p^*_{\text{out}}(\text{G}) = 1 - \mu \frac{1+\theta}{\theta} + O(\mu^2). \tag{11}$$

Therefore, DISC players almost always cooperate with both ingroup and outgroup recipients unless θ is small (i.e., $\psi = 1 - \mu/\theta + O(\mu^2)$). Because donors defect slightly more often against outgroup than ingroup recipients, weak ingroup favoritism occurs (i.e., $\rho = \mu/\theta + O(\mu^2)$).

Equations (5), (6), and (7) yield the payoff differences given by

$$\pi_{\text{ALLC}} - \pi_{\text{DISC}} = \frac{\mu}{\theta}\left[b(1-\theta) - c\right] + O(\mu^2) \tag{12}$$

and

$$\pi_{\text{ALLD}} - \pi_{\text{DISC}} = -(b - c) + O(\mu). \tag{13}$$

Therefore, the stability condition (Eq. (8)) reads

$$1 < \frac{b}{c} < \frac{1}{1-\theta}. \tag{14}$$

ALLC mutants invade DISC players if $b/c > 1/(1-\theta)$. The cooperation is stable up to a large value of b/c when ingroup interaction is frequent (i.e., large θ). ALLD mutants invade a DISC population under a trivial condition $b/c < 1$.

JG

Under JG, DISC players have the same cooperativeness toward ingroup recipients as under ST, i.e., $p^*_{\text{in}}(\text{G}) = 1 - \mu$. This result is consistent with the previous results in which JG enables perfect cooperation when a population does not possess group structure (corresponding to $M = 1$) [11, 12].

The fraction of G players in the eyes of outgroup observers is given by

$$p^*_{\mathrm{out}}(\mathrm{G}) = \frac{1}{2}. \tag{15}$$

Therefore, DISC players cooperate with outgroup recipients with probability 1/2. In contrast to the case of ST, frequent intergroup interaction considerably reduces cooperation under JG (i.e., $\psi = (1+\theta)/2 + \mu\theta$). The degree of ingroup bias under JG is given by $\rho = 1/2 - \mu$, which is independent of θ. DISC players show a significant level of ingroup favoritism, even though they simply use the reputations without intending to discriminate recipients by the group identity.

The payoff differences are given by

$$\pi_{\mathrm{ALLC}} - \pi_{\mathrm{DISC}} = -\frac{1-\theta}{2}\left[b\frac{M\theta - 1}{M-1} + c\right] + O(\mu) \tag{16}$$

and

$$\pi_{\mathrm{ALLD}} - \pi_{\mathrm{DISC}} = -\frac{1}{2}\left[b\frac{1 + (M-3)\theta + M\theta^2}{M-1} - c(1+\theta)\right] + O(\mu). \tag{17}$$

The stability condition reads

$$\begin{cases} \dfrac{(M-1)(1+\theta)}{1+(M-3)\theta + M\theta^2} < \dfrac{b}{c} < \dfrac{M-1}{1-M\theta} & \text{if } 0 \le \theta < \dfrac{1}{M}, \\ \dfrac{(M-1)(1+\theta)}{1+(M-3)\theta + M\theta^2} < \dfrac{b}{c} & \text{if } \dfrac{1}{M} \le \theta \le 1. \end{cases} \tag{18}$$

The DISC population is resistant to invasion by ALLC mutants when $\theta \ge 1/M$, i.e., when ingroup interaction occurs more frequently than in the case of unbiased random pairing. When $\theta < 1/M$ and $b/c > (M-1)/(1-M\theta)$, ALLC mutants invade the population of DISC players. When $b/c < (M-1)(1+\theta)/[1+(M-3)\theta+M\theta^2]$, ALLD mutants invade the population of DISC players. The cooperation is stable down to a small value of b/c if ingroup interaction is frequent (i.e., large θ) or the number of groups (i.e., M) is small. In the limit $M \to \infty$, Eq. (18) is reduced to $b/c > 1/\theta$, which coincides with the results obtained from a previous model with infinite groups [41].

Under both ST and JG, in particular JG, ingroup favoritism emerges. This is because the donors (equivalently, ingroup observers) and outgroup observers generally perceive different reputations of the same player due to the assignment error (see Figs. 1, 3 **(c)**, 3 **(d)**, and 3 **(e)**). For example, if a donor defects against a recipient whose reputation is B in the eyes of the donor's group members, the donor receives a G reputation from the ingroup observer. However, if the same recipient has a G reputation in the eyes of the outgroup observer, the outgroup observer assigns B to the donor under ST and JG. As another example, if a donor cooperates with a recipient whose reputation is G in the eyes of the donor's group members, the donor receives G from the ingroup observer. However, if the recipient has a B reputation in the eyes of the outgroup observer, the outgroup observer assigns B to the donor under JG. As these examples suggested, different groups may perceive the opposite reputations of the same players in a long run. Players in the same group coordinate the subjective information about a given player's reputation, whereas those in different groups do not. This discrepancy causes ingroup favoritism.

Numerical Results

We compare the theoretical results with numerical results obtained from individual-based simulations in Fig. 4. The procedure of the numerical analysis is described in Appendix A. The analytical and numerical results are sufficiently close to each other in terms of both cooperativeness (Fig. 4 **(a)**) and ingroup bias (Fig. 4 **(b)**).

We also examine the error-prone case in which donors fail to help recipients (i.e., select D when the donors intend C) with probability ϵ [9]. The numerical results for $\epsilon = 0.01$ and 0.1 are shown in Fig. 5. The error reduces cooperativeness (Fig. 5 **(a)**) and ingroup bias (Fig. 5 **(b)**) under both ST and JG (see Fig. 4 for the error-free case). Nevertheless, the results with the error are qualitatively the same as those without the error.

Mixed Assignment Rules

We have shown that JG leads to strong ingroup favoritism, whereas ST does not. To examine the transition between the two regimes, we consider an assignment rule denoted by MX, which is a mixture of JG and ST. In a one-shot game under MX, observers independently assign reputations by using JG with probability α and ST with probability $1 - \alpha$. Therefore, $\Phi_{\rm G}({\rm C}, {\rm G}) = 1 - \mu$, $\Phi_{\rm G}({\rm D}, {\rm G}) = \mu$, $\Phi_{\rm G}({\rm C}, {\rm B}) = \alpha\mu + (1 - \alpha)(1 - \mu)$, and $\Phi_{\rm G}({\rm D}, {\rm B}) = 1 - \mu$. Parameter α controls the degree of sternness with which observers assign B to donors that cooperate with B recipients. ST and JG correspond to $\alpha = 0$ and $\alpha = 1$, respectively. We numerically solve Eqs. (2) and (4) under MX.

The results under MX are shown in Fig. 6. Sternness gradually decreases cooperativeness (Fig. 6 **(a)**) and increases ingroup bias (Fig. 6 **(b)**) for different values of M and θ. The results interpolate those for ST and JG and imply that sternness promotes ingroup favoritism. The shaded parameter regions in Fig. 6 **(c)**–**(f)** indicate the values of α and b/c for which DISC residents are stable. Above (below) the shaded regions, ALLC (ALLD) mutants invade the DISC population. In all the cases, the upper and lower bounds of the stability region in terms of b/c increase with α. A decrease in M induces cooperativeness and reduces ingroup bias. A decrease in M also broadens the stability regions if θ is large. An increase in θ induces cooperativeness, reduces ingroup bias, and broadens the stability regions for the following reason. When θ is large, players are largely involved in ingroup interactions. Then, they do not suffer from a B reputation that outgroup observers may frequently attach to the donor because of the discrepancy between players' reputations perceived by different groups (see subsection JG in Results for related discussion).

Heterogeneous Assignment Rules

We have assumed that all the groups use a common reputation assignment rule. In this section, we numerically examine a case in which observers in different groups use different reputation assignment rules. We consider a situation in which m ($1 \leq m \leq M - 1$) groups use JG and $M - m$ groups use ST. The procedure of the numerical analysis is described in Appendix B.

Numerically obtained equilibria with $M = 8$ and $M = 20$ are shown in Fig. 7 **(a)** and **(b)**, respectively. As the number of JG groups (i.e., m) increases, the cooperativeness ($\psi_{\rm ST}$ and $\psi_{\rm JG}$ for ST and JG groups, respectively) decreases, and ingroup bias ($\rho_{\rm ST}$ and $\rho_{\rm JG}$ for ST and JG groups, respectively) increases. Figure 7 **(c)** and **(d)** shows the difference between the payoff to a player in a ST group and that to a player in a JG group (i.e., $\pi_{\rm JG} - \pi_{\rm ST}$) when $M = 8$ and

$M = 20$, respectively. When the benefit-to-cost ratio is small (i.e., $b = 2$), $\pi_{\mathrm{JG}} - \pi_{\mathrm{ST}}$ is positive. Therefore, if observers update their assignment rules according to an evolutionary dynamics (e.g., group competition [16]), the evolutionary dynamics would lead to a homogeneous population in which all the observers adopt JG. When the benefit-to-cost ratio is large (i.e., $b = 6$), $\pi_{\mathrm{JG}} - \pi_{\mathrm{ST}}$ is positive when m is large and negative when m is small. This implies that a homogeneous population of ST and that of JG are bistable under evolutionary dynamics. The basin of attraction for the homogeneous ST population in terms of m broadens as b increases. When the benefit-to-cost ratio takes an intermediate value (i.e., $b = 4$), the results for $M = 8$ (Fig. 7 **(c)**) and those for $M = 20$ (Fig. 7 **(d)**) are qualitatively different. For $M = 8$, $\pi_{\mathrm{JG}} - \pi_{\mathrm{ST}}$ is negative only when $m = 2$ or 3. Therefore, a stable mixture of ST and JG groups and a homogeneous population of JG are bistable. For $M = 20$, a homogeneous population of ST and that of JG are bistable.

Discussion

In the present study, we showed that ingroup favoritism emerges in a group-structured model of indirect reciprocity. In our model, players share information about reputations in each group but not across different groups. We assumed that a player's action purely depends on the coplayer's reputation; players do not refer to the group identity of the coplayers or use other types of prejudices. We also assumed that observers impartially assess ingroup and outgroup donors. Ingroup favoritism occurs under both simple standing (ST) and stern judging (JG) assignment rules. The cooperativeness is reduced by the frequent intergroup interactions, i.e., small θ. The ingroup bias is severer and the cooperativeness is smaller under JG than under ST. The parameter region for the stability of the cooperative equilibrium is larger under JG than under ST. Under ST and JG, a population of discriminators is evolutionarily stable if the probability of ingroup interaction (i.e., θ) is sufficiently large. If θ is small, the population is invaded by unconditional cooperators and unconditional defectors under ST and JG, respectively. We also studied the case in which observers may adopt different assignment rules in different groups. We found that JG would dominate ST in evolutionary settings when the benefit-to-cost ratio is small. Otherwise, the homogeneous population in which all the groups employ ST and that in which all the groups employ JG are bistable in large parameter regions.

Different mechanisms govern ingroup favoritism in our model and that observed in psychological experiments [24, 25, 26, 27, 28, 29, 30, 31, 32, 33]. In the latter, players use a cue that indicates the group identity of the coplayer and preferably cooperate with ingroup members. In our model, players do not refer to the group identity of the coplayer. They show ingroup favoritism because they perceive that outgroup members have bad reputations more often than do ingroup members.

We implemented the group structure by controlling probabilities of ingroup and outgroup interactions (i.e., θ and $1 - \theta$, respectively) and assuming the groupwise information sharing. In terms of the structure of information sharing, most previous theoretical studies of indirect reciprocity are classified into two types: public [5, 6, 8, 9, 11, 12, 15, 16, 14, 17, 41] and private [7, 10, 13, 17, 19] reputation models.

In public reputation models, all the players have access to a common information source that provides the reputation values of the players. Therefore, a donor and observer perceive the same reputation of a recipient such that they do not suffer from the discrepancy of reputations. In public reputation models without group structure of the population, ST and JG realize evolutionarily stable cooperation [11, 12]. This result is consistent with ours because, in the limit $\theta \to 1$, Eqs. (14) and (18) are reduced to a trivial condition $b/c > 1$ such that the population of discriminators is stable under ST and JG.

In private reputation models, each player individually collects others' reputations such that a reputation of a player varies between individuals. In contrast to the case of public reputation models, a homogeneous population of discriminators is invaded by unconditional cooperators in private reputation models. A mixture of discriminators and unconditional cooperators is often stable under variants of ST [7, 10, 13, 19]. Under variants of JG, a population of discriminators is invaded by unconditional defectors [17, 19] (but see Ref. [13]), or discriminators and unconditional cooperators are frequent in an island model if dispersal of offspring is confined within each island [10]. In the limit $\theta \to 0$ and $M \to \infty$, our model can be interpreted as a private reputation model. In this situation, the population of discriminators is unstable because Eqs. (14) and (18) are violated. Therefore, the results obtained from our model in this limit are consistent with the previous results.

For intermediated θ and M values, our model uses a public reputation scheme within each group and a private reputation scheme across groups. In this sense, the structure of information sharing in our model is situated between public and private reputation models.

One of the present authors previously studied a model of ingroup favoritism on the basis of indirect reciprocity [41], which we refer to as the multiple standard model. The multiple standard model and the model analyzed in the present study are different in two aspects. First, in the multiple standard model, a given player's reputation is made public to different groups such that the problem of coordination in regard to reputations among different groups does not exist. In the present model, observers in different groups may differently perceive a player's reputation, which leads to the coordination problem. Second, in the multiple standard model, observers are allowed to use different rules to assign reputations to ingroup and outgroup members. Similarly, donors may use different action selection rules toward ingroup and outgroup recipients. Then, ingroup favoritism of different degrees emerges. Consider a situation in which the action rule is of a single standard such that donors are discriminators toward both ingroup and outgroup recipients. Then, at most partial ingroup favoritism in which players always cooperate with ingroup members and partially (i.e., with probability 1/2) cooperate with outgroup members is evolutionarily stable. Consider another situation in which the action rule is of a double standard such that donors are discriminators toward ingroup members and unconditional defectors toward outgroup members. Then, perfect ingroup favoritism in which players always cooperate with ingroup members and always defect against outgroup members is evolutionarily stable. In the present model, observers use a single-standard reputation assignment rule, and donors use a single-standard action rule. Then, partial ingroup favoritism, but not perfect ingroup favoritism, can be evolutionarily stable.

Group competition models of indirect reciprocity were previously studied [15, 16]. The authors numerically examined competition between different assignment rules employed in different groups. In our terminology, they assumed that the donation game is played inside each group and that reputations are updated exclusively by ingroup observers under the public reputation scheme. They showed that JG (stern-judging in their terminology) emerges in the course of evolutionary dynamics based on group competition and individual selection. Their models and ours are fundamentally different although both studies have stressed the importance of JG. First, they assumed group competition and we did not. Second, they mainly focused on competition between different assignment rules and we did not; we only studied the special case in which observers in different groups adopt either of ST or JG. Third, we determined the possibility of ingroup favoritism and group-independent cooperation. In contrast, their model is not concerned with ingroup favoritism because interaction between a donor and recipient in different groups is not assumed.

Uchida and Sigmund analyzed competition between assignment rules by using replicator dynamics [18]. In their model, a player selected as donor uses the public information source corresponding to the assignment rule that the player adopts. For example, if the surviving assignment rules are only ST and JG (SUGDEN and KANDORI, respectively, in their terminology), there are two public information sources. Although their model is apparently a public reputation model, the players can be interpreted to belong to one of the groups defined by the assignment rule; members in each group share a common information source and use the same assignment rule. Helping a recipient having a bad reputation in the eyes of both ST and JG groups is assessed to be good by the ST group and bad by the JG group. Therefore, JG players assess ST players to be bad more often than they assess JG players. Because this tendency is strong enough, ingroup favoritism occurs in the JG group. Their model and ours are consistent with each other because, when different groups can adopt different assignment rules, both their model and ours with sufficiently many groups predict bistability between ST and JG. Their model and ours complement each other in the following respects. First, they investigated competition between assignment rules, whereas we mainly studied the case in which all the groups share an assignment rule. Second, they assumed a well-mixed population, whereas we varied the frequency of ingroup and outgroup interactions. Third, they studied competition among at most five groups (i.e., five assignment rules), whereas we assumed a general number of groups.

Conclusion

To explore the possibility of spontaneous ingroup favoritism in indirect reciprocity, we analyzed a social dilemma game in a population with group structure. We showed that the degree of ingroup bias depends on the reputation assignment rule. In particular, considerable ingroup favoritism occurs under the so-called JG assignment rule, whereby observers assign bad reputations to players helping bad players. Ingroup favoritism has been considered to be an evolutionary outcome [25, 26, 27, 29, 33]. The present work supports this general idea. To measure the dependency of ingroup bias on the assignment rule in behavioral experiments may be an interesting challenge.

Appendices

A Numerical methods in the case of the homogeneous assignment rule

We prepare a population of $N = 10^3$ DISC players divided into M groups of equal size. We consider an $N \times M$ reputation matrix, denoted by $\boldsymbol{R} = (r_{i,\ell})$, where $r_{i,\ell} \in \{\mathrm{G}, \mathrm{B}, \mathrm{U}\}$ represents the reputation of player i $(1 \leq i \leq N)$ in the eyes of the observer in group ℓ $(1 \leq \ell \leq M)$. U represents the unknown reputation. We assume that all the entries of $\boldsymbol{R}$ are equal to U in the beginning of a run. In a one-shot donation game, we randomly select a player i as donor. Then, with probability $\theta/(N/M - 1)$, we select a recipient $j (\neq i)$ that is in the donor's group. With probability $(1 - \theta)/(N - N/M)$, we select a recipient j that is in a group different from the donor's group. When determining the action, the donor refers to $r_{j,k}$, where k is the donor's group. We assume that the donor cooperates when $r_{j,k} = \mathrm{U}$. After the game, the observer in each group ℓ $(1 \le \ell \le M)$ assigns a new reputation to donor i such that $r_{i,\ell} = \mathrm{G}$ with probability $\Phi_{\mathrm{G}}(a, r_{j,\ell})$ and $r_{i,\ell} = \mathrm{B}$ with probability $1 - \Phi_{\mathrm{G}}(a, r_{j,\ell})$, where $a \in \{\mathrm{C}, \mathrm{D}\}$ is the donor's action and $\Phi_{\mathrm{G}}(a, r_{j,\ell})$ under each assignment rule is defined in Tab. 2. When $r_{j,\ell} = \mathrm{U}$, we assume that the observer uses IM; $\Phi_{\mathrm{G}}(\mathrm{U}, \mathrm{C}) = 1 - \mu$ and $\Phi_{\mathrm{G}}(\mathrm{U}, \mathrm{D}) = \mu$. We set $\mu = 0.01$.

After repeating $T = 10^5$ rounds of the donation game, we calculate the fraction of G players in group k in the eyes of the observer in group ℓ, which is given by $\Phi^*_{k,\ell}(\mathrm{G}) = \sum_{i=1;\ \text{player } i \text{ in group } k}^{N} \delta(r_{i,\ell})/(N/M)$, where $\delta(\mathrm{G}) = 1$ and $\delta(\mathrm{B}) = \delta(\mathrm{U}) = 0$. The fractions of G players in the eyes of ingroup and outgroup observers are given by $\Phi^*_{\mathrm{in}}(\mathrm{G}) = \sum_{k=1}^{M} \Phi^*_{k,k}(\mathrm{G})/M$ and $\Phi^*_{\mathrm{out}}(\mathrm{G}) = \sum_{k=1}^{M} \sum_{\ell=1,\ell\neq k}^{M} \Phi^*_{k,\ell}(\mathrm{G})/[M(M-1)]$, respectively. By substituting these quantities in Eqs. (9) and (10), we obtain ψ and ρ. We average ψ and ρ over 10^2 runs of the simulation.

B Numerical methods in the case of the heterogeneous assignment rule

To analyze heterogeneous populations, we assume that observers in groups $1, 2, \cdots, m$ adopt JG and those in groups $m+1, m+2, \cdots, M$, adopt ST. By applying the procedure explained in Appendix A, we obtain the fraction of G players in group k in the eyes of the observer in group ℓ, i.e., $\Phi^*_{k,\ell}(\mathrm{G})$. The probability that a donor in group k helps a recipient is given by

$$\psi_k = \theta \Phi^*_{k,k}(\mathrm{G}) + (1-\theta)\frac{1}{M-1}\sum_{\ell=1,\ell\neq k}^{M} \Phi^*_{\ell,k}(\mathrm{G}). \tag{19}$$

The probability that a recipient in group k is helped by a donor is given by

$$\phi_k = \theta \Phi^*_{k,k}(\mathrm{G}) + (1-\theta)\frac{1}{M-1}\sum_{\ell=1,\ell\neq k}^{M} \Phi^*_{k,\ell}(\mathrm{G}). \tag{20}$$

The ingroup bias of the players in group k is given by

$$\rho_k = \Phi^*_{k,k}(\mathrm{G}) - \frac{1}{M-1}\sum_{\ell=1,\ell\neq k}^{M} \Phi^*_{\ell,k}(\mathrm{G}). \tag{21}$$

The payoff to the players in group k is given by

$$\pi_k = -c\psi_k + b\phi_k. \tag{22}$$

The cooperativeness, ingroup bias, and payoff to the players in groups employing JG and ST are defined by $Q_{\mathrm{JG}} = \sum_{k=1}^{m} Q_k/m$ and $Q_{\mathrm{ST}} = \sum_{k=m+1}^{M} Q_k/(M-m)$, respectively, where Q represents either ψ, ρ, or π. We average these quantities over 10^2 runs for each parameter set to generate Fig. 7.

Acknowledgements

We thank Shoma Tanabe for careful reading of the manuscript. MN acknowledges the support provided through Grants-in-Aid for Scientific Research (No. 10J08999) from JSPS, Japan. NM acknowledges the support provided through Grants-in-Aid for Scientific Research (No. 23681033) and Innovative Areas "Systems Molecular Ethology"(No. 20115009) from MEXT, Japan.

References

[1] Sigmund K: **Moral assessment in indirect reciprocity**. *J Theor Biol* 2012, **299**:25–30.

[2] Pfeiffer T, Tran L, Krumme C, Rand DG: **The value of reputation**. *J R Soc Interface* 2012.

[3] Alexander RD: *The Biology of Moral Systems*. New York, NY: Aldine de Gruyter 1987.

[4] Trivers RL: **The evolution of reciprocal altruism**. *Q Rev Biol* 1971, **46**:35–37.

[5] Nowak MA, Sigmund K: **Evolution of indirect reciprocity by image scoring**. *Nature* 1998, **393**:573–577.

[6] Nowak MA, Sigmund K: **The dynamics of indirect reciprocity**. *J Theor Biol* 1998, **194**:561–574.

[7] Leimar O, Hammerstein P: **Evolution of cooperation through indirect reciprocity**. *Proc R Soc B* 2001, **268**:745–753.

[8] Panchanathan K, Boyd R: **A tale of two defectors: The importance of standing for evolution of indirect reciprocity**. *J Theor Biol* 2003, **224**:115–126.

[9] Fishman MM: **Indirect reciprocity among imperfect individuals**. *J Theor Biol* 2003, **225**:285–292.

[10] Brandt H, Sigmund K: **The logic of reprobation: Assessment and action rules for indirect reciprocation**. *J Theor Biol* 2004, **231**:475–486.

[11] Ohtsuki H, Iwasa Y: **How should we define goodness?–reputation dynamics in indirect reciprocity**. *J Theor Biol* 2004, **231**:107–120.

[12] Ohtsuki H, Iwasa Y: **The leading eight: Social norms that can maintain cooperation by indirect reciprocity**. *J Theor Biol* 2006, **239**:435–444.

[13] Takahashi N, Mashima R: **The importance of subjectivity in perceptual errors on the emergence of indirect reciprocity**. *J Theor Biol* 2006, **243**:418–436.

[14] Ohtsuki H, Iwasa Y: **Global analyses of evolutionary dynamics and exhaustive search for social norms that maintain cooperation by reputation**. *J Theor Biol* 2007, **244**:518–531.

[15] Chalub FACC, Santos FC, Pacheco JM: **The evolution of norms**. *J Theor Biol* 2006, **241**:233–240.

[16] Pacheco JM, Santos FC, Chalub FACC: **Stern-judging: A simple, successful norm which promotes cooperation under indirect reciprocity**. *PLoS Comp Biol* 2006, **2**:e178.

[17] Ohtsuki H, Iwasa Y, Nowak MA: **Indirect reciprocity provides only a narrow margin of efficiency for costly punishment**. *Nature* 2009, **457**:79–82.

[18] Uchida S, Sigmund K: **The competition of assessment rules for indirect reciprocity**. *J Theor Biol* 2010, **263**:13–19.

[19] Uchida S: **Effect of private information on indirect reciprocity**. *Phys Rev E* 2010, **82**:1–8.

[20] Wedekind C, Milinski M: **Cooperation through image scoring in humans**. *Science* 2000, **288**:850–852.

[21] Semmann D, Krambeck HJ, Milinski M: **Strategic investment in reputation**. *Behav Ecol Sociobiol* 2004, **56**:248–252.

[22] Bolton GE, Katok E, Ockenfels A: **Cooperation among strangers with limited information about reputation**. *J Public Econ* 2005, **89**:1457–1468.

[23] Sommerfeld RD, Krambeck HJ, Semmann D, Milinski M: **Gossip as an alternative for direct observation in games of indirect reciprocity**. *Proc Natl Acad Sci USA* 2007, **104**:17435–17440.

[24] Tajfel H, Billig MG, Bundy RP, Flament C: **Social categorization and intergroup behaviour**. *Eur J Soc Psychol* 1971, **1**:149–178.

[25] Yamagishi T, Jin N, Miller AS: **In-group bias and culture of collectivism**. *Asian Soc Psychol* 1998, **1**:315–328.

[26] Yamagishi T, Jin N, Kiyonari T: **Bounded generalized reciprocity: Ingroup boasting and ingroup favoritism**. *Adv Group Proc* 1999, **16**:161–197.

[27] Kiyonari T, Tanida S, Yamagishi T: **Social exchange and reciprocity: confusion or a heuristic?** *Evol Hum Behav* 2000, **21**:411–427.

[28] Goette L, Huffman D, Meier S: **The impact of group membership evidence using random assignment to real enforcement : Social groups**. *Am Econc Rev* 2006, **96**:212–216.

[29] Yamagishi T, Mifune N: **Does shared group membership promote altruism?: Fear, greed, and reputation**. *Ration Soc* 2008, **20**:5–30.

[30] Fowler JH, Kam CD: **Beyond the self: Social identity, altruism, and political participation**. *J Polit* 2008, **69**:813–827.

[31] Rand DG, Pfeiffer T, Dreber A, Sheketoff RW, Wernerfelt NC, Benkler Y: **Dynamic remodeling of in-group bias during the 2008 presidential election**. *Proc Natl Acad Sci USA* 2009, **106**:6187–6191.

[32] Güth W, Ploner M, Regner T: **Determinants of in-group bias: Is group affiliation mediated by guilt-aversion?** *J Econ Psychol* 2009, **30**:814–827.

[33] Mifune N, Hashimoto H, Yamagishi T: **Altruism toward in-group members as a reputation mechanism**. *Evol Hum Behav* 2010, **31**:109–117.

[34] Jansen VAA, van Baalen M: **Altruism through beard chromodynamics**. *Nature* 2006, **440**:663–666.

[35] Traulsen A: **Mechanisms for similarity based cooperation**. *Euro Phys J B* 2008, **63**:363–371.

[36] Antal T, Ohtsuki H, Wakeley J, Taylor PD, Nowak MA: **Evolution of cooperation by phenotypic similarity**. *Proc Natl Acad Sci USA* 2009, **106**:8597–8600.

[37] Masuda N, Ohtsuki H: **Tag-based indirect reciprocity by incomplete social information**. *Proc R Soc B* 2007, **274**:689–695.

[38] Fu F, Tarnita CE, Christakis NA, Wang L, Rand DG, Nowak MA: **Evolution of in-group favoritism**. *Sci Rep* 2012, **2**:1–6.

[39] Choi JK, Bowles S: **The coevolution of parochial altruism and war**. *Science* 2007, **318**:636–640.

[40] Konrad KA, Morath F: **Evolutionarily stable in-group favoritism and out-group spite in intergroup conflict**. *J Theor Biol* 2012, **306**:61–67.

[41] Masuda N: **Ingroup favoritism and intergroup cooperation under indirect reciprocity based on group reputation**. *J Theor Biol* 2012, **311**:8–18.

[42] Fortunato S: **Community detection in graphs**. *Phys Rep* 2010, **486**:75–174.

[43] Nowak MA, Sigmund K: **Evolution of indirect reciprocity**. *Nature* 2005, **437**:1291–1298.

Figures

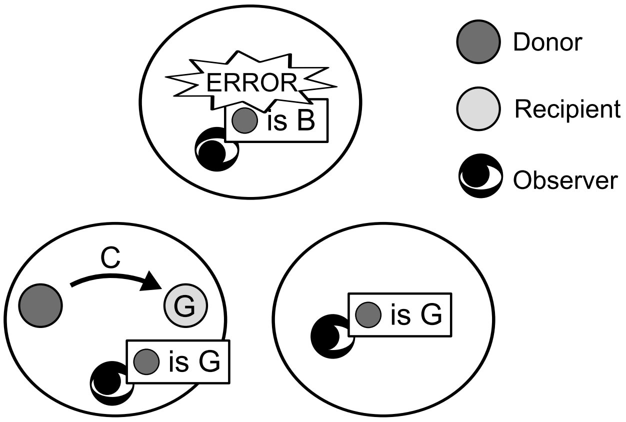
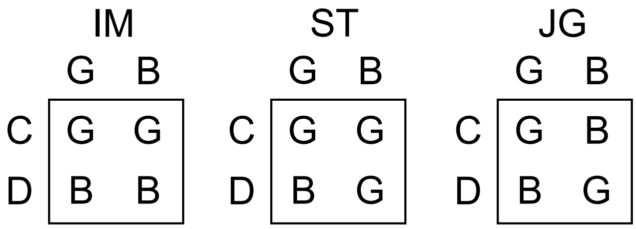


Figure 1: **Behavior of different observers in different groups** ($M = 3$).



IM

	G	B
C	G	G
D	B	B

ST

	G	B
C	G	G
D	B	G

JG

	G	B
C	G	B
D	B	G

Figure 2: **Three reputation assignment rules**. Image scoring (IM), simple standing (ST), and stern judging (JG). The rows represent the donor's actions (i.e., C and D), the columns represent the recipient's reputations (G and B), and G and B inside the boxes represent the reputations that observers assign to the donor.

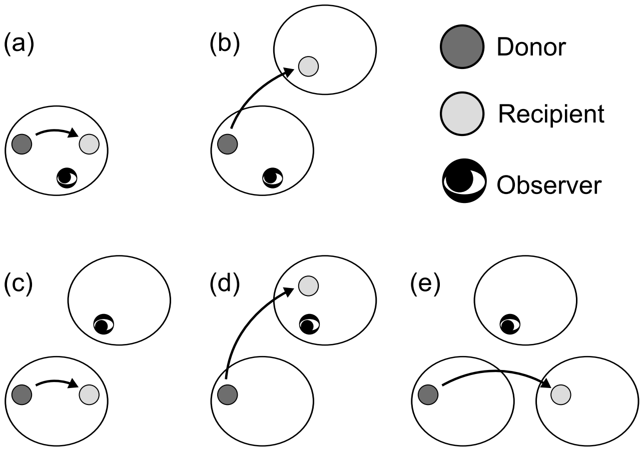


Figure 3: **Five possible situations of the reputation update**. Observations are made by ingroup observers in **(a)** and **(b)**, and by outgroup observers in **(c)**, **(d)**, and **(e)**.

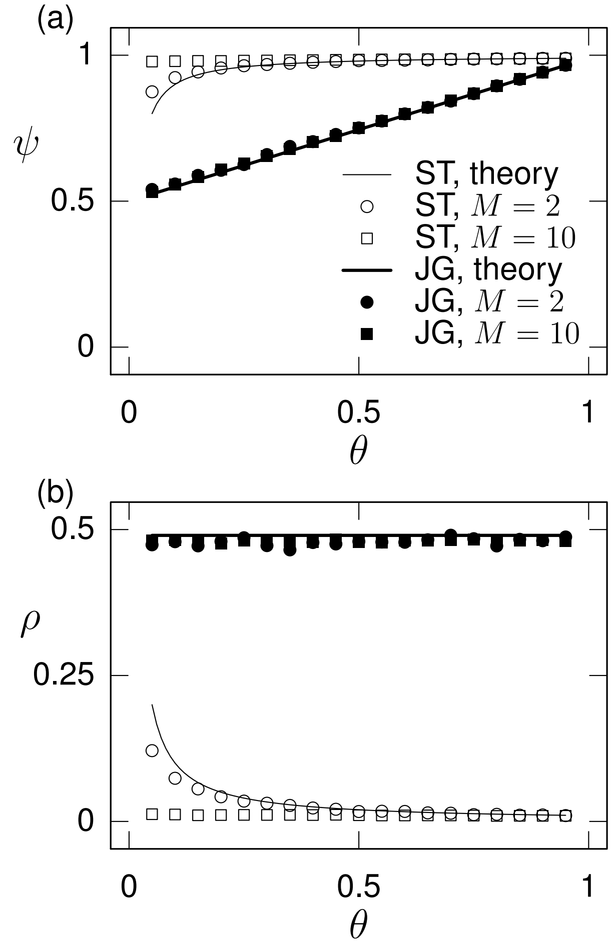


Figure 4: **Equilibria for a population of DISC players under ST and JG**. **(a)** Cooperativeness (ψ) and **(b)** ingroup bias (ρ). We vary the assignment rule (ST or JG), the number of groups ($M = 2$ or 10), and the probability of ingroup interaction (θ). The lines represent theoretical results shown in Tab. 3. The symbols represent numerical results.

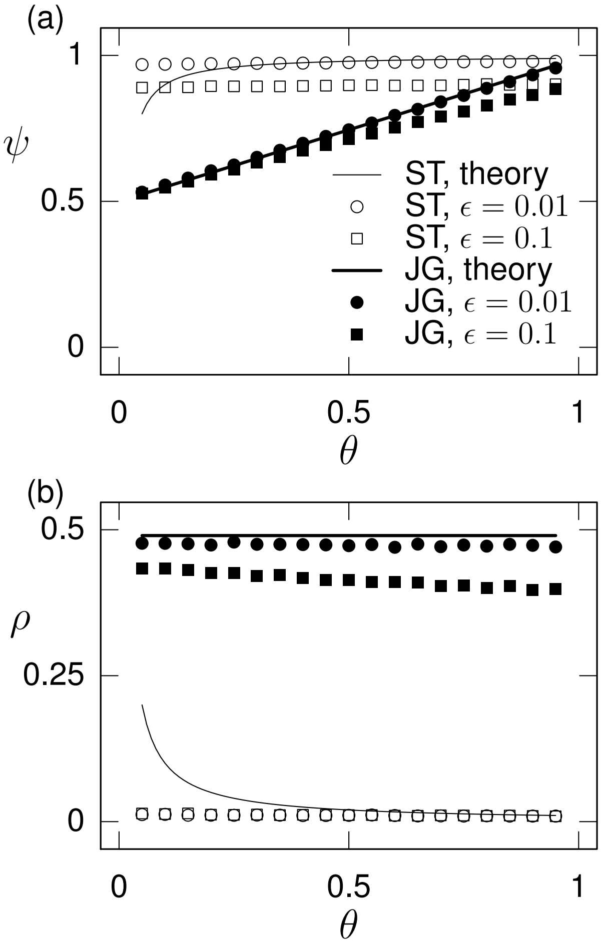


Figure 5: **Equilibria for a population of DISC players under action implementation error**. **(a)** Cooperativeness (ψ) and **(b)** ingroup bias (ρ). We fix the number of groups ($M = 10$) and vary the assignment rule (ST or JG), the probability that a donor fails to help a recipient ($\epsilon = 0.01$ or 0.1), and the probability of ingroup interaction (θ). The lines represent theoretical results when $\epsilon = 0$ and are the replicates of those shown in Tab. 3. The symbols represent numerical results.

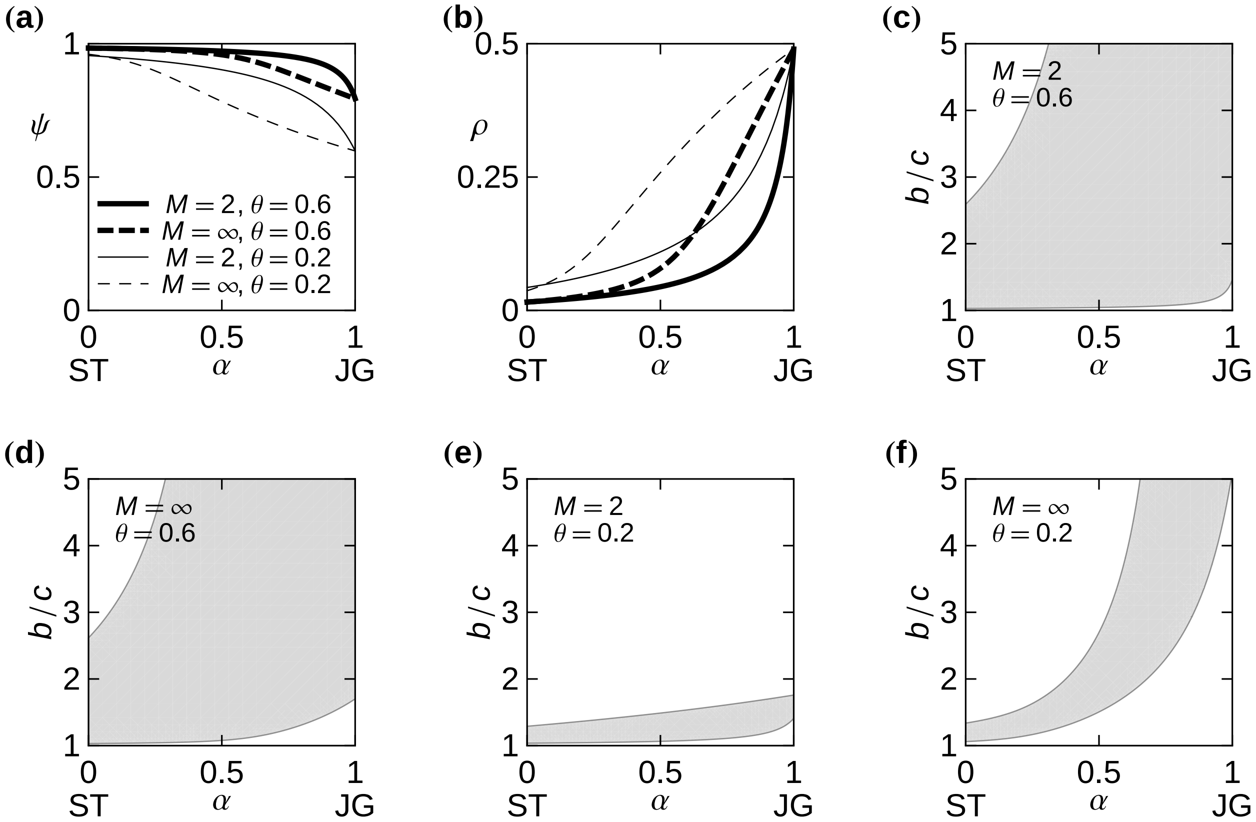


Figure 6: **Equilibria and the stability conditions for a population of DISC players under MX**. **(a)** Cooperativeness (ψ) and **(b)** ingroup bias (ρ). In **(a)** and **(b)**, we set $(M, \theta) = (2, 0.6)$, $(\infty, 0.6)$, $(2, 0.2)$, and $(\infty, 0.2)$. **(c)**–**(f)** Stability conditions. The homogeneous population of DISC players is stable in the shaded parameter regions. We set **(c)** $(M, \theta) = (2, 0.6)$, **(d)** $(M, \theta) = (\infty, 0.6)$, **(e)** $(M, \theta) = (2, 0.2)$, and **(f)** $(M, \theta) = (\infty, 0.2)$.

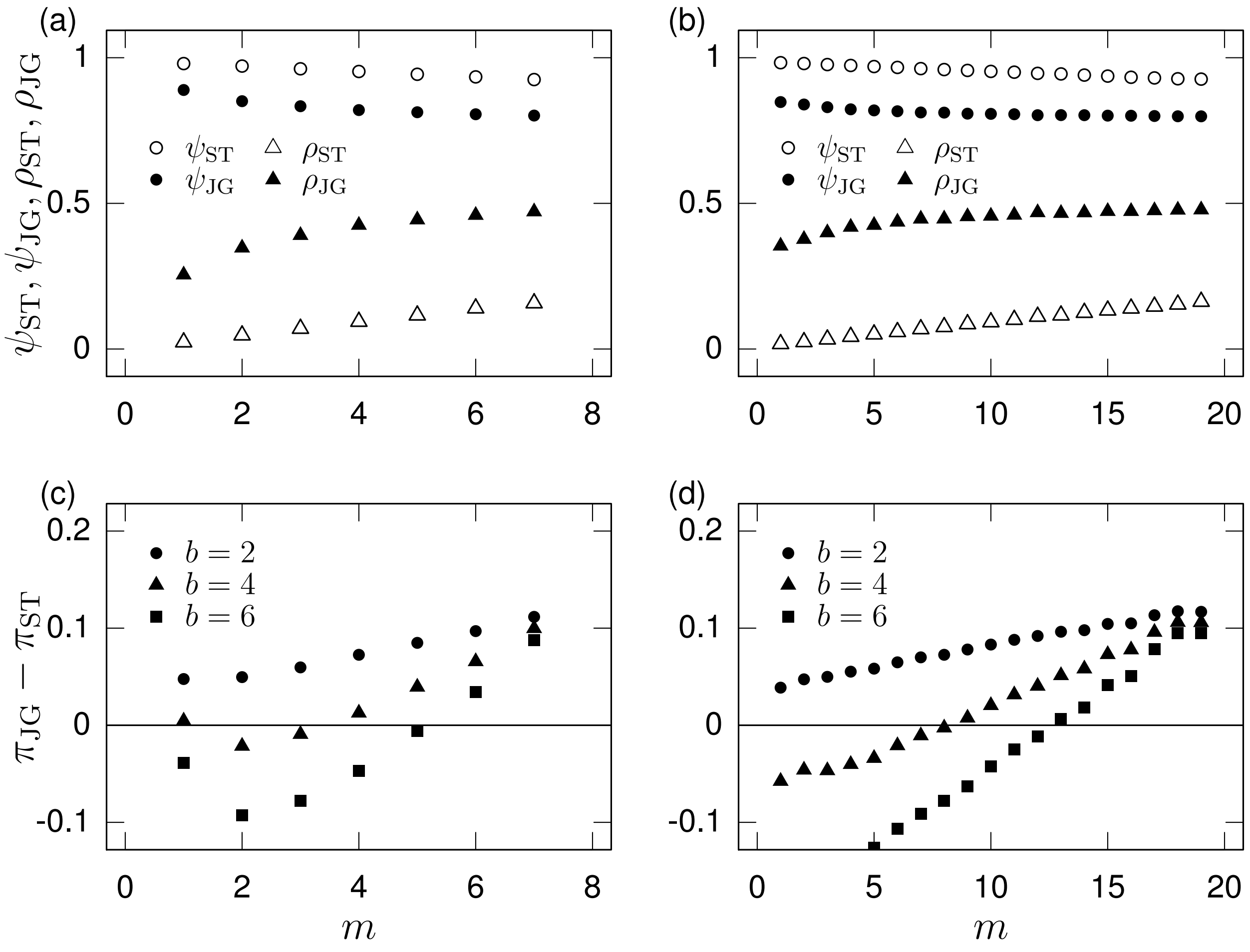


Figure 7: **Equilibria for a population of DISC players under heterogeneous assignment rules**. **(a)**, **(b)** Cooperativeness (ψ_{ST} and ψ_{JG}) and ingroup bias (ρ_{ST} and ρ_{JG}) for groups employing ST and JG. **(c)**, **(d)** Payoff difference between a player in a ST group and that in a JG group ($\pi_{\mathrm{JG}} - \pi_{\mathrm{ST}}$). We set $\theta = 0.6$ and $c = 1$. We also set $M = 8$ in **(a)** and **(c)**, $M = 20$ in **(b)** and **(d)**, and vary the number of JG groups (i.e., m) and b.

Tables

Symbol	Meaning
M	Number of groups
θ	Probability that a donor and recipient in a one-shot game are in the same group
$\boldsymbol{r} \in \{\mathrm{G}, \mathrm{B}\}^M$	Reputation vector of a player in the eyes of M observers
$p_k(\boldsymbol{r})$	Probability that a player in group k has reputation vector $\boldsymbol{r}$
$p_{-k}(\boldsymbol{r})$	Probability that a player outside group k has reputation vector $\boldsymbol{r}$
$\sigma(r) \in \{\mathrm{C}, \mathrm{D}\}$	Donor's action to a recipient having reputation $r \in \{\mathrm{G}, \mathrm{B}\}$
$\Phi_r(a, r')$	Probability that an observer assigns reputation $r \in \{\mathrm{G}, \mathrm{B}\}$ to a donor selecting action $a \in \{\mathrm{C}, \mathrm{D}\}$ to a recipient having reputation $r' \in \{\mathrm{C}, \mathrm{D}\}$
$p_{\mathrm{in}}(r)$	Probability that a player in the eyes of an ingroup observer has reputation r
$p_{\mathrm{out}}(r)$	Probability that a player in the eyes of an outgroup observer has reputation r

Table 1: **Meaning of symbols**.

Rule	$\Phi_{\mathrm{G}}(\mathrm{C}, \mathrm{G})$	$\Phi_{\mathrm{G}}(\mathrm{D}, \mathrm{G})$	$\Phi_{\mathrm{G}}(\mathrm{C}, \mathrm{B})$	$\Phi_{\mathrm{G}}(\mathrm{D}, \mathrm{B})$
IM	$1-\mu$	μ	$1-\mu$	μ
ST	$1-\mu$	μ	$1-\mu$	$1-\mu$
JG	$1-\mu$	μ	μ	$1-\mu$

Table 2: **Probability that an observer assigns G to a donor**. $\Phi_{\mathrm{G}}(a, r)$ represents the probability that a donor receives G when the donor selects action $a \in \{\mathrm{C}, \mathrm{D}\}$ and the recipient has reputation $r \in \{\mathrm{G}, \mathrm{B}\}$. The donor receives B with probability $\Phi_{\mathrm{B}}(a, r) = 1 - \Phi_{\mathrm{G}}(a, r)$.

Rule	$p^*_{\mathrm{in}}(\mathrm{G})$	$p^*_{\mathrm{out}}(\mathrm{G})$	Stability condition	ψ	ρ
IM	$\frac{1}{2}$	$\frac{1}{2}$	Unstable	$\frac{1}{2}$	0
ST	$1-\mu$	$1-\mu\frac{1+\theta}{\theta}+O(\mu^2)$	Eq. (14)	$1-\frac{\mu}{\theta}+O(\mu^2)$	$\frac{\mu}{\theta}+O(\mu^2)$
JG	$1-\mu$	$\frac{1}{2}$	Eq. (18)	$\frac{1+\theta}{2}-\mu\theta$	$\frac{1}{2}-\mu$

Table 3: **Equilibria and the stability conditions for a population of DISC players under different assignment rules**.