Indirect reciprocity in three types of social dilemmas

Mitsuhiro Nakamura[1,*] and Hisashi Ohtsuki[1]

[1]*Department of Evolutionary Studies of Biosystems,*
The Graduate University for Advanced Studies (SOKENDAI),
Hayama, Kanagawa 240-0193, Japan

(Dated: March 19, 2014)

Abstract

Indirect reciprocity is a key mechanism for the evolution of human cooperation. Previous studies explored indirect reciprocity in the so-called donation game, a special class of Prisoner's Dilemma (PD) with unilateral decision making. A more general class of social dilemma includes Snowdrift (SG), Stag Hunt (SH), and PD games, where two players perform actions simultaneously. In these simultaneous-move games, moral assessments need to be more complex; for example, how should we evaluate defection against an ill-reputed, but now cooperative, player? We examined indirect reciprocity in the three social dilemmas and identified twelve successful social norms for moral assessments. These successful norms have different principles in different dilemmas for suppressing cheaters. To suppress defectors, any defection against good players is prohibited in SG and PD, whereas defection against good players may be allowed in SH. To suppress unconditional cooperators, who help anyone and thereby indirectly contribute to jeopardizing indirect reciprocity, we found two mechanisms: indiscrimination between actions towards bad players (feasible in SG and PD) or punishment for cooperation with bad players (effective in any social dilemma). Moreover, we discovered that social norms that unfairly favour reciprocators enhance robustness of cooperation in SH, whereby reciprocators never lose their good reputation.

Keywords: evolutionary game theory; indirect reciprocity; Prisoner's Dilemma game; Snowdrift game; Stag Hunt game

* Author for correspondence: nakamuramh@soken.ac.jp

I. INTRODUCTION

In everyday life, your social image influences what you obtain. Helping someone raises your reputation in your community and others help you later when required. This is called indirect reciprocity, a key mechanism for explaining the evolution of cooperative behavior among unrelated individuals [1–3]. Indirect reciprocity based on reputation has been extensively investigated for decades through numerous theoretical studies [4–23] and experimental tests [24–31]. The global success of humans in the past was partially dependent on the establishment of indirect reciprocity, as it was used to explore for more suitable partners for effective economic exchange instead of maintaining closed transactions in inefficient relationships [4, 32].

One important feature of indirect reciprocity is that it endogenously provides an incentive for actors to reward or punish other community members, which is achieved by controlling the actors' reputations that lead to the future rewards or punishments for the actors themselves. We can imagine numerous possibilities of rules to control the reputation of actors that behave differently in various social contexts; such rules are called social norms [4, 10]. Some promising norms can stabilize cooperation in indirect reciprocity, but others cannot. Previous studies have systematically obtained successful social norms in Prisoner's Dilemma scenarios when the reputation information is well-shared in a population [10, 11], when it belongs to each individual [9], with the presence of costly punishment [13], with incomplete reputation information [19], with multiple reputation states [22], and with group-level reputations [21].

Most of the previous studies have investigated social norms for the so-called donation game, a variant of Prisoner's Dilemma with unilateral decision making [33]. In the donation game, two individuals called donor and recipient participate in and only the donor can decide whether or not to help the recipient, *i.e.*, whether to benefit the recipient by making an investment. Because the donation game focuses on the unilateral behavior of a donor, it ignores many aspects that exist in reality. One such aspect is that the donation game is merely an instance of various social dilemmas. Reputation systems would also play an important role in various simultaneous-move games such as Snowdrift, Stag Hunt, and general Prisoner's Dilemma games. In these games, social norms may depend not only on an actor's choice but also on his/her co-player's choice. For example, how should we define goodness when an actor defects against a bad co-player that unexpectedly cooperates with the actor? Should the actor's defection be justified, even if the co-player shows reformation? Moreover, individuals could infer that a focal player's reputation should be bad when the player received punishment from another player who had established a high reputaion. Can such possibility be stable in evolutionry scenarios? To the best of our knowledge, although two previous studies have investigated games other than the donation game, they have not done so exhaustively and not clarified the general characteristics of social norms for the simultaneous-move games [4, 18].

The present study is directed towards completely exploring reputation systems in simultaneous-move games that comprise more extensive social situations than those in the donation game. We discover that diverse social norms stabilize reciprocation and realize cooperative and stable populations. These successful social norms vary for different types of social dilemmas. To suppress cheating in Prisoner's Dilemma and Snowdrift games, these norms have a common characteristic such that defection against good players is regarded as bad irrespective of the co-player's action. However, in the Stag Hunt game, defection against good players may be allowed, whereas social norms that unfairly favour reciprocators are required to achieve robustness of reciprocation; under these norms, reciprocators never lose their good reputation. It is also imperative to punish unconditional cooperators that help anyone, because they blindly support cheaters [7, 8]. There are two mechanisms to restrain unconditional cooperation. One method is to avoid distinguishing between cooperation and defection towards bad players, in which case unconditional cooperators pay an extra cost of helping bad players while reciprocators do not. The other method is to regard cooperation with a bad player as a bad deed, in which case unconditional cooperators are explicitly punished. We discover that the former mechanism is feasible in Prisoner's Dilemma and Snowdrift games, whereas the latter works for all three social dilemmas.

II. MODEL

We consider a large, well-mixed population in which players from time to time play a symmetric two-player simultaneous-move game. In a one-shot game, two players are sampled from the population in a uniform random manner. Each player selects an action, which is either cooperation (C) or defection (D). There are four possible outcomes of the game for a player: both players select C (the outcome is called reward; R), the focal player selects C and his/her co-player selects D (sucker; S), the focal player selects D and his/her co-player selects C (temptation; T), and both players select D (punishment; P). The payoff matrix of the game is given by

$$\begin{array}{c} \\ \mathrm{C} \\ \mathrm{D} \end{array} \begin{array}{c} \begin{array}{cc} \mathrm{C} & \mathrm{D} \end{array} \\ \begin{bmatrix} 1 & S \\ T & 0 \end{bmatrix} \end{array}, \tag{1}$$

where the payoff of the focal player is 1, S, T, or 0 when the outcome is R, S, T, or P, respectively. Figure 1 illustrates the outcomes of competitions (*e.g.*, replicator dynamics) between cooperators and defectors for the three types of social dilemmas contained in the payoff matrix (1) [33–35]. In a two-dimensional payoff space, the region defined by $T > 1 > S > 0$ yields a Snowdrift game (SG) that has one stable internal equilibrium at which the fraction $S/(S + T - 1)$ of players are cooperators and the rest are defectors. The region $T > 1 > 0 > S$ yields a Prisoner's Dilemma game (PD) that has a unique stable equilibrium at which defectors dominate the population. It should be noted that the donation game, where the sum of the payoffs of outcomes S (one-sidedly paying cost of helping) and T (one-sidedly enjoying benefit of being helped) is always equal to the payoff of outcome R (both paying cost and enjoying benefit), is projected onto a half-line $S + T = 1$ ($T > 1$) in the payoff space (solid red line in Fig. 1); the PD game defined here is more general than the donation game. The region $1 > T > 0 > S$ yields a Stag Hunt game (SH) that has two pure stable equilibria at which cooperators and defectors each dominate the population. Because there is no dilemma when $1 > T > 0$ and $1 > S > 0$, we do not study this trivial region.

We employ a binary reputation model in which reputation states are either good (G) or bad (B) (*e.g.*, Ref. [6]; see Refs. [33, 36, 37]). In a one-shot game, each of the two players selects an action (*i.e.*, C or D), which is a response to each co-player's reputation (*i.e.*, G or B). A rule that specifies when to use which action is called an action rule, and it is denoted by a. There are four possible action rules. A reciprocator cooperates with a good co-player and defects against a bad co-player, *i.e.*, $a(\mathrm{G}) = \mathrm{C}$ and $a(\mathrm{B}) = \mathrm{D}$. An unconditional cooperator always cooperates ($a(\mathrm{G}) = a(\mathrm{B}) = \mathrm{C}$) while an unconditional defector always defects ($a(\mathrm{G}) = a(\mathrm{B}) = \mathrm{D}$). A 'contrary' player cooperates with a bad co-player and defects against a good co-player ($a(\mathrm{G}) = \mathrm{D}$ and $a(\mathrm{B}) = \mathrm{C}$). Hereafter, we denote reciprocators, unconditional cooperators, unconditional defectors, and contrary players by CD, CC, DD, and DC, respectively.

After a one-shot game, each participant of the game receives a new reputation that is determined by a social norm according to the outcome of the game (R, S, T, or P) and each co-player's reputation (G or B). Note that in our model, every member in a population has the same opinion about a player's reputation, which is attained through public information sharing [11, 36, 37]. Table I shows an example of a social norm under which a player receives a bad reputation only when he/she plays with a good co-player and the outcome is T or P, *i.e.*, whenever the player selects defection against a good co-player. This social norm was called simple standing in previous studies [12]. Because a social norm is specified by inserting G or B into the eight placeholders in a 4×2 table, there are $2^{4 \times 2} = 256$ possible norms.

We introduce errors in assessments with which a player is assigned an opposite reputation; if a player is assessed as good (bad), with a small probability μ, the player receives a bad (good) reputation [9, 10]. The models of indirect reciprocity generally consider errors not only in observers' assessments but also in players' taking actions [38]. Nevertheless, because the difference between the two kinds of errors usually does not change the reuslts qualitatively when assuming public information sharing (see, *e.g.*, Refs. [13, 20]), we only introduce errors in assessments.

III. METHODS

Our aim is to obtain desirable social norms that achieve cooperative and stable populations of reciprocators in different social dilemmas. To do so, we verify whether each candidate of the 256 social norms satisfies the following criteria in each of three social dilemmas, SG, PD, and SH.

Goodness: The population of reciprocators develops mutual cooperation except for defection caused by assessment errors.

Stability: The population of reciprocators is stable against any invasion by rare mutants (either CC, DD, or DC players).

Because the population of unconditional cooperators is stable in SH (see Fig. 1), one might wonder why we bother to need reciprocators and reputation systems to maintain cooperation.

One reason could be that reciprocators enhance robustness of cooperation. Therefore, for SH, we additionally check the following criterion.

Usefulness: The population of reciprocators is more robust against an invasion by unconditional defectors than that of unconditional cooperators.

We extend the standard methods for indirect reciprocity in the donation game regime (see, *e.g.*, Refs. [10, 19, 21]) to consider the simultaneous-move games, and introduce the above three criteria. Table II summarizes the definitions of symbols used in this section.

A. Goodness

Consider a population in which all players adopt a unique action rule denoted by a. After repeating the random matching games sufficiently many times, the population reaches an equilibrium in which the fraction of players that have good reputations, denoted by $p(\mathrm{G})$, satisfies

$$p(\mathrm{G}) = \sum_{r_{\mathrm{focal}} \in \{\mathrm{G},\mathrm{B}\}} \sum_{r_{\mathrm{co}} \in \{\mathrm{G},\mathrm{B}\}} p(r_{\mathrm{focal}}) p(r_{\mathrm{co}}) \phi(g(a(r_{\mathrm{co}}), a(r_{\mathrm{focal}})), r_{\mathrm{co}}). \tag{2}$$

The right-hand side of Eq. (2) averages the probability with which a player receives a good reputation, $\phi(g(a(r_{\mathrm{co}}), a(r_{\mathrm{focal}})), r_{\mathrm{co}})$, when the player and his/her co-player have reputations r_{focal} and r_{co}, respectively. In Eq. (2), $g(a(r_{\mathrm{co}}), a(r_{\mathrm{focal}}))$ represents the outcome of the game when the focal player and his/her co-player select actions $a(r_{\mathrm{co}})$ and $a(r_{\mathrm{focal}})$, respectively. Since the assessments involve errors that occur with probability μ, $\phi(\cdot,\cdot)$ is either $1-\mu$ or μ. $p(\mathrm{B})$ $(= 1 - p(\mathrm{G}))$ represents the fraction of bad players. By solving Eq. (2), we obtain

$$p(\mathrm{G}) = \begin{cases} \dfrac{B - \sqrt{B^2 - 4AC}}{2A} & (A \neq 0) \\ \dfrac{C}{B} & (A = 0), \end{cases} \tag{3}$$

where

$$\begin{aligned} A = {} & \phi(g(a(\mathrm{G}), a(\mathrm{G})), \mathrm{G}) + \phi(g(a(\mathrm{B}), a(\mathrm{B})), \mathrm{B}) \\ & - \phi(g(a(\mathrm{B}), a(\mathrm{G})), \mathrm{B}) - \phi(g(a(\mathrm{G}), a(\mathrm{B})), \mathrm{G}), \end{aligned} \tag{4a}$$

$$\begin{aligned} B = {} & 1 + 2\phi(g(a(\mathrm{B}), a(\mathrm{B})), \mathrm{B}) \\ & - \phi(g(a(\mathrm{B}), a(\mathrm{G})), \mathrm{B}) - \phi(g(a(\mathrm{G}), a(\mathrm{B})), \mathrm{G}), \end{aligned} \tag{4b}$$

and

$$C = \phi(g(a(\mathrm{B}), a(\mathrm{B})), \mathrm{B}). \tag{4c}$$

We are interested in whether a homogeneous population of reciprocators achieves cooperation. In a population of reciprocators (*i.e.*, CD players), the frequency of cooperation clearly equals the fraction of good players. Therefore, under a social norm, we expand the obtained $p(\mathrm{G})$ by μ, and when

$$p(\mathrm{G}) = 1 - O(\mu), \tag{5}$$

we regard the social norm as satisfying the criterion of goodness.

B. Stability

The expected payoff of a resident player in a homogeneous population of players adopting an action rule a is given by

$$f(a|a) = \sum\sum p(r_{\mathrm{focal}})p(r_{\mathrm{co}})\psi(g(a(r_{\mathrm{co}}), a(r_{\mathrm{focal}}))), \tag{6}$$

where $\psi(g(a(r_{\mathrm{co}}), a(r_{\mathrm{focal}})))$ determines a player's payoff for each outcome of the game, *i.e.*, $g(a(r_{\mathrm{co}}), a(r_{\mathrm{focal}}))$. Hereafter, we omit the ranges of the summations over r_{focal} and r_{co}.

We next consider that an infinitesimal fraction of mutant players adopting another action rule b ($\neq a$) invade the population. The fraction of mutants that have good reputations satisfies

$$q(\mathrm{G}) = \sum\sum q(r_{\mathrm{focal}})p(r_{\mathrm{co}})\phi(g(b(r_{\mathrm{co}}), a(r_{\mathrm{focal}})), r_{\mathrm{co}}). \tag{7}$$

Equation (7) yields

$$q(\mathrm{G}) = \frac{\delta - (\delta - \gamma)p(\mathrm{G})}{1 + \delta - \beta - (\alpha + \delta - \beta - \gamma)p(\mathrm{G})}, \tag{8}$$

where

$$\alpha = \phi(g(b(\mathrm{G}), a(\mathrm{G})), \mathrm{G}), \tag{9a}$$

$$\beta = \phi(g(b(\mathrm{B}), a(\mathrm{G})), \mathrm{B}), \tag{9b}$$

$$\gamma = \phi(g(b(\mathrm{G}), a(\mathrm{B})), \mathrm{G}), \tag{9c}$$

and

$$\delta = \phi(g(b(\mathrm{B}), a(\mathrm{B})), \mathrm{B}). \tag{9d}$$

The expected payoff of a mutant player is given by

$$f(b|a) = \sum\sum q(r_{\text{focal}})p(r_{\text{co}})\psi(g(b(r_{\text{co}}), a(r_{\text{focal}}))). \tag{10}$$

We define that the population of players adopting an action rule a is stable in a region of the payoff space (*i.e.*, SG, PD, or SH) if it satisfies

$$\Delta f \equiv f(b|a) - f(a|a) < 0 \quad \forall b \neq a \tag{11}$$

in all the area of the focused region. Δf is a function of μ and thus, it can be expanded as

$$\Delta f = d_0 + \mu d_1 + O(\mu^2), \tag{12}$$

where d_k represents the series coefficient of k-th order when expanded by μ. Because Δf is indeed at most of $O(\mu)$ when the population satisfies the goodness criterion, we only need to check at most d_1. We consider that $\Delta f < 0$ if

$$\begin{cases} d_0 < 0 & (d_0 \neq 0) \\ d_1 < 0 & (d_0 = 0 \text{ and } d_1 \neq 0) \end{cases} \tag{13}$$

holds true.

C. Usefulness

Because a homogeneous population of unconditional cooperators (*i.e.*, CC players) is stable in SH, even when a social norm satisfies the stability criterion, it is worse to adopt the CD action rule if the basin of attraction of CD players in competition with DD players is narrower than that of CC players. To examine this point, after detecting the social norms that satisfy the criteria of goodness and stability in SH, we numerically compare the basins of attraction of CC and CD players when they compete with DD players under those candidates. We select only the norms whereby CD players have larger basins of attraction than that of CC players in all the area of SH.

Here we consider a population that consists of players adopting either two action rules denoted by a and b. We denote by x the fraction of a-players; the fraction $1 - x$ are b-players. We also denote by $p(\text{G})$ and $q(\text{G})$ the fractions of good players within a- and b-players, respectively. Note that the fraction of good players in the entire population equals

$xp(\mathrm{G}) + (1-x)q(\mathrm{G})$. $p(\mathrm{G})$ and $q(\mathrm{G})$ are governed by the following time evolution:

$$\begin{aligned}\dot{p}(\mathrm{G}) = &-p(\mathrm{G}) + x\sum\sum p(r_{\mathrm{focal}})p(r_{\mathrm{co}})\phi(g(a(r_{\mathrm{co}}), a(r_{\mathrm{focal}})), r_{\mathrm{co}}) \\ &+ (1-x)\sum\sum p(r_{\mathrm{focal}})q(r_{\mathrm{co}})\phi(g(a(r_{\mathrm{co}}), b(r_{\mathrm{focal}})), r_{\mathrm{co}}),\end{aligned} \tag{14a}$$

and

$$\begin{aligned}\dot{q}(\mathrm{G}) = &-q(\mathrm{G}) + x\sum\sum q(r_{\mathrm{focal}})p(r_{\mathrm{co}})\phi(g(b(r_{\mathrm{co}}), a(r_{\mathrm{focal}})), r_{\mathrm{co}}) \\ &+ (1-x)\sum\sum q(r_{\mathrm{focal}})q(r_{\mathrm{co}})\phi(g(b(r_{\mathrm{co}}), b(r_{\mathrm{focal}})), r_{\mathrm{co}}).\end{aligned} \tag{14b}$$

We numerically solve $\dot{p}(\mathrm{G}) = \dot{q}(\mathrm{G}) = 0$ in Eq. (14) and obtain the equilibrium values of $p(\mathrm{G})$ and $q(\mathrm{G})$ that satisfy $\mathrm{Tr}\,\mathcal{J} < 0$ and $\det\mathcal{J} > 0$, where $\mathcal{J}$ is the Jacobian matrix of Eq. (14). Using them, the expected payoffs of a- and b-players, which depend on x, are given by

$$\begin{aligned}f(a|x) = &\, x\sum\sum p(r_{\mathrm{focal}})p(r_{\mathrm{co}})\psi(g(a(r_{\mathrm{co}}), a(r_{\mathrm{focal}}))) \\ &+ (1-x)\sum\sum p(r_{\mathrm{focal}})q(r_{\mathrm{co}})\psi(g(a(r_{\mathrm{co}}), b(r_{\mathrm{focal}}))),\end{aligned} \tag{15a}$$

and

$$\begin{aligned}f(b|x) = &\, x\sum\sum q(r_{\mathrm{focal}})p(r_{\mathrm{co}})\psi(g(b(r_{\mathrm{co}}), a(r_{\mathrm{focal}}))) \\ &+ (1-x)\sum\sum q(r_{\mathrm{focal}})q(r_{\mathrm{co}})\psi(g(b(r_{\mathrm{co}}), b(r_{\mathrm{focal}}))),\end{aligned} \tag{15b}$$

respectively.

When the competition between a- and b-players is bistable, the basin of attraction of a-players is given by $1 - x_a^*$, where x_a^* is the critical fraction of a-players at which $f(a|x_a^*) = f(b|x_a^*)$ holds true. Because the competitions between CC or CD players and DD players are indeed bistable in SH, to compare the basins, we only need to compare x_{CC}^* and x_{CD}^* under each social norm. In case of the competition between CC and DD players, we easily obtain $x_{\mathrm{CC}}^* = S/(S+T-1)$. In case of the competition between CD and DD players, we fix $a(\mathrm{G}) = \mathrm{C}$, $a(\mathrm{B}) = \mathrm{D}$, and $b(\mathrm{G}) = b(\mathrm{B}) = \mathrm{D}$ for Eqs. (14) and (15), set $\mu = 0.01$ or 0.1, and identify x_{CD}^* using the bisection method [39]. For each social norm, we check whether $x_{\mathrm{CD}}^* < x_{\mathrm{CC}}^*$ holds true for all the payoff configurations $(T, S) \in \{\epsilon, 2\epsilon, \ldots, 1-\epsilon\} \times \{-(1-\epsilon), -(1-2\epsilon), \ldots, -\epsilon\}$, where we set $\epsilon = 0.05$.

After the above numerical examination, in Appendix A, we analytically verified whether the obtained norms (shown in Fig. 2(e)) satisfy the usefulness criterion.

IV. RESULTS

We found that, among the 256 candidates, twelve social norms shown in Fig. 2 satisfy the goodness, stability, and/or usefulness criteria for at least one of SG, PD, and SH. Among these twelve norms, eight satisfy the two criteria defined for SG and PD (Fig. 2(a, b1, b2)) and six satisfy the three criteria defined for SH (Fig 2(b1, d1, d2)). The accurate conditions for the stability of reciprocators are listed in Appendix B.

A. Snowdrift and Prisoner's Dilemma games

The four social norms shown in Fig. 2(a) satisfy the two criteria for SG and PD. A sufficient condition for the stability of reciprocators under these norms is given by $T > 1$. In these four norms, the assessments of an action towards a good co-player do not depend on the co-player's action; cooperation with a good co-player is always regarded as good and defection against a good co-player is always regarded as bad. When a player encounters a bad co-player that selects cooperation (*i.e.*, outcome R or T), any action performed by the focal player is regarded as good. When a player encounters a bad co-player that selects defection (*i.e.*, outcome S or P), the assessment varies among the four norms.

The four social norms shown in Fig. 2(b1,2) also satisfy the two criteria for SG and PD. The condition for the stability of reciprocators under these four norms is given by $S < T$. These four norms are different from those shown in Fig. 2(a) with regard to only the assessment such that mutual cooperation (*i.e.*, outcome R) with a bad co-player is regarded as bad.

Figure 2(c) extracts the common features of the eight norms in Fig. 2(a, b1, b2) that are successful in SG and PD. These norms claim that cooperation (*i.e.*, outcomes R and S) and defection (*i.e.*, outcomes T and P) towards good players should be regarded as good and bad, respectively, while one-sided defection (*i.e.*, outcome T) against bad players should be regarded as good.

B. Stag Hunt game

The four social norms shown in Fig. 2(d1,2) as well as those shown in Fig. 2(b1) satisfy the three criteria for SH. It should be noted that in SH, the four norms shown in Fig. 2(b1,2) satisfy the goodness and stability criteria; however, only those in Fig. 2(b1) satisfy the usefulness criterion. A sufficient condition for the stability of reciprocators under the four norms in Fig. 2(d1,2) is given by $S < T < 3/2$, whereas the corresponding condition under the two norms in Fig. 2(b1) is given by $S < T$. The four norms in Fig. 2(d1,2) are different from those in Fig. 2(b1) with respect to the assessments such that either one-sided or mutual defection (*i.e.*, outcome T or P) against a good co-player is regarded as good.

Figure 2(e) extracts the common features of the six norms shown in Fig. 2(b1, d1, d2) that are successful in SH. These norms require that cooperation (*i.e.*, outcomes R and S) with a good player and defection (*i.e.*, outcomes T and P) against a bad player are regarded as good, *i.e.*, reciprocation should always be regarded as good. In addition, mutual cooperation (*i.e.*, outcome R) with a bad player is regarded as bad.

V. INTUITIONS

From the twelve social norms obtained, we discovered that reputation systems are based on different mechanisms to maintain indirect reciprocity. In this section, we provide explanations for how a homogeneous population of reciprocators (*i.e.*, CD players) prevents invasions by mutants that adopt the DD or DC action rules (Sec. V A) and the CC action rule (Sec. V B).

A. Universality and an exception for excluding unconditional defectors and contrary players

Let us consider an invasion event in an error-free limit (*i.e.*, $\mu \to 0$) under any successful social norm. Here, most players (residents) adopt the CD action rule and an infinitesimal fraction of players (mutants) adopt the DD or DC action rule. Because the social norm satisfies the goodness criterion, most residents have good reputations, whereas we assume that mutants have good and bad reputations with probabilities $q(\mathrm{G})$ and $q(\mathrm{B})$, respectively. In this population, a mutant is likely to play a game with a good CD resident. In the game, the DD or DC mutant selects defection because the resident is of a good reputation, whereas the CD resident selects cooperation or defection depending on the mutant's reputation. Therefore, the outcome for the mutant is T (one-sided defection) when his/her reputation is good, and P (mutual defection) when his/her reputation is bad. The expected payoff of the mutant is $q(\mathrm{G}) \cdot T + q(\mathrm{B}) \cdot 0 = q(\mathrm{G})T$, and that of the resident is clearly 1. The payoff difference is thus $\Delta f = q(\mathrm{G})T - 1$. The condition for stability against an invasion by the mutants is $\Delta f < 0$, which is rewritten as

$$q(\mathrm{G}) < \frac{1}{T}. \tag{16}$$

From Eq. (16), we see that there are two cases in which the mutants are suppressed. In one case, $q(\mathrm{G})$ is sufficiently small, *i.e.*, the reputation of mutants is effectively damaged. This policy is employed by the social norms in Fig. 2(c) that stabilize CD players in SG and PD. They have a universal principle as per which, when a player plays with a good co-player, cooperation (*i.e.*, outcome R or S) and defection (*i.e.*, outcome T or P) are regarded as good and bad, respectively. Because of this principle, once a DD or DC mutant appears in the population, he/she repeatedly encounters good players, selects defection, and receives bad reputations. Intuitively, because the temptation of defection is considerably strong in SG and PD (*i.e.*, $T > 1$), defection against a good player should be accused.

In the other case, T is smaller than 1 and the inequality (16) is satisfied by any value of $q(\mathrm{G})$. This is naturally met in SH and some of the social norms shown in Fig. 2(e) disregard the reputation of defection against a good co-player (*i.e.*, outcome T or P). Intuitively speaking, because defection against cooperation is simply irrational (*i.e.*, $T < 1$) in SH, there is no requirement to damage the reputations of unconditional defectors or contrary players as punishment. However, to satisfy the usefulness criterion in SH, the reputation of these mutants should be slightly damaged (see Appendix C); thus, the norms in Fig. 2(e) have at least one pivot that assigns a bad reputation to defection, either one-sided or mutual (*i.e.*, outcome T or P), against good players (see the '†'-ed pivots in Fig. 2(e)).

B. Diversity for excluding unconditional cooperators

In contrast, imagine that rare CC players invade a population of CD players. A CC mutant is likely to encounter a good CD resident and always select cooperation. The expected payoff of the mutant is $q(\mathrm{G}) \cdot 1 + q(\mathrm{B}) \cdot S = q(\mathrm{G}) + q(\mathrm{B})S$, and that of the resident is 1. The payoff difference is thus $\Delta f = [q(\mathrm{G}) + q(\mathrm{B})S] - 1 = -q(\mathrm{B})(1 - S)$. Because $S < 1$ holds true in all three social dilemmas, the condition for stability against an invasion by CC mutants,

$\Delta f < 0$, is

$$q(\mathrm{B}) > 0 \tag{17}$$

in the error-free limit (*i.e.*, $\mu \to 0$). Equation (17) implies that a small yet non-erroneous reduction of reputation suffices to suppress unconditional cooperators. However, by observing Fig. 2, it is evident that unconditional cooperators in most cases receive good reputations because selecting cooperation (*i.e.*, outcomes R and S) when one plays with a good co-player is always regarded as good under those norms. Since both CD and CC players generally have good reputations, the payoff difference between them is yielded by their different behaviors when they encounter rare bad players, who have erroneously received bad reputations.

In the four social norms shown in Fig. 2(a), when a CD or a CC player (both have good reputations) encounters a bad CD co-player, each selects defection and cooperation, while the CD co-player selects cooperation. As a result, both the focal CD and CC players receive good reputations. The payoff difference under these norms is, therefore, approximated by

$$\Delta f \propto 1 - T, \tag{18}$$

which yields the sufficient condition (shown before) for the stability of CD players against an invasion by CC mutants, *i.e.*, $T > 1$. During the above game sequence, the CD and CC players experience outcomes T and R, respectively, whereas their resultant reputations remain the same. Intuitively, if the temptation for defection is sufficiently large (*i.e.*, if $T > 1$) and an actor's behavior towards a bad player does not influence his/her reputation, defection is more rational than cooperation. This mechanism is feasible only in SG and PD.

In the eight social norms shown in Fig. 2(b1, b2, d1, d2), the adopted mechanism is different. When a focal (good) CD player encounters a bad CD co-player, he/she selects defection, whereas the co-player selects cooperation. As a result, the focal player maintains a good reputation. In the next round, the focal CD player encounters another good CD co-player, and mutual cooperation is achieved. The sum of payoffs in these two rounds is $T + 1$. In contrast, when a focal (good) CC player encounters a bad CD co-player, both select cooperation, and the focal CC player receives a bad reputation. In the next round, the focal, bad CC player encounters a good CD co-player. The focal player selects cooperation, whereas the co-player selects defection. As a result, the focal CC player retrieves a good reputation. The sum of payoffs in these two rounds is $1 + S$. Thus, the payoff difference between the CC and the CD players is approximated by

$$\Delta f \propto (1+S)-(T+1)=S-T, \tag{19}$$

which yields the condition (shown before) for the stability of CD players against an invasion by CC mutants, *i.e.*, $S < T$. Intuitively, these eight norms enforce defection against potentially harmful players, *i.e.*, bad players. Unconditional cooperators do not obey this enforcement and are punished for a moment. This mechanism is feasible in all three social dilemmas.

VI. DISCUSSION

A. Summary

We analyzed an extended model of indirect reciprocity in symmetric two-player simultaneous-move games that include three types of social dilemmas: Snowdrift (SG), Prisoner's Dilemma (PD), and Stag Hunt (SH) games. We showed that twelve social norms achieve cooperative and stable populations of reciprocators that exclusively cooperate with good co-players (Fig. 2). These norms possess different characteristics for providing the stability to reciprocators in different payoff structures and in excluding mutants. In SG and PD, eight norms stabilize the populations of reciprocators (Fig. 2(c)). In SH, six norms stabilize the populations of reciprocators and also enable them to secure larger basins of attraction than unconditional cooperators in competition with unconditional defectors (Fig. 2(e)). Among them, only two norms are almighty such that they satisfy all the criteria in any type of social dilemmas (Fig. 2(b1)). These two norms are the variants of the so-called Kandori social norm, which is characterized as possessing enforcement of defection against bad players and is known to exhibit strong stability in previous models [4, 14, 16].

The twelve social norms implement mechanisms in diverse manners for detecting and punishing players that do not follow reciprocation. We confirmed a principle in SG and PD for preventing an invasion by defectors; cooperation (*i.e.*, outcomes R and S) and defection (*i.e.*, outcomes T and P) towards a good player should be regarded as good and bad, respectively. This principle is identical to one of the fundamental properties in the so-called 'leading eight' social norms that have been known to stabilize indirect reciprocity in the donation game regime [10, 11]. In the norms in Fig. 2(d1,2), either one-sided or mutual defection (*i.e.*, outcome T or P) against a good player is regarded as good, and the defectors are not severely punished. This exception is only plausible in SH, because the temptation for defection is weak in SH (*i.e.*, $T < 1$).

An invasion by unconditional cooperators is a substantial risk because they indiscriminatingly help defectors and allow their indirect invasion. We summarize mechanisms for preventing invasion by unconditional cooperators as follows:

Rationality: Not discriminating between cooperation and defection towards bad players when one-sided defection is individually rational, *i.e.*, $T > 1$ (Fig. 2(a); feasible in SG and PD)

Enforcement: Unjustifying mutual cooperation with bad players (Fig. 2(b1, b2, d1, d2); feasible in SG, PD, and SH).

In previous works, the variants of the norms called standing and shunning employed the rationality mechanism, and those of the norm called Kandori employed the enforcement mechanism (see, *e.g.*, Refs. [10, 19]).

The social norms presented in Fig. 2(e) are successful in SH. They assign good reputations to players that select cooperation (defection) towards good (bad) co-players. In other words, these norms possess an unfair bias for favouring reciprocators whereby they always regard reciprocation as a good deed. This property maintains mutual cooperation among reciprocators even when there is a non-negligible fraction of other strategists, and thus, it succeeds in enlarging their basin of attraction. The other features of these norms are that their punishments of defection against good players (*i.e.*, outcomes T and P) can be milder than those in case of SG and PD, and that they have the enforcement mechanism introduced above.

B. Information use and emerging uncontrollability of reputation

For determining a focal player's reputation, social norms in our model use three sources of information, *i.e.*, the focal player's action, his/her co-player's action, and the co-player's reputation (see Tab. III). The stability and efficiency of indirect reciprocity is generally sensitive to which kinds of information are available. Early studies of indirect reciprocity in evolutionary games focused on the so-called first-order assessment, which only takes into account a focal player's past action (a_{focal} in Tab. III) for determining the player's reputation [5, 6, 40]. However, the first-order assessment is not sufficient to stabilize reciprocation except adopting special assumptions [7–9, 38, 41]. Reciprocation can be stable when the assessment uses at least two sources of information: a focal player's action and his/her co-player's reputations (a_{focal} and r_{co} in Tab. III). This is because they enable one to distinguish naïve defection (*i.e.*, defection against a good player) and defection to be justified (*i.e.*, defection against a bad player). There are a couple of reviews that explain the issue of justified defection [33, 36, 42]. It should be noted that the justified defection is not an only way to stabilize indirect reciprocity; *e.g.*, the 'shunning' social norm [12, 19, 43] and the 'tolerant scoring' [22, 44].

The availability of information introduces not only the justified defection, but also a 'gamble'. In our model, players face with a gamble in which the player's new reputation may depend on the co-player's action (a_{co} in Tab. III). This means that an actor in a game cannot fully control his/her new reputation by taking an appropriate action. Such a sort of uncontrollability has tacitly appeared in previous studies. For example, under the shunning social norm, when an actor meets a bad recipient, the actor always receives a bad reputation regardless of his/her behavior [36, 43]. In the shunning norm, an actor faces with a gamble on what kind of recipient he/she encounters. This is also true in the simple-standing norm, in which an actor always receives a good reputation when he/she by chance encounters a bad recipient [12]. In contrast to such uncontrollability in encounters, our model contains another uncontrollability in the co-player's actions. On the uncontrollability in the co-player's actions, our results have shown that successful social norms have the following characteristics:

1. In PD and SG (see Fig. 2(c)), the uncontrollability disappears when a player encounters a good co-players; the player's new reputation when the game outcome is R and S (T and P) is consistently good (bad) regardless of the co-player's action.

2. In SH (see Fig. 2(e)), the uncontrollability disappears when a player adopts reciprocation; selecting cooperation with a good co-player (outcomes R and S) or defection against a bad co-player (outcomes T and P) is consistently assessed as good.

3. Otherwise (*i.e.*, in cases when encountering a bad player in PD and SG, or when selecting cooperation (defection) toward a bad (good) co-player in SH), the gamble

can emerge.

Under a population of reciprocators using one of the successful social norms, players typically encounter good co-players and select cooperation. The uncontrollability in the co-player's actions disappears in such typical scenarios, while it remains in rare scenarios (*i.e.*, encountering bad players, selecting defection against a good player, *etc.*). This is also true for the uncontrollability in encounters in the previous studies.

In sum, our study revealed that indirect reciprocity is sometimes feasible even under social norms that include the apparently unreasonable uncontrollability in which a player can not necessarily anticipate his/her reputation by taking an appropriate action; the reputation may depend on the co-player's choice. However, in most cases under successful social norms that enable stable reciprocation, such uncertain situations are rare.

C. Comparison with the leading eight

Indirect reciprocity is also stabilized when using information about the focal player's action, the focal player's reputation, and the co-player's reputation (see the '3rd-order' column in Tab. III), under the so-called leading eight social norms in the donation game regime [10, 11]. Because the information used in the classical model in Refs. [10, 11] and that in the present model are different (see Tab. III), we cannot directly compare these two models. However, if we regard the co-player's actions C and D (a_{co} in Tab. III) as the focal player's reputations G and B (r_{focal} in Tab. III), respectively, the information use in our model corresponds with that in the classical model, and the social norms in Fig. 2(c) just agree with the leading eight. Therefore, we want to compare these two models using the above correspondence. The classical model, in contrast to ours, assumed more cognitively powerful players that use their own reputation information for their action rule. To clarify the difference between the two models, we extended our basic model to allow such intelligent players, and we found that only six of the leading eight norms survived the equilibrium selection (see Appendix D).

In Tab. IV, we show the two norms among the leading eight that failed to stabilize reciprocation in our extended model. In the classical model, the two norms succeed in stabilizing reciprocation when paired with the so-called OR strategy, with which a player defects against a bad co-player only when the player has a good reputation. Consider a game involving two bad players, both adopting OR strategy. In the game, a focal player and his/her co-player both select cooperation, because each of them has a bad reputation. However, since the co-player selects cooperation, the outcome of the game may be either R or T, and the outcome T when playing with a bad co-player results in the focal player's good reputation. Therefore, in the two norms, the focal player can enjoy a better payoff if he/she switches the action to defection when $T > 1$, which is satisfied in the donation game. On the other hand, under the corresponding two norms in the classical model, when both players have bad reputations, cooperation and defection are regarded as good and bad, respectively (see Tab. IV). Thus, the OR strategy players have no incentive to switch their actions in the same situation.

In sum, a slight difference in the manner to use information destabilizes OR strategy, which is stable in the classical indirect reciprocity model and appears only when assuming players to be more intelligent. Note that, we have only analyzed the stability of homogeneous populations of the variants of reciprocator including OR strategy; it could be possible that a mixture of OR strategists and some more defective strategists, *e.g.*, normal reciprocators, is evolutionarily stable.

D. Difference from two previous works that studied other than the donation game

Kandori was the first to study simultaneous-move games in community enforcement using reputation information [4]. His model and ours are fundamentally different in the following ways: 1) He investigated random matching games between two populations of players (*e.g.*, games between lenders and borrowers), whereas we studied random matching games between players in one population. 2) In his model, any equilibrium can be stabilized by long-term punishment (called T-period punishment) or by damaging the group-level reputation of the violator's population (called contagious equilibrium; see also Ref. [21]). They are strong devices for punishing defectors. In contrast, our model was not restricted to such strong punishments. We found that milder devices for punishment, the two mechanisms introduced above, are sufficient.

Uchida studied a Snowdrift-type donation game model [18]. He conducted a complete search on the entire combinations of third-order social norms and action rules and found that only two social norms, one, a variant of the Kandori norm and the other, called 'L4', develop cooperative and stable populations of reciprocators. The 'L4' social norm damages reputation of players that defect against good players or cooperate with bad players when their own reputations are bad. If we regard a co-player's cooperation and defection as the proxy of a focal player's good and bad reputations, respectively, then the 'L4' norm implies that the outcomes T and P when a focal player plays with a good co-player or the outcome S when a focal player plays with a bad co-player are regarded as bad. Thus, the 'L4' norm is included in the norms shown in Fig. 2(a), which indeed are successful in SG. However, in the extended model that introduces more intelligent players, they are not successful in SG (see Tab. VI(a)).

E. Limitations in the present study

In this study, we ignored the possibility that reputation is updated based on complete information about a focal player's and his/her co-player's actions and reputations, *i.e.*, fourth-order social norms (see Tab. III) [4, 11]. We also assumed that reputation information about a player is publicly shared among players, and ignored the possibility of nonpublic sharing in which players do not necessarily share reputation information, as studied in several previous works [9, 17, 20, 23]. These open questions should be explored in the future.

ACKNOWLEDGMENTS

This work was supported by JSPS KAKENHI Grant No. 13J05595 to MN and JSPS KAKENHI Grant No. 25118006 to HO.

Appendix A: Social norms in Fig. 2(e) satisfy the usefulness criterion in SH

We proof that CD players under the social norms shown in Fig. 2(e) have larger basins of attraction than CC players in competition with DD players in all the area of SH in the payoff space. Let $\Delta f(x) \equiv f(b|x) - f(a|x)$ denote the payoff difference between the players adopting action rules b and a when the frequencies of a- and b-players are given by x and $1 - x$, respectively. We assume that the action rules a and b are CD and DD, respectively.

By substituting $a(\mathrm{G}) = \mathrm{C}$ and $a(\mathrm{B}) = b(\mathrm{G}) = b(\mathrm{B}) = \mathrm{D}$ into Eq. (15), we obtain

$$\Delta f(x) = x\left[(q(\mathrm{G}) - p(\mathrm{G}))(S+T) + p(\mathrm{G})^2(S+T-1)\right] - q(\mathrm{G})S. \tag{A1}$$

Let $x^*_{\mathrm{CC}} = S/(S+T-1)$ denote the critical fraction of CC players in competition with DD players over which the CC players are advantageous than the DD players. If the basin of attraction of CD players in competition with DD players is larger than that of CC players, then $\Delta f(x^*_{\mathrm{CC}}) < 0$ holds true. The social norms in Fig. 2(e) imply in common that $\phi(\mathrm{R},\mathrm{G}) = \phi(\mathrm{S},\mathrm{G}) = \phi(\mathrm{T},\mathrm{B}) = \phi(\mathrm{P},\mathrm{B}) = 1-\mu$ and $\phi(\mathrm{R},\mathrm{B}) = \mu$. Substituting these ϕ values, $a(\mathrm{G}) = \mathrm{C}$, and $a(\mathrm{B}) = b(\mathrm{G}) = b(\mathrm{B}) = \mathrm{D}$ into Eq. (14), we solve $\dot{p}(\mathrm{G}) = \dot{q}(\mathrm{G}) = 0$, and obtain

$$\begin{cases} p(\mathrm{G}) & = 1-\mu, \\ q(\mathrm{G}) & = \dfrac{(1-\mu)\left[1 - x(1-\mu-\phi(\mathrm{P},\mathrm{G}))\right]}{2-\mu-x(1-\mu)\left[1-\phi(\mathrm{P},\mathrm{G})+\phi(\mathrm{T},\mathrm{G})\right]-(1-x)\phi(\mathrm{P},\mathrm{G})}. \end{cases} \tag{A2}$$

Substituting Eq. (A2) into Eq. (A1) at $x = x^*_{\mathrm{CC}}$, we see that in an error-free limit (*i.e.*, $\mu \to 0$),

$$\Delta f(x^*_{\mathrm{CC}}) = -x^*_{\mathrm{CC}} \frac{(1-T)\left[1-\phi_0(\mathrm{P},\mathrm{G})\right] - S\left[1-\phi_0(\mathrm{T},\mathrm{G})\right]}{(1-T)\left[2-\phi_0(\mathrm{P},\mathrm{G})\right] - S\left[1-\phi_0(\mathrm{T},\mathrm{G})+\phi_0(\mathrm{P},\mathrm{G})\right]} \tag{A3}$$

holds true, where $\phi_0(\cdot,\cdot) \in \{0,1\}$ represents $\phi(\cdot,\cdot) \in \{\mu, 1-\mu\}$ in the error-free limit. Since $1 > T > 0 > S$ holds true in SH, the denominator of the first term in the right-hand side of Eq. (A3) is clearly positive. In order to let $\Delta f(x^*_{\mathrm{CC}})$ be negative, the numerator, *i.e.*, $(1-T)\left[1-\phi_0(\mathrm{P},\mathrm{G})\right] - S\left[1-\phi_0(\mathrm{T},\mathrm{G})\right]$, needs to be positive. This is satisfied unless both $\phi_0(\mathrm{T},\mathrm{G})$ and $\phi_0(\mathrm{P},\mathrm{G})$ are equal to 1, *i.e.*, when at least one pivot is B in the assessments of defection (outcomes T and P) against good players, which indeed is met in the six norms in Fig. 2(e).

Appendix B: Accurate conditions for the stability of reciprocators

Table V lists the accurate conditions for the stability of CD players under the social norms shown in Fig. 2, which are derived from Eq. (13).

In Tab. V and hereafter, we denote a social norm in line as $r_{11}r_{21}r_{31}r_{41}r_{12}r_{22}r_{32}r_{42}$, where r_{ij} is either G, B, or '*' in row i and column j of the 4×2 table that represents a norm as seen in Fig. 2.

Appendix C: How to secure robustness of reciprocation in SH

To stabilize reciprocation in SH, in Sec. V A, we mentioned that there is no need to damage the reputation of defectors, since $T < 1$ holds true in SH. To satisfy the usefulness criterion, however, we need to do so.

Here we consider that in a population under the social norms in Fig. 2(e), the fraction x of players are reciprocators (*i.e.*, CD players) and the rest $1 - x$ are unconditional defectors (*i.e.*, DD players). A DD player, which has a good (bad) reputation with probability $q(\mathrm{G})$ ($q(\mathrm{B})$), encounters a CD player with probability x and achieves either of the outcomes T and P with probabilities $q(\mathrm{G})$ and $q(\mathrm{B})$, respectively. On the other hand, he/she encounters a DD player with probability $1 - x$ and here achieves the outcome P only. Thus, the expected payoff of the DD player is $x[q(\mathrm{G}) \cdot T + q(\mathrm{B}) \cdot 0] + (1 - x) \cdot 0 = xq(\mathrm{G})T$. In a similar manner, the expected payoff of a CD player is given by $x \cdot 1 + (1 - x)[q(\mathrm{G}) \cdot S + q(\mathrm{B}) \cdot 0] = x + (1 - x)q(\mathrm{G})S$, where it should be noted that a pair of CD players always achieve mutual cooperation (*i.e.*, outcome R) because their reputations are always good under those norms. Therefore, the payoff difference between the DD and CD players is

$$\Delta f = xq(\mathrm{G})T - [x + (1 - x)q(\mathrm{G})S] = x[q(\mathrm{G})(S + T) - 1] - q(\mathrm{G})S \tag{C1}$$

in an error-free limit (*i.e.*, $\mu \to 0$). By solving $\Delta f = 0$, we obtain the critical fraction of CD players over which they are advantageous than DD players, given by

$$x^*_{\mathrm{CD}} = \frac{q(\mathrm{G})S}{q(\mathrm{G})(S + T) - 1}. \tag{C2}$$

On the other hand, the corresponding critical fraction of CC players is given by

$$x^*_{\mathrm{CC}} = \frac{S}{S + T - 1}. \tag{C3}$$

If the basin of attraction of CD players is larger than that of CC players in competition with DD players, $x^*_{\mathrm{CD}} < x^*_{\mathrm{CC}}$ holds true. This yields the condition for satisfying the criterion of usefulness,

$$q(\mathrm{B}) > 0. \tag{C4}$$

The condition (C4) implies that at least we need to slightly reduce the reputation of DD players for securing better robustness of CD players than that of CC players. Indeed, although defection against a good player (*i.e.*, outcomes T and P) can be allowed in SH under the norms in Fig. 2(e) (see Sec. V), these norms do not completely allow such defections.

Appendix D: Equilibrium selection when assuming more intelligent players

1. The extended model

Here we assume that players are more intelligent; a player performs an action based on his/her own as well as his/her co-player's reputation. In this case, an action rule is extended as $a(r_{\mathrm{focal}}, r_{\mathrm{co}})$, where r_{focal} is the focal player's and r_{co} is the co-player's reputations. The number of possible action rules is $2^{2\times 2} = 16$. We denote the extended action rule in line by $s_{\mathrm{GG}}s_{\mathrm{GB}}s_{\mathrm{BG}}s_{\mathrm{BB}}$, where $s_{uv} = a(u, v) \in \{\mathrm{C}, \mathrm{D}\}$. For example, the action rule CDCD represents a normal reciprocator that selects cooperation and defection when his/her co-player's reputation is good and bad, respectively, irrespective of his/her own reputation.

We are interested in identifying successful pairs of reciprocating action rules and social norms that satisfy the criteria introduced in Sec. III. Among the 16 possible action rules, we consider that four action rules, CDCC, CDCD, CDDC, and CDDD, are the variants of reciprocator, because they perform reciprocation when they are of good reputation. Therefore, the number of pairs to be examined is $4 \times 256 = 1024$. We replace all the action-rule terms in Sec. III by the above extended ones, *e.g.*, $a(r_{\mathrm{co}}) \to a(r_{\mathrm{focal}}, r_{\mathrm{co}})$, and perform the same procedure except for the following three points.

Change in the goodness criterion: If players adopt action rules other than CDCD, the fraction of good players does not necessarily agree with the frequency of cooperation; it is given by

$$p(\mathrm{C}) = \sum\sum p(r_{\mathrm{focal}})p(r_{\mathrm{co}})\mathbb{1}_{\mathrm{C}}(a(r_{\mathrm{focal}}, r_{\mathrm{co}})), \tag{D1}$$

where $\mathbb{1}_{\mathrm{C}}(\cdot)$ is an indicator function by which $\mathbb{1}_{\mathrm{C}}(\mathrm{C}) = 1$ and $\mathbb{1}_{\mathrm{C}}(\mathrm{D}) = 0$. We redefine that a pair of an action rule and a social norm satisfies the criterion of goodness if

$$p(\mathrm{C}) = 1 - O(\mu) \tag{D2a}$$

and

$$\lim_{\mu\to 0} p(\mathrm{G}) > \frac{1}{2} \tag{D2b}$$

holds true. Note that the condition (D2b) is necessary in order to rule out possible pairs of the CDDC action rule and some social norms whereby a majority of players are of bad reputation but cooperative. In such a population, the CDDC players achieve mutual cooperation because they have bad reputations and thereby help bad players; here the symbols G and B actually stand for 'bad' and 'good', respectively [10].

Change in the stability criterion: In the extended model, if a pair of an action rule and a social norm satisfies the goodness criterion, the payoff difference between the mutants and residents, *i.e.*, Δf in Eq. (11), is indeed at most of $O(\mu^2)$. Thus, we expand Δf by μ as

$$\Delta f = d_0 + \mu d_1 + \mu^2 d_2, \tag{D3}$$

and if

$$\begin{cases} d_0 < 0 & (d_0 \neq 0) \\ d_1 < 0 & (d_0 = 0 \text{ and } d_1 \neq 0) \\ d_2 < 0 & (d_0 = d_1 = 0 \text{ and } d_2 \neq 0) \end{cases} \tag{D4}$$

holds true, we regard that the pair satisfies the criterion of stability.

Change in the usefulness criterion: In the extended model, a reputation dynamic in a polymorphic population (*cf.*, Eq. (14)) has possibly multiple stable equilibria, and which equilibrium to be reached depends on the initial states. Therefore, we assume that all the players have good reputations in the beginning, and numerically obtain an equilibrium reached from the initial state.

2. Results

We examined the 1024 pairs of the variants of reciprocator (either CDCC, CDCD, CDDC, or CDDD) and social norms. Unfortunately, no pair survives the equilibrium selection when we consider the entire payoff space, *i.e.*, $0 < T$ and $S < 1$ (see Fig. 1). However, mutual cooperation is Pareto efficient only when $S+T < 2$ holds true (see, *e.g.*, Ref. [34]). Narrowing the region of interest in the payoff space by adding the constraint $S + T < 2$, we identified the successful 27 pairs shown in Tab. VI.

The pairs shown in Tab. VI(a, b1, b2, c) are included in Fig. 2(a, b1, b2, d1). Paired with the CDCD action rule, *i.e.*, the normal reciprocator, the three social norms in Tab. VI(a) satisfy the goodness and stability criteria in PD; the three social norms in Tab. VI(b1,2) satisfy the goodness and stability criteria in PD and SH, whereas only those in Tab. VI(b1) satisfy the usefulness criterion for SH; the two social norms in Tab. VI(c) satisfy the goodness, stability, and usefulness criteria in SH. Figure 2 shows successful twelve social norms in the basic model, whereas Tab. VI(a, b1, b2, c) shows only eight. The lacking four pairs are CDCD-GGBBGGGB in Fig. 2(a), CDCD-GGBBBGGB in Fig. 2(b2), and CDCD-GGBGB*GG in Fig. 2(d2). In the extended model, the CDCD players are invaded by more intelligent mutants under these four norms. Moreover, the pairs in Tab. VI(a,b1,b2) are no longer stable in SG.

The 14 pairs shown in Tab. VI(d,e) satisfy the criteria of goodness, stability, and usefulness in SH. The five pairs shown in Tab. VI(f,g) satisfy the criteria of goodness and stability in SG. In these 19 pairs, the dominating action rules are CDDC or CDDD whereby a player selects defection against a good co-player when the focal player has a bad reputation, and the social norms have an assessment in common such that the outcome P is always regarded as good, irrespective of the co-player's reputation. This assessment is plausible for the two action rules. Consider that in a population of CDDC or CDDD players, a bad player is playing a game with a good co-player. Because they adopt the CDDC or CDDD action rule, both of them select defection, *i.e.*, the outcome is P, and they receive good reputations under those norms. Intuitively, a player that adopts either of the two action rules infers about the co-player's next action from his/her own reputation, and the player selects defection when he/she is of a bad reputation. Such inference is effective in SH, which requires coordination (*i.e.*, mutual cooperation or defection) between two players. In SG, players have an incentive to select an action that is different to the co-player's, *i.e.*, C with D or D against C, and this characteristic of anti-coordination tends to break the mutual cooperation. However, the social norms shown in Tab. VI(f,g) assign bad reputations to such outcomes, *i.e.*, outcomes S and T, when a focal player plays with a good co-player. This assessments change SG into a coordination problem, and therefore, the two action rules perform well.

---

[1] Trivers, R. L., 1971 The evolution of reciprocal altruism. *Q. Rev. Biol.* **46**, 35–37. (doi: 10.1086/406755).

[2] Sugden, R., 1986 *The economics of rights, co-operation and welfare.* Oxford: Blackwell.

[3] Alexander, R. D., 1987 *The biology of moral systems.* New York: Aldine de Gruyter.

[4] Kandori, M., 1992 Social norms and community enforcement. *Rev. Econ. Stud.* **59**, 63–80. (doi:10.2307/2297925).

[5] Nowak, M. A. & Sigmund, K., 1998 Evolution of indirect reciprocity by image scoring. *Nature* **393**, 573–577. (doi:10.1038/31225).

[6] Nowak, M. A. & Sigmund, K., 1998 The dynamics of indirect reciprocity. *J. Theor. Biol.* **194**, 561–574. (doi:10.1006/jtbi.1998.0775).

[7] Leimar, O. & Hammerstein, P., 2001 Evolution of cooperation through indirect reciprocity. *Proc. R. Soc. Lond. B* **268**, 745–753. (doi:10.1098/rspb.2000.1573).

[8] Panchanathan, K. & Boyd, R., 2003 A tale of two defectors: the importance of standing for evolution of indirect reciprocity. *J. Theor. Biol.* **224**, 115–126. (doi:10.1016/S0022-5193(03)00154-1).

[9] Brandt, H. & Sigmund, K., 2004 The logic of reprobation: assessment and action rules for indirect reciprocation. *J. Theor. Biol.* **231**, 475–486. (doi:10.1016/j.jtbi.2004.06.032).

[10] Ohtsuki, H. & Iwasa, Y., 2004 How should we define goodness?–reputation dynamics in indirect reciprocity. *J. Theor. Biol.* **231**, 107–120. (doi:10.1016/j.jtbi.2004.06.005).

[11] Ohtsuki, H. & Iwasa, Y., 2006 The leading eight: social norms that can maintain cooperation by indirect reciprocity. *J. Theor. Biol.* **239**, 435–444. (doi:10.1016/j.jtbi.2005.08.008).

[12] Ohtsuki, H. & Iwasa, Y., 2007 Global analyses of evolutionary dynamics and exhaustive search for social norms that maintain cooperation by reputation. *J. Theor. Biol.* **244**, 518–531. (doi: 10.1016/j.jtbi.2006.08.018).

[13] Ohtsuki, H., Iwasa, Y. & Nowak, M. A., 2009 Indirect reciprocity provides only a narrow margin of efficiency for costly punishment. *Nature* **457**, 79–82. (doi:10.1038/nature07601).

[14] Pacheco, J. M., Santos, F. C. & Chalub, F. A. C. C., 2006 Stern-judging: a simple, successful norm which promotes cooperation under indirect reciprocity. *PLoS Comput. Biol.* **2**, e178. (doi:10.1371/journal.pcbi.0020178).

[15] Takahashi, N. & Mashima, R., 2006 The importance of subjectivity in perceptual errors on the emergence of indirect reciprocity. *J. Theor. Biol.* **243**, 418–436. (doi: 10.1016/j.jtbi.2006.05.014).

[16] Uchida, S. & Sigmund, K., 2010 The competition of assessment rules for indirect reciprocity. *J. Theor. Biol.* **263**, 13–19. (doi:10.1016/j.jtbi.2009.11.013).

[17] Uchida, S., 2010 Effect of private information on indirect reciprocity. *Phys. Rev. E* **82**, 036111. (doi:10.1103/PhysRevE.82.036111).

[18] Uchida, S., 2011 Indirect reciprocity in snowdrift games. *Pacific Sci. Rev.* **12**, 190–195.

[19] Nakamura, M. & Masuda, N., 2011 Indirect reciprocity under incomplete observation. *PLoS Comput. Biol.* **7**, e1002113. (doi:10.1371/journal.pcbi.1002113).

[20] Nakamura, M. & Masuda, N., 2012 Groupwise information sharing promotes ingroup favoritism in indirect reciprocity. *BMC Evol. Biol.* **12**, 213. (doi:10.1186/1471-2148-12-213).

[21] Masuda, N., 2012 Ingroup favoritism and intergroup cooperation under indirect reciprocity based on group reputation. *J. Theor. Biol.* **311**, 8–18. (doi:10.1016/j.jtbi.2012.07.002).

[22] Tanabe, S., Suzuki, H. & Masuda, N., 2013 Indirect reciprocity with trinary reputations. *J. Theor. Biol.* **317**, 338–347. (doi:10.1016/j.jtbi.2012.10.031).

[23] Oishi, K., Shimada, T. & Ito, N., 2013 Group formation through indirect reciprocity. *Phys. Rev. E* **87**, 030801. (doi:10.1103/PhysRevE.87.030801).

[24] Wedekind, C. & Milinski, M., 2000 Cooperation through image scoring in humans. *Science* **288**, 850–852. (doi:10.1126/science.288.5467.850).

[25] Milinski, M., Semmann, D., Bakker, T. C. & Krambeck, H. J., 2001 Cooperation through indirect reciprocity: image scoring or standing strategy? *Proc. R. Soc. Lond. B* **268**, 2495–2501. (doi:10.1098/rspb.2001.1809).

[26] Bolton, G. E., Katok, E. & Ockenfels, A., 2005 Cooperation among strangers with limited information about reputation. *J. Public. Econ.* **89**, 1457–1468. (doi: 10.1016/j.jpubeco.2004.03.008).

[27] Seinen, I. & Schram, A., 2006 Social status and group norms: indirect reciprocity in a repeated helping experiment. *Euro. Econ. Rev.* **50**, 581–602. (doi:10.1016/j.euroecorev.2004.10.005).

[28] Sommerfeld, R. D., Krambeck, H.-J., Semmann, D. & Milinski, M., 2007 Gossip as an alternative for direct observation in games of indirect reciprocity. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 17435–17440. (doi:10.1073/pnas.0704598104).

[29] Engelmann, D. & Fischbacher, U., 2009 Indirect reciprocity and strategic reputation building in an experimental helping game. *Games Econ. Behav.* **67**, 399–407. (doi: 10.1016/j.geb.2008.12.006).

[30] Pfeiffer, T., Tran, L., Krumme, C. & Rand, D. G., 2012 The value of reputation. *J. R. Soc. Interface* **9**, 2791–2797. (doi:10.1098/rsif.2012.0332).

[31] Yoeli, E., Hoffman, M., Rand, D. G. & Nowak, M. A., 2013 Powering up with indirect reciprocity in a large-scale field experiment. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 10424–10429. (doi:10.1073/pnas.1301210110).

[32] Greif, A., 1989 Reputation and coalitions in medieval trade: evidence on the Maghribi traders. *J. Econ. Hist.* **49**, 857–882. (doi:10.1017/S0022050700009475).

[33] Sigmund, K., 2010 *The calculus of selfishness.* Princeton: Princeton University Press.

[34] Macy, M. W. & Flache, A., 2002 Learning dynamics in social dilemmas. *Proc. Natl. Acad. Sci. U.S.A.* **99**, 7229–7236. (doi:10.1073/pnas.092080099).

[35] Santos, F. C., Pacheco, J. M. & Lenaerts, T., 2006 Evolutionary dynamics of social dilemmas in structured heterogeneous populations. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 3490–3494. (doi: 10.1073/pnas.0508201103).

[36] Nowak, M. A. & Sigmund, K., 2005 Evolution of indirect reciprocity. *Nature* **437**, 1291–1298. (doi:10.1038/nature04131).

[37] Sigmund, K., 2012 Moral assessment in indirect reciprocity. *J. Theor. Biol.* **299**, 25–30. (doi:10.1016/j.jtbi.2011.03.024).

[38] Fishman, M. A., 2003 Indirect reciprocity among imperfect individuals. *J. Theor. Biol.* **225**, 285–292. (doi:10.1016/S0022-5193(03)00246-7).

[39] Weber, H.-J. & Arfken, G. B., 2003 *Essential Mathematical Methods for Physicists.* San Diego: Academic Press.

[40] Takahashi, N., 2000 The Emergence of Generalized Exchange. *Am. J. Sociol.* **105**, 1105–1134.

[41] Brandt, H. & Sigmund, K., 2005 Indirect reciprocity, image scoring, and moral hazard. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 2666–2670. (doi:10.1073/pnas.0407370102).

[42] Brandt, H., Ohtsuki, H., Iwasa, Y. & Sigmund, K., 2007 A survey of indirect reciprocity. In *Math. Ecol. Environ. Sci.* (eds. Y. Takeuchi, Y. Iwasa & K. Sato), number 1971 in Biological and Medical Physics, Biomedical Engineering, pp. 21–49. Berlin, Heidelberg: Springer Berlin Heidelberg. (doi:10.1007/978-3-540-34428-5_3).

[43] Takahashi, N. & Mashima, R., 2003 The emergence of indirect reciprocity: Is the standing strategy the answer?

[44] Berger, U., 2011 Learning to cooperate via indirect reciprocity. *Games Econ. Behav.* **72**, 30–37. (doi:10.1016/j.geb.2010.08.009).

[45] Suzuki, S. & Akiyama, E., 2005 Reputation and the evolution of cooperation in sizable groups. *Proc. R. Soc. Lond. B* **272**, 1373–1377. (doi:10.1098/rspb.2005.3072).

[46] Brandt, H. & Sigmund, K., 2006 The good, the bad and the discriminator–errors in direct and indirect reciprocity. *J. Theor. Biol.* **239**, 183–194. (doi:10.1016/j.jtbi.2005.08.045).

[47] Suzuki, S. & Akiyama, E., 2008 Evolutionary stability of first-order-information indirect reciprocity in sizable groups. *Theor. Popul. Biol.* **73**, 426–436. (doi:10.1016/j.tpb.2007.12.005).

[48] Masuda, N. & Nakamura, M., 2012 Coevolution of trustful buyers and cooperative sellers in the trust game. *PLoS One* **7**, e44169. (doi:10.1371/journal.pone.0044169).

[49] Panchanathan, K. & Boyd, R., 2004 Indirect reciprocity can stabilize cooperation without the second-order free rider problem. *Nature* **432**, 499–502. (doi:10.1038/nature02978).

[50] Suzuki, S. & Akiyama, E., 2007 Evolution of indirect reciprocity in groups of various sizes and comparison with direct reciprocity. *J. Theor. Biol.* **245**, 539–552. (doi: 10.1016/j.jtbi.2006.11.002).

[51] Uchida, S. & Sasaki, T., 2013 Effect of assessment error and private information on stern-judging in indirect reciprocity. *Chaos, Sol. Frac.* **56**, 175–180. (doi: 10.1016/j.chaos.2013.08.006).

[52] Chalub, F. A. C. C., Santos, F. C. & Pacheco, J. M., 2006 The evolution of norms. *J. Theor. Biol.* **241**, 233–240. (doi:10.1016/j.jtbi.2005.11.028).

[53] Panchanathan, K., 2011 Two wrongs don't make a right: the initial viability of different assessment rules in the evolution of indirect reciprocity. *J. Theor. Biol.* **277**, 48–54. (doi: 10.1016/j.jtbi.2011.02.009).

[54] Martinez-Vaquero, L. A. & Cuesta, J. A., 2013 Evolutionary stability and resistance to cheating in an indirect reciprocity model based on reputation. *Phys. Rev. E* **87**, 052810. (doi: 10.1103/PhysRevE.87.052810).

FIGURE CAPTIONS

FIG. 1. Three types of social dilemmas. In the payoff space spanned by T and S, the game defined by the payoff matrix (1) is the Snowdrift game (SG) when $T > 1 > S > 0$ (green region), the Prisoner's Dilemma game (PD) when $T > 1 > 0 > S$ (red region), and the Stag Hunt game (SH) when $1 > T > 0 > S$ (yellow region). The standard donation game is on the solid red line ($S + T = 1$ ($T > 1$)). Schematic diagrams inside these regions represent dynamics in competitions between cooperators (C) and defectors (D). Arrows represent the direction of evolution. Solid and hollow circles represent stable and unstable rest points, respectively.

FIG. 2. Surviving social norms. The symbol '*' indicates a placeholder to be replaced by two patterns: G or B. The symbol '†' indicates another placeholder to be replaced by three vertical patterns: G and B, B and G, or B and B. Thus, each table represents $2^n \times 3^m$ norms where n and m are the numbers of '*' and '†' in the table, respectively. The Venn diagram indicates stability in different social dilemmas: SG (green), PD (red), and SH (yellow). **(a)** Social norms that are stable in SG and PD. **(b1,2)** Norms that are stable in SG, PD, and SH. Only those in b1 satisfy the usefulness criterion for SH. **(c)** Common characteristics of a, b1, and b2 that are successful in SG and PD. **(d1,2)** Norms that are stable and meet the usefulness criterion for SH. **(e)** Common characteristics of b1, d1, and d2 that are successful in SH. Note that all the norms here satisfy the goodness criterion.

TABLES

TABLE I. An example of a social norm. The rows represent outcomes of a game (R, S, T or P), the columns represent a co-player's reputations (G or B), and G and B in each pivot represent the reputations that a focal player receives.

	G	B
R	G	G
S	G	G
T	B	G
P	B	G

TABLE II. Meaning of symbols.

symbol	meaning
$a(r) \in \{\mathrm{C}, \mathrm{D}\}$	Resident player's action in response to his/her co-player that have reputation r.
$b(r) \in \{\mathrm{C}, \mathrm{D}\}$	Mutant player's action in response to his/her co-player that have reputation r.
$p(r)$	Fraction of resident players that have reputation r.
$q(r)$	Fraction of mutant players that have reputation r.
$g(u, v) \in \{\mathrm{R}, \mathrm{S}, \mathrm{T}, \mathrm{P}\}$	Outcome of a game when a focal player and his/her co-player select actions u and v, respectively.
$\phi(g, r) \in \{1 - \mu, \mu\}$	Probability that a focal player receives a good reputation when the outcome of the game is g and his/her co-player has reputation r.
$\psi(g) \in \{1, S, T, 0\}$	Payoff to a focal player when the outcome of the game is g.

TABLE III. Information use in social norms. First-, second-, and third-order social norms have been studied previously. The columns of 'information use' indicate whether to use the information of a focal player's action (a_{focal}), a focal player's reputation (r_{focal}), a co-player's action (a_{co}), and/or a co-player's reputation (r_{co}). The column of 'justified defection' indicates the availability of justified defection. The columns of 'uncontrollability' indicate the possibility of the uncontrollability of reputation in encounters (r_{co}) and in the co-player's actions (a_{co}).

norm class	information use				justified	uncontrollability		previous studies
	a_{focal}	r_{focal}	a_{co}	r_{co}	defection	r_{co}	a_{co}	
1st-order	✓	-	-	-	-	-	-	Refs. [5, 6, 38, 41, 44–48]
2nd-order	✓	-	-	✓	✓	✓	-	Refs. [12, 13, 15–23, 49–51]
3rd-order	✓	✓	-	✓	✓	✓	-	Refs. [2, 4, 7–11, 14, 52–54]
our model	✓	-	✓	✓	✓	✓	✓	-
4th-order	✓	✓	✓	✓	?	?	?	-

TABLE IV. The two social norms among the leading eight, which failed to stabilize reciprocation in our extended model. The left and right tables show these corresponding norms in our model and in the classical model studied in Refs. [10, 11], respectively. The symbol '*' indicates a placeholder to be replaced by two patterns: G or B. In the right table, the columns (GG, GB, BG, or BB) indicate that a focal player and his/her co-player in a game have both good, good and bad, bad and good, or both bad reputations, respectively.

	G	B
R	G	*
S	G	G
T	B	G
P	B	B

$\Longleftrightarrow$

	GG	GB	BG	BB
C	G	*	G	G
D	B	G	B	B

TABLE V. Conditions for the stability of reciprocators under the social norms in Fig. 2.

panel	social norm	stability condition
(a)	GGBBGGGG	$(T = 1 \wedge S < 0) \vee T > 1$
	GGBBGGGB	$T > 1$
	GGBBGBGG	$(T = 1 \wedge S < 1/2) \vee T > 1$
	GGBBGBGB	$(T = 1 \wedge S < 0) \vee T > 1$
(b1,2)	GGBBB*G*	$S < T$
(d1,2)	GGGBB*GG	$(T < 3/2 \wedge S < T) \vee (T = 3/2 \wedge 1/6 < S < 3/2)$
	GGBGB*GG	$(T < 2 \wedge S < T) \vee (T = 2 \wedge 1/2 < S < 2)$

TABLE VI. Surviving social norms when assuming more intelligent players.

group	action rule - social norm	stability			usefulness
		SG	PD	SH	SH
(a)	CDCD - GGBBG*GG	-	✓	-	-
	CDCD - GGBBGBGB	-	✓	-	-
(b1)	CDCD - GGBBB*GG	-	✓	✓	✓
(b2)	CDCD - GGBBBBGB	-	✓	✓	-
(c)	CDCD - GGGBB*GG	-	-	✓	✓
(d)	CDDC - GGBGG**G	-	-	✓	✓
	CDDC - GGBGB*BG	-	-	✓	✓
(e)	CDDD - GGBG***G	-	-	✓	✓
(f)	CDDC - GBBGGBBG	✓	-	✓	-
(g)	CDDD - GBBG*B*G	✓	-	✓	-