The Evolution of Preferences in a Haystack Model with Finite Populations

Ryoji Sawa*

Faculty of Engineering, Information and Systems, University of Tsukuba

August 22, 2022

Abstract We consider a finite population analogue of the haystack model of Herold (2012). Players repeatedly and randomly break into groups, and play an extensive form game. In the game, Player 1 (players assigned the position of Player 1) either cooperates or defects. Player 2 observes Player 1's action, he either rewards or does not reward Player 1 if Player 1 cooperates, and he either punishes or does not punish if Player 1 defects. Their action choice is determined by their own preference. We consider the stochastic dynamics of preferences that evolve according to their average payoffs. We characterize a condition under which the state where players adopt the punisher preference, with which they punish Player 1 who defects, is the long-run equilibrium. We also show that expanding the choice of punishments may weaken the stability of such cooperative state.

Keywords: Evolutionary games; Stochastic stability; Evolution of preferences; Group selection; Haystack model.

JEL Classification Numbers: C72, C73.

*Address: Tennoudai 1-1-1, Tsukuba, Ibaraki, 305-8573 Japan, e-mail: *rsawa@sk.tsukuba.ac.jp*. The author is grateful to seminar participants at the 2022 Asia Meeting of the Econometric Society, East and South East Asia for helpful comments. We acknowledge financial support from the Japan Society for the Promotion of Science (JSPS KAKENHI Grant Numbers 18K12740 and 22H00826), Nomura Foundation (Grants in Social Science), KIER Foundation, and Tokyo Center for Economic Research (TCER).

1 Introduction

People often cooperate even when cooperation does not yield any material gain. It is known from theoretical biology that evolution favors cooperative behavior. For example, the theory of kin selection explain evolution of cooperation among relatives (Hamilton, 1964a,b). The theories of direct and indirect reciprocity explain evolution of cooperation in settings where people can engage in long-term interactions (Trivers, 1971) or build a reputation (Alexander, 1987; Nowak and Sigmund, 1998a,b). However, there are patterns of cooperation that cannot be explained by those theories. For example, people may cooperate with genetically unrelated people, with people they will not meet again, and in settings where reputations are not built. Fehr and Gächter (2002) report that the altruistic punishment can explain such pattern of cooperation, and people actually engaged in the altruistic punishment in their experiments. It, however, raised a new question: why people engage in such altruistic and costly punishment?

An answer to the question is offered by theory of group selection — individuals cluster into groups within which social interaction takes place. Herold (2012) shows that cooperative preference can evolve even with the weakest form of group selection, which assumes no assortative matching and no repeated interaction within groups. He considers a continuum of players who, in each time period, form small groups and play a two-player extensive-form game within their groups. In the game, Player 1 (players assigned the position of Player 1) either cooperates or defects. Player 2 observes Player 1's action. He either rewards or does not reward Player 1 if Player 1 cooperates, and he either punishes or does not punish if Player 1 defects. The strategy ('does not reward', 'does not punish') is the dominant strategy for Player 2 though he prefers Player 1 to cooperate. At the end of each period, groups are dissolved and new groups are formed in the next period. Herold (2012) employs the indirect evolutionary approach pioneered by Güth and Yaari (1992) and Güth (1995) — individual behavior is driven by preferences, while evolutionary success of preferences is driven by objective fitness payoffs. Players with the 'self-intersted' preference choose the dominant strategy, while those with some reciprocal preference choose to reward cooperative behavior (the 'rewarder' preference) or to punish non-cooperative behavior (the 'punisher' preference). Herold (2012) shows that the state where all players adopt the punisher preference is uniquely evolutionary stable under certain conditions.

This study extends the analysis of Herold (2012) by examining the long-run equilibria. Herold (2012) employs the deterministic process of preference evolution, where the predictions are generally dependent on initial states and are viewed as short-run (or mid-run) outcomes. We employ the stochastic evolutionary game theory, pioneered by Foster and Young (1990), Kandori et al. (1993), and Young (1993). In contrast to Herold (2012), we assume a finite population and persistent stochastic perturbation on the evolutionary process. The process with the perturbation takes into consideration the infrequent but possible transitions among equilibria, and it can refine the prediction of the deterministic process by assessing the robustness of equilibria. The most robust ones against the perturbation are called long-run equilibria or stochastically stable states. Some of our results are consistent to Herold (2012). Under similar conditions to Herold (2012), the all-punisher state, where all players adopt the punisher preference, is the unique long-run equilibrium (Remark 5.3). Besides, we characterize a more precise and weaker condition under which the same result holds (Proposition 3.8). Roughly speaking, if escaping from the all-punisher state is more unlikely than escaping from the all-self-interested state, then the all-punisher state is the unique long-run equilibrium. Other reciprocal preferences, e.g. 'rewarder', do not constitute a long-run equilibrium. However, they may play the role of a catalyst for the evolution of cooperation by making it more likely that the process moves into the all-punisher state. We further consider settings with more than one punishment option. Ironically, such a richer set of punishment options may not always be in favor of the all-punisher state (Proposition 4.2).

We model preference evolution using stochastic imitative dynamics (see Binmore and Samuelson 1997; Binmore et al. 1995; Sandholm 2012 for example). In each time period, one player is chosen for reproduction with a probability proportional to her fitness. With a probability close to one, her preference is inherited to the offspring. With a small probability, the offspring mutates, that is, she randomly adopts some preference. After reproduction, one randomly chosen player dies and is replaced by the offspring. We examine the stationary distribution in the double limit — the small mutation rate and the large population size — in order to characterize the long-run equilibria. The double limit approach, in general, faces difficulties when applying to settings with more than two alternatives (see Sawa (2021b) and Staudigl et al. (2021) for some recent developments). We overcome the difficulties by employing stochastic imitative dynamics and the technique of Imhof et al. (2005) and Fudenberg and Imhof (2006, 2008).

Related literature

We offer a brief literature review. It is found that in public goods experiments people often engage in punishment (see Fehr and Gächter 2000). If those who free ride are punished, cooperation will emerge. This finding raised a question known as the second-order free rider problem — who will bear the cost of punishing free riders. Fehr and Gächter

(2002) offer an answer to the question. Their experimental result shows that people engage in altruistic punishment — people have a tendency to punish free riders even if it yields no material benefits for them. That is, we seemingly have innate preferences that resolve the second-order public goods problem. Andreoni et al. (2003) reported a similar observation. They conducted experiments of settings that are similar to the game of Herold (2012). Half of the subjects were assigned to be proposers, and the other half to be responders. Subjects were randomly and anonymously paired to play the game with the condition that no two subjects were paired more than once in order to avoid any repeated interaction effects. Andreoni et al. (2003) reported that subjects often engaged in (altruistic) punishments and/or rewards whenever those options were available. See also Section 3 of Chaudhuri (2011) for a survey on experimental studies of public goods experiments with punishments.

The literature finds several key factors impacting emergence of preference that induces cooperative behavior, e.g. altruistic punishment. Some of the most notable factors are group selection (with random matching), assortative matching, direct reciprocity, and indirect reciprocity. Our paper falls into the literature on group selection. There are many studies that consider preference evolution in the literature on evolutionary game theory. We name only a few of the works and refer readers to Alger and Weibull (2019) and Section 3 of Newton (2018) for more comprehensive surveys.

An evolutionary process is said to have a group selection mechanism when individuals are randomly divided into groups, social interaction takes place within the groups, and preferences/strategies evolve according to payoffs from the social interaction.[1] The evolution of a strategy depends not only on its within-group advantage but also on its between-group advantage. Studies on group selection that are closely related to ours are, for example, Schaffer (1988), Huck and Oechssler (1999), Sethi and Somanathan (2001), and Gintis (2000). Schaffer (1988) proposes a static evolutionary stability concept for games with a finite population and a given group size. Instead of such a static concept, we consider the dynamic stability of equilibria using stochastic dynamics. Huck and Oechssler (1999) consider evolution of preferences for punishing unfair offers in the ultimatum game. They consider a finite population of players who are partitioned into subgroups and examine the long-run equilibrium, or the stochastically stable state. Players engage in long-term interaction with other players within their subgroups, but payoffs are compared across the entire population to determine the evolutionary success of preferences. The differences between ours and Huck and Oechssler (1999) are that we

[1] To distinguish from the assortative matching mechanism, we define group selection as a mechanism that forms groups completely randomly.

consider an extensive game where the responder can both punish and reward the proposer, and we consider the limit of the large population size, with which we show that the set of equilibria can be further refined. Sethi and Somanathan (2001) consider deterministic evolutionary dynamics of preference evolution where players randomly form groups and play an aggregative game, which is similar to a public goods game, within groups. They characterize conditions under which the state all players adopt the self-interested preference is unstable and conditions under which the state all players adopt the reciprocal preference is stable. Our difference is that we consider global stability of monomorphic states in the sense that the long-run equilibrium does not depend on the initial state, whereas Sethi and Somanathan (2001) consider local stability of monomorphic states. Gintis (2000) considers an evolutionary model where players randomly form groups, which may randomly dissolve, and play a repeated public goods game. They show a condition under which the reciprocal preference evolves. Their model is qualitatively different because players' preferences evolve according to the average payoff of members of their group in Gintis (2000), whereas players' preferences evolve according to the average payoff of each preference in ours.

The assortative matching can also explain evolution of cooperative behavior. A matching mechanism is assortative if people of similar preferences/actions are likely matched with one another. It is known that cooperation arises in evolutionary processes with such matching mechanism. See Bergstrom (1995, 2003) for example. For the indirect evolutionary approach with assortative matching mechanisms, see Section 4.2 of Alger and Weibull (2019). With the presence of (unobserved) players' preference types, the extent to which cooperative behavior emerges depends on the assortative matching mechanism. See, for example, Alger and Weibull (2013, 2016) for assortative matching according to types; Bilancini et al. (2018) for assortative matching according to actions; Newton (2017) for type-specific assortative matching; and Wu (2020) for assortative matching according to observable labels that are correlated with unobserved preferences. Some studies find that partner choice works similarly to assortative matching; see Fujiwara-Greve and Okuno-Fujiwara (2009), and Izquierdo et al. (2010), for example. See Bergstrom (2002) for a related literature on the evolution of social behavior with assortative/nonassortative group formation mechanisms.

The theories of direct reciprocity can explain evolution of cooperation in long-term interactions. See Mailath and Samuelson (2006) for a comprehensive review. See Guttman (2003) for the indirect evolutionary approach with finitely repeated games. The theories of indirect reciprocity can explain evolution of cooperation in larger groups where people can build a reputation. See Ellison (1994); Fudenberg and Levine (1989); Kandori (1992);

Takahashi (2010) for example. We consider settings that are not a focus of these theories; players are randomly matched (to form small groups), only engage in short-term interactions, and do not build a reputation.

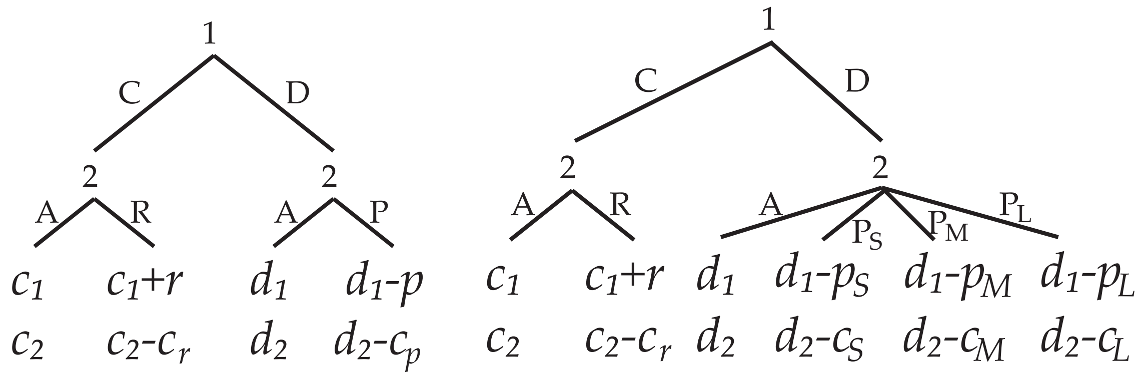
We also note that we only assume imperfect observation of others' preferences. The literature on evolutionary game theory finds that natural selection leads to altruistic or reciprocal preferences when players can perfectly observe each other's preferences. See Robson (1990), Güth and Yaari (1992), Possajennikov (2000), Heifetz et al. (2007a,b) for example. We relax the assumption of perfect observation and instead assume that players only observe the distribution of preferences in their groups. Under imperfect observation, cooperative behavior (or efficient equilibria) may not emerge. See Dekel et al. (2007), and Herold and Kuzmics (2009) for the effect of the the degree to which preferences are observed. The haystack model where players form subgroups can be a key factor for cooperation to emerge under imperfect observation. It is also known that altruistic or reciprocal preferences can emerge under imperfect observation when players can build a reputation. See Berger (2011); Berger and Grüne (2016); Ohtsuki and Iwasa (2004, 2006) for example. We do not assume that players can form a reputation in order to focus on the effect of the haystack model. Some studies consider models that the degree to which preferences are observed is endogenous. See Heller and Mohlin (2019) for this line of research.

The paper is organized as follows. Section 2 describes the model. The long-run outcomes of the haystack model are characterized in Section 3. We extend the model by expanding the punishment choice and analyze the long-run outcomes in Section 4. Section 5 offers a discussion on the design of the reward/punishment scheme in order to sustain cooperation. Section 6 concludes.

2 Model

2.1 Games and dynamics

There is a finite population of players, denoted by $\mathcal{N} = \{1, \ldots, N\}$. In each discrete time period, the players randomly form groups of constant size, called *haystacks*, and then they play an extensive form game $\mathcal{G}$ within their groups. Game $\mathcal{G}$ is a two-player two-stage extensive from game. Player 1 (the proposer role) chooses either to cooperate (C) or to defect (D). Player 2 (the responder role) observes Player 1's action, and chooses to reward (R), to punish (P), or to abstain from doing either (A). The available actions to Player 2 depend on settings, which will be introduced shortly. Let $\mathcal{O}$ denote the set of



(a) Basic setting

(b) SML setting

Figure 1: Settings

terminal nodes of game $\mathcal{G}$. We call $o \in \mathcal{O}$ an outcome. The payoff function $\mathcal{F} : \mathcal{O} \to \mathbb{R}^2$ maps each outcome to fitness that players obtain.

Figure 1a shows the game considered in Herold (2012, Setting 3). We call it *the basic setting*. The fitness payoffs are such that $c_1, c_2, d_1, d_2, c_r, c_p > 0$, $d_1 - p < c_1 < d_1 < c_1 + r$, and $d_2 < c_2 - c_r < c_2$. The Player 1's action that is optimal for Player 2 is C. However, the optimal action for Player 1 is D unless Player 2 rewards cooperation or punishes defection. Both actions are costly for Player 2, and it is optimal for Player 2 to abstain after observing Player 1's action — a social dilemma.

We extend the basic setting to a richer set of punishment options: P_S, P_M, and P_L. We call this setting *the SML setting*. It is illustrated in Figure 1b. We assume that the parameters are such that $c_1 > d_1 - p_S$, $p_S \le p_M \le p_L$, $c_S \le c_M \le c_L$. The first two conditions imply that all three punishments are severe enough to induce cooperative behavior. Punishment P_S is the softest punishment with the smallest cost of enforcement, P_M is a moderate punishment, and P_L is the most severe punishment with the largest cost. Note that in both settings, the dominant action of Player 2 is to play A in every decision node. We analyze the basic setting in Section 3, and the SML setting in Section 4.

In each discrete time period, players randomly form groups of constant size. Let $2h \in \{2, 4, 6, \ldots\}$ denote the size of a group. The number of groups formed in each period is $\frac{N}{2h}$. We assume that $\frac{N}{2h}$ is an integer. For each group, h of players are (randomly) assigned the position of Player 1, and the other h of players are assigned that of Player 2. Every player in a position plays game $\mathcal{G}$ with each of players in the opposite position. We assume that players can not observe other players' preferences but they know the distribution of preferences of players in the opposite position. At the end of each period, preferences are updated according to a birth-death process. One player is randomly drawn. She is

Table 1: Definition and notation for simplified preferences in the basic setting

Preference	Notation	Definition
Homo economicus	E	$\theta(C) = A, \theta(D) = A$
Rewarder	$\mathcal{R}$	$\theta(C) = R, \theta(D) = A$
Punisher	$\mathcal{P}$	$\theta(C) = A, \theta(D) = P$
Moral	$\mathcal{M}$	$\theta(C) = R, \theta(D) = P$

given an opportunity to update her preference. The update follows an imitative protocol described in the next section.

We assume that payoffs are generic such that Player 1 has a unique best response in any haystack. For example, in the basic setting, $kc_1 + (h-k)(c_1+r) \neq md_1 + (h-m)(d_1-p)$ for all $k, m \in \{0, \ldots, h\}$.

2.2 Preference dynamics

We adopt the indirect evolutionary approach (see Güth and Yaari 1992 and Güth and Kliemt 1998). Preferences determine players' actions in game $\mathcal{G}$, which in turn determine their fitness. Players' preferences evolve according to an evolutionary process.

Every player has two preferences: one for the position of each role. Following Herold (2012), we assume that preferences in the position of Player 1 do not vary among players; if a player is assigned the position of Player 1, they simply choose the action that maximizes their expected fitness. While, preferences in the position of Player 2 determine which action players will choose (regardless of the expected fitness) when they are assigned the position of Player 2. We focus on the evolution of preferences in the position of Player 2 since they are the player who has reciprocal actions. For the analysis, it is enough to consider the following simplified preferences for Player 2.

Definition 2.1. *A simplified preference of Player 2 is a mapping from each Player 1's action to an action of Player 2.*

Let Θ denote the set of simplified preferences.Definitions and notations for simplified preferences in the basic setting are shown in Table 1. For example, preference θ is the *homo economicus* preference, denoted by E, if $\theta(C) = A$ and $\theta(D) = A$. We sometimes call this preference the *self-interested* preference.

Preferences evolve according to a discrete-time, finite-state imitation dynamic with mutations. The states in the state space $\mathcal{X}$ represent the fraction of players adopting each preference:

$$\mathcal{X} = \left\{ (x_1, x_2, \ldots, x_{|\Theta|}) \mid x_i \in \{0, \tfrac{1}{N}, \ldots, 1\},\ x_1 + \ldots + x_{|\Theta|} = 1 \right\}.$$

The dynamic can be described by a Markov chain. The parameter $\varepsilon > 0$ denotes the mutation rate. Let $P_0(x, x')$ denote the transition probability from x to x' in the Markov chain without mutations. This Markov chain is called *the unperturbed dynamic*. We impose the following assumption on the unperturbed dynamic.

Assumption 2.2. *(i) $P_0(x, x') = 0$ for all $x, x' \in \mathcal{X}$ such that $x_i = 0$ and $x'_i > 0$.*
(ii) $P_0(x, x') > 0$ for all $x, x' \in \mathcal{X}$ such that $0 < x_i < 1$, $x'_i = x_i + \frac{1}{N}$, $x'_j = x_j - \frac{1}{N}$ for some $i, j \in \Theta$ and $x_h = x'_h$ for all $h \notin \{i, j\}$.

The first statement is the property that extinct preferences are never reintroduced. The second one is the property that any preference has a chance to be imitated if it is adopted by some player.

Let $P_\varepsilon(x, x')$ denote the transition probability from x to x' in the Markov chain with the mutation rate ε. We call the Markov chain *the ε-perturbed dynamic*. Let x^i denote the state all players adopt preference i, called the all-i state. Let $x^{i/j}$ denote the state where all players adopt preference i except one player, who adopts j. We impose the following assumption on the ε-perturbed dynamic.

Assumption 2.3. $\lim_{\varepsilon \to 0} \frac{P_\varepsilon(x^i, x^{i/j})}{\exp(-\varepsilon^{-1})} = \mu_{ij} > 0$ *for all $i, j \in \Theta$, $i \neq j$.*

The same assumption is imposed in Fudenberg and Imhof (2006). Without loss of generality, we assume that $\mu_{ij} \leq 1$ for all $i, j \in \Theta$ with $i \neq j$. This mutation model implies that the probability that a mutant invades a homogeneous state is $O(\exp(-\varepsilon^{-1}))$.

The next example illustrates a potential difference on fitness payoffs between the haystack model and the model without haystacks. In the model without haystacks where players interact with every player in the opposite position, self-interested players always earn (weakly) higher fitness payoffs than players with other preferences do. In the haystack model, players with a reciprocal preference may earn higher fitness payoffs.

Example 1 (haystack and fitness payoffs)**.** Consider the basic setting with $c_1 = 0$, $c_2 = 5$, $d_1 = 1$, $d_2 = 0$, $r = 2$, $c_r = 1$, $p = 2$, and $c_p = 1$. These parameter values are the same as in the example of Herold (2012). Suppose that there are 12 players, and that groups are formed as in the left illustration of Figure 2. In Figure 2, preference θ in the position of Player 1 denotes the fitness-maximizing preference. Preference $\mathcal{R}$ denote the rewarder preference, i.e., they play R and A against Player 1's C and D, respectively. Preference E

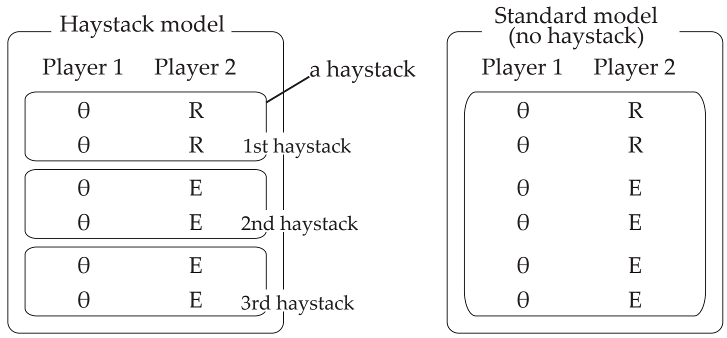


Figure 2: The haystack model and the standard model

denotes the self-interested preference, i.e., they always play A. In the left illustration of Figure 2, both players in the position of Player 2 adopt R in the first haystack, and those in the second and third haystacks adopt E. The players in the position of Player 1 in the first haystack will choose C since their cooperation will be always rewarded by Player 2. While, those players in the second and third haystacks will choose D since their action is neither rewarded nor punished. Players with the rewarder preference earn the fitness payoffs of 4 (on average), while the players with the self-interested preference earn the fitness payoffs of 0.

The right illustration of Figure 2 shows the model without haystacks. The fitness payoffs in the model without subgroups are different from ones in the haystack model. If a player in the position of Player 1 were to choose C, that action would be rewarded only two out of six interactions. Since $6d_1 > 6c_1 + 2r$, players in the position of Player 1 will choose D. Then, all players in the position of Player 2 earn the fitness payoffs of 0.

2.3 The most stochastically stable state

All homogeneous states are absorbing in the unperturbed dynamic. Our main focus is to characterize the states that are the most likely in the ε-perturbed dynamic as the mutation rate goes to zero. Let $\pi_\varepsilon(\cdot)$ denote the stationary distribution of the ε-perturbed dynamic; $\pi_\varepsilon(x)$ is the mass of x in the stationary distribution. Let $\pi_0(x) = \lim_{\varepsilon\to 0} \pi_\varepsilon(x)$ for all $x \in \mathcal{X}$. As discussed shortly, every homogeneous state has positive mass in $\pi_0(\cdot)$. However, some state becomes more likely than the other states when the population size is large. In the view of this observation, we define the most likely states as follows.

Definition 2.4. *A state x is the* most stochastically stable (MSS) *if* $\lim_{N\to\infty} \pi_0(x) > 0$.

In the remainder of this section, we consider the limiting stationary distribution π_0 for the ε-perturbed dynamic satisfying Assumption 2.3. The analysis follows Fudenberg and Imhof (2006). Recall that x^i denotes the state all players adopt preference i and $x^{i/j}$ denotes the state all players adopt preference i except one player, who adopts j. Let ρ_{ij} denote the probability that the unperturbed dynamic starting from $x^{i/j}$ will be absorbed in x^j. Define a $|\Theta| \times |\Theta|$ matrix Λ by

$$\Lambda_{ij} = \mu_{ij}\rho_{ij} \quad \forall j \neq i, \qquad\qquad \Lambda_{ii} = 1 - \sum_{j \neq i} \mu_{ij}\rho_{ij}.$$

We have the following lemma for the matrix Λ.

Lemma 2.5. *There is a unique vector $\lambda \in \mathbb{R}_+^{|\Theta|}$ such that $\lambda\Lambda = \lambda$, $\lambda_1 + \ldots + \lambda_{|\Theta|} = 1$, and $\lambda_1 > 0, \ldots, \lambda_{|\Theta|} > 0$.*

We omit the formal proof. Roughly speaking, Λ can be interpreted as a stochastic matrix, i.e. $\sum_{j \in \Theta} \Lambda_{ij} = 1$ for all $i \in \Theta$. This stochastic matrix is irreducible and aperiodic, and thus it has a unique eigenvector associated with eigenvalue 1.

The next theorem characterizes the limiting stationary distribution. The proof is omitted since it is implied by Fudenberg and Imhof (2006, Theorem 1).

Theorem 2.6. *$\pi_0(x^i) = \lambda_i$ for all $i \in \Theta$.*

Here is a brief intuition. The matrix Λ can be interpreted as a stochastic matrix. When mutations are rare, the behavior of the preference dynamic can be approximately described by a Markov chain on the homogeneous states, $\{x^i\}_{i \in \Theta}$, with the transition matrix Λ. For Λ, λ is the stationary distribution (Lemma 2.5). The limiting stationary distribution of the ε-perturbed dynamic coincides with λ as shown in Theorem 2.6. Note that the vector λ depends on the population size N. In the sections to come, we examine λ in the limit of the large population size to characterize the MSS states.

3 Analysis for the basic setting

3.1 Imitative dynamics

Theorem 2.6 allows us to restrict attention to transitions along the edges of the state space — transitions between the homogeneous states via one mutation. We define the restricted state space $\hat{\mathcal{X}}$ as follows:

$$\hat{\mathcal{X}} = \{x \in \mathcal{X} : x_i + x_j = 1 \text{ for some } i, j \in \Theta\}.$$

$\hat{\mathcal{X}}$ is the set of states where at most two sorts of preferences are adopted by players. Let $\hat{\mathcal{X}}^{\infty} = \{x \in [0,1]^{|\Theta|} : \sum_{\theta \in \Theta} x_\theta = 1, x_i + x_j = 1 \text{ for some } i, j \in \Theta\}$, that is, an extension of $\hat{\mathcal{X}}$ to continuous space.

Our analysis is focused on the preference evolution under imitative dynamics. Let $\sigma_{ij}^N : \hat{\mathcal{X}} \to [0,1]$ be a function that maps each population state in $\hat{\mathcal{X}}$ to the probability that a player adopting preference $i \in \Theta$ switches to preference j (without perturbation). That is, $P_0(x, x') = \sigma_{ij}^N(x)$ for all $x \in \hat{\mathcal{X}}$ with $x_i, x_j > 0$, $x'_i = x_i - \frac{1}{N}$, $x'_j = x_j + \frac{1}{N}$. It takes the form:

$$\sigma_{ij}^N(x) = \begin{cases} \frac{Nx_j}{N-1} r_{ij}(x) & \forall x \in \hat{\mathcal{X}} \text{ such that } x_i + x_j = 1, \\ 0 & \textit{otherwise}, \end{cases}$$

where $r_{ij} : \hat{\mathcal{X}}^{\infty} \to [r_{\min}, 1]$ is Lipschitz continuous and bounded away from zero ($r_{\min} > 0$). Using known results on birth-death processes (see Fudenberg and Imhof 2006, Example 2), the probabilities ρ_{ij} are computed as follows:

$$\rho_{ij} = \frac{1}{1 + \sum_{k=1}^{N-1} \prod_{m=1}^{k} \frac{x_j(m)\sigma_{ji}^N(x(m))}{x_i(m)\sigma_{ij}^N(x(m))}}, \tag{1}$$

where $x(m) \in \hat{\mathcal{X}}$ is such that $x_i(m) = 1 - \frac{m}{N}$ and $x_j(m) = \frac{m}{N}$. In the equation, note that $x_i(m)\sigma_{ij}^N(x(m))$ is the probability that an i-player is (randomly) selected to switch preference and she switches to preference j.

Recall that $\overline{\mathcal{F}}_i(x)$ denotes the expected fitness payoffs of players with preference i when the state is x. We impose the following assumption throughout the paper. In words, a preference with higher expected payoffs has a higher imitation rate.

Assumption 3.1. *For all $x \in \hat{\mathcal{X}}^{\infty}$ with $x_i, x_j > 0$, $r_{ij}(x) > r_{ji}(x)$ if $\overline{\mathcal{F}}_j(x) > \overline{\mathcal{F}}_i(x)$, and $r_{ij}(x) = r_{ji}(x)$ if $\overline{\mathcal{F}}_j(x) = \overline{\mathcal{F}}_i(x)$.*

The expected (or average) fitness payoffs of preference i for $x \in \hat{\mathcal{X}}$ with $x_i, x_j > 0$ for some $i, j \in \Theta$ are computed as

$$\overline{\mathcal{F}}_i(x) = h \sum_{k=\underline{k}}^{\overline{k}} \mathcal{F}_{i,j}(k, h-k) \ \Pr(i = k, j = h-k | x, i),$$

where $\underline{k} = \max\{1, h - Nx_j\}$, $\overline{k} = \min\{h, Nx_i\}$. $\mathcal{F}_{i,j}(k, h-k)$ is the fitness payoffs that a player with preference i in the position of Player 2 receives when a haystack has k of

i-players and $h-k$ of j-players in the position of Player 2. Let $\Pr(i=k, j=h-k|x,i)$ denote the probability that a haystack consists of k of i-players and $h-k$ of j-players given that the state is x and one player in the haystack is an i-player. This is given by

$$\Pr(i=k, j=h-k|x,i) = \frac{\frac{(Nx_i-1)!}{(k-1)!(Nx_i-k)!}\frac{Nx_j!}{(h-k)!(Nx_j-(h-k))!}}{\frac{(N-1)!}{(h-1)!(N-h)!}}$$

For a homogeneous state x with $x_i = 1$, the expected fitness payoffs are given by $\overline{\mathcal{F}}_i(x) = h\mathcal{F}_{i,j}(h,0)$ for any $j \neq i$. In the limit of the large population size N, the expected payoffs are computed as follows:

$$\overline{\mathcal{F}}_i^{\infty}(x) = h\sum_{k=0}^{h} \mathcal{F}_{i,j}(k, h-k)\,\Pr^{\infty}(i=k, j=h-k|x,i),$$
$$\text{where}\quad \Pr^{\infty}(i=k, j=h-k|x,i) = \frac{(h-1)!}{(k-1)!(h-k)!}(x_i)^{k-1}(x_j)^{h-k}.$$

Note that $\Pr^{\infty}(\cdot)$ is the probability mass function of a binomial distribution. The number of players in a haystack follows the binomial distribution with parameters $h-1$ (as the sample size) and x_i (as the success rate) in the large population limit.

Example 2 (Frequency-dependent Moran process)**.** Recall that $\overline{\mathcal{F}}_i(x)$ denotes the expected payoff of preference $i \in \Theta$ when the population state is x. Assume that fitness payoffs of all outcomes are positive. An imitative dynamic is called a frequency-dependent Moran process if the imitation rate function $r_{ij}(\cdot)$ is given by

$$r_{ij}(x) = \overline{\mathcal{F}}_j(x).$$

This process is used in many studies in the evolutionary game literature, e.g. Nowak et al. (2004), Fudenberg et al. (2006), and Hashimoto and Aihara (2009).

3.2 Characterization of the MSS states

Define

$$\phi_{ij}(y_j) = \ln\frac{r_{ji}(x)}{r_{ij}(x)} \qquad\qquad \forall y_j \in [0,1],$$

where $x \in \hat{\mathcal{X}}^{\infty}$ is such that $x_j = y_j$ and $x_i = 1 - y_j$. Since $r_{ij}(\cdot)$ for all $i, j \in \Theta$ is Lipschitz continuous and bounded away from zero, $\phi_{ij}(\cdot)$ is Lipschitz continuous and bounded for all $i, j \in \Theta$. For what follows, let $\hat{\mathcal{X}}_{ij}^{\infty} = \{y \in \hat{\mathcal{X}}^{\infty} : y_i + y_j = 1, y_i, y_j > 0\}$ for all $i, j \in \Theta$ with $i \neq j$.

In order to characterize the MSS states, it suffices to analyze the stationary distribution of the Markov chain with the state space Θ and the transition matrix Λ, where the transition probability from the all-i state to the all-j state is $\mu_{ij}\rho_{ij}$ for $i, j \in \Theta$ (Theorem 2.6). For $i, j \in \Theta$ with $i \neq j$, let

$$\beta_{ij} = \max_{t\in[0,1]} \int_0^t \phi_{ij}(z)dz. \tag{2}$$

Corollary 3.2 below shows that β_{ij} can be a measure of the unlikeliness of the transition from the all-i state to the all-j state. We call β_{ij} the *cost* of the transition. Corollary 3.2 is immediate from Lemma A.1, which is presented and proved in the Appendix.

Corollary 3.2.

$$\lim_{N\to\infty} \frac{1}{N} \ln \mu_{ij}\rho_{ij} = -\beta_{ij} \qquad\qquad \forall i, j \in \Theta, i \neq j.$$

The transition probability between two homogeneous states is roughly approximated as $M \exp(-N\beta_{ij})$ for some constant $M > 0$. With this result, we can compare the likeliness of different transitions, that is, $\lim_{N\to\infty}(\mu_{ij}\rho_{ij})/(\mu_{hl}\rho_{hl}) = 0$ if $\beta_{ij} > \beta_{hl}$ for all $i, j, h, l \in \Theta$.

The next lemmas characterize properties of the payoff structure of the basic setting. Those will be found useful in assessing the likeliness of the transition between two homogeneous states.

Lemma 3.3. *For $i, j \in \{E, \mathcal{P}\}$ with $i \neq j$, there exists x_{ij}^* such that $\overline{\mathcal{F}}_j^{\infty}(x) - \overline{\mathcal{F}}_i^{\infty}(x) > 0$ for all $x \in \hat{\mathcal{X}}_{ij}^{\infty}$ with $x_j > x_{ij}^*$, and $\overline{\mathcal{F}}_j^{\infty}(x) - \overline{\mathcal{F}}_i^{\infty}(x) < 0$ for all $x \in \hat{\mathcal{X}}_{ij}^{\infty}$ with $x_j < x_{ij}^*$. Furthermore, $x_{E\mathcal{P}}^* = 0$ and $x_{\mathcal{P}E}^* = 1$ if $p > h(d_1 - c_1)$. Otherwise, $x_{E\mathcal{P}}^*, x_{\mathcal{P}E}^* \in (0, 1)$.*

Lemma 3.4. *For $i, j \in \{E, \mathcal{R}\}$ with $i \neq j$, there exists $x_{ij}^* \in [0, 1]$ such that $\overline{\mathcal{F}}_j^{\infty}(x) - \overline{\mathcal{F}}_i^{\infty}(x) < 0$ for all $x \in \hat{\mathcal{X}}_{ij}^{\infty}$ with $x_j > x_{ij}^*$, and $\overline{\mathcal{F}}_j^{\infty}(x) - \overline{\mathcal{F}}_i^{\infty}(x) > 0$ for all $x \in \hat{\mathcal{X}}_{ij}^{\infty}$ with $x_j < x_{ij}^*$.*

Lemma 3.5. *For $i \in \{\mathcal{R}, \mathcal{M}\}$, $\overline{\mathcal{F}}_{\mathcal{P}}^{\infty}(x) - \overline{\mathcal{F}}_i^{\infty}(x) > 0$ for all $x \in \hat{\mathcal{X}}_{i\mathcal{P}}^{\infty}$.*

Lemma 3.6. *Let $i = E$ and $j = \mathcal{M}$. If $r + p < h(d_1 - c_1)$, there exists some $x_{ij}^* \in (0, 1]$ such that $\overline{\mathcal{F}}_j^{\infty}(x) - \overline{\mathcal{F}}_i^{\infty}(x) < 0$ for all $x \in \hat{\mathcal{X}}_{ij}^{\infty}$ with $x_j < x_{ij}^*$.*
If $r + p > h(d_1 - c_1)$, there exists $x_{ij}^* \in (0, 1)$ such that $\overline{\mathcal{F}}_j^{\infty}(x) - \overline{\mathcal{F}}_i^{\infty}(x) < 0$ for all $x \in \hat{\mathcal{X}}_{ij}^{\infty}$ with $x_j > x_{ij}^*$, and $\overline{\mathcal{F}}_j^{\infty}(x) - \overline{\mathcal{F}}_i^{\infty}(x) > 0$ for all $x \in \hat{\mathcal{X}}_{ij}^{\infty}$ with $x_j < x_{ij}^*$.

Lemmas 3.3–3.6 give characterizations of β_{ij}. Lemma 3.3 implies that for $i, j \in \{E, \mathcal{P}\}$ with $i \neq j$, $\beta_{ij}, \beta_{ji} > 0$ if $p < h(d_1 - c_1)$. That is, both transitions — the all-E state to/from the all-$\mathcal{P}$ state — are unlikely. To see this, note that if $p < h(d_1 - c_1)$, then Player 1 chooses D in a haystack where only one player in the Player 2 position adopts preference $\mathcal{P}$ and the others adopt preference E. When the proportion of preference $\mathcal{P}$ is small, it is likely for players with preference $\mathcal{P}$ to find themselves in such a haystack and punish Player 1 by sacrificing their own payoff. This implies that preference $\mathcal{P}$ can not invade to the all-E state, or $\beta_{E\mathcal{P}} > 0$. For $\beta_{\mathcal{P}E}$, it is always hard for preference E to invade to the all-$\mathcal{P}$ state. Thus, $\beta_{\mathcal{P}E} > 0$. If $p > h(d_1 - c_1)$, then Player 1 chooses C in any haystack with at least one player who punishes. Players with preference $\mathcal{P}$ can always induce cooperation, and thus their average payoff is higher than that of preference E, i.e., $\beta_{E\mathcal{P}} = 0$.

For $i, j \in \{E, \mathcal{R}\}$ with $i \neq j$, Lemma 3.4 implies that at least either $\beta_{E\mathcal{R}} = 0$ or $\beta_{\mathcal{R}E} = 0$ holds, i.e., either transition is likely. To see this, suppose that $hc_1 + r > hd_1$, that is, Player 1 chooses C if there is at least one rewarder. If a player mutates to $\mathcal{R}$ in the all-E state, then the average payoff of the player is $c_2 - c_r$, while that of preference-E players is approximately d_2. Since $c_2 - c_r > d_2$, players with preference $\mathcal{R}$ increase until they reach the state, say state($*$), where the payoffs of the two preferences are the same. In a similar logic, preference E can invade to the all-$\mathcal{R}$ state. If a player mutates to E in the all-$\mathcal{R}$ state, the player earns a higher payoff than players with preference $\mathcal{R}$ do, and players with preference E increases until they reach state($*$). If the all-$\mathcal{R}$ state is reached from state($*$) more likely than the all-E state, $\beta_{E\mathcal{R}} = 0$. Otherwise, $\beta_{\mathcal{R}E} = 0$.

For $i \in \{\mathcal{R}, \mathcal{M}\}$, $\beta_{i\mathcal{P}} = 0$. Lemma 3.5 implies that $\phi_{i\mathcal{P}}(t)$ is strictly negative for all $t \in [0, 1]$. Then, the integral attains its maximum at $t = 0$. This also implies that $\beta_{\mathcal{P}i} > 0$ for $i \in \{\mathcal{R}, \mathcal{M}\}$.

For $i, j \in \{E, \mathcal{M}\}$, Lemma 3.6 implies that $\beta_{E\mathcal{M}} > 0$ if $r + p < h(d_1 - c_1)$. Otherwise, either $\beta_{E\mathcal{M}} = 0$ or $\beta_{\mathcal{M}E} = 0$ holds. The intuition of the latter is similar to that of Lemma 3.4.

To characterize the MSS states, we view Θ as a set of nodes and make a couple of definitions. A directed edge is an ordered pair $(\theta, \hat{\theta}) \in \Theta \times \Theta$ with $\theta \neq \hat{\theta}$. A path from θ to $\hat{\theta}$ is a sequence of directed edges connecting θ to $\hat{\theta}$ (with no repeated nodes), for example, $\{(\theta, \theta_1), (\theta_1, \theta_2), \ldots, (\theta_t, \hat{\theta})\}$ with $\theta_i \neq \theta_j$ for $i \neq j$. A tree with root $\theta \in \Theta$ is a set of directed edges with three properties: (i) no outgoing edge from θ, (ii) exactly one outgoing edge for every $\hat{\theta} \neq \theta$, and (iii) a unique path from each $\hat{\theta} \neq \theta$ to θ. Let τ_θ be an arbitrary θ-tree. Define the cost of τ_θ as

$$C(\tau_\theta) = \sum_{(i,j) \in \tau_\theta} \beta_{ij}.$$

It is the sum of the "unlikeliness" of the edges in τ_θ. Let Υ_θ denote the set of θ-trees for each $\theta \in \Theta$. Using $C(\tau_\theta)$, define

$$C_\theta = \min_{\tau_\theta \in \Upsilon_\theta} C(\tau_\theta), \quad C^* = \min_{\theta \in \Theta} C_\theta,$$
$$\Theta^* = \{\theta \in \Theta : C_\theta = C^*\}.$$

We call C_θ the cost of state θ. Θ^* is the set of states that minimize the cost among all states. The next result refines the set of long-run equilibria characterized by Theorem 2.6. Every MSS state must be in Θ^*.

Theorem 3.7. $\lim_{N\to\infty} \sum_{\theta \in \Theta^*} \pi_0(\theta) = 1$.

The next proposition is induced by Theorem 3.7. It characterizes a sufficient condition for the all-$\mathcal{P}$ state to be the unique long-run equilibrium; it is the unique MSS state if escaping from the all-$\mathcal{P}$ state to any other state is harder than that from the all-E state.

Proposition 3.8. *If* $\min\{\beta_{\mathcal{P}E}, \beta_{\mathcal{P}\mathcal{R}}, \beta_{\mathcal{P}\mathcal{M}}\} > \min\{\beta_{E\mathcal{P}}, \beta_{E\mathcal{R}}, \beta_{E\mathcal{M}}\}$*, then the all-*$\mathcal{P}$ *state is the unique MSS state.*

We illustrate the intuition of Proposition 3.8 and omit the formal proof. Lemma 3.5 implies that the cost of escaping from the all-$\mathcal{R}$ or all-$\mathcal{M}$ state to the all-$\mathcal{P}$ state is zero: $\beta_{\mathcal{R}\mathcal{P}} = \beta_{\mathcal{M}\mathcal{P}} = 0$. This implies that $C_{\mathcal{P}} = \min\{\beta_{E\mathcal{P}}, \beta_{E\mathcal{R}}, \beta_{E\mathcal{M}}\}$. While, C_E must be at least weakly greater than the cost of escaping from the all-$\mathcal{P}$ state. This is because any E-tree must include an edge $(\mathcal{P}, \cdot)$, i.e. an edge emanating from $\mathcal{P}$. Then, $C_E \geq \min\{\beta_{\mathcal{P}E}, \beta_{\mathcal{P}\mathcal{R}}, \beta_{\mathcal{P}\mathcal{M}}\} > 0$. Thus, the inequality in Proposition 3.8 implies $C_{\mathcal{P}} < C_E$, or the unique stochastic stability of the all-$\mathcal{P}$ state.

We remark that the all-$\mathcal{R}$ and all-$\mathcal{M}$ states cannot be MSS. As discussed above, $\beta_{\mathcal{R}\mathcal{P}} = \beta_{\mathcal{M}\mathcal{P}} = 0$. Lemmas 3.3 and 3.5 imply that $\beta_{\mathcal{P}E}, \beta_{\mathcal{P}\mathcal{R}}, \beta_{\mathcal{P}\mathcal{M}} > 0$. Then, the cost of the all-$\mathcal{P}$ state is strictly lower than that of the all-$\mathcal{R}$ or all-$\mathcal{M}$ state. To see this, observe that for any $\mathcal{R}$-tree (or $\mathcal{M}$-tree), we can construct a strictly more cost efficient $\mathcal{P}$-tree by removing an edge emanating from $\mathcal{P}$ and adding an edge from $\mathcal{R}$ (or $\mathcal{M}$) to $\mathcal{P}$. Thus, the all-$\mathcal{P}$ state is the only candidate for MSS states that sustain cooperation. If the all-$\mathcal{P}$ state is not MSS, then the all-E state must be the unique MSS state.

Example 3 (MSS states and haystack sizes)**.** Consider the basic setting with $c_1 = 3$, $c_2 = 8$, $d_1 = 4$, $d_2 = 3$, $r = 2$, $c_r = 1$, $p = 2$, and $c_p = 1$. The parameter values are the same as in Herold (2012), except that we add 3 to c_1, c_2, d_1, d_2. The addition does not change

Table 2: The unlikeliness measure β_{ij} for transitions between homogeneous states

(a) $h = 20$

β_{ij}		j: E	$\mathcal{R}$	$\mathcal{P}$	$\mathcal{M}$
i	E	–	**0.032**	0.124	0.127
	$\mathcal{R}$	0	–	0	0
	$\mathcal{P}$	**0.019**	0.133	–	0.133
	$\mathcal{M}$	0	0	0	–

(b) $h = 16$

β_{ij}		j: E	$\mathcal{R}$	$\mathcal{P}$	$\mathcal{M}$
i	E	–	**0.022**	0.111	0.112
	$\mathcal{R}$	0	–	0	0
	$\mathcal{P}$	**0.027**	0.133	–	0.133
	$\mathcal{M}$	0	0	0	–

the equilibrium of the game but ensures positive fitness payoffs. Assume the frequency-dependent Moran process with $r_{ij}(x) = \overline{\mathcal{F}}_j(x)$ for all $i, j \in \Theta$. Then, the function $\phi_{ij}(\cdot)$ is written as follows:

$$\phi_{ij}(x) = \ln \frac{\overline{\mathcal{F}}_i^{\infty}(x)}{\overline{\mathcal{F}}_j^{\infty}(x)} \qquad\qquad \forall x \in \hat{\mathcal{X}}^{\infty}.$$

Table 2 shows the unlikeliness measure β_{ij} in Eq.(2) for transitions between homogeneous states. The haystack size is 20, or $h = 20$, in Table 2(a), and it is 16 in Table 2(b). The table shows the dependence of the MSS states on the haystack size. A player with a reciprocal preference, $\mathcal{R}$, $\mathcal{P}$, or $\mathcal{M}$, more likely finds themselves in a haystack where Player 1 chooses C than a player with preference E does. This is because they can be the pivotal player who induces cooperation. When a reciprocal player is the pivotal player, they earn higher fitness payoffs than if they were self-interested. However, this effect becomes insignificant as the haystack size becomes large. The non-cooperative convention, or the all-E state, tends to be the MSS state when the haystack size is large.

For $h = 20$, $\min\{\beta_{\mathcal{P}E}, \beta_{\mathcal{P}\mathcal{R}}, \beta_{\mathcal{P}\mathcal{M}}\} = 0.019 < 0.032 = \min\{\beta_{E\mathcal{P}}, \beta_{E\mathcal{R}}, \beta_{E\mathcal{M}}\}$. (The minimum values are in bold in Table 2.) The hypothesis of Proposition 3.8 is not satisfied, and the all-E state is the unique MSS state for this case. For $h = 16$, the hypothesis holds, or $\min\{\beta_{\mathcal{P}E}, \beta_{\mathcal{P}\mathcal{R}}, \beta_{\mathcal{P}\mathcal{M}}\} = 0.027 > 0.022 = \min\{\beta_{E\mathcal{P}}, \beta_{E\mathcal{R}}, \beta_{E\mathcal{M}}\}$. Proposition 3.8 implies that the all-P state is the unique MSS state. A numerical computation confirms that the all-P state is the unique MSS state for $h \leq 17$, while the all-E state is the unique MSS for $h \geq 18$.

Note that in Example 3, preference $\mathcal{R}$ plays the role of a catalyst in the spread of preference $\mathcal{P}$. As suggested by Lemma 3.4, the average payoff of preference $\mathcal{R}$ is strictly higher than that of preference E when the fraction of preference E is sufficiently large, that is, preference $\mathcal{R}$ can easily invade to the all-E state. This sometimes makes $\beta_{E\mathcal{R}}$ smaller as in Table 2, and stimulates the evolution of reciprocal preference $\mathcal{P}$ via the de-

Table 3: Definition and notation for simplified preferences in the SML setting

Preference	Notation	Definition
Homo economicus	E	$\theta(C) = A, \theta(D) = A$
Rewarder	$\mathcal{R}$	$\theta(C) = R, \theta(D) = A$
Moral	$\mathcal{M}$	$\theta(C) = R, \theta(D) = P_S$
Punisher S	$\mathcal{P}_S$	$\theta(C) = A, \theta(D) = P_S$
Punisher M	$\mathcal{P}_M$	$\theta(C) = A, \theta(D) = P_M$
Punisher L	$\mathcal{P}_L$	$\theta(C) = A, \theta(D) = P_L$

tour: all-E $\rightarrow$ all-$\mathcal{R}$ $\rightarrow$ all-$\mathcal{P}$. One can confirm that if preference $\mathcal{R}$ is not available, then $\min\{\beta_{\mathcal{P}E}, \beta_{\mathcal{P}\mathcal{M}}\} < \min\{\beta_{E\mathcal{P}}, \beta_{E\mathcal{M}}\}$ and the all-E state is the MSS state even for $h = 16$.

4 Analysis for the SML setting

4.1 Characterization under uniform-order mutations

We consider the SML setting, illustrated in Figure 1b. The difference from the basic setting is that the SML setting has a richer set of punishment alternatives. We examine the influence of expanding the choice of punishments on the long-run equilibrium. The result will show that expanding the punishment choice may paradoxically lead to non-cooperative long-run equilibrium. Recall that Θ denotes the set of simplified preferences. The set of simplified preferences for the SML setting is shown in Table 3.

We refer the set of the all-$\mathcal{P}_S$, all-$\mathcal{P}_M$, and all-$\mathcal{P}_L$ states as the all-$\mathcal{P}$ set. This is a set of states where Player 1 always chooses C to avoid punishment. We assume uniform-order mutations as in Section 3. The MSS states in this setting are also characterized by Theorem 3.7. The next lemma offers the key observation in this setting.

Lemma 4.1. *$\beta_{ij} = 0$ for all $i \in \{\mathcal{R}, \mathcal{M}, \mathcal{P}_S, \mathcal{P}_M, \mathcal{P}_L\}$ and $j \in \{\mathcal{P}_S, \mathcal{P}_M, \mathcal{P}_L\}$ with $i \neq j$.*

Proof of Lemma 4.1. Lemma 3.5 and its proof imply that $\beta_{ij} = 0$ for all $i \in \{\mathcal{R}, \mathcal{M}\}$ and $j \in \{\mathcal{P}_S, \mathcal{P}_M, \mathcal{P}_L\}$. To show that $\beta_{ij} = 0$ for $i, j \in \{\mathcal{P}_S, \mathcal{P}_M, \mathcal{P}_L\}$, it suffices to prove that $\overline{\mathcal{F}}_i^{\infty}(x) - \overline{\mathcal{F}}_j^{\infty}(x) = 0$ for all $x \in \{y \in \hat{\mathcal{X}}^{\infty} : y_i + y_j = 1,\ y_i, y_j > 0\}$ for all $i, j \in \{\mathcal{P}_S, \mathcal{P}_M, \mathcal{P}_L\}$ with $i \neq j$. Then, Assumption 3.1 implies that $\phi_{ij}(t) = 0$ for all $t \in [0, 1]$.

Fix $x \in \{y \in \hat{\mathcal{X}}^{\infty} : y_i + y_j = 1,\ y_i, y_j > 0\}$ for $i, j \in \{\mathcal{P}_S, \mathcal{P}_M, \mathcal{P}_L\}$. Observe that players in Player 1's position will be punished at least by p_S in every match if they choose D. Since $c_1 > d_1 - p_k$ for all $k \in \{S, M, L\}$, players in Player 1's position choose C. Thus $\overline{\mathcal{F}}_i^{\infty}(x) = \overline{\mathcal{F}}_j^{\infty}(x)$ for all $i, j \in \{\mathcal{P}_S, \mathcal{P}_M, \mathcal{P}_L\}$. $\square$

Lemma 4.1 shows that the cost of transitions among states in the all-$\mathcal{P}$ set is zero. During the transition between states in the all-$\mathcal{P}$ set, all players in Player 2's position receive the same payoff c_2 regardless of their preference since players in Player 1's position choose C to avoid punishment. Then, for example, from any state where players have either preference $\mathcal{P}_S$ or $\mathcal{P}_M$, the process can go either way: the all-$\mathcal{P}_S$ or all-$\mathcal{P}_M$ state. An important implication is that if some state in the all-$\mathcal{P}$ set is MSS, then all states in the all-$\mathcal{P}$ set are MSS as well.[2]

We say that a set of states is the unique MSS set if all states in the set are MSS and any state outside the set is not MSS. The next proposition characterizes a sufficient condition for the all-$\mathcal{P}$ set to be the unique MSS set. The cooperative states will be long-run equilibria if the cost of escaping from the all-$\mathcal{P}$ set to any other state is greater than that of escaping from the all-E state.

Proposition 4.2.
If $\min_{i \in \{S,M,L\}} \min\{\beta_{\mathcal{P}_i E}, \beta_{\mathcal{P}_i \mathcal{R}}, \beta_{\mathcal{P}_i \mathcal{M}}\} > \min\{\beta_{E\mathcal{P}_S}, \beta_{E\mathcal{P}_M}, \beta_{E\mathcal{P}_L}, \beta_{E\mathcal{R}}, \beta_{E\mathcal{M}}\}$*, then the all-*$\mathcal{P}$ *set is the unique MSS set.*

We sketch the proof of Proposition 4.2 and omit the formal one. Lemma 4.1 implies that the cost of any state in the all-$\mathcal{P}$ set equals the escaping cost from the all-E state, i.e., $C_{\mathcal{P}_i} = \min\{\beta_{E\mathcal{P}_S}, \beta_{E\mathcal{P}_M}, \beta_{E\mathcal{P}_L}, \beta_{E\mathcal{R}}, \beta_{E\mathcal{M}}\}$. This is because for any state in the all-$\mathcal{P}$ set, the transition cost is positive only from the all-E state to it, and the transition cost is zero from all the other states. Thus, the right-hand side in the inequality in Proposition 4.2 is the cost of states in the all-$\mathcal{P}$ set. It is easy to see that the left-hand side in the inequality is the escaping cost from states in the all-$\mathcal{P}$ set. The cost of the all-E state is weakly greater than that cost. Thus, if the inequality holds, the cost of states in the all-$\mathcal{P}$ set is minimum among the all-θ states for $\theta \in \Theta$.

Proposition 4.2 shows the mixed effect of expanding the punishment choice. On the one hand, more punishment alternatives weakly decrease the escaping cost from the all-E state, or the right-hand side of the inequality, and thus strengthen the stability of the cooperative equilibrium. For example, introducing $\mathcal{P}_M$ to $\Theta = \{E, \mathcal{R}, \mathcal{P}_S, \mathcal{M}\}$, i.e. adding the moderate punishment, reduces the escaping cost from the all-E state if $\beta_{E\mathcal{P}_M} < \min\{\beta_{E\mathcal{R}}, \beta_{E\mathcal{P}_S}, \beta_{E\mathcal{M}}\}$. This effect generally helps the all-$\mathcal{P}$ set be the unique MSS. On the other hand, more punishment alternatives may reduce the escaping cost from the all-$\mathcal{P}$ set. For example, consider that $\beta_{\mathcal{P}_M E} < \min\{\beta_{\mathcal{P}_S E}, \beta_{\mathcal{P}_S \mathcal{R}}, \beta_{\mathcal{P}_S \mathcal{M}}\}$. Then, adding $\mathcal{P}_M$ makes easier to escape from the all-$\mathcal{P}$ set. That is, the process can move away

[2] Lemma 4.1 shows that one mutation is enough to upset a state in the all-$\mathcal{P}$ set and move the process toward another. Then, use Lemma 4 of Nöldeke and Samuelson (1993) to obtain this implication.

Table 4: β_{ij} for haystack size $h = 16$ in settings A and B. The entries in the first four rows and columns are common among the two settings, while the entries for $\mathcal{P}_M$ and $\mathcal{P}_L$ are only for setting B.

β_{ij}		j					
		E	$\mathcal{R}$	$\mathcal{P}_S$	$\mathcal{M}$	$\mathcal{P}_M$	$\mathcal{P}_L$
i	E	–	0.014	0.072	0.047	0.058	0.035
	$\mathcal{R}$	0	–	0	0	0	0
	$\mathcal{P}_S$	0.015	0.061	–	0.061	0	0
	$\mathcal{M}$	0	0	0	–	0	0
	$\mathcal{P}_M$	0.012	0.061	0	0.061	–	0
	$\mathcal{P}_L$	0.01	0.061	0	0.061	0	–

from the all-$\mathcal{P}$ set by a detour through the all-$\mathcal{P}_M$ state into the all-E state. This may destabilize the cooperative equilibrium. We explore it more in the next section.

4.2 Cooperation stability of introducing more punishments

We further discuss the influence of expanding the punishment choice on the stability of cooperative equilibria. The main point is that even strictly more efficient punishment alternatives may not always help sustain the cooperative equilibrium and may sometimes destabilize it. This seemingly counter-intuitive phenomenon arises due to the second-order free rider problem. Free riders may take advantage of more efficient punishments. To make the point clearly, we consider an extreme case: $c_S = c_M = c_L$ and $p_S < p_M < p_L$. Intuitively, punishments M and L are strictly better than punishment S. They can enforce a more severe punishment than punishment S at exactly the same cost. However, even introducing such perfectly more efficient punishments may potentially destabilize the cooperative MSS state.

Example 4. Consider the following two settings, A and B. Setting A is the basic setting with $c_1 = 12$, $c_2 = 17$, $d_1 = 13$, $d_2 = 12$, $r = 2$, $c_r = 1$, $p_S = 1.01$, and $c_{p_S} = 1$. Setting B is the SML setting where punishments M and L with $p_M = 1.2$, $p_L = 1.7$, and $c_{p_M} = c_{p_L} = 1$ are added to setting A. Note that $c_S = c_M = c_L$ and $p_S < p_M < p_L$. Assume the frequency-dependent Moran process with $r_{ij}(x) = \overline{\mathcal{F}}_j(x)$ for all $i, j \in \Theta$.

Table 4 shows β_{ij} for each pair of preferences for the haystack size being 16 in settings A and B. Observe that $\min\{\beta_{\mathcal{P}_S E}, \beta_{\mathcal{P}_S \mathcal{R}}, \beta_{\mathcal{P}_S \mathcal{M}}\} = 0.015$ and $\min\{\beta_{E\mathcal{P}_S}, \beta_{E\mathcal{R}}, \beta_{E\mathcal{M}}\} = 0.014$. Proposition 3.8 implies that the all-$\mathcal{P}_S$ state is the unique MSS state in setting A. However, in setting B, the all-$\mathcal{P}$ set is not MSS. Proposition 4.2 does not apply since $\min\{\beta_{\mathcal{P}_M E}, \beta_{\mathcal{P}_L E}\} < \min\{\beta_{E\mathcal{P}_S}, \beta_{E\mathcal{P}_M}, \beta_{E\mathcal{P}_L}, \beta_{E\mathcal{R}}, \beta_{E\mathcal{M}}\}$. In fact, the all-$E$ state is the unique

MSS state in setting B.

Why such cost efficient punishments may weaken the stability of cooperation? The rationale behind this result is that severe punishments benefit a type of free riders who don't punish, as known as the second-order free riders. This allows preference E to invade the all-$\mathcal{P}$ set more easily in setting B than it does in setting A. To see this, consider an invasion by preference E to the all-$\mathcal{P}_i$ state for $i \in \{S, M, L\}$. When there is a small fraction of preference E, the most likely haystack is that (i) all players have preference $\mathcal{P}_i$, and the second most likely one is that (ii) all but one players have preference $\mathcal{P}_i$ and one player has E. For the invasion to the all-$\mathcal{P}_S$ state, Player 1 chooses D in case (ii). Since the punishment is very mild, action D is the best response when there is at least one player who does not punish. This implies that players with preference E cannot free-ride players with preference $\mathcal{P}_S$ and earn c_2. Then, the average payoff of players with preference E is $d_2 = 12$, while the average payoff of players with $\mathcal{P}_S$ is close to $c_2 = 17$ when the fraction of preference-E players is small. This makes harder the invasion by preference E.

On the other hand, for the invasion to the all-$\mathcal{P}_i$ state for $i \in \{M, L\}$, Player 1 chooses C in case (ii) to avoid punishment. If players with preference E is in such a haystack, they can free-ride players with preference $\mathcal{P}_M$ or $\mathcal{P}_L$ and earn c_2. The average payoff of players with preference E is close to $c_2 = 17$, i.e. they can evolve relatively easily, when the fraction of preference-E players is small. Thus, more cost-efficient punishments ironically make easier the invasion by preference E.

5 Discussion: design of a reward-punishment scheme

Propositions 3.8 and 4.2 suggest that the MSS states may depend on the effectiveness of rewards and punishments. A well-designed reward-punishment scheme may help us sustain cooperative conventions where all players (in the Player 1's position) choose to cooperate. We discuss such a scheme that improves the robustness of the all-$\mathcal{P}$ state. To simplify the analysis, we focus on the basic setting and consider the best choice of r, p, c_r, and c_p.

Assume that feasible schemes for rewards and punishments are defined by the cost functions $c_r : [\underline{r}, \overline{r}] \to \mathbb{R}_+$ and $c_p : [\underline{p}, \overline{p}] \to \mathbb{R}_+$. The function $c_r(t)$ maps each reward t to the cost associated with giving t to Player 1. That is, if Player 1 plays C and Player 2 rewards it with t, they receive $c_1 + t$ and $c_2 - c_r(t)$, respectively. The function $c_p(t)$ maps each punishment t to the cost associated with deducting t from Player 1's payoffs. We assume that $c_r(\cdot)$ and $c_p(\cdot)$ are weakly increasing. The lower bounds $\underline{r}$ and $\underline{p}$ are such that $c_1 + \underline{r} > d_1$ and $c_1 > d_1 - \underline{p}$. Those bounds guarantee that players in the position of

Player 1 will cooperate if all players in the Player 2's position are rewarders or punishers. The upper bounds $\overline{r}, \overline{p}$ are such that $c_2 - c_r(\overline{r}) > d_2$ and $\overline{p} < \infty$. Define

$$\mathcal{R}^* = \left\{ r \in [\underline{r}, \overline{r}] : \int_0^1 \phi_{E\mathcal{R}}(z)dz \leq 0 \right\}, \qquad \mathcal{P}^* = \left\{ p \in [\underline{p}, \overline{p}] : \int_0^1 \phi_{\mathcal{P}E}(z)dz > 0 \right\},$$

$$\beta^*_{E\mathcal{R}} = \min_{r\in[\underline{r},\overline{r}]} \int_0^1 \phi_{E\mathcal{R}}(z)dz, \qquad \beta^*_{\mathcal{P}E} = \max_{p\in[\underline{p},\overline{p}]} \max_{t\in[0,1]} \int_0^t \phi_{\mathcal{P}E}(z)dz,$$

$$\beta^*_{E\mathcal{M}} = \min_{r\in[\underline{r},\overline{r}] p\in[\underline{p},\overline{p}]} \min_{t\in[0,1]} \int_0^t \phi_{E\mathcal{M}}(z)dz.$$

Note that $\phi_{E\mathcal{R}}(\cdot)$ does not depend on the punishment level p, and $\phi_{\mathcal{P}E}(\cdot)$ does not depend on the reward level r. Thus, $\mathcal{R}^*$ is defined independently from the punishment p, and $\mathcal{P}^*$ is so from the reward r.

In part of the next result, we assume the following property. In words, the payoff-ratios of preferences are preserved in the ratios of imitation rates of preferences. The frequency-dependent Moran process in Example 2 satisfies this property.

Definition 5.1 (Payoff-ratio monotonicity)**.** *An imitative dynamic satisfies the* Payoff-Ratio Monotonicity *(PRM) if its imitative rate function satisfies that*

$$\frac{\overline{\mathcal{F}}_j(x)}{\overline{\mathcal{F}}_i(x)} > \frac{\overline{\mathcal{F}}_m(\hat{x})}{\overline{\mathcal{F}}_l(\hat{x})} \Leftrightarrow \frac{r_{ij}(x)}{r_{ji}(x)} > \frac{r_{lm}(\hat{x})}{r_{ml}(\hat{x})} \qquad \forall i,j,l,m \in \Theta, \;\; x, \hat{x} \in \hat{\mathcal{X}}^\infty.$$

The next proposition shows conditions under which we can or cannot sustain cooperation.

Proposition 5.2. *For all $(r, p) \in \{\mathcal{R}^* \times [\underline{p}, \overline{p}]\}$, the all-$\mathcal{P}$ state is the unique MSS state in any imitative dynamics.*

For all $p \in \mathcal{P}^*$, the all-$\mathcal{P}$ state is the unique MSS state in any imitative dynamics satisfying PRM.

If $\mathcal{R}^* = \mathcal{P}^* = \varnothing$, $\beta^*_{\mathcal{P}E} < \min\{\beta^*_{E\mathcal{R}}, \beta^*_{E\mathcal{M}}\}$, and $\int_0^1 \phi_{\mathcal{P}E}(z)dz < 0$ for all $p \in [\underline{p}, \overline{p}]$, then the all-E state is the unique MSS state for all $r \in [\underline{r}, \overline{r}]$, and all $p \in [\underline{p}, \overline{p}]$.

The first claim says that if $\mathcal{R}^* \neq \varnothing$, then we can sustain cooperation under a reward-punishment scheme with some $r \in \mathcal{R}^*$ and any $p \in [\underline{p}, \overline{p}]$. The punishment level p does not matter for this case. To see this, observe that $\beta_{E\mathcal{R}} = 0$ for all $r \in \mathcal{R}^*$. Players could likely move away from the all-E state toward the all-$\mathcal{R}$ state, from which players could likely move to the all-$\mathcal{P}$ state.

For $\mathcal{P}^* \neq \varnothing$, the PRM property guarantees that the all-$\mathcal{P}$ state is the unique MSS state. For any punishment level $p \in \mathcal{P}^*$, $\beta_{\mathcal{P}E} > \beta_{E\mathcal{P}}$ always holds. It is not certain if

$\min\{\beta_{\mathcal{PR}}, \beta_{\mathcal{PM}}\} > \min\{\beta_{E\mathcal{R}}, \beta_{E\mathcal{M}}\}$ holds, which is part of the hypothesis of Proposition 3.8. The PRM property ensures that the inequality holds.

The last claim of the proposition shows a condition under which players end up in the non-cooperative convention in the long run for any choice of (r, p).

Remark 5.3. There are conditions on lower- or upper- bounds that guarantee that $\mathcal{R}^* \neq \emptyset$ or $\mathcal{P}^* \neq \emptyset$. If $\underline{r} < \frac{h}{h-1}(d_1 - c_1)$, then $[\underline{r}, \frac{h}{h-1}(d_1 - c_1)) \subset \mathcal{R}^*$. This observation corresponds to Herold (2012, Remark 1). If $r < \frac{h}{h-1}(d_1 - c_1)$, then players in the Player 1's position will cooperate only if all players in the Player 2's position are rewarders. Since the existence of any of self-interested players always destroy cooperation, self-interested players cannot free-ride, and thus cannot outperform rewarders.

As for $\mathcal{P}^*$, if $\overline{p} > h(d_1 - c_1)$, then $(h(d_1 - c_1), \overline{p}] \subset \mathcal{P}^*$. It corresponds to Herold (2012, Remark 2). If $p > h(d_1 - c_1)$, a single punisher in a group can induce cooperation. Thus, punishers always outperform self-interested players, who fail to induce cooperation when the other players in their group are also self-interested.

Proof of Proposition 5.2. For the first claim, observe that Lemma 3.4 implies that $\beta_{ER} = \max\{0, \int_0^1 \phi_{ER}(z)dz\}$. Then, $\beta_{E\mathcal{R}} = 0$ for all $r \in \mathcal{R}^*$. Proposition 3.8 together with the fact that $\min\{\beta_{\mathcal{P}E}, \beta_{\mathcal{PR}}, \beta_{\mathcal{PM}}\} > 0$ implies the first claim.

For $\mathcal{P}^*$, Lemma 3.3 implies that $\beta_{\mathcal{P}E} = \int_0^{x^*_{\mathcal{P}E}} \phi_{\mathcal{P}E}(z)dz$ and $\beta_{E\mathcal{P}} = \int_0^{x^*_{E\mathcal{P}}} \phi_{E\mathcal{P}}(z)dz = \int_{x^*_{\mathcal{P}E}}^1 -\phi_{\mathcal{P}E}(z)dz$. This implies that $\beta_{\mathcal{P}E} - \beta_{E\mathcal{P}} = \int_0^1 \phi_{\mathcal{P}E}(z)dz$, which further implies that $\beta_{\mathcal{P}E} > \beta_{E\mathcal{P}}$ for all $p \in \mathcal{P}^*$. We can also show that for any imitative dynamic with PRM, $\beta_{\mathcal{PR}} = \beta_{\mathcal{PM}} > \beta_{E\mathcal{R}}$. Fix $y \in \hat{\mathcal{X}}^\infty$ such that $y_{\mathcal{P}} + y_{\mathcal{R}} = 1$. Let $\hat{y}$ be such that $\hat{y}_{\mathcal{P}} = y_{\mathcal{P}}$ and $\hat{y}_{\mathcal{M}} = y_{\mathcal{R}}$. Since Player 1 chooses C in both y and $\hat{y}$, it is easy to see that $\overline{\mathcal{F}}_{\mathcal{R}}(y) = \overline{\mathcal{F}}_{\mathcal{M}}(\hat{y})$ and $\overline{\mathcal{F}}_{\mathcal{P}}(y) = \overline{\mathcal{F}}_{\mathcal{P}}(\hat{y})$. The PRM property guarantees that $\frac{r_{\mathcal{RP}}(y)}{r_{\mathcal{PR}}(y)} = \frac{r_{\mathcal{MP}}(\hat{y})}{r_{\mathcal{PM}}(\hat{y})}$, which further implies that $\beta_{\mathcal{PR}} = \beta_{\mathcal{PM}}$. For $\beta_{\mathcal{PR}} > \beta_{E\mathcal{R}}$, fix $\xi \in \hat{\mathcal{X}}^\infty$ such that $\xi_{\mathcal{P}} + \xi_{\mathcal{R}} = 1$. Let $\hat{\xi}$ be such that $\hat{\xi}_{\mathcal{R}} = \xi_{\mathcal{R}}$ and $\hat{\xi}_E = \xi_{\mathcal{P}}$. Observe that

$$\frac{\overline{\mathcal{F}}_{\mathcal{P}}(\xi)}{\overline{\mathcal{F}}_{\mathcal{R}}(\xi)} = \frac{c_2}{c_2 - c_r} > \frac{p(\hat{\xi})c_2 + (1 - p(\hat{\xi}))d_2}{\hat{p}(\hat{\xi})(c_2 - c_r) + (1 - \hat{p}(\hat{\xi}))d_2} \geq \frac{\overline{\mathcal{F}}_E(\hat{\xi})}{\overline{\mathcal{F}}_{\mathcal{R}}(\hat{\xi})}, \tag{3}$$

where $p(\hat{\xi})$ denotes the probability that Player 1 chooses C in state $\hat{\xi}$ conditional on that one player in Player 2's position has preference E, and $\hat{p}(\hat{\xi})$ denotes that probability conditional on that one player in Player 2's position has preference $\mathcal{R}$. Note that $p(\hat{\xi}) \leq \hat{p}(\hat{\xi})$. Inequality (3) together with the PRM property implies that $\beta_{\mathcal{PR}} > \beta_{E\mathcal{R}}$. With $\beta_{\mathcal{P}E} > \beta_{E\mathcal{P}}$ and $\beta_{\mathcal{PR}} = \beta_{\mathcal{PM}} > \beta_{E\mathcal{R}}$, Proposition 3.8 implies the second claim.

If $\mathcal{R}^* = \mathcal{P}^* = \emptyset$, then $\beta_{E\mathcal{R}} > 0$ and $\beta_{\mathcal{P}E} \leq \beta_{E\mathcal{P}}$ for all $(r, p) \in [\underline{r}, \overline{r}] \times [\underline{p}, \overline{p}]$. If $\beta^*_{\mathcal{P}E} < \min\{\beta^*_{E\mathcal{R}}, \beta^*_{E\mathcal{M}}\}$, then $\beta_{\mathcal{P}E} < \min\{\beta_{E\mathcal{R}}, \beta_{E\mathcal{M}}\}$ for all $(r, p) \in [\underline{r}, \overline{r}] \times [\underline{p}, \overline{p}]$. Observe also that $\beta_{\mathcal{P}E} < \beta_{E\mathcal{P}}$ if $\int_0^1 \phi_{PE}(z)dz < 0$. Then, the claim is implied by Theorem 3.7. $\square$

6 Concluding remarks

We have considered an extension of the haystack model. The two main contributions are that (i) we characterize a sufficient condition under which adopting reciprocal preferences is the long-run equilibrium, and that (ii) we figure out the effect of expanding the punishment choice. For the latter contribution, it is intriguing that more cost-efficient punishments may weaken the robustness of the cooperative equilibrium.

A next step in future research is to replace the extensive form game $\mathcal{G}$ with the Nash demand game or the contract game. It is known that behavioral biases significantly affect the long-run behavior in the Nash demand game (see Khan 2021, 2022; Sawa 2021a), and in the contract game (see Hwang et al. 2018; Naidu et al. 2010). Behavioral biases are fixed in the models of those works. Matching evolution of preferences with the works mentioned above may lead to a new insight — a characterization of preferences and behavior that simultaneously emerge in the long run in the bargaining setting.

Another step we are considering taking is to employ the best response dynamics in the haystack model. A difficulty with it is that the technique of Fudenberg and Imhof (2006) is not available for the best response dynamics. One may inevitably employ some technique developed for the best response dynamics, e.g. Arigapudi (2020a,b); Sandholm (2010a); Sandholm and Staudigl (2016, 2018); Sawa (2012, 2021b).

Appendix A

A.1 Proofs for Section 3

Let $\Pr(k|x,i) = \Pr^\infty(i = k, j = h-k|x,i)$ throughout the Appendix. $\Pr(k|x,i)$ is the probability that an i-player is in a haystack with k of i-players (including herself) and $h-k$ of j-players in Player 2's position given that the state is x. Recall that $\hat{\mathcal{X}}^\infty_{ij} = \{y \in \hat{\mathcal{X}}^\infty : y_i + y_j = 1, y_i, y_j > 0\}$ for all $i, j \in \Theta$ with $i \neq j$.

Proof of Lemma 3.3. We prove the claim for the case that $i = E$ and $j = \mathcal{P}$. The proof for the other case, $i = \mathcal{P}$ and $j = E$, is similar. Let $x \in \hat{\mathcal{X}}^\infty_{E\mathcal{P}}$. Let k^* denote the minimum integer such that players in Player 1's position will cooperate if there are k^* of P-players and $h-k^*$ of E-players (in Player 2's position) in the haystack.

Observe that

$$\overline{\mathcal{F}}_{\mathcal{P}}^{\infty}(x) - \overline{\mathcal{F}}_{E}^{\infty}(x) = \Pr(k^*|x,\mathcal{P})(c_2 - d_2) + \sum_{k=1}^{k^*-1} \Pr(k|x,\mathcal{P})((d_2 - c_p) - d_2)$$
$$= c_p \Pr(k^*|x,\mathcal{P}) \left\{ \frac{c_2 - d_2}{c_p} - \sum_{k=1}^{k^*-1} \frac{\Pr(k|x,\mathcal{P})}{\Pr(k^*|x,\mathcal{P})} \right\}.$$

Note that $c_p \Pr(k^*|x,\mathcal{P})$ is strictly positive for all $x \in \hat{\mathcal{X}}_{E\mathcal{P}}^{\infty}$. The sum in the last term is strictly decreasing in $x_{\mathcal{P}}$. To see this, observe that

$$\frac{\Pr(k|x,\mathcal{P})}{\Pr(k^*|x,\mathcal{P})} = \frac{\frac{(h-1)!}{(k-1)!(h-k)!}}{\frac{(h-1)!}{(k^*-1)!(h-k^*)!}} \frac{(x_{\mathcal{P}})^{k-1}(x_E)^{h-k}}{(x_{\mathcal{P}})^{k^*-1}(x_E)^{h-k^*}}$$
$$= \frac{(k^*-1)!(h-k^*)!}{(k-1)!(h-k)!} \left(\frac{1-x_{\mathcal{P}}}{x_{\mathcal{P}}}\right)^{k^*-k}.$$

Furthermore, $\lim_{x_{\mathcal{P}}\to 0} \frac{\Pr(k|x,\mathcal{P})}{\Pr(k^*|x,\mathcal{P})} = \infty$ and $\lim_{x_{\mathcal{P}}\to 1} \frac{\Pr(k|x,\mathcal{P})}{\Pr(k^*|x,\mathcal{P})} = 0$ for all $k < k^*$. Since $0 < (c_2 - d_2)/c_p < \infty$, $\overline{\mathcal{F}}_{\mathcal{P}}^{\infty}(x) - \overline{\mathcal{F}}_{E}^{\infty}(x)$ is negative for sufficiently small $x_{\mathcal{P}}$ and positive for sufficiently large $x_{\mathcal{P}}$. Then the claim follows.

If $p > h(d_1 - c_1)$, then $k^* = 1$. This implies that $\sum_{k=1}^{k^*-1} = 0$. Thus, $\overline{\mathcal{F}}_{\mathcal{P}}^{\infty}(x) - \overline{\mathcal{F}}_{E}^{\infty}(x) > 0$ for all x with $x_{\mathcal{P}} > 0$. □

Proof of Lemma 3.4. We prove the claim for the case that $i = E$ and $j = \mathcal{R}$. The proof for the other case is similar. Let $x \in \hat{\mathcal{X}}_{E\mathcal{R}}^{\infty}$. Let k^{**} denote the minimum integer such that players in Player 1's position will cooperate when there are k^{**} of $\mathcal{R}$-players and $h - k^{**}$ of E-players (in Player 2's position) in the haystack. Observe that

$$\overline{\mathcal{F}}_{\mathcal{R}}^{\infty}(x) - \overline{\mathcal{F}}_{E}^{\infty}(x) = \Pr(k^{**}|x,\mathcal{R})((c_2 - c_r) - d_2) + \sum_{k=k^{**}+1}^{h} \Pr(k|x,\mathcal{R})((c_2 - c_r) - c_2)$$
$$= c_r \Pr(k^{**}|x,\mathcal{R}) \left\{ \frac{c_2 - d_2}{c_r} - 1 - \sum_{k=k^{**}+1}^{h} \frac{\Pr(k|x,\mathcal{R})}{\Pr(k^{**}|x,\mathcal{R})} \right\}.$$

Note that $c_r \Pr(k^{**}|x,\mathcal{R})$ is strictly positive for all $x \in \hat{\mathcal{X}}_{E\mathcal{R}}^{\infty}$. The sum in the last term is strictly increasing in $x_{\mathcal{R}}$ since each summand is written as follows:

$$\frac{\Pr(k|x,\mathcal{R})}{\Pr(k^{**}|x,\mathcal{R})} = \frac{(k^{**}-1)!(h-k^{**})!}{(k-1)!(h-k)!} \left(\frac{x_{\mathcal{R}}}{1-x_{\mathcal{R}}}\right)^{k-k^{**}}.$$

It is increasing in $x_{\mathcal{R}}$ for all $k > k^{**}$. Furthermore, $\lim_{x_{\mathcal{R}} \to 0} \frac{\Pr(k|x,\mathcal{R})}{\Pr(k^{**}|x,\mathcal{R})} = 0$ and $\lim_{x_{\mathcal{R}} \to 1} \frac{\Pr(k|x,\mathcal{R})}{\Pr(k^{**}|x,\mathcal{R})} = \infty$. Since $1 < (c_2 - d_2)/c_r < \infty$, the claim follows. □

Proof of Lemma 3.5. We prove the claim for $i = \mathcal{R}$. The other case can be similarly proved. Observe that for all $x \in \hat{\mathcal{X}}^{\infty}_{\mathcal{RP}}$, players in Player 1's position always cooperate regardless of the distribution of preferences in Player 2's position. Then, $\overline{\mathcal{F}}^{\infty}_{\mathcal{P}}(x) - \overline{\mathcal{F}}^{\infty}_{\mathcal{R}}(x) = c_2 - (c_2 - c_r) = c_r > 0$ for all such x. □

Proof of Lemma 3.6. Let $x \in \hat{\mathcal{X}}^{\infty}_{E\mathcal{M}}$. Let k_* denote the minimum integer such that players in Player 1's position will cooperate if there are k_* of $\mathcal{M}$-players and $h - k_*$ of E-players (in Player 2's position) in the haystack. Observe that $k_* > 1$ if $r + p < h(d_1 - c_1)$.

Observe that

$$\overline{\mathcal{F}}^{\infty}_{\mathcal{M}}(x) - \overline{\mathcal{F}}^{\infty}_{E}(x) = \Pr(k_*|x,\mathcal{M}) \left\{ (c_2 - c_r - d_2) - c_p \sum_{k=1}^{k_*-1} \frac{\Pr(k|x,\mathcal{M})}{\Pr(k_*|x,\mathcal{M})} \right.$$
$$\left. - c_r \sum_{k=k_*+1}^{h} \frac{\Pr(k|x,\mathcal{M})}{\Pr(k_*|x,\mathcal{M})} \right\}.$$

Similarly to the proofs of Lemmas 3.3 and 3.4, we can show that $\lim_{x_{\mathcal{M}} \to 0} \frac{\Pr(k|x,\mathcal{M})}{\Pr(k_*|x,\mathcal{M})} = \infty$ for $k < k_*$, and $\lim_{x_{\mathcal{M}} \to 0} \frac{\Pr(k|x,\mathcal{M})}{\Pr(k_*|x,\mathcal{M})} = 0$ for $k > k_*$. Then, if $k_* > 1$, $\overline{\mathcal{F}}^{\infty}_{\mathcal{M}}(x) - \overline{\mathcal{F}}^{\infty}_{E}(x) < 0$ for all sufficiently small $x_{\mathcal{M}}$.

If $r + p > h(d_1 - c_1)$, then $k_* = 1$. Note that $\sum_{k=1}^{k_*-1} \frac{\Pr(k|x,\mathcal{M})}{\Pr(k_*|x,\mathcal{M})} = 0$ for this case. We can prove the claim similarly to the proof of Lemma 3.4. □

The next lemma is the key lemma for Corollary 3.2 and Theorem 3.7. Corollary 3.2 is immediate from it.

Lemma A.1. *Let $t^*_{ij} \in [0,1]$ be such that $\beta_{ij} = \int_0^{t^*_{ij}} \phi_{ij}(z)dz$ for $i, j \in \Theta$ with $i \neq j$.*

(i) If $t^*_{ij} \in (0,1)$, then there are constants $0 < m < M < \infty$ such that for all $N \geq 2$, $m \leq \sqrt{N} \exp(N\beta_{ij})\rho_{ij} \leq M$.

(ii) If $t^*_{ij} \in \{0,1\}$, then there are constants $0 < m < M < \infty$ such that for all $N \geq 2$, $m \leq \exp(N\beta_{ij})\rho_{ij} \leq M$.

Proof of Lemma A.1. We prove claim (i) of the lemma. It is similar to the proof of Lemma 2 in Fudenberg et al. (2006). Recall Eq.(1). Observe that ρ_{ij} can be expressed as

$$\frac{1}{\rho_{ij}} = 1 + \sum_{k=1}^{N-1} \prod_{m=1}^{k} \frac{x_j(m)\sigma^N_{ji}(x(m))}{x_i(m)\sigma^N_{ij}(x(m))} = 1 + \sum_{k=1}^{N-1} \prod_{m=1}^{k} \frac{\frac{m}{N}\frac{N-m}{N-1} r_{ji}(x(m))}{\frac{N-m}{N}\frac{m}{N-1} r_{ij}(x(m))}$$

$$= 1 + \sum_{k=1}^{N-1} \exp\left(\sum_{m=1}^{k} \phi_{ij}\left(\frac{m}{N}\right)\right),$$

where recall that $\phi_{ij}(x) = \ln \frac{r_{ji}(x)}{r_{ij}(x)}$, and $x(m)$ is such that $x_j(m) = \frac{m}{N}$ and $x_i(m) = \frac{N-m}{N}$. Let $\psi(x) = \int_0^x \phi_{ij}(z)dz$, and $K_1 = \sup_{t\in[0,1]} |\phi'_{ij}(t)|$. The mean value theorem implies that, for all $m = 1, \ldots, N-1$,

$$\begin{aligned}
&\int_{(m-1)/N}^{m/N} \phi_{ij}(z)dz = \frac{1}{N}\phi_{ij}(\hat{t}) && \text{for some } \hat{t} \in \left[\frac{m-1}{N}, \frac{m}{N}\right], \\
\Rightarrow \quad &\int_{(m-1)/N}^{m/N} \phi_{ij}(z)dz \geq \frac{1}{N}\left(\phi_{ij}\left(\frac{m}{N}\right) - \frac{K_1}{N}\right), \\
\Rightarrow \quad &\phi_{ij}\left(\frac{m}{N}\right) \leq N \int_{(m-1)/N}^{m/N} \phi_{ij}(z)dz + \frac{K_1}{N}. && (4)
\end{aligned}$$

This yields that for all $k = 1, \ldots, N-1$,

$$\exp\left(\sum_{m=1}^{k} \phi_{ij}\left(\frac{m}{N}\right)\right) \leq \exp\left[N\psi\left(\frac{k}{N}\right) + K_1\right]. \tag{5}$$

Let $K_2 = \sup_{t\in[0,1]} |\phi_{ij}(t)|$. Observe that

$$\begin{aligned}
\exp\left[N\psi\left(\frac{k}{N}\right)\right] &= N \int_{(k-1)/N}^{k/N} \exp\left[N\psi\left(\frac{k}{N}\right)\right] d\xi \\
&\leq N \int_{(k-1)/N}^{k/N} \exp\left[N\left(\psi(\xi) + \frac{1}{N}K_2\right)\right] d\xi \\
&= Ne^{K_2} \int_{(k-1)/N}^{k/N} \exp\left[N\psi(\xi)\right] d\xi. && (6)
\end{aligned}$$

The inequality comes from $\psi'(t) = \phi_{ij}(t)$ for $t \in [0,1]$ and that $\exp(\cdot)$ is an increasing function. Combining Eq.(4)–(6), we have that

$$\frac{1}{\rho_{ij}} \leq 1 + Ne^{K_1+K_2} \int_0^1 \exp\left[N\psi(\xi)\right] d\xi. \tag{7}$$

Similarly to that, we can show that

$$\frac{1}{\rho_{ij}} \geq 1 + Ne^{-K_1-K_2} \int_0^1 \exp\left[N\psi(\xi)\right] d\xi. \tag{8}$$

Combining the two inequalities above, we obtain

$$\frac{1}{1+Ne^{K_1+K_2}\int_0^1 \exp\left[N\psi(\xi)\right]d\xi} \le \rho_{ij} \le \frac{1}{Ne^{-K_1-K_2}\int_0^1 \exp\left[N\psi(\xi)\right]d\xi}. \tag{9}$$

Recall that $\beta_{ij} = \max_{t\in[0,1]} \int_0^t \phi_{ij}(z)dz = \int_0^{t_{ij}^*} \phi_{ij}(z)dz$.[3] It implies that $\beta_{ij} = \psi(t_{ij}^*)$. Observe also that $\psi''(t) = \phi'_{ij}(t)$. Then the Laplace's method for approximating integrals yields that[4]

$$\lim_{N\to\infty} \frac{\int_0^1 \exp\left[N\psi(\xi)\right]d\xi}{N^{-1/2}\exp(N\beta_{ij})} = \sqrt{-\frac{2\pi}{\phi'_{ij}(t_{ij}^*)}} \equiv T_{ij}.$$

Fix $\xi > 0$ such that $T_{ij} - \xi > 0$. Then, for a sufficiently large $\bar{N}$, there exist some constants $\tilde{m}, \tilde{M} \in [T_{ij} - \xi, T_{ij} + \xi]$ such that

$$\tilde{m}N^{-1/2}\exp(N\beta_{ij}) < \int_0^1 \exp[N\psi(\xi)]d\xi < \tilde{M}N^{-1/2}\exp(N\beta_{ij}) \qquad \forall N \ge \bar{N}.$$

Using the above inequality, the expression (9) can be rewritten as[5]

$$\frac{1}{1+\tilde{M}e^{K_1+K_2}} \le N^{1/2}\exp(N\beta_{ij})\rho_{ij} \le \frac{1}{\tilde{m}e^{-K_1-K_2}} \qquad \forall N \ge \bar{N}. \tag{10}$$

Then, there are constants $0 < m_{ij} < M_{ij} < \infty$ such that $m_{ij} \le N^{1/2}\exp(N\beta_{ij})\rho_{ij} \le M_{ij}$ for all $N \ge 2$.

The proof of claim (ii) is omitted. It can be proved in a similar way using the Laplace's method for approximating integrals at the boundary (see Section 4.3 of De Bruijn 1981 and Lemma 2 of Fudenberg and Imhof 2008). □

Proof of Theorem 3.7. Recall that we analyze the stationary distribution of the Markov chain where the transition probability from the all-i state to the all-j state is $\mu_{ij}\rho_{ij}$ for $i, j \in \Theta$. We follow the discussion in Sandholm (2010b, Section 12.A.1). Recall that Υ_θ is the set of θ-trees for $\theta \in \Theta$. Define a vector $v \in \mathbb{R}_+^{|\Theta|}$ as

$$v_\theta = \sum_{\tau_\theta \in \Upsilon_\theta} \prod_{(i,j)\in\tau_\theta} \mu_{ij}\rho_{ij}.$$

[3]If $\operatorname{argmax}_{t\in[0,1]} \psi(t)$ is not a singleton, choose t_{ij}^* that maximizes $-2\pi/\phi'_{ij}(t_{ij}^*)$.

[4]See, for example, Section 4.2 of De Bruijn (1981) for the Laplace's method.

[5]For the left inequality to hold, we need $N^{1/2}\exp(N\beta_{ij}) \ge 1$. It is satisfied if $\bar{N}$ is sufficiently large.

It is well-known that v is a positive multiple of the stationary distribution of the Markov chain. Freidlin and Wentzell (1998) show that this result is useful to compute the limiting stationary distribution. Corollary 3.2 with Lemma A.1 implies that for each $\tau_\theta \in \Upsilon_\theta$,

$$\lim_{N\to\infty} \frac{1}{N} \ln \left(\prod_{(i,j)\in\tau_\theta} \mu_{ij}\rho_{ij} \right) = -C(\tau_\theta).$$

This further implies that

$$\lim_{N\to\infty} \frac{1}{N} \ln \left(\sum_{\tau_\theta\in\Upsilon_\theta} \prod_{(i,j)\in\tau_\theta} \mu_{ij}\rho_{ij} \right) = -C_\theta.$$

Recall that x^i is the all-i state and $\pi_0(x^i)$ is the probability mass on the all-i state in the limiting stationary distribution. Freidlin and Wentzell (1998)'s technique implies that $\pi_0(x^i)/\pi_0(x^j) = v_i/v_j$. As N becomes large, the ratio of the probability mass of one state to another can be approximated as

$$\frac{\pi_0(x^i)}{\pi_0(x^j)} \propto \exp(-N(C_i - C_j)).$$

Thus, $\lim_{N\to\infty} \pi_0(x^i)/\pi_0(x^j) = 0$ if $C_i > C_j$. The claim follows. □

References

Alexander, R. D., 1987, *The Biology of Moral Systems*: Taylor & Francis Inc.

Alger, I. and J. W. Weibull, 2013, "Homo moralis–preference evolution under incomplete information and assortative matching," *Econometrica* 81, No. 6, 2269–2302.

——— 2016, "Evolution and kantian morality," *Games and Economic Behavior* 98, 56–67.

——— 2019, "Evolutionary models of preference formation," *Annual Review of Economics* 11, No. 1, 329–354.

Andreoni, J., W. Harbaugh, and L. Vesterlund, 2003, "The carrot or the stick: Rewards, punishments, and cooperation," *American Economic Review* 93, No. 3, 893–902.

Arigapudi, S., 2020a, "Exit from equilibrium in coordination games under probit choice," *Games and Economic Behavior* 122, 168–202.

——— 2020b, "Transitions between equilibria in bilingual games under logit choice," *Journal of Mathematical Economics* 86, 24–34.

Berger, U., 2011, "Learning to cooperate via indirect reciprocity," *Games and Economic Behavior* 72, No. 1, 30–37.

Berger, U. and A. Grüne, 2016, "On the stability of cooperation under indirect reciprocity with first-order information," *Games and Economic Behavior* 98, 19–33.

Bergstrom, T. C., 1995, "On the evolution of altruistic ethical rules for siblings," *American Economic Review* 85, No. 1, 58–81.

——— 2002, "Evolution of social behavior: Individual and group selection," *Journal of Economic Perspectives* 16, No. 2, 67–88.

——— 2003, "The algebra of assortative encounters and the evolution of cooperation," *International Game Theory Review* 5, No. 3, 211–228.

Bilancini, E., L. Boncinelli, and J. Wu, 2018, "The interplay of cultural intolerance and action-assortativity for the emergence of cooperation and homophily," *European Economic Review* 102, 1–18.

Binmore, K. and L. Samuelson, 1997, "Muddling through: Noisy equilibrium selection," *Journal of Economic Theory* 74, 235–265.

Binmore, K., L. Samuelson, and R. Vaughan, 1995, "Musical chairs: Modeling noisy evolution," *Games and Economic Behavior* 11, 1–35.

Chaudhuri, A., 2011, "Sustaining cooperation in laboratory public goods experiments: a selective survey of the literature," *Experimental Economics* 14, No. 1, 47–83.

De Bruijn, N. G., 1981, *Asymptotic Methods in Analysis*: Dover Publications.

Dekel, E., J. C. Ely, and O. Yilankaya, 2007, "Evolution of preferences," *Review of Economic Studies* 74, No. 3, 685–704.

Ellison, G., 1994, "Cooperation in the prisoner's dilemma with anonymous random matching," *Review of Economic Studies* 61, No. 3, 567–588.

Fehr, E. and S. Gächter, 2000, "Fairness and retaliation: The economics of reciprocity," *American Economic Review* 90, No. 4, 980–994.

——— 2002, "Altruistic punishment in humans," *Nature* 415, 137–140.

Foster, D. P. and H. P. Young, 1990, "Stochastic evolutionary game dynamics," *Theoretical Population Biology* 38, 219–232.

Freidlin, M. I. and A. D. Wentzell, 1998, *Random perturbations of dynamical systems*: Springer Verlag, New York, 2nd edition.

Fudenberg, D. and L. A. Imhof, 2006, "Imitation processes with small mutations," *Journal of Economic Theory* 131, 251–262.

——— 2008, "Monotone imitation dynamics in large populations," *Journal of Economic Theory* 140, 229–245.

Fudenberg, D. and D. K. Levine, 1989, "Reputation and equilibrium selection in games with a patient player," *Econometrica* 57, No. 4, 759–778.

Fudenberg, D., M. A. Nowak, C. Taylor, and L. A. Imhof, 2006, "Evolutionary game dynamics in finite populations with strong selection and weak mutation," *Theoretical Population Biology* 70, No. 3, 352–363.

Fujiwara-Greve, T. and M. Okuno-Fujiwara, 2009, "Voluntarily separable repeated prisoner's dilemma," *Review of Economic Studies* 76, No. 3, 993–1021.

Gintis, H., 2000, "Strong reciprocity and human sociality," *Journal of Theoretical Biology* 206, No. 2, 169–179.

Güth, W., 1995, "An evolutionary approach to explaining cooperative behavior by reciprocal incentives," *International Journal of Game Theory* 24, No. 4, 323–344.

Güth, W. and H. Kliemt, 1998, "The indirect evolutionary approach:: Bridging the gap between rationality and adaptation," *Rationality and Society* 10, No. 3, 377–399.

Güth, W. and M. E. Yaari, 1992, "Explaining reciprocal behavior in simple strategic games: An evolutionary approach," in U. Witt ed. *Explaining Process and Change: Approaches to Evolutionary Economics*: University of Michigan Press, 23–34.

Guttman, J. M., 2003, "Repeated interaction and the evolution of preferences for reciprocity," *Economic Journal* 113, No. 489, 631–656.

Hamilton, W. D., 1964a, "The genetical evolution of social behaviour. I," *Journal of Theoretical Biology* 7, No. 1, 17–52.

——— 1964b, "The genetical evolution of social behaviour. II," *Journal of Theoretical Biology* 7, No. 1, 1–16.

Hashimoto, K. and K. Aihara, 2009, "Fixation probabilities in evolutionary game dynamics with a two-strategy game in finite diploid populations," *Journal of Theoretical Biology* 258, No. 4, 637–645.

Heifetz, A., C. Shannon, and Y. Spiegel, 2007a, "The dynamic evolution of preferences," *Economic Theory* 32, No. 2, 251–286.

——— 2007b, "What to maximize if you must," *Journal of Economic Theory* 133, No. 1, 31–57.

Heller, Y. and E. Mohlin, 2019, "Coevolution of deception and preferences: Darwin and nash meet machiavelli," *Games and Economic Behavior* 113, 223–247.

Herold, F., 2012, "Carrot or stick? the evolution of reciprocal preferences in a haystack model," *American Economic Review* 102, No. 2, 914–940.

Herold, F. and C. Kuzmics, 2009, "Evolutionary stability of discrimination under observability," *Games and Economic Behavior* 67, No. 2, 542–551.

Huck, S. and J. Oechssler, 1999, "The indirect evolutionary approach to explaining fair allocations," *Games and Economic Behavior* 28, No. 1, 13–24.

Hwang, S.-H., W. Lim, P. Neary, and J. Newton, 2018, "Conventional contracts, intentional behavior and logit choice: Equality without symmetry," *Games and Economic Behavior* 110, 273–294.

Imhof, L. A., D. Fudenberg, and M. A. Nowak, 2005, "Evolutionary cycles of cooperation and defection," *Proceedings of the National Academy of Sciences* 102, No. 31, 10797–10800.

Izquierdo, S. S., L. R. Izquierdo, and F. Vega-Redondo, 2010, "The option to leave: Conditional dissociation in the evolution of cooperation," *Journal of Theoretical Biology* 267, No. 1, 76–84.

Kandori, M., 1992, "Social norms and community enforcement," *Review of Economic Studies* 59, No. 1, 63–80.

Kandori, M., G. J. Mailath, and R. Rob, 1993, "Learning, mutation, and long run equilibria in games," *Econometrica* 61, 29–56.

Khan, A., 2021, "Evolutionary stability of behavioural rules in bargaining," *Journal of Economic Behavior & Organization* 187, 399–414.

——— 2022, “Expected utility versus cumulative prospect theory in an evolutionary model of bargaining,” *Journal of Economic Dynamics and Control*, p. 104332.

Mailath, G. J. and L. Samuelson, 2006, *Repeated Games and Reputations*: Oxford University Press Inc.

Naidu, S., S.-H. Hwang, and S. Bowles, 2010, “Evolutionary bargaining with intentional idiosyncratic play,” *Economics Letters* 109, No. 1, 31–33.

Newton, J., 2017, “The preferences of homo moralis are unstable under evolving assortativity,” *International Journal of Game Theory* 46, No. 2, 583–589.

——— 2018, “Evolutionary game theory: A renaissance,” *Games* 9, No. 2, p. 31.

Nöldeke, G. and L. Samuelson, 1993, “An evolutionary analysis of backward and forward induction,” *Games and Economic Behavior* 5, 425–454.

Nowak, M. A. and K. Sigmund, 1998a, “The dynamics of indirect reciprocity,” *Journal of Theoretical Biology* 194, No. 4, 561–574.

——— 1998b, “Evolution of indirect reciprocity by image scoring,” *Nature* 393, No. 6685, 573–577.

Nowak, M. A., A. Sasaki, C. Taylor, and D. Fudenberg, 2004, “Emergence of cooperation and evolutionary stability in finite populations,” *Nature* 428, No. 6983, 646–650.

Ohtsuki, H. and Y. Iwasa, 2004, “How should we define goodness? — reputation dynamics in indirect reciprocity,” *Journal of Theoretical Biology* 231, No. 1, 107–120.

——— 2006, “The leading eight: Social norms that can maintain cooperation by indirect reciprocity,” *Journal of Theoretical Biology* 239, No. 4, 435–444.

Possajennikov, A., 2000, “On the evolutionary stability of altruistic and spiteful preferences,” *Journal of Economic Behavior & Organization* 42, No. 1, 125–129.

Robson, A. J., 1990, “Efficiency in evolutionary games: Darwin, nash and the secret handshake,” *Journal of Theoretical Biology* 144, No. 3, 379–396.

Sandholm, W. H., 2010a, “Orders of limits for stationary distributions, stochastic dominance, and stochastic stability,” *Theoretical Economics* 5, 1–26.

——— 2010b, *Population Games and Evolutionary Dynamics*: MIT Press, 1st edition.

——— 2012, "Stochastic imitative game dynamics with committed agents," *Journal of Economic Theory* 147, No. 5, 2056–2071.

Sandholm, W. H. and M. Staudigl, 2016, "Large deviations and stochastic stability in the small noise double limit," *Theoretical Economics* 11, 279–355.

——— 2018, "Sample path large deviations for stochastic evolutionary game dynamics," *Mathematics of Operations Research* 43, No. 4, 1348–1377.

Sawa, R., 2012, "Mutation rates and equilibrium selection under stochastic evolutionary dynamics," *International Journal of Game Theory* 41, 489–496.

——— 2021a, "A prospect theory nash bargaining solution and its stochastic stability," *Journal of Economic Behavior & Organization* 184, 692–711.

——— 2021b, "Stochastic stability in the large population and small mutation limits for coordination games," *Journal of Dynamics & Games*. Forthcoming.

Schaffer, M. E., 1988, "Evolutionarily stable strategies for a finite population and a variable contest size," *Journal of Theoretical Biology* 132, No. 4, 469–478.

Sethi, R. and E. Somanathan, 2001, "Preference evolution and reciprocity," *Journal of Economic Theory* 97, No. 2, 273–297.

Staudigl, M., S. Arigapudi, and W. H. Sandholm, 2021, "Large deviations and stochastic stability in population games," *Journal of Dynamics & Games*. Forthcoming.

Takahashi, S., 2010, "Community enforcement when players observe partners' past play," *Journal of Economic Theory* 145, No. 1, 42–62.

Trivers, R. L., 1971, "The evolution of reciprocal altruism," *Quarterly Review of Biology* 46, No. 1, 35–57.

Wu, J., 2020, "Labelling, homophily and preference evolution," *International Journal of Game Theory* 49, No. 1, 1–22.

Young, H. P., 1993, "The evolution of conventions," *Econometrica* 61, 57–84.