Diverse Behavior Patterns in a Symmetric Society with Voluntary Partnerships*

Takako Fujiwara-Greve†
Dept. of Economics
Keio University

Masahiro Okuno-Fujiwara‡
Dept. of Economics
Musashino University

December 29, 2016

Abstract: Voluntarily Separable Repeated Prisoner's Dilemma (Fujiwara-Greve and Okuno-Fujiwara, 2009) describes a large, mobile society where players can easily find a new partner but also can end the partnership unilaterally without information flows to future partners. Although there is no Nash equilibrium in which all players cooperate from the beginning of each new partnership, we show that this model is rich in its own way: there is a continuum of weakly evolutionarily stable distributions, in which many partnerships eventually achieve long-term cooperation, the periods it takes to reach long-term cooperation can vary from one to any finite number, and in the meantime any action combination sequence can be played. The equilibria give a rationale to diverse modes of behavior including tolerance of the partner's non-cooperation for some periods. No personalized punishment is needed to achieve long-term cooperation. The diversity of strategies in the society itself is the incentive device. (148 words.)

JEL: C73.

Keywords: diversity, tolerance, evolution, voluntary partnerships, repeated Prisoner's Dilemma.

---

*We are grateful to Nobue Suzuki for her cooperation. We also thank Darrel Duffie, Hugo Sonnenschein, Bruno Strulovici, Bob Wilson, Klaus Ritzberger, Akihiko Matsui, Tadashi Sekiguchi, Takuo Sugaya, Chen-Lai Yang, Henrich Greve, and seminar participants at AMES Singapore 2013, Keio University, Waseda University, Ryukoku University, Washington University in St. Louis, University of Wisconsin (Madison), Northwestern University, Stanford University and University of Chicago for helpful comments on the versious earlier drafts. This work is supported by JSPS KAKENHI Grant Numbers (C) No. 21530171 and (B) No. 24330064. The usual disclaimer applies.

†Corresponding author. Mailing address: Dept. of Economics, Keio University, 2-15-45 Mita, Minatoku, Tokyo 108-8345 JAPAN. Email: takakofg@econ.keio.ac.jp

‡Email: okuno.fujiwara@gmail.com

1 Introduction

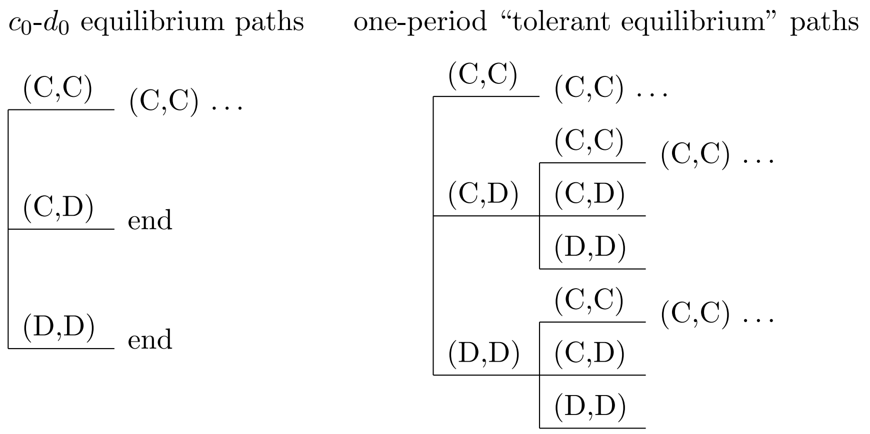
Voluntarily repeated games with a Prisoner's Dilemma type stage game, formulated and analyzed in, e.g., Datta (1996), Ghosh and Ray (1996), Kranton (1996a), Carmichael and Macleod (1997), Eeckhout (2006), Fujiwara-Greve and Okuno-Fujiwara (2009), Rob and Yang (2010), and Mcadams (2011)[1], are natural models to describe a large, anonymous society in which meeting a new partner is random but continuation of a partnership is a mutual choice by the partners. The main problem is how to achieve mutually cooperative partnerships when each player can unilaterally end the partnership without information flow to new partners. Cooperation from the beginning of new partnerships by all agents is never a Nash equilibrium. If all players cooperate with a stranger, one can defect and run away in every partnership to earn the highest one-shot payoff every period.

Most literature of the voluntarily repeated Prisoner's Dilemma focused on symmetric-strategy, trust-building/gradual cooperation equilibria to cope with the lack of personalized punishments. If all partnerships in the society require some periods of non-cooperation/low levels of cooperation before shifting to a full cooperation, deviations towards a cooperative partner are discouraged by this "community punishment", because one-shot deviation gain is offset by a sufficiently long phase of trust-building in the next partnership. The gradual increase is also important in symmetric equilibria to keep the continuation payoff increasing to motivate the partners to stay.

We now think that in a very large society with limited information, such coordination on the initial trust-building by all players is unlikely. In our previous paper, Fujiwara-Greve et al. (2015) (henceforth GOS), we thus proposed a simple two-strategy equilibrium in which no coordination is needed. The "fundamentally asymmetric" equilibrium is as follows. The cooperative c_0-strategy always chooses the cooperative action C in the Prisoner's Dilemma but keeps the partnership if and only if (C,C) is observed. The myopic d_0-strategy always chooses the selfish D action and ends the partnership regardless of the history within the partnership. GOS showed that the long-run average payoff of the c_0-strategy is nonlinear in its share in the matching pool, while that of the d_0-strategy is linear in the share of the c_0-strategy, so that the two strategies can earn the same payoff at some ratio (see Figure 4 below). The cooperation incentive is provided by the existence of the myopic defectors in the matching pool, and no coordination to make a community punishment is necessary.

[1]See also the dynamic analyses by Schumacher (2013) and Izquierdo et al. (2014), focusing on stationary or Markov strategies.

Figure 1: Internalization of the matching pool distribution



In this paper, we show that the voluntary nature of the partnerships gives us a huge variety of asymmetric equilibria by *internalization of the matching pool distribution* when the partnership is about to end in the fundamentally asymmetric equilibrium. The idea is as follows. Consider the moment when a pair of players is about to end the partnership. This is an on-path behavior when a c_0-player meets a d_0-player or two d_0-players meet. The continuation payoff for each partner is the payoff to start in the matching pool. Instead of ending the partnership, if the players can reproduce the matching pool strategy distribution as the continuation strategy distribution, they can choose to stay, and this "tolerant" strategy combination is also an equilibrium (see Figure 1).

Using this idea, we show that the Voluntarily Separable Repeated Prisoner's Dilemma (henceforth VSRPD) of Fujiwara-Greve and Okuno-Fujiwara (2009) is "rich" in equilibrium modes of behavior. There is a continuum of equilibria in which (i) each player uses a pure strategy, (ii) a lot of (but not all) partnerships eventually achieve long-term cooperation (and the cooperative partnerships gradually accumulate), (iii) the periods it takes to achieve long-term cooperation varies from one to any finite number, and (iv) in the meantime, any action combination sequence can be played. This is the meaning of the title of the paper. The internalization logic is only possible in the voluntary partnership setup. Note also that all these equilibria have the same average payoff (which is the equilibrium payoff of the c_0-d_0 equilibrium). These features highlight how the voluntary framework makes a difference from ordinary repeated Prisoner's Dilemma. The incentive to cooperate and/or tolerate the (C,D) action combination is provided by the diversity of the equilibrium strategies itself. Tolerant players are hoping to be matched with someone who will eventually cooperate with them, with an intention to either establish a long-term cooperative partnership, or exploit the cooperator. When there are similar tolerant players in the society, this is an equilibrium behavior.

As refinements of Nash equilibrium, we consider static evolutionary stability concepts, based on the Evolutionarily Stable Strategy (ESS) introduced by Maynard Smith and Price (1973). Although the VSRPD model describes a natural economic situation of large, anonymous markets, it is too much to assume that each player can rationally choose actions at each information set. Consider, for example, the decision of whether to end the current partnership or not. A player needs to have a consistent belief on not only the continuation strategy by the current opponent but also the strategy distribution in the matching pool after this period. The strategy distribution in the matching pool evolves stochastically due to the random death process and also depends on the details of the strategies in the society which determine the voluntary termination of pairs. Thus the rational equilibrium approach would impose an unrealistically strong assumption on the belief formation. Evolutionary approach does not require individual optimization but still gives refinements of Nash equilibria and can be interpreted as a long-run outcome of boundedly rational agents' strategy adjustment processes.

The diverse-behavior equilibria satisfy an evolutionary stability, such that any mutant is outperformed by some equilibrium strategy, when mutant distributions are within a "sufficiently random distribution" set.[2] Random mutations that put positive probabilities over a wide range of strategies are often assumed in evolutionary models such as Kandori et al. (1993). However, the outcome of random mutations is quite different. In the finite population, coordination game context in Kandori et al. (1993), random mutations eventually lead to coordination on a pure strategy by sufficiently many mutants to upset an equilibrium. By contrast, in our infinite population, Prisoner's Dilemma framework, random mutation means that cooperative mutants, who can invade the incumbents by secret handshake (Robson, 1990), are always exploited by other mutants, and the mutant distributions "self-defeat".

A weaker richness result holds for stronger stability concepts. Namely, the c_0-d_0 equilibrium satisfies many extensions of the ESS to the VSRPD model, when mutant distributions are a little more random than the ones for the diverse-behavior equilibria. Since there are new matches in the society in every period, any action combination in the Prisoner's Dilemma are observed in the society every period, under the two-strategy equilibrium.

[2]Since the stage game is an extensive form, the ordinary ESS does not exist, and GOS showed that the c_0-d_0-equilibrium is not neutrally stable.

	C	D
C	c, c	ℓ, g
D	g, ℓ	d, d

Table 1: Prisoner's Dilemma: $g > c > d > \ell$ and $2c \geqq g + \ell$.

The rest of the paper is organized as follows. In Section 2 we describe the model, stability notions, and the fundamentally asymmetric equilibrium of GOS. Section 3 gives an example of two-period richness result of equilibrium behaviors, to give the intuition. In Section 4, we prove the evolutionary stabilities of the diverse-behavior equilibria and the fundamentally asymmetric equilibrium of GOS. Section 5 concludes the paper.

2 Model and Preliminaries

2.1 Voluntarily Separable Repeated Prisoner's Dilemma

The model of *Voluntarily Separable Repeated Prisoner's Dilemma* (VSRPD) introduced by Fujiwara-Greve and Okuno-Fujiwara (2009) (henceforth Greve-Okuno) is as follows. There is a continuum population of homogeneous players of measure 1, over the infinite, discrete time horizon. At the beginning of each period, players are either matched with a partner from the previous period or without a partner. Those without a partner enter a random matching process and form pairs[3] to play the following extensive form game.

Newly matched players have no verifiable knowledge[4] of the past action history of each other (the no-information-flow assumption), and they play the ordinary two-action symmetric Prisoner's Dilemma of Table 1. The actions in the Prisoner's Dilemma are observable only to the current partners. After observing the actions in the Prisoner's Dilemma, the partners simultaneously choose whether to keep the partnership (action k) or to end it (action e). A partnership dissolves if at least one partner chooses action e. In addition, at the end of a period, each player may exit from the society for some exogenous reason (which we call a "death") with probability $1 - \delta$, where $0 < \delta < 1$. If a player dies, a new player enters into the society, keeping the population size constant. Players who lost the partner for some reason as well as newly born players enter the matching pool in the next period. This justifies the no-information-flow assumption, because

[3]For simplicity and following Greve-Okuno, we assume that a player finds a new partner for sure. This assumption makes cooperation most difficult.

[4]In the continuum population, even if players remember their personal histories and try to communicate with others, it is unlikely that they can give information to a positive measure of other players. Therefore effectively there is no information flow across different partnerships.

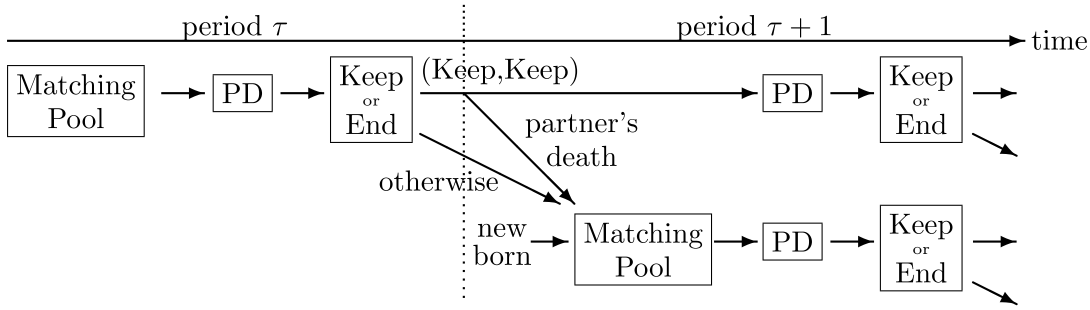


Figure 2: Timeline of the VSRPD

the players in the matching pool have different backgrounds. In sum, a partnership continues if and only if both partners choose action k and do not die. In this case the same partners play the Prisoner's Dilemma in the next period, skipping the matching process. The game continues this way *ad infinitum.* The timeline of VSRPD is depicted in Figure 2.

The one-shot payoffs in the Prisoner's Dilemma are in Table 1, where $g > c > d > \ell$ and $2c \geqq g+\ell$. The latter makes the symmetric pure-action profile (C,C) more efficient than alternating (C,D) and (D,C).

The game continues with probability δ from an individual player's point of view, hence δ is the effective discount factor of a player. (However, even if both partners choose to keep each other, the partnership continues with probability δ^2 only.)

Under the no-information-flow assumption, we focus on match-independent strategies[5] that only depend on the period $t = 1, 2, \ldots$ within a partnership (not the calendar time in the whole game) and the private history of actions within a partnership. Let $H_t := [\{C, D\} \times \{C, D\}]^{t-1}$ be *the set of partnership histories*[6] *at the beginning of* $t \geqq 2$ and let $H_1 := \{\emptyset\}$.

Definition 1. A *pure strategy* s of VSRPD consists of $(s_{1t}, s_{2t})_{t=1}^{\infty}$ where:
$s_{1t} : H_t \to \{C, D\}$ specifies an action choice $s_{1t}(h_t) \in \{C, D\}$ given the partnership history $h_t \in H_t$, and
$s_{2t} : H_t \times \{C, D\}^2 \to \{k, e\}$ specifies whether to keep or to end the partnership, depending on the partnership history $h_t \in H_t$ and the current period action profile.

The set of pure strategies of VSRPD is denoted as $\mathbf{S}$ and the set of all strategy distributions in the population is denoted as $\Delta(\mathbf{S})$. A pure strategy can be viewed as a degenerate strategy

[5]The continuum population implies that "contagious" strategies used in Kandori (1992) and Ellison (1994) cannot make an impact on a positive measure of the population and hence are irrelevant.

[6]The first coordinate is the player's action. The relevant histories on which partners can condition their actions are the action combinations in the Prisoner's Dilemma only, because the continuation decision history must be (k, k) throughout.

distribution and thus belongs to $\Delta(\mathbf{S})$, and with a slight abuse of notation, we write $s \in \mathbf{S}$ in both sense as a pure strategy and a distribution in $\Delta(\mathbf{S})$ that puts the mass of 1 on s.

For simplicity, we assume that each player is endowed with a pure strategy, and occasionally changes to another pure strategy. This is without loss of generality. In a continuum population, the same distribution of pairs are generated by (i) all players using the same individually-mixed strategy $p \in \Delta(S)$, and (ii) all players are playing a pure strategy but the fraction of players using s is $p(s)$.[7] Hence we adopt a simpler formulation of (ii).

We investigate evolutionary stability of **stationary** (pure-strategy) distributions **in the matching pool**. Although the strategy distribution in the matching pool may be different from the distribution in the entire society, if the former is stationary, the distribution of various "states" of matches (corresponding to possible private histories) is also stationary, thanks to the stationary death process.[8] Since each player is born into the random matching pool, the life-time payoff is determined by the strategy distribution in the matching pool.

2.2 Stability Concepts

The expected lifetime average payoff of a player endowed with a pure strategy $s \in \mathbf{S}$ at the time of entering the matching pool consisting of a stationary strategy distribution p (with finite or countable support) is written as $v(s;p)$ and is derived as follows.[9] Let $T(s,s')$ be the planned length of the partnership when the player with strategy s meets a player with strategy s'. The expected length of this match is then

$$L(s,s') = 1 + \delta^2 + \cdots + \delta^{2(T(s,s')-1)}.$$

Let the total expected payoff for the s-player in the (s,s') pair be $V(s,s')$. Then the expected long-run payoff when s-player is in the matching pool with the stationary distribution p (with up to the countable support) is recursively formulated as follows.

$$V(s;p) = \sum_{s' \in supp(p)} p(s') \Big[V(s,s') + [\delta(1-\delta)\{1 + \delta^2 + \cdots + \delta^{2(T(s,s')-2)}\} + \ \delta^{2(T(s,s')-1)} \cdot \delta] V(s;p) \Big].$$

[7]See Sun (2006) and Duffie and Sun (2012) for the foundation of the Law of Large Numbers.

[8]See Greve-Okuno footnote 7 for the details. For specific strategies, e.g., c_T- and $D^T_{d_0}$-strategies, we can prove that any stationary distribution in the matching pool exists consistently with the model.

[9]For more details, see Greve-Okuno. The probabilistic foundation to the dynamic process is given by Duffie et al. (2016). Their "mutation" should be interpreted as changes of each player's "states" which is a combination of whether the player is a newborn or not and the strategy (s)he has.

By rearrangements, we have the average payoff for s-player facing a stationary distribution p in the matching pool:

$$v(s;p) := \frac{\sum_{s' \in supp(p)} p(s')V(s,s')}{\sum_{s' \in supp(p)} p(s')L(s,s')}. \tag{1}$$

Notice that, in general, the average payoff is **not linear** in the share distribution p.

Definition 2. A stationary strategy distribution in the matching pool $p \in \Delta(\mathbf{S})$ is a *Nash equilibrium* if, for all $s \in supp(p)$ and all $s' \in \mathbf{S}$,

$$v(s;p) \geqq v(s';p).$$

From the evolutionary perspective in a continuum population, a Nash equilibrium is a robust distribution against single-strategy and measure-zero mutants. It is the weakest stability notion in terms of the allowed measure of mutants (zero instead of a positive measure).

Local stability requires that the incumbent distribution is robust against a small positive measure of mutant distributions, which use only (a subset of) the incumbent strategies[10] and have a higher weight on a strategy than the incumbent distribution does. (See Figure 4 below.) For any $p \in \Delta(\mathbf{S})$ define such distributions:

$$Q(p) = \{q \in \Delta(\mathbf{S}) \mid supp(q) \subseteq supp(p),\ \exists\ s' \in supp(p);\ q(s') > p(s')\}.$$

For any $q \in Q(p)$, let $S_p(q) = \{s' \in supp(p) \mid q(s') > p(s')\}$ be the (non-empty) set of strategies with increased shares in q as compared to p.

Definition 3. A stationary strategy distribution in the matching pool $p \in \Delta(\mathbf{S})$ is a *locally stable Nash equilibrium* if,
(i) p is a Nash equilibrium; and
(ii) for any $q \in Q(p)$, there exists $\bar{\epsilon} \in (0,1)$ such that, for any $\epsilon \in (0,\bar{\epsilon})$, any $s' \in S_p(q)$, and any $s \in supp(p) \setminus supp(S_p(q))$,

$$v(s;(1-\epsilon)p + \epsilon \cdot q) \geqq v(s';(1-\epsilon)p + \epsilon \cdot q),$$

with the strict inequality for at least one $\tilde{s} \in supp(p) \setminus \mathrm{supp}(S_p(q))$.

Definition 3 is an extension of the Local Stability of GOS, which corresponds to the case that q puts a mass of 1 on some $s' \in supp(p)$. A monomorphic Nash equilibrium trivially satisfies

[10] Thus the name "local" is about the possible strategies that mutants can use, or the "direction" of perturbations in $\Delta(\mathbf{S})$. Each of our stability concept also requires a sufficiently small "distance" of perturbations.

Definition 3, because $Q(p)$ is empty, but not Definition 4 of GOS, for just a technical reason. For polymorphic Nash equilibria, Definition 3 is stronger than Definition 4 of GOS, because the equilibrium must be robust against any $q \in Q(p)$, not just a point-mass distribution.

Let us expand the set of possible mutant strategy distributions, and to make the stability concepts flexible, we define evolutionary stability concepts **with respect to** the set of potential mutant distributions. The larger the set of feasible mutant strategy distributions is, the stronger the stability concept becomes. This is similar to the comparison of the size of the basin of attraction among equilibria, as in Bendor and Swistak (1997).

The strongest stability that we can hope for is the following. (Because the set of mutant distributions M can be a continuum, it is too much to require a uniform invasion barrier, that is to have $\bar{\epsilon}$ for any $q \in M$.)

Definition 4. A stationary strategy distribution p in the matching pool is a *Strongly Evolutionarily Stable Distribution under mutants within* $M(\subset \Delta(\mathbf{S}))$ (p is a Strong-ESD(M)) if
(i) p is a locally stable Nash equilibrium,
(ii) for any $q \in M$, there exists $\bar{\epsilon} \in (0,1)$ such that for any $\epsilon \in (0, \bar{\epsilon})$ and any $s' \in supp(q) \setminus supp(p)$,

$$\forall s \in supp(p), \;\; v(s; (1-\epsilon)p + \epsilon \cdot q) > v(s'; (1-\epsilon)p + \epsilon \cdot q).$$

The condition (ii) requires that all incumbent strategy outperform all mutant strategy head to head. Hence in any monotone dynamic[11], all mutant strategies would disappear and the condition (i) restores the balance among incumbents.

However, it many cases we need weaker concepts.

Definition 5. A stationary strategy distribution p (with up to the countable support) in the matching pool is a *Mean Stable Distribution under mutants within* $M(\subset \Delta(\mathbf{S}))$ (p is an MSD(M)) if
(i) p is a locally stable Nash equilibrium,
(ii') for any $q \in M$, there exists $\bar{\epsilon} \in (0,1)$ such that for any $\epsilon \in (0, \bar{\epsilon})$,

$$\sum_{s \in \mathbf{S}} p(s) \cdot v(s; (1-\epsilon)p + \epsilon \cdot q) > \sum_{s' \in \mathbf{S}} q(s') \cdot v(s'; (1-\epsilon)p + \epsilon \cdot q).$$

[11] A possible underlying process is as follows. Divide the infinite time horizon into infinitely many sub-horizons, which are also infinite horizons. Each sub-horizon is a "medium"-run, where the population adjusts to yield a stationary post-entry distribution $(1-\epsilon)p + \epsilon \cdot q$ in the matching pool. After that, in the "long"-run over the infinitely many sub-horizons, the strategy distribution evolves according to $v(s; (1-\epsilon)p + \epsilon \cdot q)$.

The condition (ii') requires that the incumbent mean of the long-run average payoff is strictly greater than that of any mutant distribution within M. From the perspectives of the mean dynamics[12] (e.g., Sandholm, 2010), the strict inequality implies that the mutants' share would decline. The ESS (Maynard Smith and Price, 1973 and Maynard Smith, 1982) concept corresponds to the case when the stage game is a normal form, $M = \Delta(\mathbf{S})$ and v is the one-shot payoff function u.

For strategy distributions containing a lot of (possibly countably many) pure strategies, we have to settle with a differently weaker stability concept.

Definition 6. A stationary strategy distribution p in the matching pool is a *Weakly Evolutionarily Stable Distribution under mutants within* $M(\subset \Delta(\mathbf{S}))$ (p is a WESD(M)) if
(i') p is a Nash equilibrium,
(iii) for any $q \in M$ and any $s' \in supp(q) \setminus supp(p)$, there exist $\tilde{s} \in supp(p)$ and $\bar{\epsilon} \in (0,1)$ such that for any $\epsilon \in (0, \bar{\epsilon})$,

$$v(\tilde{s}; (1-\epsilon)p + \epsilon \cdot q) > v(s'; (1-\epsilon)p + \epsilon \cdot q).$$

The condition (iii) requires that each mutant strategy is strictly outperformed by some incumbent strategy head to head. A slightly stronger stability requires a "uniform winner" in the incumbent distribution.

Definition 7. A stationary strategy distribution p in the matching pool is *Stable with a Uniform Winner under mutants within* $M(\subset \Delta(\mathbf{S}))$ (p is SUW(M)) if
(i) p is a locally stable Nash equilibrium,
(iii') there exists $s^* \in supp(p)$ such that for any $q \in M$ and any $s' \in supp(q) \setminus supp(p)$, there exists $\bar{\epsilon} \in (0,1)$ such that for any $\epsilon \in (0, \bar{\epsilon})$,

$$v(s^*; (1-\epsilon)p + \epsilon \cdot q) > v(s'; (1-\epsilon)p + \epsilon \cdot q).$$

Clearly, if p is SUW(M) then it is a WESD(M). If p is a Strong-ESD(M), then it is SUW(M), and the converse is also true if p is a monomorphic distribution. However, WESD(M) and MSD(M) are independent. Figure 3 illustrates the relationship among the various stability concepts and a preview of our results. Recall that the larger M means a stronger stability. Hence there is no clear order of stability between the c_0-d_0 equilibrium and the diverse-behavior equilibria.

[12] Izquierdo et al. (2014) gives an explicit mean dynamic within Markov strategies to support the fundamentally asymmetric equilibrium of GOS.

Figure 3: Relationship among stability concepts and our results

$\Leftarrow$ if p is monomorphic.

Strong-ESD (M) $\Rightarrow$	SUW(M)	$\Rightarrow$ WESD(M)
(c_0-d_0 satisfies for $D(\delta)$.)	(c_0-d_0 satisfies for $M(\delta)$.)	(All in $\bar{P}_n$ satisfy for $M_\infty(\delta)$.)
$\Searrow$		
	MSD(M) (c_0-d_0 satisfies for $M_+(\delta)$.)	$M_+(\delta) \subsetneq M(\delta) \subsetneq M_\infty(\delta)$ $D(\delta) \subsetneq M(\delta)$

It is possible to define the above stability concepts with the weak inequality, similar to the neutrally stable distribution in Greve-Okuno. However, there is no Nash equilibrium in which all players cooperate from the first period of new partnerships (Greve-Okuno, Lemma 1), thus we cannot hope for an evolutionary folk theorem similar to that of the ordinary repeated Prisoner's Dilemma (see Bendor and Swistak, 1997 and Garcìa and van Veelen, 2016) even at the level of neutrally stable strategies. Allowing the weak inequality may only enlarge the sufficient set of mutant distributions to include "neutral" strategies (those with the same play paths as one of the existing strategies', even if they enter the population by a positive measure) and the boundary of the set for the strict inequality case. Instead, with the strict inequality, we can safely interpret that mutant distributions in the sufficient set would be expelled.

2.3 Fundamentally Asymmetric Equilibrium

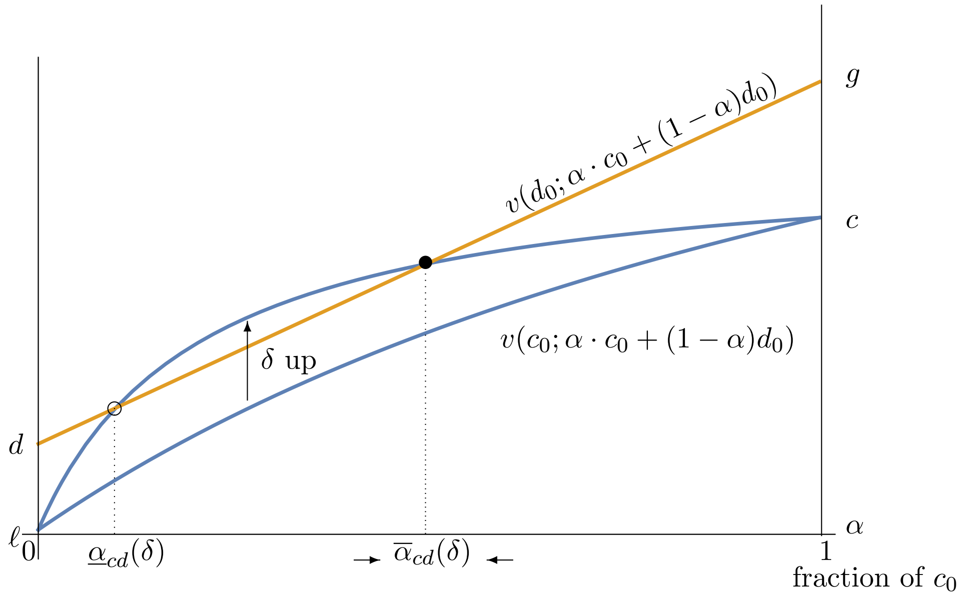
GOS showed the existence of the *fundamentally asymmetric* equilibrium consisting of conditional cooperators and myopic defectors.

Definition 8. Let c_0-*strategy* be a strategy as follows: in any period $t = 1, 2, \dots$ of a partnership, play C and keep the partnership if and only if (C, C) is observed in that period.
Let d_0-*strategy* be a strategy as follows: in any period $t = 1, 2, \dots$ of a partnership, play D and end the partnership for any action combination in that period.

The c_0-strategy is similar to the C-trigger strategy in the ordinary Folk Theorem (e.g., Fudenbeg and Maskin, 1986) with ending the partnership as a punishment. The d_0-strategy is the most myopic strategy but constitutes a Nash equilibrium trivially.[13] Since we have a new definition of local stability, we extend the result of GOS and Izquierdo et al. (2014).

[13]See Greve-Okuno Section 2.3, where it is called the $\tilde{d}$-strategy.

Figure 4: Locally stable c_0-d_0 equilibrium



Lemma 1. *There exists* $\underline{\delta}_{c_0 d_0} \in (0,1)$ *such that* $\delta > \underline{\delta}_{c_0 d_0}$ *if and only if there exists a unique* $\overline{\alpha}_{cd}(\delta) \in (0,1)$*, such that the bimorphic distribution* $\overline{\alpha}_{cd}(\delta)c_0 + \{1 - \overline{\alpha}_{cd}(\delta)\}d_0$ *is the unique locally stable Nash equilibrium with the support* $\{c_0, d_0\}$*.*

Proof. See Appendix.

The equilibrium share ratio $\overline{\alpha}_{cd}(\delta)$ is going to be important, and we call it **the C-D ratio**. The locally stable one exists uniquely for any $\delta > \underline{\delta}_{c_0 d_0}$.

The intuition of Lemma 1 is depicted in Figure 4. The average payoff functions of the c_0- and the d_0-strategy in the population of $\alpha c_0 + (1-\alpha)s_D$, where s_D is any strategy that plays D in the first period of a match, are computed as follows:

$$v(c_0; \alpha \cdot c_0 + (1-\alpha)s_D) = \frac{\alpha \cdot \frac{c}{1-\delta^2} + (1-\alpha)\ell}{\alpha \cdot \frac{1}{1-\delta^2} + 1 - \alpha}; \tag{2}$$

$$v(d_0; \alpha \cdot c_0 + (1-\alpha)s_D) = \alpha \cdot g + (1-\alpha)d. \tag{3}$$

(Note that these average payoffs only depend on the share of the c_0-strategy.) The average payoff of the c_0-strategy is concave in its share α, while that of the d_0-strategy is linear. The concavity is due to the voluntary nature of partnerships. As the survival rate δ increases, the average payoff of the c_0-player increases for any α, because the c_0-pairs last longer. Hence, the average payoff function of the c_0-strategy becomes more concave as δ increases. At some point, the average payoff functions of the two strategies have two intersections and the one with the larger share of the c_0-strategy is locally stable as illustrated in Figure 4.[14] Lemma 5 of GOS proves that if $v(c_0; \alpha c_0 + (1-\alpha)d_0) = v(d_0; \alpha c_0 + (1-\alpha)d_0)$, then the common payoff v^* satisfies the Best Reply Condition in Greve-Okuno with the strict inequality:

$$g + \delta \frac{v^*}{1-\delta} < \frac{c}{1-\delta^2} + \frac{\delta(1-\delta)}{1-\delta^2} \cdot \frac{v^*}{1-\delta}.$$

The LHS is the payoff of one-period deviation from the c_0-strategy when the partner is also the c_0-strategy, and the RHS is the payoff from following the c_0-strategy. Since all other one-period deviations have even lower payoff, the two-strategy combination is a Nash equilibrium.

3 Example: Diverse behaviors for two periods

To give an intuition of how the internalization of the matching pool distribution works, we give the class of equilibria in which all action combinations over two periods are generated, as illustrated in Figure 1 in the Introduction.

Definition 9. *Let the C_{c_0}-strategy be a strategy such that*

$t = 1$ *(Tolerant phase): Play C and keep for any observation;*

$t \geqq 2$ *(Commitment phase): Play the c_0-strategy as the continuation strategy for any observation in $t = 1$.*

Let the C_{d_0}-strategy be a strategy such that

$t = 1$*: Play C and keep for any observation;*

$t \geqq 2$*: Play the c_0-strategy as the continuation strategy if (C, C) was observed in $t = 1$, and play the d_0-strategy if (C, D) was observed in $t = 1$.*

Let the D_{c_0}-strategy be a strategy such that

$t = 1$*: Play D and keep for any observation;*

$t \geqq 2$*: Play the c_0-strategy as the continuation strategy for any observation in $t = 1$.*

Let the D_{d_0}-strategy be a strategy such that

$t = 1$*: Play D and keep for any observation;*

$t \geqq 2$*: Play the d_0-strategy as the continuation strategy for any observation in $t = 1$.*

We call them the *one-period tolerant* strategies. They choose "keep" action in the tolerant phase of one period regardless of the action combination, and in the second period of a match,

[14] For later references we denote the smaller payoff-equivalent share by $\underline{\alpha}_{cd}(\delta)$.

they commit to one of the c_0- or the d_0-strategy, unless (C,C) is established in $t = 1$, in which case they follow the c_0-strategy. Let $\tilde{S}_0^1 = \{c_0, d_0, C_{c_0}, C_{d_0}, D_{c_0}, D_{d_0}\}$ be the set of degenerate (0-period) and one-period tolerant strategies. Although $\tilde{S}_0^1$ does not exhaust all combinations of action plans in two periods, they are sufficient to reproduce the c_0-d_0 equilibrium distribution in the second period, when (C,D) or (D,D) is played in the first period of a match. For example, if a C_{c_0}-player experienced (C,D) and (keep, keep) in the first period of a match, the possible strategies by the partner belongs to $\{D_{c_0}, D_{d_0}\}$. If the share of D_{c_0} in $\{D_{c_0}, D_{d_0}\}$ is the C-D ratio in the matching pool, then the continuation strategy combination is the c_0-d_0 equilibrium distribution from the viewpoint of the C_{c_0}-player.

A generic element of $\tilde{S}_0^1$ is written as X_{y_0} where $X \in \{\emptyset, C, D\}$ is the tolerant phase action and $y_0 \in \{c_0, d_0\}$ is the commitment strategy unless (C,C) is established during the tolerant phase, with the convention that $\emptyset_{c_0} = c_0$ and $\emptyset_{d_0} = d_0$.

3.1 Average payoff function decomposition

For notational simplicity, for each $y_0 \in \{c_0, d_0\}$, we write $V_{cd}(y_0; \alpha) := V(y_0; \alpha \cdot p_{c_0} + (1-\alpha)p_D)$, $L_{cd}(y_0; \alpha) := L(y_0; \alpha \cdot p_{c_0} + (1-\alpha)p_D)$ and $v_{cd}(y_0; \alpha) := v(y_0; \alpha \cdot p_{c_0} + (1-\alpha)p_D)$, where p_{c_0} is any strategy distribution in the "C-start strategies" in $\tilde{S}_0^1$, $\{c_0, C_{c_0}, C_{d_0}\}$ and p_D is any strategy distribution consisting of strategies that play D in $t = 1$. The average payoffs of the c_0- and the d_0-strategies only depend on this structure.

We first show that the average payoff functions of all strategies in $\tilde{S}_0^1$ decompose into a weighted sum of $v_{cd}(c_0; \alpha)$'s and $v_{cd}(d_0; \alpha)$'s. To show that, we express a strategy distribution $p \in \Delta(\tilde{S}_0^1)$ by the "relative ratio form" as follows.

$$
\begin{aligned}
p =& p(\{c_0, C_{c_0}, C_{d_0}\})\Big[\frac{p(c_0)}{p(\{c_0, C_{c_0}, C_{d_0}\})} \cdot c_0 \\
&+ \big[1 - \frac{p(c_0)}{p(\{c_0, C_{c_0}, C_{d_0}\})}\big] \cdot \big[\frac{p(C_{c_0})}{p(\{C_{c_0}, C_{d_0}\})} \cdot C_{c_0} + \frac{p(C_{d_0})}{p(\{C_{c_0}, C_{d_0}\})} \cdot C_{d_0}\big]\Big] \\
&+ p(\{d_0, D_{c_0}, D_{d_0}\})\Big[\frac{p(d_0)}{p(\{d_0, D_{c_0}, D_{d_0}\})} \cdot d_0 \\
&+ \big[1 - \frac{p(d_0)}{p(\{d_0, D_{c_0}, D_{d_0}\})}\big] \cdot \big[\frac{p(D_{c_0})}{p(\{D_{c_0}, D_{d_0}\})} \cdot D_{c_0} + \frac{p(D_{d_0})}{p(\{D_{c_0}, D_{d_0}\})} \cdot D_{d_0}\big]\Big]. \qquad (4)
\end{aligned}
$$

To simplify the notation, we sort the strategies in $\tilde{S}_0^1$ with respect to initial actions: $C_+ := \{c_0, C_{c_0}, C_{d_0}\}$, $D_+ := \{d_0, D_{c_0}, D_{d_0}\}$, $CC_+ := \{C_{c_0}\}$, $CD_+ := \{C_{d_0}\}$, $DC_+ := \{D_{c_0}\}$, and $DD_+ := \{D_{d_0}\}$. The last four sets are singletons within $\tilde{S}_0^1$ but will contain many strategies as we enlarge the scope of tolerant strategies in the main analysis. The idea is that a set X_+ contains strategies

Table 2: Play paths among $\tilde{S}_0^1$ strategies

you\partner	c_0	C_{c_0}	C_{d_0}	d_0	D_{c_0}	D_{d_0}
c_0	(C,C) ...	(C,C) ...	(C,C) ...	(C,D)	(C,D)	(C,D)
C_{c_0}	(C,C) ...	(C,C) ...	(C,C) ...	(C,D)	(C,D), (C,C) ...	(C,D), (C,D)
C_{d_0}	(C,C) ...	(C,C) ...	(C,C) ...	(C,D)	(C,D), (D,C)	(C,D), (D,D)
d_0	(D,C)	(D,C)	(D,C)	(D,D)	(D,D)	(D,D)
D_{c_0}	(D,C)	(D,C), (C,C) ...	(D,C),(C,D)	(D,D)	(D,D), (C,C) ...	(D,D), (C,D)
D_{d_0}	(D,C)	(D,C), (D,C)	(D,C),(D,D)	(D,D)	(D,D), (D,C)	(D,D), (D,D)

which start with action X in a match. A set XY_+ contains strategies which start with action X in $t = 1$, keep regardless of the observation, and choose action Y in $t = 2$ unless (C,C) is established in $t = 1$ (in which case it plays the c_0-strategy). Furthermore, let $C{*}_+ = CC_+ \cup CD_+$ and $D{*}_+ = DC_+ \cup DD_+$.

Then (4) becomes

$$\begin{aligned} p = p(C_+)\Big[&p(c_0 \mid C_+) \cdot c_0 \\ &+ \{1 - p(c_0 \mid C_+)\}\big\{p(CC_+|C{*}_+) \cdot C_{c_0} + p(CD_+|C{*}_+) \cdot C_{d_0}\big\}\Big] \\ + p(D_+)\Big[&p(d_0 \mid D_+) \cdot d_0 \\ &+ \{1 - p(d_0 \mid D_+)\}\big\{p(DC_+|D{*}_+) \cdot D_{c_0} + p(DD_+|D{*}_+) \cdot D_{d_0}\big\}\Big]. \end{aligned} \tag{5}$$

By (2) and (3), the average payoffs of the c_0- and the d_0-strategy only depend on the share $p(C_+)$ of the C-start strategies.

$$v(c_0; p) = v_{cd}(c_0; p(C_+));$$
$$v(d_0; p) = v_{cd}(d_0; p(C_+)).$$

Next, consider the one-period tolerant C_{c_0}-strategy. As Table 2 shows, if it meets any partner using a strategy in C_+, they establish (C,C) in $t = 1$ and continue (C,C) as long as the partners live. If (C,D) is observed in $t = 1$, the match continues to $t = 2$ if the partner's strategy belongs to $D_+ \setminus \{d_0\}$ and both partners survive. This probability is $\delta^2 p(D_+)[1 - p(d_0|D_+)]$. In $t = 2$, (C,C)... obtains if the partner has the D_{c_0}-strategy (with the conditional probability $p(DC_+|D{*}_+)$) and, otherwise, (C,D) occurs and the C_{c_0}-player ends the match. Hence,

$$
\begin{aligned}
V(C_{c_0};p) = p(C_+)\frac{c}{1-\delta^2} + [1-p(C_+)]\cdot \ell \\
&+ \delta^2 p(D_+)[1-p(d_0|D_+)]\cdot \left[p(DC_+|D*_+)\cdot \frac{c}{1-\delta^2} + [1-p(DC_+|D*_+)]\cdot \ell\right] \\
&= V_{cd}(c_0;p(C_+)) + \delta^2 p(D_+)[1-p(d_0|D_+)]\cdot V_{cd}(c_0;p(DC_+|D*_+)).
\end{aligned}
$$

Similarly,

$$
V(C_{d_0};p) = V_{cd}(c_0;p(C_+)) + \delta^2 p(D_+)[1-p(d_0|D_+)]\cdot V_{cd}(d_0;p(DC_+|D*_+)).
$$

The expected payoff of a D-start one-period tolerant strategy has more terms because the play path diverges after (D,C) and (D,D) (see Table 2).

$$
\begin{aligned}
V(D_{c_0};p) &= V_{cd}(d_0;p(C_+)) + \delta^2 p(C_+)[1-p(c_0|C_+)]\cdot V_{cd}(c_0;p(CC_+|C*_+)) \\
&\quad + \delta^2 p(D_+)[1-p(d_0|D_+)]\cdot V_{cd}(c_0;p(DC_+|D*_+); \\
V(D_{d_0};p) &= V_{cd}(d_0;p(C_+)) + \delta^2 p(C_+)[1-p(c_0|C_+)]\cdot V_{cd}(d_0;p(CC_+|C*_+)) \\
&\quad + \delta^2 p(D_+)[1-p(d_0|D_+)]\cdot V_{cd}(d_0;p(DC_+|D*_+)).
\end{aligned}
$$

To simplify all the above payoff formulas into one general formula, for each $X \in \{C, D\}$ (the first period action by the relevant player), define

$$
\rho_p(Z_1;X) = \begin{cases} p(C_+)[1-p(c_0|C_+)]\cdot 1_{X=D} & \text{if } Z_1 = C \\ p(D_+)[1-p(d_0|D_+)] & \text{if } Z_1 = D, \end{cases}
$$

where $1_{X=D}$ is the indicator function which takes the value 1 if $X = D$ and 0 if $X = C$. Also, we simplify the conditional probability as

$$
\pi_p(XY_+) := p(XY_+|X*_+).
$$

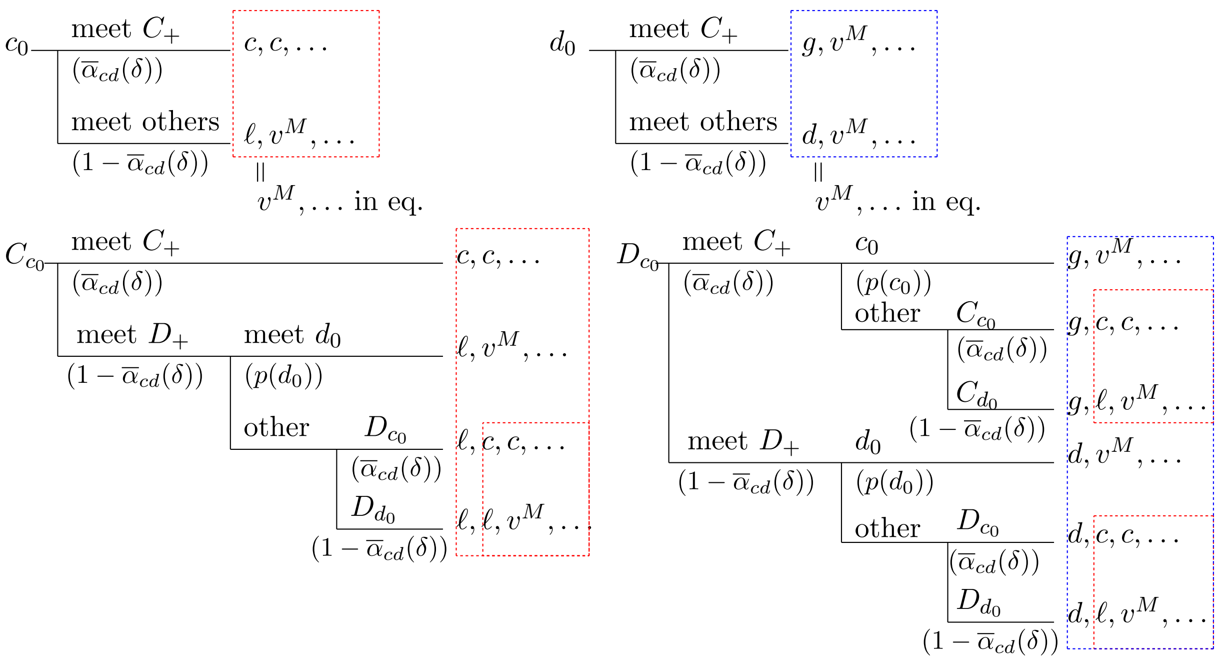
Then, for any one-period tolerant strategy X_{y_0}, its expected payoff facing the matching pool distribution p is written as simply as

$$
V(X_{y_0};p) = V_{cd}(x_0;p(C_+)) + \delta^2 \sum_{Z_1=C,D} \rho_p(Z_1;X)V_{cd}(y_0;\pi_p(Z_1C_+)), \tag{6}
$$

where $x_0 = c_0$ (resp. d_0) if $X = C$ (resp. D). Analogously, the expected length of the partnerships that a one-period tolerant player experiences can be written as

$$
L(X_{y_0};p) = L_{cd}(x_0;p(C_+)) + \delta^2 \sum_{Z_1=C,D} \rho_p(Z_1;X)L_{cd}(y_0;\pi_p(Z_1C_+)).
$$

Figure 5: Payoff-Equivalence of c_0-d_0 and some strategies in $\bar{p} \in \bar{P}_1$



Therefore, the average payoff is

$$\begin{aligned} v(X_{y_0};p) &= \frac{V_{cd}(x_0;p(C_+)) + \delta^2 \sum_{Z_1=C,D} \rho_p(Z_1;X) V_{cd}(y_0;\pi_p(Z_1C_+))}{L_{cd}(x_0;p(C_+)) + \delta^2 \sum_{Z_1=C,D} \rho_p(Z_1;X) L_{cd}(y_0;\pi_p(Z_1C_+))} \\ &= \frac{\frac{V_{cd}(x_0;p(C_+))}{L_{cd}(x_0;p(C_+))} + \delta^2 \sum_{Z_1=C,D} \rho_p(Z_1;X) \frac{L_{cd}(y_0;\pi_p(Z_1C_+))}{L_{cd}(x_0;p(C_+))} v_{cd}(y_0;\pi_p(Z_1C_+))}{1 + \delta^2 \sum_{Z_1=C,D} \rho_p(Z_1;X) \frac{L_{cd}(y_0;\pi_p(Z_1C_+))}{L_{cd}(x_0;p(C_+))}} \\ &= v_{cd}(x_0;p(C_+)) \\ &\quad + \delta^2 \sum_{Z_1=C,D} \rho_p(Z_1;X) \frac{L_{cd}(y_0;\pi_p(Z_1C_+))}{L(X_{y_0};p)} \Big[v_{cd}(y_0;\pi_p(Z_1C_+)) - v_{cd}(x_0;p(C_+)) \Big]. \end{aligned} \quad (7)$$

3.2 Nash equilibrium and the two-period richness result

Remark 1. *Take any $\delta \in (\underline{\delta}_{c_0 d_0}, 1)$. Any $p \in \Delta(\tilde{S}_0^1)$ such that $p(C_+) = \pi_p(CC_+) = \pi_p(DC_+) = \overline{\alpha}_{cd}(\delta)$ is a payoff-equivalent Nash equilibrium to $\overline{\alpha}_{cd}(\delta) \cdot c_0 + \{1 - \overline{\alpha}_{cd}(\delta)\} d_0$.*

Proof. Because $p(C_+) = \overline{\alpha}_{cd}(\delta)$, $v_{cd}(c_0;\overline{\alpha}_{cd}(\delta)) = v_{cd}(d_0;\overline{\alpha}_{cd}(\delta))$. In view of (7), $v(X_{y_0};p) = v_{cd}(c_0;\overline{\alpha}_{cd}(\delta))$ or $v_{cd}(d_0;\overline{\alpha}_{cd}(\delta))$ as well (because the second component is 0). Lemma 1 then implies that the common payoff satisfies the Best Reply Condition so that p is a Nash equilibrium. □

Remark 1 shows a continuum of Nash equilibria because the shares $p(c_0)$ and $p(d_0)$ are arbitrary. Only the relative shares of the classes of the strategies of the form Z_1C_+ (which are $C_+ = \{c_0, C_{c_0}, C_{d_0}\}$, $CC_+ = \{C_{c_0}\}$, and $DC_+ = \{D_{c_0}\}$ in $\tilde{S}_0^1$) matter. See Figure 5 for an intuition of the payoff equivalence.

Although Remark 1 encompasses the two-strategy distribution of c_0 and d_0 (when $p(c_0) = p(d_0) = 1$), the four strategy distribution of the one-perion tolerant strategies only (when $p(c_0) = p(d_0) = 0$), and a continuum of six strategy distributions, we focus on the set of Nash equilibria within $\tilde{S}_0^1$ such that all strategies are present:

$$\bar{P}_1 := \left\{\bar{p} \in \Delta(\tilde{S}_0^1) \mid \bar{p}(C_+) = \pi_{\bar{p}}(CC_+) = \pi_{\bar{p}}(DC_+) = \overline{\alpha}_{cd}(\delta), \;\; \bar{p}(c_0|C_+),\; \bar{p}(d_0|D_+) \in (0,1)\right\}.$$

The subscript 1 indicates that the distributions have up to one-period tolerance. Any strategy distribution in $\bar{P}_1$ generates all action profile sequences for two periods (with the condition that if (C,C) is established, the partners keep the partnership and play C again), as in Figure 1.

To state this formally, for any $n = 1, 2, \ldots$, let the set of n-period action profile sequences with the **cooperative property** as follows.

$$A_{co}^n := \{((x_1, x_1'), \ldots, (x_n, x_n')) \in [\{C, D\}^2]^n \mid$$
$$(x_m, x_m') = (C, C) \Rightarrow (x_{m+1}, x_{m+1}') = (C, C)\}.$$

For any pair of pure strategies (s, s'), let the on-path action profile sequence generated by the pair be

$$\mathbf{x}(s, s') = ((x_1(s, s'), x_1'(s, s')), \ldots, (x_{T(s,s')}(s, s'), x_{T(s,s')}'(s, s'))) \in [\{C, D\}^2]^{T(s,s')},$$

where $T(s, s')$ is the planned length of the partnership (s, s'). (However, the path continues with probability δ^2 each period.)

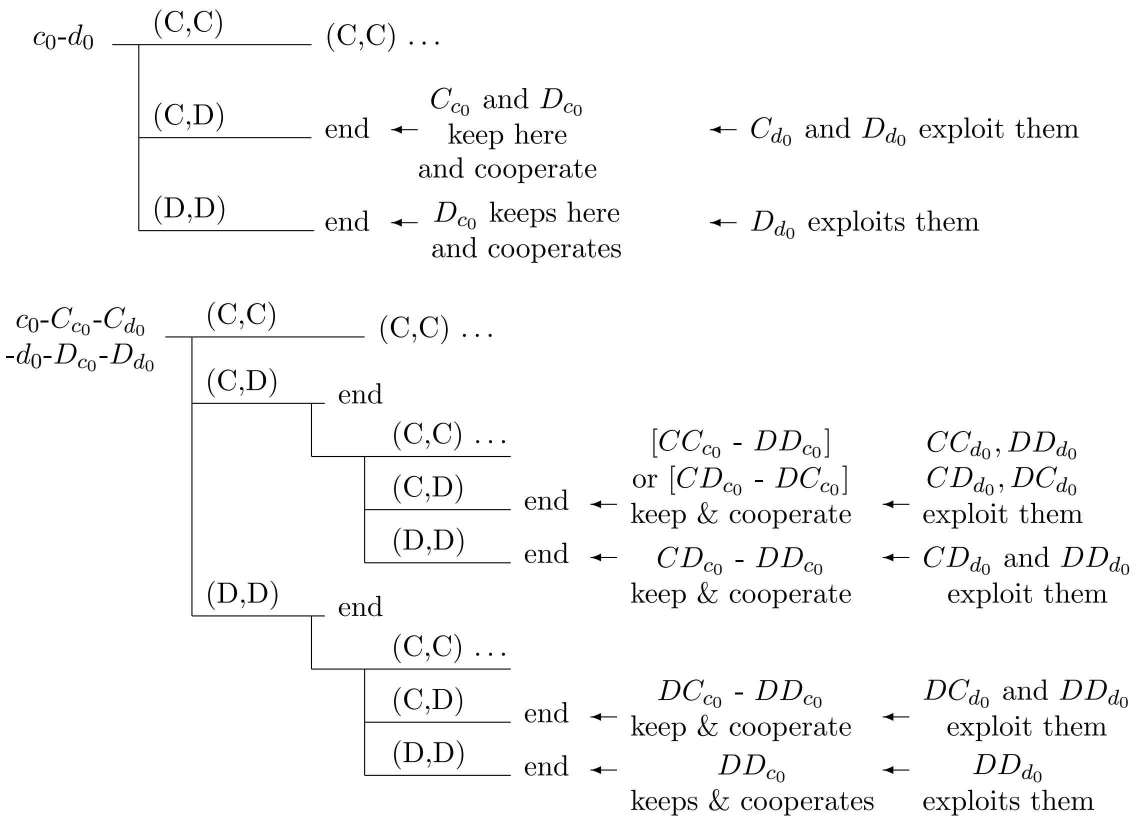
Remark 2. *For any $\delta \in (\underline{\delta}_{c_0 d_0}, 1)$, there exists a class of Nash equilibria $\bar{P}_1$ such that, for any $((x_1, x_1'), (x_2, x_2')) \in A_{co}^2$ and any $\bar{p} \in \bar{P}_1$, there are $s, s' \in supp(\bar{p})$ such that $((x_1, x_1'), (x_2, x_2'))$ is the first two period action profiles in $\mathbf{x}(s, s')$.*

3.3 Evolutionary stability

Since the stage game is an extensive form game, it is desirable to investigate whether the above strategy combinations satisfy a stronger stability than Nash equilibrium. As we discussed in the Introduction, we consider evolutionary stability as a refinement, to avoid excessive assumptions about each player's belief formation after each private history.

First, notice that any strategy that do not create a new play path by entering the population with a positive measure cannot overtake the population. Hence we exclude them from consideration. Moreover, any strategy that defects after establishing (C,C) on the play path would do strictly

Figure 6: Secret-Handshake and Exploitation



worse than the strategy continuing (C,C) thereafter, because any $\bar{p} \in \bar{P}_1$ satisfies the Best Reply Condition with the strict inequality. Thus we do not consider those strategies, either.

Therefore relevant mutant strategies are the ones that do "secret handshake" by keeping the partnership when incumbents end it and cooperating among them, and the ones that exploit such secret-handshakers, such as C_{c_0} and C_{d_0} (resp. D_{c_0} and D_{d_0}) towards the c_0-strategy (resp. the d_0-strategy). The secret-handshake and exploitation can take a further hierarchical structure. (See Figure 6.) To keep the example simple, let us only add the following two-period tolerant strategies into consideration.

Definition 10. *For any* $\mathbf{X} = (X_1, X_2) \in \{C, D\}^2$ *and any* $y_0 \in \{c_0, d_0\}$*, the two-period tolerant* $\mathbf{X}_{y_0}$*-strategy is a strategy such that*

$t = 1$*: play* $X_1 \in \{C, D\}$ *and keep for any observation:*

$t = 2$*: if (C,C) is observed in the previous period, play the* c_0*-strategy as the continuation strategy, and otherwise play* $X_2 \in \{C, D\}$ *and keep for any observation:*

$t \geqq 3$*: if (C,C) is observed in the previous period, play the* c_0*-strategy as the continuation strategy, and otherwise play the* y_0*-strategy as the continuation strategy.*

Note that if X_t is not C, then the condition "if (C,C) is observed ..." is to be ignored. All two-period tolerant strategies are either a secret-handshaker of some shorter-period tolerant strategy or an exploiter of a cooperative two-period tolerant strategy. For example, the CD_{c_0}-strategy is a secret-handshake version of the C_{d_0}-strategy, and the CD_{d_0}-strategy is an exploiter of the CD_{d_0}-strategy. Therefore, they are "effective" mutant strategies which create new play paths in the population and may earn higher payoffs than that of the zero- and one-period tolerant strategies.

Let the set of $0, 1, 2$-period tolerant strategies be

$$\tilde{S}_0^2 = \{\mathbf{X}_{y_0} \in S \mid \mathbf{X} \in \{\emptyset, C, D\} \cup \{C, D\}^2, \quad y_0 \in \{c_0, d_0\}\}.$$

Again, we sort the strategies with respect to the initial action sequences. For example, in $\tilde{S}_0^2$, the strategies which play C in $t = 1$ and $t = 2$ for sure are $CC_+ = \{C_{c_0}, CC_{c_0}, CC_{d_0}\}$. (Note that the C_{c_0}-strategy also plays C twice for sure with any partner.) In general, define $XY_+ :=$ $\{X_{y_0}, XY_{c_0}, XY_{d_0}\}$ and $XYZ_+ := \{XY_{z_0}\}$, for any $X, Y, Z \in \{C, D\}$ with the convention that $y_0 = c_0$ (resp. d_0) if $Y = C$ (resp. D) and $z_0 = c_0$ (resp. d_0) if $Z = C$ (resp. $Z = D$). Let also $XY{*}_+ = XYC_+ \cup XYD_+$. We also simplify the notation of the conditional probabilities. For any $k = 1, 2, \ldots$ and any initial k action sequence of $Z_1 \cdots Z_k \in \{C, D\}^k$, the conditional probability that the partner has a strategy in the set $Z_1 \cdots Z_k C_+$ after the partner's action sequence of $(Z_1, \ldots, Z_k)$ is

$$\pi_p(Z1 \cdots Z_k C_+) := p(Z_1 \cdots Z_k C_+ | Z_1 \cdots Z_k {*}_+).$$

The relative-share form of a mutant distribution $q \in \Delta(\tilde{S}_0^2)$ can be inductively written as follows.

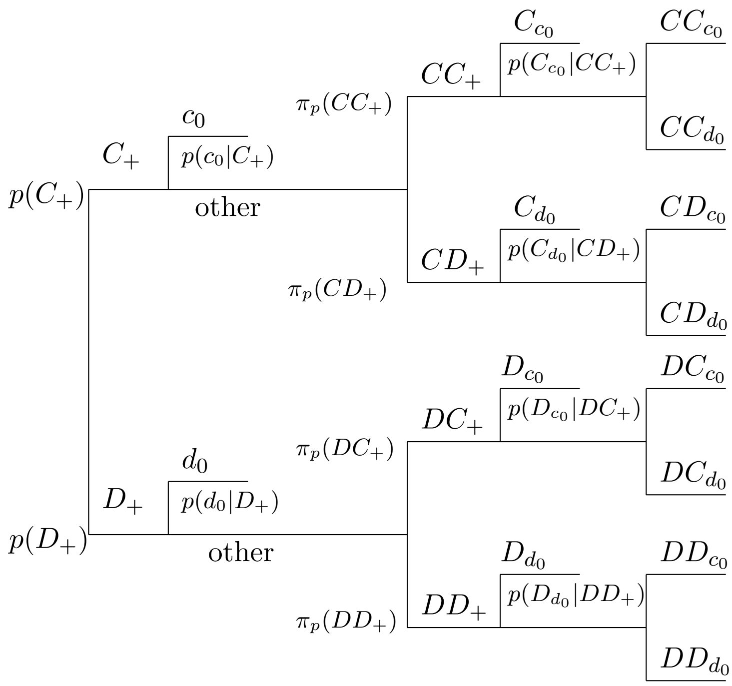
$$\begin{aligned} q = q(C_+)\Big[&q(c_0 \mid C_+) \cdot c_0 \\ &+ \{1 - q(c_0 \mid C_+)\}\big\{\pi_q(CC_+) \cdot q\,|_{CC} + \pi_q(CD_+) \cdot q\,|_{CD}\big\}\Big] \\ + q(D_+)\Big[&q(d_0 \mid D_+) \cdot d_0 \\ &+ \{1 - q(d_0 \mid D_+)\}\big\{\pi_q(DC_+) \cdot q\,|_{DC} + \pi_q(DD_+) \cdot q\,|_{DD}\big\}\Big], \end{aligned} \tag{8}$$

where $q\,|_{XY}$ is a distribution of tolerant strategies starting with the initial action sequence XY;

$$\begin{aligned} q\,|_{XY} = &q(X_{y_0} \mid X{*}_+) \cdot X_{y_0} \\ &+ \{1 - q(X_{y_0} \mid X{*}_+)\}\Big\{\pi_q(XYC_+) \cdot XY_{c_0} + \pi_q(XYD_+) \cdot XY_{d_0}\Big\}. \end{aligned}$$

This inductive definition (see Figure 7 for an intuition) will be useful when we consider arbitrary periods of tolerance later.

Figure 7: Relative share structure of two-period tolerant strategy distributions



To formulate the average payoff of a two-period tolerant strategy $\mathbf{X}_{y_0}$ (facing a stationary distribution p in the matching pool) with the initial action sequence $\mathbf{X} = (X_1 X_2)$, we introduce one more notation. For any observation of the partner's action history Z_1, Z_2, and your own action X_2 in $t = 2$, we define the probability that the payoff sequence branches out as follows.

$$\rho_p(Z_1 Z_2; X_2) = \begin{cases} p(Z_1 C_+)\{1 - p(Z_{1c_0} \mid Z_1 C_+)\} \cdot 1_{X_2 = D} & \text{if } Z_2 = C \\ p(Z_1 D_+)\{1 - p(Z_{1d_0} \mid Z_1 D_+)\} & \text{if } Z_2 = D. \end{cases}$$

Then we can write the expected payoff of a two-period tolerant strategy $\mathbf{X}_{y_0}$ facing a stationary distribution p in the matching pool as

$$\begin{aligned} V(\mathbf{X}_{y_0}; p) = {} & V_{cd}(x_{10}; p(C_+)) \\ & + \delta^2 \sum_{Z_1 = C,D} \rho_p(Z_1; X_1) V_{cd}(x_{20}; \pi_p(Z_1 C_+)) \\ & + \delta^4 \sum_{Z_1 = C,D} \rho_p(Z_1; X_1) \sum_{Z_2 = C,D} \rho_p(Z_1 Z_2; X_2) V_{cd}(y_0; \pi_p(Z_1 Z_2 C_+)) \end{aligned}$$

with the convention that $x_{t0} = c_0$ (resp. d_0) if $X_t = C$ (resp. D). To compare with the one-period tolerant strategy X_{y_0}, consider a two-period tolerant strategy XY_{w_0} (where XY is the initial action sequence). When the stationary strategy distribution in the matching pool is p, its average payoff is written as follows. (See also Figure 7.)

$$
\begin{aligned}
v(XY_{w_0};p) = {} & v_{cd}(x_0;p(C_+)) \\
& + \delta^2 \sum_{Z_1=C,D} \rho_p(Z_1;X) \frac{L_{cd}(y_0;\pi_p(Z_1C_+))}{L(XY_{w_0};p)} [v_{cd}(y_0;\pi_p(Z_1C_+)) - v_{cd}(x_0;p(C_+))] \\
& + \delta^4 \sum_{Z_1=C,D} \rho_p(Z_1;X) \sum_{Z_2=C,D} \rho_p(Z_1Z_2;Y) \\
& \qquad \cdot \frac{L_{cd}(w_0;\pi_p(Z_1Z_2C_+))}{L(XY_{w_0};p)} [v_{cd}(w_0;\pi_p(Z_1Z_2C_+)) - v_{cd}(x_0;p(C_+))]. \qquad (9)
\end{aligned}
$$

To explain (9), the first component $v_{cd}(x_0;p(C_+))$ is the payoff of playing the initial action X. The second component is the sum of extra payoffs by playing Y in $t = 2$ after the first period histories (X,C) and (X,D), not counted in $v_{cd}(x_0;p(C_+))$. If $X = C$, the payoff after the history (X,C) is already included in $v_{cd}(x_0;p(C_+))$, and hence it is not present in this component. The third component is the sum of extra payoffs after four possible histories over the two periods, $((X,Z_1),(Y,Z_2))$, not counted in $v_{cd}(x_0;p(C_+))$ and $v_{cd}(y_0;\pi_p(Z_1C_+))$.

Note that the formula (9) embeds the average payoff of a one-period tolerant strategy (7) in the first two components. Hence it is easy to compare the average payoffs of X_{y_0}-strategy and XY_{w_0}-strategy. When the incumbent distribution is $\bar{p} \in \bar{P}_1$ and mutant distributions have up to two-period tolerant strategies in the support, in view of the definition of WESD, it suffices to show that a two-period tolerant strategy XY_{w_0} has less post-entry payoff than that of the one-period tolerant strategy X_{y_0}.

Remark 3. Take any $\delta \in (\underline{\delta}_{c_0d_0}, 1)$. Let

$$M_2(\delta) := \left\{ q \in \Delta(\tilde{S}_0^2) \mid \ q(Z_1Z_2C_+|Z_1Z_2*_+) < \overline{\alpha}_{cd}(\delta), \ \forall Z_1, Z_2 \in \{C,D\} \right\}.$$

Fix any $\bar{p} \in \bar{P}_1$ and any mutant distribution $q \in M_2(\delta)$. For any $\epsilon \in (0,1)$, let the post-entry distribution be $p^{PE} = (1-\epsilon)\cdot\bar{p} + \epsilon \cdot q$.

(i) For any $Z_1, Z_2 \in \{C,D\}$, there exists $\overline{\epsilon}_1 \in (0,1)$ such that, for any $\epsilon \in (0,\overline{\epsilon}_1)$,

$$p^{PE}(Z_1Z_2C_+|Z_1Z_2*_+) < p^{PE}(C_+), \text{ or } \ \pi_{p^{PE}}(Z_1Z_2C_+) < p^{PE}(C_+).$$

(ii) For any $Z_1, Z_2 \in \{C,D\}$, there exists $\overline{\epsilon}_2 \in (0,1)$ such that, for any $\epsilon \in (0,\overline{\epsilon}_2)$,

$$v_{cd}(c_0;\pi_{p^{PE}}(Z_1Z_2C_+)) < v_{cd}(d_0;p^{PE}(C_+)).$$

(iiI) For each $XY_{w_0} \in \tilde{S}_0^2$, there exists $X_{y_0} \in \tilde{S}_0^1$ and $\overline{\epsilon} \in (0,1)$ such that for any $\epsilon \in (0,\overline{\epsilon})$,

$$v(XY_{w_0};(1-\epsilon)\cdot\bar{p} + \epsilon\cdot q) < v(X_{y_0};(1-\epsilon)\cdot\bar{p}+\epsilon\cdot q).$$

Hence $\bar{p}$ is a WESD($M_2(\delta)$)).

Proof. See Appendix.

The key of the proof is that post-entry shares of two-period tolerant strategies with the c_0-strategy in $t = 3$ are the same as the mutants', so that **they do not depend on** ϵ, that is

$$p^{PE}(Z_1Z_2C_+|Z_1Z_2*_+) = q(Z_1Z_2C_+|Z_1Z_2*_+), \text{ or } \pi_{p^{PE}}(Z_1Z_2C_+) = \pi_q(Z_1Z_2C_+), \quad (10)$$

because the class of $Z_1Z_2*_+$ contains only mutants. Since $q \in M_2(\delta)$, we have that

$$p^{PE}(Z_1Z_2C_+|Z_1Z_2*_+) < \overline{\alpha}_{cd}(\delta),$$

while the post-entry share of the C-start class converges to the C-D ratio as $\epsilon \to 0$, i.e.,

$$\lim_{\epsilon \to 0} p^{PE}(C_+) = \overline{\alpha}_{cd}(\delta).$$

This implies (i). Then, for any $\epsilon \in (0, \overline{\epsilon}_1)$,

$$v_{cd}(c_0; \pi_{p^{PE}}(Z_1Z_2C_+)) < v_{cd}(c_0; p^{PE}(C_+)),$$
$$v_{cd}(d_0; \pi_{p^{PE}}(Z_1Z_2C_+)) < v_{cd}(d_0; p^{PE}(C_+)).$$

Local stability and $\pi_{p^{PE}}(Z_1Z_2C_+) < \overline{\alpha}_{cd}(\delta)$ imply also that

$$v_{cd}(d_0; \pi_{p^{PE}}(Z_1Z_2C_+)) < v_{cd}(c_0; \pi_1^{PE}(c_0)).$$

Hence it remains to prove (ii) in order to show that the third component of (9) is negative (which implies (iii)).

To prove (ii), we use the fact that $\pi_{p^{PE}}(Z_1Z_2C_+)$ does not depend on ϵ again. Because $\pi_{p^{PE}}(Z_1Z_2C_+) < \overline{\alpha}_{cd}(\delta)$, there exists $\hat{\alpha}$ such that $\pi_{p^{PE}}(Z_1Z_2C_+) < \hat{\alpha} < \overline{\alpha}_{cd}(\delta)$ and

$$v_{cd}(d_0; \pi_{p^{PE}}(Z_1Z_2C_+)) = v_{cd}(c_0; \hat{\alpha}).$$

This $\hat{\alpha}$ is also independent of ϵ. Hence for sufficiently small ϵ, (ii) holds. (See Figure 8 in the Appendix.) Therefore the third component of (9) is negative for all $\epsilon < \min\{\overline{\epsilon}_1, \overline{\epsilon}_2\} =: \overline{\epsilon}$.

An interpretation of $M_2(\delta)$ is that the mutant distributions are "sufficiently random" so that the weight on $\{Z_1Z_{2c_0}\}$ is not so high (to reach the C-D ratio) within $\{Z_1Z_{2c_0}\} \cup \{Z_1Z_{2d_0}\}$, for any $Z_1, Z_2 \in \{C, D\}$. In other words, if mutant distributions are sufficiently spread-out among all tolerant strategies, they belong to $M_2(\delta)$.

We cannot strengthen the stability, or enlarge M, because the boundary of $M_2(\delta)$ includes mutant distributions consisting only of two-period tolerant strategies and having exactly $\pi_q(Z_1 Z_2 C_+) = \overline{\alpha}_{cd}(\delta)$ for all $Z_1, Z_2 \in \{C, D\}$. They have the same post-entry payoff to any strategy in $\bar{p} \in \bar{P}_1$ so that $\bar{p}$ is no longer a WESD.

4 Diverse Equilibrium Behaviors

4.1 Tolerant Equilibria

To generalize, we consider k-period tolerant strategies which play a certain sequence of actions and keep the partnership, until either (C,C) is established (then play the c_0-strategy thereafter) or the k-period is over (then play one of the c_0- or the d_0-strategy).

Definition 11. *For any $k = 1, 2, \ldots$, any k-sequence of Prisoner's Dilemma actions $\mathbf{X} \in \{C, D\}^k$, and any $y_0 \in \{c_0, d_0\}$, the k-period tolerant*[15] *$\mathbf{X}_{y_0}$-strategy is a strategy such that*
$t = 1$: play $X_1 \in \{C, D\}$ and keep for any action combination;
$2 \leqq t \leqq k$: if (C,C) is observed in the previous period, play the c_0-strategy as the continuation strategy, and otherwise play $X_t \in \{C, D\}$ and keep for any observation;
$t = k + 1$: if (C,C) is observed in the previous period, play the c_0-strategy as the continuation strategy, and otherwise play the y_0-strategy as the continuation strategy.

Recall that we also interpret the c_0- and the d_0-strategy as the degenerate (0-period) tolerant strategies. The set of m- to n-period tolerant strategies ($m < n$) is

$$\tilde{S}_m^n = \{\mathbf{X}_{y_0} \in S \mid \mathbf{X} \in \{C, D\}^k,\ k = m, m+1, \ldots, n,\ y_0 \in \{c_0, d_0\}\},$$

and the set of all tolerant strategies is denoted as $\tilde{S}_0^\infty$.

The set $\tilde{S}_0^\infty$ is constructed to cover any initial action sequence for arbitrary finite periods, and, for each element $\mathbf{X}_{y_0} \in \tilde{S}_0^\infty$, there is a secret-handshake strategy $\mathbf{X}Y_{c_0} \in \tilde{S}_0^\infty$ (where $Y = C$ if $y_0 = c_0$ and $Y = D$ if $y_0 = d_0$), and an exploiter $\mathbf{X}Y_{d_0} \in \tilde{S}_0^\infty$ of the secret-handshake strategy as well. Both of these are effective mutant strategies, just like the D_{c_0}-strategy is an effective mutant towards the d_0-strategy but it can be exploited by the D_{d_0}-strategy. In other words, the set $\tilde{S}_0^\infty$ is constructed by adding (one-more-period tolerant) strategies which add a new play path in the society, when they enter by a positive measure. Hence all elements (with $k > 0$) in $\tilde{S}_0^\infty$

[15] In particular, if $\mathbf{X} = (D, \ldots, D)$ and $y_0 = c_0$, it is a tolerant version of k-period **trust-building strategy** in Greve-Okuno.

have different average payoff formula (see the proof of Proposition 1) and are non-trivial mutant strategies against some strategy in $\tilde{S}_0^\infty$. Moreover, the set $\tilde{S}_0^\infty$ does not contain strategies such that, even if they enter by a positive measure, no new play path is generated in the society. Such strategies can earn the same payoff as the strategies in $\tilde{S}_0^\infty$ but never overtake the population.

We use an inductive form to describe any strategy distribution $p \in \Delta(\tilde{S}_0^\infty)$ as follows.

$$\begin{aligned} p = p(C_+)&\Big[p(c_0|C_+)\cdot c_0 + [1-p(c_0|C_+)]\Big\{\pi_p(CC_+)\cdot p\,|_{CC} + \pi_p(CD_+)\cdot p\,|_{CD}\,(\pi)\Big\}\Big] \\ + p(D_+)&\ \Big[p(d_0|D_+)\cdot d_0 + [1-p(d_0|D_+)]\Big\{\pi_p(DC_+)\cdot p\,|_{DC} + \pi_p(DD_+)\cdot p\,|_{DD}\Big\}\Big], \end{aligned} \tag{11}$$

where $p\,|_{XY}$ is a distribution starting with a one-period tolerant strategy X_{y_0} such that

$$\begin{aligned} p\,|_{XY} = p(X_{y_0}|X{*_+})\cdot X_{y_0} &+ \{1-p(X_{y_0}|X{*_+})\} \\ &\cdot\Big\{\pi_p(XYC_+)\cdot p\,|_{XYC} + \pi_p(XYD_+)\cdot p\,|_{XYD}\,(\pi)\Big\}, \end{aligned}$$

and $p\,|_{XYZ}$ is a distribution starting with a two-period tolerant strategy XY_{z_0} of the same form

$$\begin{aligned} p\,|_{XYZ} = p(XY_{z_0}|XY{*_+})\cdot XY_{z_0} &+ \{1-p(XY_{z_0}|XY{*_+})\} \\ &\cdot\Big\{\pi_p(XYZC_+)\cdot p\,|_{XYZC} + \pi_p(XYZD_+)\cdot p\,|_{XYZD}\Big\}, \end{aligned}$$

and so on. (Recall Figure 7.) This formulation makes payoff decomposition of each strategy and payoff comparison among different strategies easy.

By a generalization of Remark 1, for each $T = 1, 2, \ldots$, any $p \in \Delta(\tilde{S}_0^T)$ such that

$$p(C_+) = \overline{\alpha}_{cd}(\delta) \text{ and } \pi_p(\mathbf{X}C_+) = \overline{\alpha}_{cd}(\delta),\ \forall \mathbf{X} \in \{C, D\}^k,\ k = 1, \ldots, T$$

is a payoff-equivalent Nash equilibrium to the fundamentally asymmetric equilibrium $\overline{\alpha}_{cd}(\delta)\cdot c_0 + \{1-\overline{\alpha}_{cd}(\delta)\}d_0$. (See the proof of Proposition 1.) In particular, the class that all T-period tolerant strategies are present generates all action combination sequences over $T+1$ periods:

$$\begin{aligned} \bar{P}_T := \Big\{\bar{p} \in \Delta(\tilde{S}_0^T) \mid \bar{p}(C_+) = \overline{\alpha}_{cd}(\delta),\ \pi_{\bar{p}}(\mathbf{X}C_+) = \overline{\alpha}_{cd}(\delta), \\ 0 < \bar{p}(\mathbf{X}_{y_0}|\mathbf{X}{*_+}) < 1, \forall \mathbf{X} \in \{C, D\}^k,\ \forall k = 1, 2, \ldots, T,\ \forall y_0 \in \{c_0, d_0\}\Big\}. \end{aligned}$$

We call this set the **diverse-behavior equilibrium class** of T-periods (DBEC-T) generated by the c_0-d_0 equilibrium.

Proposition 1. (i) For any $\delta \in (\underline{\delta}_{c_0 d_0}, 1)$, any $T < \infty$, and any $\bar{p} \in \bar{P}_T$, $\bar{p}$ is a Nash equilibrium and, for any $s \in supp(\bar{p})$,

$$v(s; \bar{p}) = v(c_0; \overline{\alpha}_{cd}(\delta)c_0 + \{1-\overline{\alpha}_{cd}(\delta)\}\cdot d_0) = v(d_0; \overline{\alpha}_{cd}(\delta)c_0 + \{1-\overline{\alpha}_{cd}(\delta)\}\cdot d_0).$$

(ii) For any $\delta \in (\underline{\delta}_{c_0 d_0}, 1)$, define

$$M_\infty(\delta) := \Big\{ q \in \Delta(\tilde{S}_0^\infty) \mid \pi_q(\mathbf{X}C_+) < \overline{\alpha}_{cd}(\delta), \quad \forall \mathbf{X} \in \{C, D\}^k, \ k = 1, 2, \ldots, \Big\}.$$

Then for any $\delta \in (\underline{\delta}_{c_0 d_0}, 1)$ and any $T < \infty$, each $\bar{p} \in \bar{P}_T$ is a WESD($M_\infty(\delta)$) and generates all $T+1$-period action combination sequences in A_{co}^{T+1}.

Proof. See Appendix.

The proof of Proposition 1 (ii) is a generalization of that of Remark 3. It shows that for any mutant $T+k$-period tolerant strategy (with $k = 1, 2, \ldots$), there is a T-period tolerant strategy in $\bar{P}_T$, which plays the same action sequence for the first T periods, shifts to the c_0- (if $X_{T+1} = C$) or the d_0-strategy (if $X_{T+1} = D$), and outperforms the mutant.

Proposition 1 encompasses countably many DBEC-T's, where each DBEC-T is a continuum of WESD($M_\infty(\delta)$)'s, since $\bar{p}(c_0), \bar{p}(d_0), \bar{p}(X_{y_0} \mid X*_+), \ldots$ can vary arbitrarily. Overall, we have shown that it is weakly evolutionarily stable that (i) a lot of pairs eventually achieve long-term cooperation until a random death, (ii) the period it takes to establish long-term cooperation varies from 1 to any finite number, and (iii) in the meantime any action combination sequence can be played. Moreover, the WESD's are all payoff-equivalent to each other as well as to the c_0-d_0 equilibrium.

The set $M_\infty(\delta)$ of "sufficiently random" mutant distributions has a similar structure as the set M_2 for the two-period example, but the property that the mutant distributions should not have too high weights on $\mathbf{X}_{c_0}$-strategies is weaker, since there are always longer-period tolerant strategies in $\tilde{S}_0^\infty$. In a large society, it is plausible that mutant strategies are not concentrated on particular type of strategies. The set $M_\infty(\delta)$ includes distributions that put a positive probability on all tolerant strategies. In a finite population, coordination game context like Kandori et al. (1993), such a random mutation leads eventual coordination. In our context, the random mutation makes coordination difficult among secret-handshakers by the presence of exploiters, and thus mutants are "self-defeated".

As δ increases, $\overline{\alpha}_{cd}(\delta)$ increases and $M_\infty(\delta)$ becomes larger. This means that with higher survival rates, any equilibrium $\bar{p}$ in the DBEC's is more stable because even relatively coordinated mutant distributions (those with high weights on $\mathbf{X}_{c_0}$-strategies) cannot de-stabilize the incumbent distribution. The logic is that higher $\overline{\alpha}_{cd}(\delta)$ increases the payoff of the "base" strategies, c_0 and d_0, which makes it easy for them to outperform mutants.

One might wonder if WESD concept is too weak. To show that the concept of WESD selects among Nash equilibria, denote a k-period tolerant $\mathbf{X}_{d_0}$-strategy in which only D is played in the first k periods by $D^k_{d_0}$ $(k = 0, 1, \ldots)$. We can also consider the $D^\infty_{d_0}$-strategy, which never initiates termination of a match and keeps playing D. Each of these strategies played by all players trivially constitutes a Nash equilibrium for any $\delta \in (0, 1)$.

Remark 4. For any $\delta \in (0, 1)$,

(i) for each $k < \infty$, the monomorphic Nash equilibrium consisting of the $D^k_{d_0}$-strategy is not a WESD($\{q \in \Delta(\{D^{k+1}_{c_0}, D^{k+1}_{d_0}\}) \mid q(D^{k+1}_{c_0}) > 0\}$); and

(ii) the monomorphic Nash equilibrium of the $D^\infty_{d_0}$-strategy is not a WESD($\{q \in \Delta(\{c_0, d_0\}) \mid q(c_0) > 0\}$).

Proof. See Appendix.

Hence always-D type Nash equilibria are not stable even in a very weak form of WESD with only two-strategy mutant distributions are allowed to emerge in the population.

Finally, recall that all equilibria in $\bar{P}_T$ for any T are payoff-equivalent to the c_0-d_0 equilibrium. GOS showed that, under some payoff condition, the c_0-d_0 equilibrium is more efficient than any Nash equilibria consisting of "trust-building" strategies (which are essentially $D^k_{c_0}$-strategies[16]).

Remark 5. (GOS, Proposition 4) For any Prisoner's Dilemma such that $g - c < (c - d)^2/(c - \ell)$, there exists $\hat{\delta} \in [\underline{\delta}_{c_0 d_0}, 1)$ such that for any $\delta \in (\hat{\delta}, 1)$, any Nash equilibrium distribution $q \in \Delta(\{D^k_{c_0} \mid k = 1, 2, \ldots\})$, any $s \in \{c_0, d_0\}$ and any $s' \in supp(q)$, $v_{cd}(s; \overline{\alpha}_{cd}(\delta)) > v(s'; q)$.

Therefore, if the deviation gain $g - c$ is not so large in the Prisoner's Dilemma, any diverse-behavior equilibrium is also more efficient than all trust-building equilibria.

4.2 Other stability concepts for weaker richness result

The fundamentally asymmetric equilibrium of the c_0- and the d_0-strategy itself generates any action combination of (C,C), (C,D)/(D,C), and (D,D) every period in the society. Therefore in a weaker sense the equilibrium behaviors are rich. Since there are only two strategies in the fundamentally asymmetric equilibrium, it satisfies stronger evolutionary stabilities. (This subsection is also an extension of the stability analysis of GOS.)

[16]However, the original definition of the trust-building strategies in Greve-Okuno is "intolerant" in the sense that they keep the partnership if and only if (D,D) is played in the first k periods of a partnership. This intolerance makes a stability difference, see Concluding Remarks.

Proposition 2. For any $\delta \in (\underline{\delta}_{c_0 d_0}, 1)$, let

$$M(\delta) := \{q \in \Delta(\tilde{S}_0^{\infty}) \mid \sup_{\mathbf{X} \in \{C,D\}^k,\ k=1,2,\ldots} \pi_q(\mathbf{X}C_+) < \overline{\alpha}_{cd}(\delta)\}.$$

Then $p^* = \overline{\alpha}_{cd}(\delta) \cdot c_0 + \{1 - \overline{\alpha}_{cd}(\delta)\} d_0$ is SUW($M(\delta)$).

Proof. See Appendix.

Proposition 2 shows that there is a stable distribution with a uniform winner against any mutant, generated in a slightly smaller set $M(\delta)(\subsetneq M_\infty(\delta))$, such that (i') in any pair, if (C,C) is achieved once, then the partners continue to play (C,C) until a random death, and (ii') in any period, any action combination of the Prisoner's Dilemma is played by a positive fraction of the society. In the aggregate, the society has all action combinations in every period, in both Propositions 1 and 2.

If we require additionally that mutant distributions must include some non-degenerate tolerant strategy, which is not a strong requirement, the c_0-d_0 equilibrium is a Mean Stable Distribution.

Proposition 3. For any $\delta \in (\underline{\delta}_{c_0 d_0}, 1)$, let

$$M_+(\delta) := \big\{q \in \Delta(\tilde{S}_0^{\infty}) \mid \sup_{\mathbf{X} \in \{C,D\}^k,\ k=1,2,\ldots} \pi_q(\mathbf{X}C_+) < \overline{\alpha}_{cd}(\delta),$$
$$q(\mathbf{X}_{y_0}) > 0,\ \exists \mathbf{X}_{y_0} \in \tilde{S}_1^{\infty}\big\}.$$

Then $p^* = \overline{\alpha}_{cd}(\delta) \cdot c_0 + \{1 - \overline{\alpha}_{cd}(\delta)\} d_0$ is a MSD($M_+(\delta)$).

Proof. See Appendix.

Finally, if we exclude mutants using the C-start tolerant strategies, so that mutant strategies are only of the form $D^k_{w_0}$, then the c_0-d_0 equilibrium is a Strong-ESD for some set. Let

$$S_D = \{\mathbf{X}_{y_0} \in \tilde{S}_0^{\infty} \mid \mathbf{X} = D^k,\quad k = 0, 1, \ldots,\ w_0 \in \{c_0, d_0\}\}.$$

This set consists of c_0, d_0, trust-building strategies $D^k_{c_0}$ for $k = 1, 2, \ldots$, and their exploiters $D^k_{d_0}$. These were the focus of GOS.

Corollary 1. For any $\delta \in (\underline{\delta}_{c_0 d_0}, 1)$, let

$$D(\delta) := \{q \in \Delta(S_D) \mid q(C_+) \leqq \overline{\alpha}_{cd}(\delta),\ \sup_{k=1,2,\ldots} \pi_q(D^k C_+) < \overline{\alpha}_{cd}(\delta)\}.$$

Then $p^* = \overline{\alpha}_{cd}(\delta) \cdot c_0 + \{1 - \overline{\alpha}_{cd}(\delta)\} d_0$ is a Strong-ESD($D(\delta)$).

Proof. See Appendix.

In summary, the weaker richness result that any action combination of (C,C), (C,D)/(D,C), and (D,D) is played every period with a positive fraction in the society is sustained in a variety of stability notions. However, because all of such stabilities require a smaller mutant distribution set than that of the DBEC's, we cannot say that the c_0-d_0 equilibrium is more stable than the equilibria in DBEC's.

5 Concluding Remarks

We have shown that the C-D ratio generates an immense variety of payoff-equivalent equilibria, which admit all action combination histories over an arbitrary length of time, when mutant distributions are sufficiently spread-out. The sufficiently random mutant distributions, which do not concentrate on the class of strategies of the form $\mathbf{X}_{c_0}$, are plausible in a large, anonymous society we consider. Therefore, the VSRPD is rich in equilibrium modes of behavior. We also clarified different stabilities among the payoff-equivalent equilibria.

Diverse behaviors in a homogeneous population are not rare observations. Many markets and societies admit a variety of long-term behaviors. We have given a rationale such that some players are trying to coordinate at some point and they can tolerate mis-coordination for a while, and others are trying to exploit the former and tolerate mis-coordination as well. In our laboratory experiment (Okuno-Fujiwara et al., 2016), we observed that subjects tolerated mis-coordination often up to two periods.

Tolerance, not to terminate a partnership even if mis-coordination in the Prisoner's Dilemma happens, can be useful in another way. The original definition of trust-building strategies of k-periods in Greve-Okuno was to play D for $t \leqq k$ and keep if and only if (D,D) is observed, and then to shift to the c_0-strategy. For sufficiently large k, the monomorphic distribution of the k-period trust-building strategy is neutrally stable, when mutants are pure-strategy distributions in $\mathbf{S}$ (Greve-Okuno, Proposition 1). However, an "intolerant" trust-building strategy distribution can be invaded by two-tolerant-strategy mutant distributions. For example, the one-period intolerant trust-building strategy (denoted as c_1) can be neutrally stable among pure-strategy mutants (see Remark 2 of GOS) but is de-stabilized by any mutant distribution of the form $\beta \cdot C_{c_0} + (1-\beta)D_{c_0}$ with $\beta > 0$ in the sense that, for any $\epsilon \in (0,1)$,

$$v(c_1; (1-\epsilon)c_1 + \epsilon\{\beta \cdot C_{c_0} + (1-\beta)D_{c_0}\}) < v(D_{c_0}; (1-\epsilon)c_1 + \epsilon\{\beta \cdot C_{c_0} + (1-\beta)D_{c_0}\}).$$

This is because the intolerant c_1-strategy terminates a match with C_{c_0} to miss the opportunity to cooperate from the second period on.

Let us also mention other advantages of our equilibria. Clearly, players need no precise coordination to play the c_0-d_0 equilibrium. By contrast, a monomorphic trust-building equilibrium requires that the length of the initial trust-building periods should be common knowledge or a group convention, but the source of such knowledge or convention is unclear. The DBEC's have many modes of behavior inside, and thus we can also interpret that players are not very coordinated and yet reach an equilibrium.

There are may interesting future research directions. An important extension is a two population model of firms and workers, to make a closed model of efficiency wage theory (e.g., Okuno, 1981[17] and Shapiro and Stiglitz, 1984). If there is an equilibrium with the fundamentally asymmetric strategy distribution on the worker side, i.e., cooperative workers and myopic workers co-exist, it gives a further rationale to equilibrium unemployment in a homogeneous worker population (i.e., in the absence of adverse selection).

In the context of *social games*, the VSRPD approach is a first step towards the research of endogenous network formation with consideration of within-network strategic behavior. There is a large literature of network formation (see for example, Jackson, 2008), but the strategic behavior within a network and dynamic change of the network are usually separately analyzed.[18] We showed that a huge variety of pairwise cooperative networks (between $\mathbf{X}_{c_0}$-players) and non-networking d_0-players[19] can co-exist in the society over the long horizon. This also implies that it is not guaranteed that all agents in the society end up in a (long-term) network (cf. Cho and Matsui, 2012).

Appendix: Proofs

Proof of lemma 1. GOS essentially showed the following.

There exists $\underline{\delta}_{c_0 d_0} \in (0,1)$ such that $\delta > \underline{\delta}_{c_0 d_0}$ if and only if there exists a unique $\overline{\alpha}_{cd}(\delta) \in (0,1)$,

[17] The English version is Okuno-Fujiwara (1987), but we cite the Japanese version to show that it precedes Shapiro and Stiglitz (1984).

[18] An exception is Immorlica et al. (2014). Restricting their attention to "consistent strategies" (similar to our c_0- and d_0-strategies that play the same action every period), they investigate dynamic formation of networks and show co-existence of cooperators and defectors.

[19] The exploiters, $\mathbf{X}_{d_0}$-players, can be an intermediate type of short-term networking players.

such that the bimorphic distribution $p^* = \overline{\alpha}_{cd}(\delta)c_0 + \{1 - \overline{\alpha}_{cd}(\delta)\}d_0$ is a Nash equilibrium and

$$\exists \overline{\epsilon}_{c_0} \in (0,1),\ \forall \epsilon \in (0, \overline{\epsilon}_{c_0}),\quad v(d_0; (1-\epsilon)p^* + \epsilon \cdot c_0) > v(c_0; (1-\epsilon)p^* + \epsilon \cdot c_0), \tag{12}$$

$$\exists \overline{\epsilon}_{d_0} \in (0,1),\ \forall \epsilon \in (0, \overline{\epsilon}_{d_0}),\quad v(c_0; (1-\epsilon)p^* + \epsilon \cdot d_0) > v(d_0; (1-\epsilon)p^* + \epsilon \cdot d_0). \tag{13}$$

It suffices to prove the condition (ii) of Definition 3. Since we have only two pure strategies in the support of p^*, any $q \in Q(p^*)$ has either $q(c_0) > \overline{\alpha}_{cd}(\delta)$ or $q(c_0) < \overline{\alpha}_{cd}(\delta)$.

For any $q \in Q(p^*)$ such that $q(c_0) > \overline{\alpha}_{cd}(\delta)$, $S_{p^*}(q) = \{c_0\}$. The inequality (12) and the property that the average payoffs of the c_0- and d_0-strategy are increasing in the share of the c_0-strategy imply that there exists $\overline{\epsilon}_{c_0} \in (0,1)$ such that for any $\epsilon \in (0, \overline{\epsilon}_{c_0})$,

$$\begin{aligned} &v(d_0; (1-\epsilon)p^* + \epsilon \cdot q) \geqq v(d_0; (1-\epsilon)p^* + \epsilon \cdot c_0) \\ &> v(c_0; (1-\epsilon)p^* + \epsilon \cdot c_0) \geqq v(c_0; (1-\epsilon)p^* + \epsilon \cdot q). \end{aligned}$$

Similarly, for any $q \in Q(p^*)$ such that $q(c_0) < \overline{\alpha}_{cd}(\delta)$, $S_{p^*}(q) = \{d_0\}$ and (13) implies that there exists $\overline{\epsilon}_{d_0} \in (0,1)$ such that for any $\epsilon \in (0, \overline{\epsilon}_{d_0})$,

$$\begin{aligned} &v(c_0; (1-\epsilon)p^* + \epsilon \cdot q) \geqq v(c_0; (1-\epsilon)p^* + \epsilon \cdot d_0) \\ &> v(d_0; (1-\epsilon)p^* + \epsilon \cdot d_0) \geqq v(d_0; (1-\epsilon)p^* + \epsilon \cdot q). \end{aligned}$$

Finally take $\overline{\epsilon} := \min\{\overline{\epsilon}_{c_0}, \overline{\epsilon}_{d_0}\}$. □

Proof of Remark 3. (i) Take any two-period tolerant strategy $\mathbf{Z}_{c_0}$ which belongs to *supp*$(q) \setminus$ *supp*$(\bar{p})$. The first two period actions are Z_1, Z_2. Because $q \in M_2(\delta)$,

$$\pi_{p^{PE}}(Z_1 Z_2 C_+) = \pi_q(Z_1 Z_2 C_+) < \overline{\alpha}_{cd}(\delta).$$

On the other hand, the post-entry share of the C-start satisfies (regardless of whether $p(C_+) \leqq \overline{\alpha}_{cd}(\delta)$ or not)

$$\lim_{\epsilon \to 0} \pi_{p^{PE}}(C_+) = \lim_{\epsilon \to 0} [(1-\epsilon)\overline{\alpha}_{cd}(\delta) + \epsilon \cdot q(C_+)] = \overline{\alpha}_{cd}(\delta).$$

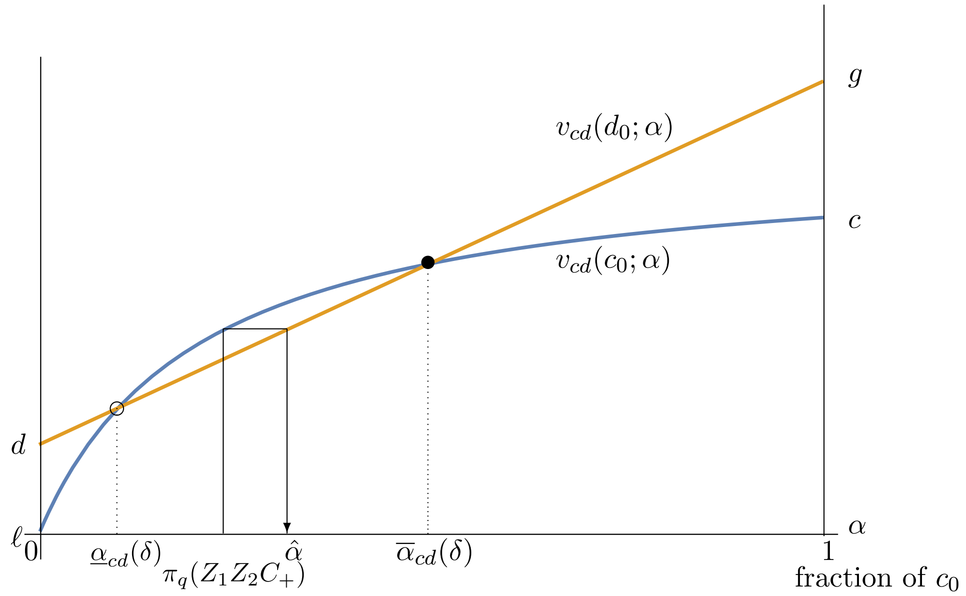
Hence there exists $\overline{\epsilon}_1 \in (0,1)$ such that

$$\pi_{p^{PE}}(Z_1 Z_2 C_+) < \pi_{p^{PE}}(C_+),\ \forall \epsilon \in (0, \overline{\epsilon}_1).$$

(ii) Case 1: If $\pi_q(Z_1 Z_2 C_+) (= \pi_{p^{PE}}(Z_1 Z_2 C_+)) \leqq \underline{\alpha}_{cd}(\delta)$, then

$$v_{cd}(c_0; \pi_{p^{PE}}(Z_1 Z_2 C_+)) \leqq v_{cd}(c_0; \underline{\alpha}_{cd}(\delta)) = v_{cd}(d_0; \underline{\alpha}_{cd}(\delta)) < v_{cd}(d_0; \overline{\alpha}_{cd}(\delta)).$$

Figure 8: Local stability and the existence of $\hat{\alpha}$



(See Figure 8.) As ϵ approaches to 0, $v_{cd}(d_0;\pi_{p^{PE}}(C_+)) \to v_{cd}(d_0;\overline{\alpha}_{cd}(\delta))$. Hence there exists $\overline{\epsilon}_2 \in (0,1)$ such that for any $\epsilon \in (0,\overline{\epsilon}_2)$,

$$v_{cd}(c_0;\pi_{p^{PE}}(Z_1Z_2C_+)) < v_{cd}(d_0;p^{PE}(C_+)).$$

Case 2: Suppose that $\pi_q(Z_1Z_2C_+) > \underline{\alpha}_{cd}(\delta)$. Since $\pi_q(Z_1Z_2C_+) < \overline{\alpha}_{cd}(\delta)$ by $q \in M_2(\delta)$,

$$v_{cd}(c_0;\pi_{p^{PE}}(Z_1Z_2C_+)) < v_{cd}(c_0;\overline{\alpha}_{cd}(\delta)) = v_{cd}(d_0;\overline{\alpha}_{cd}(\delta)).$$

By the Intermediate Value Theorem, there exists a constant $\hat{\alpha} < \overline{\alpha}_{cd}(\delta)$ such that

$$v_{cd}(c_0;\pi_{p^{PE}}(Z_1Z_2C_+)) = v_{cd}(c_0;\pi_q(Z_1Z_2C_+)) = v_{cd}(d_0;\hat{\alpha}).$$

(See Figure 8.) Again, $v_{cd}(d_0;\pi_{p^{PE}}(C_+)) \to v_{cd}(d_0;\overline{\alpha}_{cd}(\delta))$ as $\epsilon \to 0$ and $\hat{\alpha} < \overline{\alpha}_{cd}(\delta)$ imply that there exists $\overline{\epsilon}_2 \in (0,1)$ such that for any $\epsilon \in (0,\overline{\epsilon}_2)$, $v_{cd}(d_0;\hat{\alpha}) < v_{cd}(d_0;p^{PE}(C_+))$ so that

$$v_{cd}(c_0;\pi_{p^{PE}}(Z_1Z_2C_+)) < v_{cd}(d_0;p^{PE}(C_+)).$$

(iii) We show that the last component of (9) is negative for any $w_0, x_0 \in \{c_0, d_0\}$.

If $w_0 = x_0$, (i) implies that there exists $\overline{\epsilon}_1 \in (0,1)$ such that for any $\epsilon \in (0,\overline{\epsilon}_1)$,

$$v_{cd}(c_0;\pi_{p^{PE}}(Z_1Z_2C_+)) < v_{cd}(c_0;p^{PE}(C_+)),\ v_{cd}(d_0;\pi_{p^{PE}}(Z_1Z_2C_+)) < v_{cd}(d_0;p^{PE}(C_+)).$$

If $w_0 = d_0$, $x_0 = c_0$, and $\underline{\alpha}_{cd}(\delta) < \pi_{p^{PE}}(Z_1Z_2C_+) < \overline{\alpha}_{cd}(\delta))$, then the local stability and (i) imply that for any $\epsilon \in (0,\overline{\epsilon}_1)$,

$$v_{cd}(d_0;\pi_{p^{PE}}(Z_1Z_2C_+)) < v_{cd}(c_0;\pi_{p^{PE}}(Z_1Z_2C_+)) < v_{cd}(c_0;p^{PE}(C_+)).$$

If $w_0 = d_0$, $x_0 = c_0$, and $\pi_{p^{PE}}(Z_1Z_2C_+) \leqq \underline{\alpha}_{cd}(\delta)$, then for any $\epsilon \in (0, \bar{\epsilon}_1)$ (see Figure 8),

$$v_{cd}(d_0; \pi_{p^{PE}}(Z_1Z_2C_+)) \leqq v_{cd}(d_0; \underline{\alpha}_{cd}(\delta)) = v_{cd}(c_0; \underline{\alpha}_{cd}(\delta)) < v_{cd}(c_0; p^{PE}(C_+)).$$

Finally, (ii) implies that when $w_0 = c_0$ and $x_0 = d_0$, the last component is negative for any $\epsilon \in (0, \bar{\epsilon}_2)$. Let $\bar{\epsilon} = \min\{\bar{\epsilon}_1, \bar{\epsilon}_2\}$. Then we have that $v(XY_{w_0}; p^{PE}) < v(X_{y_0}; p^{PE})$, $\forall \epsilon \in (0, \bar{\epsilon})$. □

PROOF OF PROPOSITION 1. (i) By an analogous derivation as that of (9), for any $t = 1, 2, \ldots, T$, the average payoff of a t-period tolerant strategy $\mathbf{X}_{y_0} \in supp(\bar{p})$ with the initial t-period action sequence $\mathbf{X} = X_1 \cdots X_t$ is decomposed as follows. For each $k = 1, \ldots, t$, let $x_{k0} = c_0$ if $X_k = C$ and $x_{k0} = d_0$ if $X_k = D$.

$$\begin{aligned}
v(\mathbf{X}_{y_0}; \bar{p}) = & \, v_{cd}(x_{10}; p(C_+)) \\
& + \delta^2 \sum_{Z_1 = C,D} \rho_{\bar{p}}(Z_1; X_1) \frac{L_{cd}(x_{20}; \pi_{\bar{p}}(Z_1C_+))}{L(\mathbf{X}_{y_0}; \bar{p})} [v_{cd}(x_{20}; \pi_{\bar{p}}(Z_1C_+)) - v_{cd}(x_{10}; \bar{p}(C_+))] \\
& + \delta^4 \sum_{Z_1 = C,D} \rho_{\bar{p}}(Z_1; X_1) \sum_{Z_2 = C,D} \rho_{\bar{p}}(Z_1Z_2; X_2) \\
& \quad \cdot \frac{L_{cd}(x_{30}; \pi_{\bar{p}}(Z_1Z_2C_+))}{L(\mathbf{X}_{y_0}; \bar{p})} [v_{cd}(x_{30}; \pi_{\bar{p}}(Z_1Z_2C_+)) - v_{cd}(x_{10}; \bar{p}(C_+))] \\
& + \cdots + \delta^{2t} \sum_{Z_1 = C,D} \rho_{\bar{p}}(Z_1; X_1) \cdots \sum_{Z_t = C,D} \rho_{\bar{p}}(Z_1 \cdots Z_t; X_t) \\
& \quad \cdot \frac{L_{cd}(y_0; \pi_{\bar{p}}(Z_1 \cdots Z_tC_+))}{L(\mathbf{X}_{y_0}; \bar{p})} [v_{cd}(y_0; \pi_{\bar{p}}(Z_1 \cdots Z_tC_+)) - v_{cd}(x_{10}; \bar{p}(C_+))].
\end{aligned}$$

For any $\bar{p} \in \bar{P}_T$, $\bar{p}(C_+) = \overline{\alpha}_{cd}(\delta)$ and $\pi_{\bar{p}}(\mathbf{Z}C_+) = \overline{\alpha}_{cd}(\delta)$ for any $\mathbf{Z} \in \{C, D\}^k$ for any $k = 1, 2, \ldots, T$. Hence all subtraction terms in the above formula is zero, and the C-D ratio implies that $v(\mathbf{X}_{y_0}; \bar{p}) = v_{cd}(c_0; \overline{\alpha}_{cd}(\delta)) = v_{cd}(d_0; \overline{\alpha}_{cd}(\delta))$.

(ii) Fix any $\delta \in (\underline{\delta}_{c_0d_0}, 1)$, any $T < \infty$, and any $\bar{p} \in \bar{P}_T$. Take any $q \in M_\infty(\delta)$ and let $p^{PE} = (1 - \epsilon)\bar{p} + \epsilon \cdot q$. Any strategy in $supp(q) \setminus supp(\bar{p})$ is a $T + k$-period tolerant strategy with some $k \geqq 1$. It suffices to prove that for any $T + k$-period tolerant strategy $\mathbf{X}_{y_0}$ (where $\mathbf{X} = X_1 \cdots X_{T+k}$ is the sequence of initial $T + k$ period actions), there exists a T-period tolerant strategy which outperforms the $\mathbf{X}_{y_0}$-strategy for sufficiently small ϵ. Note that any T-period tolerant strategy belongs to the support of $\bar{p}$.

Given a $T + k$-period tolerant strategy $\mathbf{X}_{y_0}$, for each $t = 1, 2, \ldots, T + k$, let $x_{t0} = c_0$ (resp. d_0)

if $X_t = C$ (resp. D). The average payoff is decomposed as follows.

$$
\begin{aligned}
v(\mathbf{X}_{y_0}; p^{PE}) = {} & v_{cd}(x_{10}; p(C_+)) \\
& + \delta^2 \sum_{Z_1=C,D} \rho_{p^{PE}}(Z_1; X_1) \frac{L_{cd}(x_{20}; \pi_{p^{PE}}(Z_1 C_+))}{L(\mathbf{X}_{y_0}; \bar{p})} [v_{cd}(x_{20}; \pi_{p^{PE}}(Z_1 C_+)) - v_{cd}(x_{10}; \bar{p}(C_+))] \\
& + \delta^4 \sum_{Z_1=C,D} \rho_{p^{PE}}(Z_1; X_1) \sum_{Z_2=C,D} \rho_{p^{PE}}(Z_1 Z_2; X_2) \\
& \quad \cdot \frac{L_{cd}(x_{30}; \pi_{p^{PE}}(Z_1 Z_2 C_+))}{L(\mathbf{X}_{y_0}; p^{PE})} [v_{cd}(x_{30}; \pi_{p^{PE}}(Z_1 Z_2 C_+)) - v_{cd}(x_{10}; p^{PE}(C_+))] \\
& + \cdots + \delta^{2T} \sum_{Z_1=C,D} \rho_{p^{PE}}(Z_1; X_1) \cdots \sum_{Z_T=C,D} \rho_{p^{PE}}(Z_1 \cdots Z_T; X_T) \\
& \quad \cdot \frac{L_{cd}(y_0; \pi_{p^{PE}}(Z_1 \cdots Z_T C_+))}{L(\mathbf{X}_{y_0}; p^{PE})} [v_{cd}(x_{T+1\,0}; \pi_{p^{PE}}(Z_1 \cdots Z_T C_+)) - v_{cd}(x_{10}; p^{PE}(C_+))] \\
& + \delta^{2(T+1)} \sum_{Z_1=C,D} \rho_{p^{PE}}(Z_1; X_1) \cdots \sum_{Z_{T+1}=C,D} \rho_{p^{PE}}(Z_1 \cdots Z_{T+1}; X_{T+1}) \\
& \quad \cdot \frac{L_{cd}(x_{T+2\,0}; \pi_{p^{PE}}(Z_1 \cdots Z_{T+1} C_+))}{L(\mathbf{X}_{y_0}; p)} [v_{cd}(x_{T+2\,0}; \pi_{p^{PE}}(Z_1 \cdots Z_{T+1} C_+)) - v_{cd}(x_{10}; p^{PE}(C_+))] \\
& + \cdots + \delta^{2(T+k)} \sum_{Z_1=C,D} \rho_{p^{PE}}(Z_1; X_1) \cdots \sum_{Z_{T+k}=C,D} \rho_{p^{PE}}(Z_1 \cdots Z_{T+k}; X_{T+k}) \\
& \quad \cdot \frac{L_{cd}(y_0; \pi_{p^{PE}}(Z_1 \cdots Z_{T+k} C_+))}{L(\mathbf{X}_{y_0}; p)} [v_{cd}(y_0; \pi_{p^{PE}}(Z_1 \cdots Z_{T+k} C_+)) - v_{cd}(x_{10}; p^{PE}(C_+))].
\end{aligned}
$$

As the average payoff of a two-period tolerant strategy XY_{w_0} embeds that of the one-period tolerant strategy X_{y0}, the above embeds the average payoff of the T-period tolerant strategy $\mathbf{X'}_{w_0}$ such that $\mathbf{X'} = X_1 \cdots X_T$ and $w_0 = c_0$ (resp. d_0) if $X_{T+1} = C$ (resp. D) in the first $T+1$ terms. Hence it suffices to show that the last k components of the above formula are negative, or

$$
\begin{aligned}
& v_{cd}(x_{T+n\,0}; \pi_{p^{PE}}(Z_1 \cdots Z_{T+n-1} C_+)) - v_{cd}(x_{10}; p^{PE}(C_+)) < 0, \;\; n = 2, 3, \ldots, k, \\
& v_{cd}(y_0; \pi_{p^{PE}}(Z_1 \cdots Z_{T+k} C_+)) - v_{cd}(x_{10}; p^{PE}(C_+)) < 0.
\end{aligned}
$$

Recall that $supp(\bar{p})$ includes only up to T-period tolerant strategies. Hence the post-entry relative shares of the $Z_1 \cdots Z_{T+n-1} C_+$-class (for $n = 2, 3, \ldots, k$) are mutant's share:

$$
\pi_{p^{PE}}(Z_1 \cdots Z_{T+n-1} C_+) = \pi^q(Z_1 \cdots Z_{T+n-1} C_+).
$$

Hence, by an analogous argument as that of Remark 3 (and because there are only finite varieties of $Z_1 \cdots Z_{T+n-1} C_+$-classes), there exists $\bar{\epsilon} \in (0, 1)$ such that for any $\epsilon \in (0, \bar{\epsilon})$, all the above are negative. □

Proof of remark 4. (i) Take an arbitrary $k < \infty$ and consider any mutant distribution $q = \gamma \cdot D_{c_0}^{k+1} + (1-\gamma) D_{d_0}^{k+1}$ such that $\gamma > 0$. Let $p^{PE} = (1-\epsilon) D_{d_0}^{k} + \epsilon q$. Clearly, $v(D_{d_0}^{k}; p^{PE}) = d$, while

$$v(D_{d_0}^{k+1}; p^{PE}) = \frac{(1+\delta^2+\cdots+\delta^{2T})d + \epsilon\delta^{2(T+1)}\{\gamma \cdot g + (1-\gamma)d\}}{(1+\delta^2+\cdots+\delta^{2T}) + \epsilon\delta^{2(T+1)}} \tag{14}$$

$$= d + \frac{\epsilon\delta^{2(T+1)}\{\gamma \cdot g + (1-\gamma)d - d\}}{(1+\delta^2+\cdots+\delta^{2T}) + \epsilon\delta^{2(T+1)}} > d, \quad \text{because } \gamma > 0.$$

(ii) Consider any mutant distribution $\gamma \cdot c_0 + (1-\gamma) d_0$ such that $\gamma > 0$. Let $p^{PE} = (1-\epsilon) D_{d_0}^{\infty} + \epsilon q$. By computation,

$$v(D_{d_0}^{\infty}; p^{PE}) = \frac{(1-\epsilon)\frac{d}{1-\delta^2} + \epsilon\{\gamma \cdot g + (1-\gamma)d\}}{(1-\epsilon)\frac{1}{1-\delta^2} + \epsilon}$$

$$= d + \frac{\epsilon}{(1-\epsilon)\frac{1}{1-\delta^2} + \epsilon}\gamma(g-d).$$

On the other hand,

$$v(d_0; p^{PE}) = (1-\epsilon)d + \epsilon\{\gamma \cdot g + (1-\gamma)d\} = d + \epsilon\gamma(g-d).$$

Note that the denominator of the coefficient of the second term of $v(D_{d_0}^{\infty}; p^{PE})$ is

$$(1-\epsilon)\frac{1}{1-\delta^2} + \epsilon = (1-\epsilon) + (1-\epsilon)\frac{\delta^2}{1-\delta^2} + \epsilon > 1.$$

Hence $v(d_0; p^{PE}) > v(D_{d_0}^{\infty}; p^{PE})$. □

Proof of proposition 2. All non-degenerate tolerant strategies ($\mathbf{X}_{y_0}$ with $\mathbf{X} \in \{C, D\}^k$, $k = 1, 2, \ldots$) are mutants. Take any T-period tolerant mutant strategy $\mathbf{X}_{y_0}$ and a stationary distribution p in the matching pool. Its average payoff embeds $v_{cd}(x_{10}; p(C_+))$, where $x_{10} = c_0$ (resp. d_0) if $X_1 = C$ (resp. D) as follows.

$$\begin{aligned}
& v(\mathbf{X}_{y_0}; p) \\
&= v_{cd}(x_{10}; p(C_+)) \\
&\quad + \delta^2 \sum_{Z_1 = C,D} \rho_p(Z_1; X_1) \frac{L_{cd}(x_{20}; \pi_p(Z_1 C_+))}{L(\mathbf{X}_{y_0}; p)} [v_{cd}(x_{20}; \pi_p(Z_1 C_+)) - v_{cd}(x_{10}; p(C_+))] \\
&\quad + \delta^4 \sum_{Z_1 = C,D} \rho_p(Z_1; X_1) \sum_{Z_2 = C,D} \rho_p(Z_1 Z_2; X_2) \\
&\qquad\qquad \cdot \frac{L_{cd}(x_{30}; \pi_p(Z_1 Z_2 C_+))}{L(\mathbf{X}_{y_0}; p)} [v_{cd}(x_{30}; \pi_p(Z_1 Z_2 C_+)) - v_{cd}(x_{10}; p(C_+))] \\
&\quad + \cdots + \delta^{2T} \sum_{Z_1 = C,D} \rho_p(Z_1; X_1) \cdots \sum_{Z_T = C,D} \rho_p(Z_1 \cdots Z_T; X_T) \\
&\qquad \cdot \frac{L_{cd}(w_0; \pi_p(Z_1 \cdots Z_T C_+))}{L(\mathbf{X}_{y_0}; p)} [v_{cd}(w_0; \pi_p(Z_1 \cdots Z_T C_+)) - v_{cd}(x_{10}; p(C_+))].
\end{aligned} \tag{15}$$

Take an arbitrary $q \in M(\delta)$ and let $p^{PE} = (1-\epsilon)p^* + \epsilon \cdot q$. Denote $\alpha^* := \sup_{\mathbf{X} \in \{C,D\}^k,\ k=1,2,\ldots} \pi_q(\mathbf{X}C_+)$, which is strictly less than $\overline{\alpha}_{cd}(\delta)$. Hence there exists $\overline{\epsilon}_q \in (0,1]$ such that for any $\epsilon \in (0, \overline{\epsilon}_q)$,

$$\max\{\underline{\alpha}_{cd}(\delta), \alpha^*\} < (1-\epsilon)\overline{\alpha}_{cd}(\delta) + \epsilon \cdot q(C_+)(= p^{PE}(C_+)).$$

Since $v_{cd}(w_0; \alpha)$ is increasing in α for both $w_0 = c_0, d_0$, for any $\epsilon \in (0, \overline{\epsilon}_q)$ and any $\mathbf{Z}_{c_0}$ with $\mathbf{Z} \in \{C, D\}^k$, $k = 1, 2, \ldots$,

$$v_{cd}(w_0; \pi_{p^{PE}}(\mathbf{Z}C_+)) \leqq v_{cd}(w_0; \alpha^*) < v_{cd}(w_0; p^{PE}(C_+)),\ \forall\ w_0 = c_0, d_0.$$

Step 1: For any $\epsilon \in (0, \overline{\epsilon}_q)$, any $k = 1, 2, \ldots,$ and any $\mathbf{Z} \in \{C, D\}^k$,

$$v_{cd}(d_0; \pi_{p^{PE}}(\mathbf{Z}C_+)) < v_{cd}(c_0; p^{PE}(C_+)).$$

Proof of Step 1: Fix $\epsilon \in (0, \overline{\epsilon}_q)$ and $\mathbf{Z} \in \{C, D\}^k$ for an arbitrary $k = 1, 2, \ldots$.

Case 1: $\pi_{p^{PE}}(\mathbf{Z}C_+) \leqq \underline{\alpha}_{cd}(\delta)$. (See Figure 8.) Since $v_{cd}(d_0; \alpha)$ is increasing in α, and at $\underline{\alpha}_{cd}(\delta)$, the c_0- and the d_0-strategy have the same average payoff,

$$v_{cd}(d_0; \pi_{p^{PE}}(\mathbf{Z}C_+)) \leqq v_{cd}(d_0; \underline{\alpha}_{cd}(\delta)) = v_{cd}(c_0; \underline{\alpha}_{cd}(\delta)) < v_{cd}(c_0; p^{PE}(C_+)).$$

Case 2: $\underline{\alpha}_{cd}(\delta) < \pi_{p^{PE}}(\mathbf{Z}C_+) \leqq \alpha^*$. By the local stability,

$$v_{cd}(d_0; \pi_{p^{PE}}(\mathbf{Z}C_+)) < v_{cd}(c_0;\ \pi_{p^{PE}}(\mathbf{Z}C_+)) < v_{cd}(c_0; p^{PE}(C_+)).$$

This completes the proof of Step 1. //

Step 2: There exists $\overline{\epsilon}_{\alpha^*} \in (0, 1)$ such that for any $\epsilon \in (0, \overline{\epsilon}_{\alpha^*})$, any $k = 1, 2, \ldots,$ and any $\mathbf{Z} \in \{C, D\}^k$,

$$v_{cd}(c_0; \pi_{p^{PE}}(\mathbf{Z}C_+)) < v_{cd}(d_0; p^{PE}(C_+)).$$

Proof of Step 2:

Case 1: $v_{cd}(c_0; \alpha^*) \leqq v_{cd}(d_0; \alpha^*)$.

Then

$$v_{cd}(c_0; \pi_{p^{PE}}(\mathbf{Z}C_+)) \leqq v_{cd}(c_0; \alpha^*) \leqq v_{cd}(d_0; \alpha^*) < v_{cd}(d_0; p^{PE}(C_+))$$

for any $\epsilon \in (0, \overline{\epsilon}_q)$. Hence let $\overline{\epsilon}_{\alpha^*} = \overline{\epsilon}_q$.

Case 2: $v_{cd}(d_0; \alpha^*) < v_{cd}(c_0; \alpha^*)$.

Note that $v_{cd}(d_0; \overline{\alpha}_{cd}(\delta)) = v_{cd}(c_0; \overline{\alpha}_{cd}(\delta))$. By the Intermediate Value Theorem, there exists $\hat{\alpha} \in (\alpha^*, \overline{\alpha}_{cd}(\delta))$ such that

$$v_{cd}(d_0; \hat{\alpha}) = v_{cd}(c_0; \alpha^*).$$

Thus, there exists $\overline{\epsilon}_{\alpha^*} \in (0,1)$ such that

$$\hat{\alpha} < (1-\epsilon)\overline{\alpha}_{cd}(\delta) + \epsilon \cdot q(C_+)(= p^{PE}(C_+)),\ \forall \epsilon \in (0, \overline{\epsilon}_{\alpha^*}).$$

Then

$$v_{cd}(c_0; \pi_{p^{PE}}(\mathbf{Z}C_+)) \leqq v_{cd}(c_0; \alpha^*) = v_{cd}(d_0; \hat{\alpha}) < v_{cd}(d_0; p^{PE}(C_+)),\ \forall \epsilon \in (0, \overline{\epsilon}_{\alpha^*}).$$

This completes the proof of Step 2. //

Finally, let $\overline{\epsilon} := \min\{\overline{\epsilon}_q, \overline{\epsilon}_{\alpha^*}\} \in (0,1)$. Then for any $\epsilon \in (0, \overline{\epsilon})$, all the subtraction terms in (15) are negative in the post-entry distribution, so that

$$v(\mathbf{X}_{y0}; p^{PE}) < v_{cd}(x_{10}; p^{PE}(C_+)) = v(x_{10}; p^{PE}),\ \forall \epsilon \in (0, \overline{\epsilon}).$$

Since x_{10} is either c_0 or d_0, if $v(c_0; p^{PE}) \geqq v(d_0; p^{PE})$, then the c_0-strategy is the uniform winner and vice versa. □

PROOF OF PROPOSITION 3. The social mean of the average payoffs under the incumbent distribution is

$$\sum_{s \in \mathbf{S}} p^*(s) \cdot v(s; p^{PE}) = \overline{\alpha}_{cd}(\delta) \cdot v(c_0; p^{PE}) + \{1 - \overline{\alpha}_{cd}(\delta)\} \cdot v(d_0; p^{PE}).$$

This is because p^* has only two strategies in the support.

To compute the mean of the average payoffs of the mutant distributions, recall first that the mutant distribution is inductively written (recall (11)):

$$\begin{aligned} q = q(C_+)&\Big[q(c_0|C_+) \cdot c_0 + [1 - q(c_0|C_+)]\Big\{\pi_q(CC_+) \cdot q\ |_{CC} + \pi_q(CD_+) \cdot q\ |_{CD}\ (\pi)\Big\}\Big] \\ + q(D_+)&\ \Big[q(d_0|D_+) \cdot d_0 + [1 - q(d_0|D_+)]\Big\{\pi_q(DC_+) \cdot q\ |_{DC} + \pi_q(DD_+) \cdot q\ |_{DD}\Big\}\Big], \end{aligned} \tag{16}$$

where $q\ |_{XY}$ is a distribution starting with a one-period tolerant strategy X_{y_0} such that

$$\begin{aligned} q\ |_{XY} = q(X_{y_0}|X*_+) \cdot X_{y_0} + \{1 - q(X_{y_0}|X*_+)\} \\ \cdot \Big\{\pi_q(XYC_+) \cdot q\ |_{XYC} + \pi_q(XYD_+) \cdot q\ |_{XYD}\ (\pi)\Big\}, \end{aligned}$$

and so on.

The first component in (16) is the C-start class and the second terms is the D-start class. In the proof of Proposition 2, we have shown that, there exists $\overline{\epsilon} \in (0,1)$ such that, for any $\epsilon \in (0, \overline{\epsilon})$, all C-start (resp. D-start) non-degenerate tolerant strategies have less post-entry average payoff than that of the c_0-strategy (resp. the d_0-strategy). By the replacement of the average payoffs of the tolerant strategies with that of $v(c_0; p^{PE})$ and $v(d_0; p^{PE})$ (and there is at least one such strategy in $q \in M_+(\delta)$), the social mean under the mutant distribution satisfies

$$\sum_{s \in \mathbf{S}} q(s) \cdot v(s; p^{PE}) < q(C_+) \cdot v(c_0; p^{PE}) + (1 - q(C_+)) \cdot v(d_0; p^{PE}).$$

Case 1: $q(C_+) \geqq \overline{\alpha}_{cd}(\delta)$.

Then $p^{PE}(C_+) = (1-\epsilon)\overline{\alpha}_{cd}(\delta) + \epsilon \cdot q(C_+) \geqq \overline{\alpha}_{cd}(\delta)$ for any $\epsilon \in (0, \overline{\epsilon})$. In this case, $v(d_0; p^{PE}) \geqq v(c_0; p^{PE})$ (for any $\epsilon \in (0,1)$) by the local stability and also the weight on the d_0-strategy is weakly larger $(1 - q(C_+)) \leqq (1 - \overline{\alpha}_{cd}(\delta))$ in p^*. Hence (for any $\epsilon \in (0,1)$),

$$\begin{aligned} &\sum_{s \in \mathbf{S}} q(s) \cdot v(s; p^{PE}) \\ <& q(C_+) \cdot v(c_0; p^{PE}) + (1 - q(C_+)) \cdot v(d_0; p^{PE}) \\ \leqq& \overline{\alpha}_{cd}(\delta) \cdot v(c_0; p^{PE}) + \{1 - \overline{\alpha}_{cd}(\delta)\} \cdot v(d_0; p^{PE}) = \sum_{s \in \mathbf{S}} p^*(s) \cdot v(s; p^{PE}). \end{aligned}$$

Case 2: $q(C_+) < \overline{\alpha}_{cd}(\delta)$.

In this case, for sufficiently small $\epsilon \in (0, \overline{\epsilon})$, $v(d_0; p^{PE}) < v(c_0; p^{PE})$, and the weight on the c_0-strategy is also higher in p^*. Again,

$$\begin{aligned} &\sum_{s \in \mathbf{S}} q(s) \cdot v(s; p^{PE}) \\ <& q(C_+) \cdot v(c_0; p^{PE}) + (1 - q(C_+)) \cdot v(d_0; p^{PE}) \\ <& \overline{\alpha}_{cd}(\delta) \cdot v(c_0; p^{PE}) + \{1 - \overline{\alpha}_{cd}(\delta)\} \cdot v(d_0; p^{PE}) = \sum_{s \in \mathbf{S}} p^*(s) \cdot v(s; p^{PE}). \end{aligned}$$

□

Proof of corollary 1. There exists $\overline{\epsilon} \in (0,1)$ such that for any $\epsilon \in (0, \overline{\epsilon})$, $p^{PE}(C_+)$ is close to $\overline{\alpha}_{cd}(\delta)$ so that by the local stability,

$$v_{cd}(c_0; p^{PE}) > v_{cd}(d_0; p^{PE}).$$

Since there are only D-start mutants in $D(\delta)$, by the same argument as the proof of Proposition 2, there exists $\overline{\epsilon}_D \in (0,1)$ such that for any $\epsilon \in (0, \overline{\epsilon}_D)$,

$$v_{cd}(d_0; p^{PE}) > v(D^k_{w_0}; p^{PE}), \ \forall w_0 \in \{c_0, d_0\}, \ \forall k = 1, 2, \ldots.$$

Hence both incumbents, c_0 and d_0, have strictly greater post-entry average payoff than all mutants for any $\epsilon \in (0, \min\{\overline{\epsilon}, \overline{\epsilon}_D\})$. □

References

Bendor, J. and Swistak, P. (1997). The Evolutionary Stability of Cooperation. *American Political Science Review* **91** 290–307.

Carmichael, L. and MacLeod, B. (1997). Gift Giving and the Evolution of Cooperation. *International Economic Review* **38** 485–509.

Cho, I. and Matsui, A. (2012). A Dynamic Foundation of the Rawlsian Maximin Criterion. *Dynamic Games and Applications* **2** 51–70.

Datta, S. (1996). Building Trust. Manuscript. London School of Economics.

Duffie, D. and Sun,Y. (2012). The Exact Law of Large Numbers for Independent Random Matching. *Journal of Economic Theory* **147** 1105–1139.

Duffie, D., Quao, L. and Sun,Y. (2016). Dynamic Directed Random Matching. Manuscript. Stanford University and National University of Singapore.

Eeckhout, J. (2006). Minorities and Endogenous Segregation. *Review of Economic Studies* **73** 31–53.

Ellison, G. (1994). Cooperation in the Prisoner's Dilemma with Anonymous Random Matching. *Review of Economic Studies* **61** 567–588.

Fudenberg, D. and Maskin, E. (1986). The Folk Theorem in Repeated Games with Discounting or with Incomplete Information. *Econometrica* **54** 533–554.

Fudenberg, D. and Maskin, E. (1990). Evolution and Cooperation in Noisy Repeated Games. *American Economic Review* **80** 274–279.

Fujiwara-Greve, T. and Okuno-Fujiwara, M. (2009). Voluntarily Separable Repeated Prisoner's Dilemma. *Review of Economic Studies* **76** 993–1021.

Fujiwara-Greve, T., Okuno-Fujiwara, M. and Suzuki, N. (2015). Efficiency may Improve when Defectors Exist. *Economic Theory* **60** 423–460.

Garcìa, J. and van Veelen, M. (2016). In and Out of Equilibrium I: Evolution of Strategies in Repeated Games with Discounting. *Journal of Economic Theory* **161** 161–189.

Ghosh, P. and Ray, D. (1996). Cooperation in Community Interaction without Information Flows. *Review of Economic Studies* **63** 491–519.

Immorlica, N., Lucier, B. and Rogers, B. (2014). Cooperation in Anonymous Dynamic Social Networks. Manuscript. Microsoft Research and Washington University in St. Louis.

Izquierdo, L., Izquierdo, S. and Vega-Redondo, F. (2014) Leave and Let Leave: A Sufficient Condition to Explain the Evolutionary Emergence of Cooperation. *Journal of Economic Dynamics and Control* **46** 91–113.

Jackson, M. (2008). *Social and Economic Networks*. Princeton University Press, New Jersey, USA.

Kandori, M. (1992). Social Norms and Community Enforcement. *Review of Economic Studies* **59** 63–80.

Kandori, M., Mailath, G. and Rob, R. (1993). Learning, Mutation and Long Run Equilibria in Games. *Econometrica* **61** 29–56.

Kim, Y. (1994). Evolutionarily Stable Strategies in the Repeated Prisoner's Dilemma. *Mathematical Social Sciences* **28** 167–197.

Kranton, R. (1996). The Formation of Cooperative Relationships. *Journal of Law, Economics & Organization* **12** 214–233.

Kranton, R. (1996). Reciprocal Exchange: A Self-Sustaining System. *American Economic Review* **86** 830–851.

McAdams, D. (2011). Performance and Turnover in a Stochastic Partnership. *American Economic Journal: Microeconomics* **3** 107–142.

Maynard Smith, J. (1982). *Evolution and the Theory of Games*. Camgridge University Press, Cambridge, UK.

Maynard Smith, J. and Price, G. (1973). The Logic of Animal Conflict. *Nature* **246** 15–18.

Okuno, M. (1981). Labor Incentives and the 'Japanese' Labor Market. *Contemporary Economics (Kikan Gendai Keizai)* **Winter** 150–162. (In Japanese.)

Okuno-Fujiwara, M. (1987). Monitoring Cost, Agency Relationship, and Equilibrium Modes of Labor Contract. *Japanese and International Economies* **1** 147–167.

Okuno-Fujiwara, M., Nishimura, N., Suzuki, N. and Fujiwara-Greve, T. (2016) Voluntary Partnerships, Cooperation, and Coordination. In preparation.

Rob, R. and Yang, H. (2010). Long-term Relationships as Safeguards. *Economic Theory* **43** 143–166.

Robson, A. (1990). Efficiency in Evolutionary Games: Darwin, Nash and the Secret Handshake. *Journal of Theoretical Biology* **144** 379–396.

Samuelson, L. (1997). *Evolutionary Games and Equilibrium Selection.* MIT Press, Cambridge, MA.

Samuelson, L. and Zhang, J. (1992). Evolutionary Stability in Asymmetric Games. *Journal of Economic Theory* **57** 363–391.

Sandholm, W. (2010). *Population Games and Evolutionary Dynamics.* MIT Press, Cambridge, MA.

Schumacher, H. (2013) Imitating Cooperation and the Formation of Long-term Relationships. *Journal of Economic Theory* **148** 409–417.

Selten, R. (1983). Evolutionary Stability in Extensive Two-Person Games. *Mathematical Social Sciences* **5** 269–363.

Shapiro, C. and Stiglitz, J. (1984). Equilibrium Unemployment as a Worker Discipline Device. *American Economic Review* **74** 433–444.

Sun, Y. (2006). The Exact Law of Large Numbers via Fubini Extension and Characterization of Insurable Risks. *Journal of Economic Theory* **126** 31–69.

Vesly, F. and Yang, C. (2012). Breakup, Secret Handshake and Neutral Stability in Repeated Prisoner'S Dilemma with Option to Leave: A Note. SSRN No. 2179126.