DISEI - Università degli Studi di Firenze

Working Papers - Economics

A Prisoners' Dilemma with Incomplete Information on the Discount Factors

Elena M. Parilina and Alessandro Tampieri

Working Paper N. 21/2018

DISEI, Università degli Studi di Firenze
Via delle Pandette 9, 50127 Firenze, Italia
www.disei.unifi.it

A Prisoners' Dilemma with Incomplete Information on the Discount Factors*

Elena M. Parilina† and Alessandro Tampieri‡

October 15, 2018

Abstract

This paper analyses a prisoners' dilemma where players' discount factor is private information. We consider an infinitely repeated game where two states of the world may occur. According to her own discount factor, a player chooses a cooperative behaviour in both states (patient), in none of the states (impatient) or in one state only (mildly patient). The presence of different states of the world affects the strategic role of beliefs. A mildly patient player has an incentive in "pretending" to be patient, which increases with the competitor's belief that the player is patient. Interestingly, this effect prevents or delays cooperative equilibria to occur when the belief in patience is strong.

JEL codes: C73, D43, L13.

Keywords: Bayesian games; two-phases game; Markov perfect equilibrium.

*We are grateful to Alessandro Gioffré, Jon Hamilton, Nicolas Klein, Flavio Menezes, Simone Meraglia, Dimitri Paolini, Jose Rodriguez-Neto, Artem Sedakov, Tristan Tomala, Georges Zaccour, and the seminar audiences at GERAD seminar (2018) and the Symposium on Dynamic Games and Applications (2018) for helpful comments. The usual disclaimer applies. The work of the first author was supported by the Shandong Province "Double-Hundred Talent Plan" (No. WST2017009).

†Saint Petersburg State University, 7/9 Universitetskaya nab., Saint Petersburg 199034, Russia; School of Mathematics and Statistics, Qingdao University, Qingdao 266071, PR China; Institute of Applied Mathematics of Shandong, Qingdao 266071, PR China; e.parilina@spbu.ru.

‡Department of Economics and Business, University of Florence and CREA, University of Luxembourg, Via delle Pandette 9, Florence 50127, Italy: alessandro.tampieri@unifi.it.

1 Introduction

Several strategic situations involve interactions where there is uncertainty about the competitor's patience. A relevant example is collusion, which can be sustained with limited information if players are sufficiently patient (Mailath and Samuelson 2006). In this context, common wisdom suggests that the belief on the competitor's reliability plays a crucial role to reach a cooperative solution. In this paper we study a case where, on the contrary, a strong belief on the competitor's patience may prevent or delay the emergence of cooperation.

We analyse an infinitely repeated prisoners' dilemma where a player's discount factor is private information, and where two states of the world randomly occur over time. The difference among states reflect a difference in terms of payoffs. In particular, in one state (the "good state") cooperation is more sustainable than in the other (the "bad state"). Moreover, the payoffs in the different states of the world are such that a not too patient player may collude in one state of the world and deviate in the other. This entails the presence of potentially 3 class of players:[1] patient, adopting a cooperative behaviour in both states, impatient (never cooperative), and the mildly patient, who cooperates in the good state but deviates in the bad state.

The game evolves in two phases. The first is the learning phase, where each player determines the competitor's class. The second is the cooperation phase, where cooperation may emerge. Along the paper, we first describe the game with public information. Second, we introduce incomplete information by assuming one learning phase. This exercise has the advantage to highlight the features of this game in a simple setting. Finally, we let the learning phase be determined at equilibrium.

The assumption of different states influences the strategic role of beliefs. Suppose that, in the learning phase, a patient player has a strong belief that the competitor is patient too. Suppose also the competitor is in fact mildly patient, she pretends to be patient and she is believed. Then the patient player will agree to play a cooperative strategy in the cooperative phase, which entails cooperation in both states; but then, the mildly patient competitor will deviate in the bad state. Therefore, the mildly patient competitor has an incentive in pretending to be patient. Throughout the paper, we will refer to this effect as "faking patience".

[1] Throughout the paper, a player's type is determined by its discount factor, while a player's class depends on whether her discount factor lies above or below certain thresholds.

We show that the faking patience effect increases with the player's belief about the competitor's patience. In turn, since players rationally predict this kind of behaviour, a strong belief in patience surprisingly will not lead to a fully cooperative equilibrium. In particular, in the simplifying case with one learning phase period, the equilibrium strategy will exhibit cooperation in the good state and non-cooperation in the bad state of the world. Conversely with endogenous learning phase, this effect delays the beginning of the cooperation phase, and again it is stronger the higher the belief on the competitor's patience.

Related literature: In the theory of repeated games, incomplete information has been examined in several elements of the game (Mailath and Samuelson, 2006). Strands of the literature analysed uncertainty about technology (Athey and Bagwell, 2008, *inter alia*) payoffs (Peski, 2014, 2008, Fudenberg and Yamamoto, 2011, 2010, Wiseman, 2005, *inter alia*) or actions (Abreu *et al.*, 1990, Fudenberg *et al.*, 1994, Ghosh and Ray, 1996, Kandori and Obara, 2006, *inter alia*).

The present paper is related mainly on those contributions that focused on uncertainty about the competitor's discount factor. In Watson (2002) and (1999), players are in a partnership, and in each period choose the level of interaction among each other and whether to cooperate. The level of interaction can be seen, for instance, as an investment in a joint project. At equilibrium, players "start small" (i.e., make a low investment) to learn about the rival's patience. Harrington and Zhao (2012) examine tacit collusion in a deterministic, infinitely repeated prisoners' dilemma where a player's discount factor is private information. Given the presence of only one payoff state, a player can be patient or impatient.

This paper is also related to Rotemberg and Saloner (1986), Haltiwanger and Harrington (1991) and Bagwell and Staiger (1997), who investigate the relationship between collusion and the business cycle. These papers examine collusive pricing in markets with demand shocks and wonder if and when collusion is procyclical or countercyclical. With respect to these contributions, a change of states may be interpreted as a demand shock. However, our analysis focuses on states that give different incentives in collusion, while the movement of the economy is set aside. Hence, one cannot tell the relationship between collusion and the business cycles from our results. Finally, unlike the present paper, in this literature the discount factor is public information.

The remainder of the paper is organised as follows. Section 2 introduces the model. Section 3 develops the analysis where discount factors are unknown. Section 4 shows the cooperative results. Section 5 proposes a numerical example to illustrate the theoretical results, while Section 6 concludes. All formal proofs can be found in the Appendix.

2 The model

2.1 Preliminaries

We consider an infinitely-repeated prisoners' dilemma with two states $\{s_1, s_2\}$ and two players $\{i, j\}$. Time is discrete and, in any time period $t = 1, 2, \ldots$, one of two states of nature can be realized. Each state is represented by a one-shot Prisoners' Dilemma. The action set of any player ν in any state s is $\mathbf{X} = \{C, D\}$. Action C stands for "to cooperate", while action D is "to deviate". The payoff function of any player $\nu \in \{i, j\}$ in state s is $u^s : \mathbf{X} \times \mathbf{X} \to \mathbb{R}$, given by the payoff matrix for state $s \in \{s_1, s_2\}$:

$$\begin{array}{ccc|c|c|} & & \multicolumn{2}{c}{j} \\ & & \multicolumn{1}{c}{C} & \multicolumn{1}{c}{D} \\ \cline{3-4} i & C & a^s; a^s & c^s; b^s \\ \cline{3-4} & D & b^s; c^s & d^s; d^s \\ \cline{3-4} \end{array} \tag{1}$$

Payoffs are symmetric among players, but change according to the state of the world.

We imply the standard assumptions about Prisoners' dilemma:

$$b^s > a^s > d^s > c^s,$$

and

$$2a^s > b^s + c^s$$

for any $s \in \{s_1, s_2\}$. The latter condition entails that, if players maximize the sum of their payoffs, they prefer the action profile (C, C) to profiles (D, C) or (C, D) in any state. This condition is not strictly necessary but aims at focusing on players who try to sustain (C, C) in each time period (Harrington and Zhao, 2012). We assume perfect monitoring: all players observe the occurring state and the history of any time period.

The following assumption ensures that, in the one-shot game, individual deviation from action profile (C, C) is more profitable in state $s = s_2$ than in state $s = s_1$.

Assumption 1 *Let* $b^{s_2} - a^{s_2} \geqslant b^{s_1} - a^{s_1}$.

The game may start with any initial state from $\{s_1, s_2\}$. The transition from a particular state does not depend on time period and the action profile realized in the state. The probability that the game transits from state s to state s_1 (s_2) is equal to π^s $(1 - \pi^s)$. A transition from a state to another one may be interpreted as a shock in the economy.

2.2 Publicly known discount factors

First, we consider the game where players' discount factor is public information. This analysis is convenient to later define the players' classes according to their intrinsic degree of patience (see Section 2.2.3). Moreover, the strategy profiles for the game with complete information are also used in the second phase of the Bayesian game, when players act as if they know the discount factor of their competitors (see Section 3 for details).

2.2.1 Strategy profiles

In this section we consider some behaviour strategy profiles which will be examined in section 2.2.2 and the conditions for which these strategy profiles are subgame perfect. We focus the analysis on pure strategies. The behaviour strategy profile is denoted by $\sigma = (\sigma_\nu : \nu \in \{i, j\})$, where strategy σ_ν determines player ν's action for any time period and any state depending on the history of the stage.

We restrict our analysis considering three behaviour strategy profiles in which players (i) play action D in any state forever (non-cooperative strategy profile $\sigma_{\mathbf{n}}$), (ii) cooperate only in state s_1, deviate in state s_2 and transit to playing action D forever if they observe a deviation from the described behaviour in the history (semi-cooperative strategy profile $\sigma_{\mathbf{sc}}$), (iii) cooperate by playing action C in any state and transit to playing action D forever if they observe deviation in the history (cooperative strategy profile $\sigma_{\mathbf{c}}$). The formal definition of these strategy profiles is given in Appendix A.

2.2.2 Expected payoffs

Let δ_ν denote the discount factor of a player ν's payoff. We are interested in finding conditions for which the pure strategy profiles described above are subgame perfect. The strategy profile is subgame perfect if, for any time period and any state, a vector of restricted strategies form the Nash equilibrium in the subgame.

A player's discounted payoff when a strategy profile σ is implemented is

$$V(\sigma, \delta_\nu) = \sum_{t=1}^{\infty} \delta_\nu^{t-1} \Pi^{t-1} U_{\nu,t}^s(\sigma),$$

where $V(\sigma, \delta_\nu) = (V^{s_1}(\sigma, \delta_\nu), V^{s_2}(\sigma, \delta_\nu))'$, $U_{\nu,t}^s(\sigma) = (u_{\nu,t}^{s_1}(\sigma), u_{\nu,t}^{s_2})'(\sigma)$, with $()'$ representing the transpose matrix, $u_{\nu,t}^s(\sigma)$ is the payoff of the player ν in time period t and state s, corresponding to the strategy profile σ, and

$$\Pi = \begin{pmatrix} \pi^{s_1} & 1 - \pi^{s_1} \\ \pi^{s_2} & 1 - \pi^{s_2} \end{pmatrix}.$$

We also define vector $\mathbf{p}^s = (\pi^s, 1 - \pi^s)$ for every $s \in \{s_1, s_2\}$. We denote the discounted payoffs of player ν in equilibria $\sigma_{\mathbf{n}}$, $\sigma_{\mathbf{sc}}$ and $\sigma_{\mathbf{c}}$ as $V_{\mathbf{n}}^s(\delta_\nu)$, $V_{\mathbf{sc}}^s(\delta_\nu)$ and $V_{\mathbf{c}}^s(\delta_\nu)$, respectively, where subscripts $\mathbf{n}$, $\mathbf{sc}$ and $\mathbf{c}$ stand for "non-cooperative", "semi-cooperative" and "cooperative" equilibrium, while superscript $s \in \{s_1, s_2\}$ indicates the state of the game in the first period. In Appendix A we provide the formal derivation and we prove the following preliminary result.

Lemma 1 *For any $\delta_\nu \in (0, 1)$, $V_{\mathbf{c}}(\delta_\nu) > V_{\mathbf{sc}}(\delta_\nu) > V_{\mathbf{n}}(\delta_\nu)$.*

We are now in a position to examine the critical value of δ for which each strategy profile is a subgame perfect equilibrium. We also define

$$\delta_1^* \equiv \frac{\Delta_1 - \pi^{s_1}(b^{s_1} - d^{s_1}) - (1 - \pi^{s_2})(b^{s_1} - a^{s_1})}{2(\pi^{s_2} - \pi^{s_1})(b^{s_1} - d^{s_1})}, \tag{2}$$

$$\delta_2^* \equiv \frac{\Delta_2 - \pi^{s_2}(a^{s_1} - d^{s_1}) - (1 - \pi^{s_2})(b^{s_2} - d^{s_2}) - \pi^{s_1}(b^{s_2} - a^{s_2})}{2(\pi^{s_2} - \pi^{s_1})(b^{s_2} - d^{s_2})} \tag{3}$$

where $\delta_1^* \in (0, 1)$, $\delta_2^* \in (0, 1)$, and

$$\begin{aligned}
\Delta_1 &= \left((\pi^{s_1}(b^{s_1} - d^{s_1}) + (1 - \pi^{s_2})(b^{s_1} - a^{s_1}))^2 + 4(b^{s_1} - a^{s_1})(\pi^{s_2} - \pi^{s_1})(b^{s_1} - d^{s_1})\right)^{1/2}, \\
\Delta_2 &= \left((\pi^{s_2}(a^{s_1} - d^{s_1}) + (1 - \pi^{s_2})(b^{s_2} - d^{s_2}) + \pi^{s_1}(b^{s_2} - a^{s_2}))^2 \right. \\
&\quad \left. + 4(b^{s_2} - a^{s_2})(\pi^{s_2} - \pi^{s_1})(b^{s_2} - d^{s_2})\right)^{1/2}.
\end{aligned}$$

The next proposition summarises the conditions on the discount factors for which each particular strategy profile is a subgame perfect Nash equilibrium (SPNE).

Proposition 1 *Let Assumption 1 hold and player $\nu \in \{i, j\}$ have a discount factor δ_ν which is public information. The cooperative strategy profile is SPNE iff $\delta_\nu \geqslant \delta_2^*$ for any $\nu \in \{i, j\}$. A semi-cooperative strategy profile is SPNE iff $\delta_\nu \geqslant \delta_1^*$ for any $\nu \in \{i, j\}$. A non-cooperative strategy profile is SPNE for any $\delta_\nu \in (0, 1)$, $\nu \in \{i, j\}$.*

Depending on the parameters of the model, there are either $\delta_1^* < \delta_2^*$ or $\delta_1^* \geqslant \delta_2^*$. If $\delta_1^* < \delta_2^*$, there are three possible combinations of the players' i and j discount factors:

- $\delta_i \in (0, \delta_1^*)$, $\delta_j \in (0, 1)$: only $\sigma_{\mathbf{n}}$ is SPNE;
- $\delta_i \in [\delta_1^*, \delta_2^*)$, $\delta_j \in [\delta_1^*, 1)$: both $\sigma_{\mathbf{n}}$ and $\sigma_{\mathbf{sc}}$ are SPNE, but $V_{\mathbf{sc}}(\delta_\nu) > V_{\mathbf{n}}(\delta_\nu)$ by Lemma 1;
- $\delta_i \in [\delta_1^*, 1)$, $\delta_j \in [\delta_1^*, 1)$: strategy profiles $\sigma_{\mathbf{n}}$, $\sigma_{\mathbf{sc}}$, $\sigma_{\mathbf{c}}$ are SPNE. Again by Lemma 1, the payoffs in cooperative strategy profile are the largest ones.

If $\delta_1^* \geqslant \delta_2^*$, the interval $(\delta_1^*, 1]$, where the semi-cooperative strategy profile is SPNE, is contained into $(\delta_2^*, 1]$, where also the cooperative strategy profile is SPNE. Hence, by Lemma 1, a semi-cooperative strategy profile is never played because the players' payoffs in $\sigma_{\mathbf{sc}}$ are less than in $\sigma_{\mathbf{c}}$. In this case, the analysis will be the same as in Harrington and Zhao (2012) where, in the interval $(0, \delta_2^*)$, only $\sigma_{\mathbf{n}}$ is SPNE (unique in the set $\{\sigma_{\mathbf{n}}, \sigma_{\mathbf{sc}}, \sigma_{\mathbf{c}}\}$), while in interval $(\delta_2^*, 1)$ the equilibrium with the largest players' payoffs is $\sigma_{\mathbf{c}}$.[2] We then focus on the first case, and we assume the following:

Assumption 2 *Let $\delta_2^* > \delta_1^*$.*

2.2.3 Player's classes

Based on Proposition 1 and Assumption 2, we define the players' classes according to their discount factors, as follows.

[2] To see this, consider three options of combination of the players i and j's discount factors:

- $\delta_i \in (0, \delta_2^*)$, $\delta_j \in (0, 1)$: only $\sigma_{\mathbf{n}}$ is SPNE.
- $\delta_i \in [\delta_2^*, \delta_1^*)$, $\delta_j \in [\delta_2^*, 1)$: both $\sigma_{\mathbf{n}}$ and $\sigma_{\mathbf{c}}$ are SPNE. But in the cooperative equilibrium the players' payoffs are larger than in the non-cooperative equilibrium.
- $\delta_i \in [\delta_1^*, 1)$, $\delta_j \in [\delta_1^*, 1)$: strategy profiles $\sigma_{\mathbf{n}}$, $\sigma_{\mathbf{sc}}$, $\sigma_{\mathbf{c}}$ are SPNE. The equilibrium with the largest players' payoffs is $\sigma_{\mathbf{c}}$.

Definition 1 *A player ν belongs to class ℓ_ν and there are three classes $\ell_\nu \in \{I, M, P\}$:*

i. I (impatient), whose discount factor is denoted as $\delta_\nu = \delta_I$ and satisfies $\delta_I \in (0, \delta_1^*)$;

ii. M (mildly patient), whose discount factor is denoted as $\delta_\nu = \delta_M$ and satisfies $\delta_M \in [\delta_1^*, \delta_2^*)$;

iii. P (patient), whose discount factor is denoted as $\delta_\nu = \delta_P$ and satisfies $\delta_P \in [\delta_2^*, 1)$.

The cooperative strategy profile is SPNE if and only if both players are of class P. Otherwise, a player who is not of class P will deviate from the cooperative strategy profile because the deviation is profitable. A semi-cooperative strategy profile is SPNE if (i) both players are of class P, (ii) both players are of class M or (iii) if one is of class P and the other is of class M. If at least one of two players is of class I, then neither a cooperative nor a semi-cooperative strategy profile is SPNE.

3 Unknown discount factors

We now turn the analysis on the case where a player's discount factor is private information. Possibly, this game may exhibit several classes of equilibria. We focus our analysis on some equilibria for the game which consist on two phases (Harrington and Zhao, 2012).

The first phase is *learning*, where the players' discount factors are private information and players try to recognize the competitor's class. In this phase players' strategies are Markovian: they are based on beliefs on the competitor's class, and not on the game history. The second phase is *cooperation*, where any player uses a behaviour strategy from the set $\{\sigma_{\mathbf{c},\nu}, \sigma_{\mathbf{sc},\nu}, \sigma_{\mathbf{n},\nu}\}$. The strategy chosen during the cooperation phase is determined by the beliefs on the competitor's class at the last period of the learning phase.

3.1 Learning phase: strategies and rules for updating beliefs

In this section, we describe how the process of learning the competitor's class takes place. In time period t, a player believes the other player to be of class P with probability α_t, to be of class M with probability β_t, and to be of class I with probability $\gamma_t = 1 - \alpha_t - \beta_t$. We define the strategies for every $t = 1, \dots, T$: the player chooses her strategy in period t based on beliefs α_t and β_t. We assume that the initial beliefs about the other player's class in period 1 are given and known, and $\alpha_1 \in (0, 1)$, $\beta_1 \in (0, 1)$ are known and $\alpha_1 + \beta_1 \in (0, 1)$.

We denote player ν's strategy in the learning phase as ψ_ν and the set of player ν's strategies in the learning phase is $\Psi(\ell_\nu)$ which depends on the player's class $\ell_\nu \in \{I, M, P\}$. The set of Markovian strategies of a player of class P is $\Psi(P) = \{q_t^s, t = 1, \ldots, T, s = s_1, s_2\}$, where $q_t^s : [0,1]\times[0,1] \to [0,1]$ is a function of α_t and β_t and corresponds to the probability of choosing action C in state s in time period t. Conversely, the set of Markovian strategies of a player of class M is $\Psi(M) = \{r_t^s, t = 1, \ldots, T, s = s_1, s_2\}$. Notice that $r_t^{s_1} : [0,1]\times[0,1] \to [0,1]$ is a function of α_t and β_t and amounts to the probability of choosing action C in state s_1 in time period t. For state s_2, $r_t^{s_2} = 0$ for any t, i.e. the strategy $r_t^{s_2}$ prescribes player of class M to choose action D in state s_2 with probability 1 in any time period. Finally, we define the set of strategies for an impatient player ν, $\ell_\nu = I$. The set of Markovian strategies of the I player in the learning phase is $\Psi(I) = \{z_t^s, t = 1, \ldots, T, s = s_1, s_2\}$, where $z_t^s : [0,1] \times [0,1] \to 0$ for any period t and any state s. Strategy z_t^s prescribes player of class I to choose action D in any state $s = s_1, s_2$ with probability 1 in any time period.

The player ν's strategy in the learning phase is $\psi(\ell_\nu)$ and determines a probability of choosing action C in any time period t and any state s. It depends on player's class ℓ_ν as follows:

$$\psi(\ell_\nu) = \begin{cases} q_t^s \in [0,1] & \text{if } \ell_\nu = P, \quad s = s_1, s_2; \\ r_t^s \in [0,1] & \text{if } \ell_\nu = M, \quad s = s_1; \\ r_t^s = 0 & \text{if } \ell_\nu = M, \quad s = s_2; \\ z_t^s = 0 & \text{if } \ell_\nu = I, \quad s = s_1, s_2. \end{cases}$$

Based on the definition of the players' strategies in the learning phase, we may state that, if a player chooses action C in state s_2, she has revealed her class as P, because she is the only class who may choose action C in state s_2 with positive probability. Conversely, if a player chooses action C in state s_1, she may be identified as a P or M class because only players of these two classes may choose action C in state s_1 with positive probability.

We use Bayes rule to update beliefs α_t and β_t over time, $t = 2, \ldots, T+1$. The rule of defining beliefs for period $t+1$ depends on the state s which appeared in period t. If state s is realized at period t, then in the next period $t+1$, the belief that the competitor is of class P is

$$\alpha_{t+1} = \begin{cases} \alpha_{t+1}^{s_1}, & \text{if } s = s_1, \\ \alpha_{t+1}^{s_2}, & \text{if } s = s_2. \end{cases}$$

The same rule applies to β_{t+1} and γ_{t+1}.

First, consider the updating rule for state $s = s_1$. If a player chooses C in period t, she is identified as class[3]

$$\begin{cases} I & \text{with prob. } \gamma_{t+1}^{s_1} = 0; \\ P & \text{with prob. } \alpha_{t+1}^{s_1} = \dfrac{\alpha_t q_t^{s_1}}{\alpha_t q_t^{s_1} + \beta_t r_t^{s_1}}; \\ M & \text{with prob. } \beta_{t+1}^{s_1} = 1 - \alpha_{t+1}^{s_1} = \dfrac{\beta_t r_t^{s_1}}{\alpha_t q_t^{s_1} + \beta_t r_t^{s_1}}. \end{cases} \tag{4}$$

If a player chooses D in state $s = s_1$, she is identified as class[4]

$$\begin{cases} I & \text{with prob. } \gamma_{t+1}^{s_1} = 1 - \alpha_{t+1}^{s_1} - \beta_{t+1}^{s_1} = \dfrac{1 - \alpha_t - \beta_t}{1 - \alpha_t q_t^{s_1} - \beta_t r_t^{s_1}}; \\ P & \text{with prob. } \alpha_{t+1}^{s_1} = \dfrac{\alpha_t (1 - q_t^{s_1})}{1 - \alpha_t q_t^{s_1} - \beta_t r_t^{s_1}}; \\ M & \text{with prob. } \beta_{t+1}^{s_1} = 1 - \alpha_{t+1}^{s_1} = \dfrac{\beta_t (1 - r_t^{s_1})}{1 - \alpha_t q_t^{s_1} - \beta_t r_t^{s_1}}. \end{cases} \tag{5}$$

Consider next the updating rule for state s_2. As already discussed, if a player chooses C, she is identified as class P with probability 1. On the other hand, if a player chooses D, she is identified as class[5]

$$\begin{cases} I & \text{with prob. } \gamma_{t+1}^{s_2} = 1 - \alpha_{t+1}^{s_2} - \beta_{t+1}^{s_2} = \dfrac{1 - \alpha_t - \beta_t}{1 - \alpha_t q_t^{s_2}}; \\ P & \text{with prob. } \alpha_{t+1}^{s_2} = \dfrac{\alpha_t (1 - q_t^{s_2})}{1 - \alpha_t q_t^{s_2}}; \\ M & \text{with prob. } \beta_{t+1}^{s_2} = \dfrac{\beta_t}{1 - \alpha_t q_t^{s_2}}. \end{cases} \tag{6}$$

3.2 Cooperation phase

We now determine how players choose strategies in the cooperation phase. The strategy of player ν in a cooperation phase is a mapping from a player's class and beliefs on the other player's class by the end of the learning phase to the set $\{\sigma_{\mathbf{c},\nu}, \sigma_{\mathbf{sc},\nu}, \sigma_{\mathbf{n},\nu}\}$. Let T be the last period of the learning phase, which we assume finite.[6]

At the end of the learning phase, the beliefs that the competitor is of class P or M are

[3] The probabilities $\alpha_{t+1}^{s_1}$ and $\beta_{t+1}^{s_1}$ are defined if $\alpha_t q_t^{s_1} + \beta_t r_t^{s_1} \neq 0$. In the case when $\alpha_t q_t^{s_1} + \beta_t r_t^{s_1} = 0$, it is impossible to observe action C in state $s = s_1$.

[4] The probabilities $\alpha_{t+1}^{s_1}$, $\beta_{t+1}^{s_1}$ and $\gamma_{t+1}^{s_1}$ are defined if $\alpha_t q_t^{s_1} + \beta_t r_t^{s_1} \neq 1$. In the case when $\alpha_t q_t^{s_1} + \beta_t r_t^{s_1} = 1$, it is impossible to observe action D in state $s = s_1$.

[5] The probabilities $\alpha_{t+1}^{s_2}$, $\beta_{t+1}^{s_2}$ and $\gamma_{t+1}^{s_2}$ are defined if $\alpha_t q_t^{s_2} \neq 1$. In the case when $\alpha_t q_t^{s_2} = 1$ ($\alpha_t = q_t^{s_2} = 1$), it is impossible to observe action D in state $s = s_2$.

[6] Here we do not consider the case when the learning phase last forever.

α_{T+1} and β_{T+1}, respectively. Thus the strategy of player ν is a function of class ℓ_ν and beliefs α_{T+1}, β_{T+1} such that:

$$\sigma_\nu(\ell_\nu, \alpha_{T+1}, \beta_{T+1}) = \begin{cases} \sigma_{\mathbf{c},\nu}, & \text{if } \ell_\nu = P, \quad \alpha_{T+1} = 1, \\ \sigma_{\mathbf{sc},\nu}, & \text{if } \{\ell_\nu = P, \quad \alpha_{T+1} + \beta_{T+1} = 1, \quad \alpha_{T+1} \neq 1\} \\ & \text{or } \{\ell_\nu = M, \quad \alpha_{T+1} + \beta_{T+1} = 1\}, \\ \sigma_{\mathbf{n},\nu}, & \text{if } \{\ell_\nu \in \{P, M\}, \quad \alpha_{T+1} + \beta_{T+1} < 1\} \\ & \text{or } \{\ell_\nu = I\}. \end{cases} \tag{7}$$

From the rule (7) it follows that the impatient player always chooses strategy $\sigma_{\mathbf{n},\nu}$, a mildly patient player chooses a semi-cooperative strategy $\sigma_{\mathbf{sc},\nu}$ if he has a belief that the competitor is not an impatient player, and he chooses non-cooperative strategy $\sigma_{\mathbf{n},\nu}$ otherwise. A patient player chooses cooperative strategy $\sigma_{\mathbf{c},\nu}$ only if her belief that the competitor is patient equals to one. She chooses semi-cooperative strategy $\sigma_{\mathbf{sc},\nu}$ if believes that the competitor is not an impatient player, but her belief that the competitor of class P does not equal to one. Otherwise, a patient player chooses the non-cooperative strategy $\sigma_{\mathbf{n},\nu}$.

3.3 Payoff and equilibrium concept

The payoff of player $\nu = i, j$ whose class is $\ell_\nu \in \{I, M, P\}$ is the sum of her payoffs in the two phases of the game and it is a function of her class, initial beliefs and the strategies of player i and j in the learning phase:[7]

$$\begin{aligned} \Phi_\nu(\psi_i, \psi_j, \alpha_1, \beta_1 | \ell_\nu) = & \sum_{t=1}^{T} \delta_\nu^{t-1} \Pi^{t-1} U_\nu(\psi_{i,t}, \psi_{j,t}) \\ & + \delta_\nu^T \Pi^T V(\delta_\nu, (\sigma_i(\ell_i, \alpha_{T+1}, \beta_{T+1}), \sigma_j(\ell_j, \alpha_{T+1}, \beta_{T+1}))), \end{aligned} \tag{8}$$

where $\psi_\nu \in \Psi(\ell_\nu)$. The first part in the RHS of (8) is the payoff during the learning phase, while the second part is the payoff during the cooperation phase. In (8), $\sigma_i(\ell_i, \alpha_{T+1}, \beta_{T+1})$ and $\sigma_j(\ell_j, \alpha_{T+1}, \beta_{T+1})$ are defined by (7).

We are now equipped to define the equilibrium concept of the game. The equilibrium concept is perfect Bayesian equilibrium (Fudenberg and Tirole, 1991). In a perfect Bayesian

[7] We omit players' strategies in a cooperation phase as the arguments of the function because they are uniquely defined by rule (7) given strategies ψ_i, ψ_j.

equilibrium for each period $t = 1, \ldots, T$ and any history of this period, the continuation strategies are a Bayesian equilibrium for the continuation game. Formally, we define

Definition 2 *A strategy profile $\psi^* = (\psi_i^*, \psi_j^*)$ is a perfect Bayesian equilibrium if for each player $\nu \in \{i, j\}$, any class $\ell_\nu \in \{M, P\}$, any initial beliefs $\alpha_1 \in (0,1)$, $\beta_1 \in (0,1)$, $\alpha_1 + \beta_1 \in (0,1)$ and each strategy $\psi_\nu \in \Psi(\ell_\nu)$, the inequality*

$$\Phi_\nu(\psi^*, \alpha_1, \beta_1 | \ell_\nu) \geqslant \Phi_\nu\left((\psi_\nu, \psi_{-\nu}^*), \alpha_1, \beta_1 | \ell_\nu\right) \tag{9}$$

holds.

The strategy set of a player in the two-phase equilibrium consists of the strategy in the learning and in the cooperation phases. The solution concept is close to Markov Perfect Bayesian Equilibrium (MPBE) with the following modification. The strategy of any player is Markovian[8] only during the learning phase when players' classes are not common knowledge, and in the cooperation phase players' strategies are behaviour from the set $\{\sigma_{\mathbf{c},\nu}, \sigma_{\mathbf{sc},\nu}, \sigma_{\mathbf{n},\nu}\}$, as described in (7). To avoid confusion, we use the name of the solution concept as Partial Markov Perfect Bayesian Equilibrium[9] (PMPBE) given in Harrington and Zhao (2012).

4 Cooperative outcomes

In this section we characterize equilibria according to which at the second phase of the game the cooperative or semi-cooperative strategy profiles may occur (depending on the players' classes). For completeness, the equilibria yielding a non-cooperative outcome are outlined in Appendix E. The rule of choosing the strategy in the cooperation phase is given by (7). In the first part of the section, we consider the case where the learning phase is limited to one period. Albeit this restriction is strong, it allows us to highlight some features of the equilibria that may be then found, in the second part, in the more general version where the length of the learning phase is endogenously determined.

[8] The Markov property is that the strategy in the learning phase in any time period t depends only on the beliefs on the competitor's class, and does not depend on the time period and the history.

[9] It is called "partial" because, in the cooperation phase, the strategies are not Markovian but they are behaviour once the learning phase ends up.

4.1 One-period learning phase

In this section we limit the length of the learning phase T to one time period. We present first this simplifying case for expositional purposes, as it helps highlighting the role of beliefs in this problem. Qualitatively similar results are obtained when we relax this assumption, but the analysis and the equilibrium conditions are more cumbersome. Nonetheless, the exogenous duration of the learning phase may be dictated by external conditions, e.g., the learning period may be costly and players are restricted by short duration of a learning phase.

We find the conditions when a Partially Markov Perfect Bayesian Equilibrium exists in the learning phase. We sort the equilibria by the type of equilibria adopted in the learning phase. For convenience, thresholds A_1, A_2, A_3 and A_4 are defined in Appendix C.

Proposition 2 *Consider a Prisoners' dilemma with the learning phase lasts for one time period. Let Assumptions 1 and 2 be satisfied, then there exists the following equilibria:*

1. *If the initial state is $s = s_1$:*

   1.i $(q_1^{s_1}, r_1^{s_1}) = (1, 0)$ *is a PMPBE for* $\alpha_1 \in [A_1; A_2]$.

   1.ii $(q_1^{s_1}, r_1^{s_1}) = (1, 1)$ *is a PMPBE for* $\alpha_1 + \beta_1 \geqslant A_3$.

2. *If the initial state is $s = s_2$:*

   2.i $(q_1^{s_2}) = 1$ *is a PMPBE for* $\alpha_1 \geqslant A_4$.

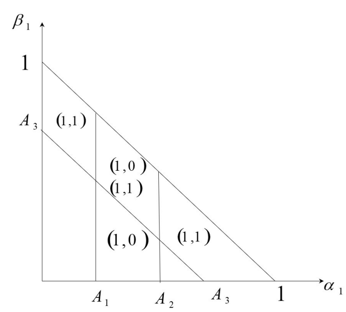

Figure 1: Equilibria with initial state $s = s_1$.

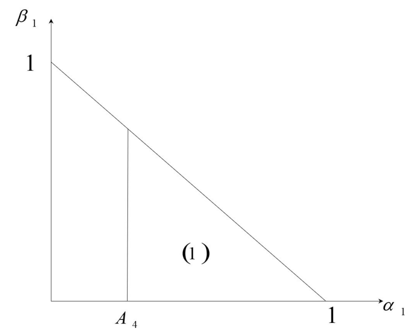

Figure 2: Equilibria with initial state $s = s_2$.

Figures 1 and 2 depict the regions of PMPBE for initial state s_1 and s_2, respectively, in the space of initial beliefs (α_1, β_1). The rule of updating beliefs in (4) and (5) helps to understand the strategy profile in the cooperation phase. Suppose, for instance, that the game starts with state s_1 and profile $(q_1^{s_1}, r_1^{s_1}) = (1, 0)$ is chosen. If action C is observed, the updated beliefs are $\alpha_2 = 1$, $\beta_2 = \gamma_2 = 0$, thus it is possible to recognize the competitor's class as P. Hence, if two players of class P meet, the equilibrium $(q_1^{s_1}, r_1^{s_1}) = (1, 0)$ leads to the cooperative strategy profile $\sigma_{\mathbf{c}}$ in the cooperation phase. If even one of the two players is not of class P, equilibrium $(q_1^{s_1}, r_1^{s_1}) = (1, 0)$ implies that the non-cooperative strategy profile $\sigma_{\mathbf{n}}$ will be implemented in the cooperation phase. Indeed, since players belonging to the mildly patient and the impatient class adopt the same strategy, a patient player cannot recognize from the learning phase if the competitor is a mildly patient one, thus the semi-cooperative strategy is never used in the cooperation phase.

When the game starts at state s_1 and profile $(q_1^{s_1}, r_1^{s_1}) = (1, 1)$ is implemented, i.e., players of class M and P cooperate with probability 1, the beliefs of a competitor's class after observing C are:

$$\alpha_2 = \frac{\alpha_1}{\alpha_1 + \beta_1}, \quad \beta_2 = \frac{\beta_1}{\alpha_1 + \beta_2}, \quad \gamma_2 = 0.$$

In this case there are positive probabilities that the competitor is either P or M. Hence the strategy of the players P or M during the second phase is semi-cooperative one $\sigma_{\mathbf{sc},\boldsymbol{\nu}}$ according to the rule (7), which allows cooperation in future states s_1 and deviation in future states s_2. This result emerges as a player does not recognize whether the competitor is of class P or M.

When the game starts with state s_2, a player of class M cannot be identified, since she deviates. Hence the belief β_1 does not play any role in determining the equilibrium. However, a class P competitor is identified with certainty. Hence if the players are both P, they choose cooperative strategies $\sigma_{\mathbf{c},\boldsymbol{\nu}}$ in the cooperation phase.

The next corollary compares the equilibrium payoffs in the parameter ranges where multiple equilibria occur, as a possible refinement in the equilibrium choice.

Corollary 1 *Suppose the game starts from state $s = s_1$, and α_1, β_1 satisfy the conditions: $\alpha_1 \in [A_1, A_2]$ and $\alpha_1 + \beta_1 \in [A_3, 1]$. Then the payoff of a class M player in equilibrium $(q_1^{s_1}, r_1^{s_1}) = (1, 1)$ is always not less than his payoff in equilibrium $(q_1^{s_1}, r_1^{s_1}) = (1, 0)$.*

Corollary 1 suggests equilibrium $(1, 1)$ as a refinement of multiple equilibria in state s_1, and the pure strategy over the mixed one in state s_2. In particular for state s_1, this result intuitively suggests that, when the beliefs that the competitor is P or M are similar, it is unlikely to reach a result of full cooperation. Indeed, the outcome is a semi-cooperative strategy profile in the cooperation phase.

The next proposition summarises some comparative statics on the equilibrium payoffs with respect to beliefs.

Proposition 3 *The equilibrium payoffs of classes P and M players are increasing functions of α_1. The payoffs of classes P and M players in equilibrium $(q_1^{s_1}, r_1^{s_1}) = (1, 1)$ are increasing functions of β_1.*

Proposition 3, together with Proposition 2, state a surprising result: a strong belief that the competitor is of class P does not lead to a cooperative strategy profile (fully collusive equilibrium) in the cooperation phase. This is immediately evident by looking at Figure 1. A very high α_1 gives a strong incentive to an M class player to fake patience, that is, it induces to act as a P class in order to lure the competitor to choose a cooperative strategy in the second phase. Indeed, if a cooperative strategy is played by a player of class P and state s_2 occurs at some period, then the M class player would deviate from cooperation, thus tricking her competitor. Given that players are aware of the "faking patience" effect, a semi-cooperative equilibrium occurs: cooperation in state s_1, non-cooperation in state s_2.

4.2 Endogenous learning phase

In this section we generalise the previous results by endogenising the duration of the learning phase. Several equilibria emerge: in what follows, we aim at showing that the faking patience effect may occur for some configurations. We restrict our attention to those strategies that allow to identify the class of any player in the shortest number of periods.

4.2.1 Initial state $s = s_1$

The natural structure of strategy profiles satisfying our requirement is the following. In the first time period, players of both class P and M use strategy C to be sure if there is an I player participating in the game. Then, if there are no I players, the game transits to state s_1 or s_2 in which players of class P and M use different strategies to be revealed in time period 2, i.e., their classes will be identified with probability 1. Now we find the conditions under which the described strategy profiles are PMPBE. Further we consider the strategy profile according to which in initial state s_1 players of classes P and M use different strategies in $t = 1$.

Assume that, in the first period, players of classes P and M adopt strategies $q_1^{s_1} = r_1^{s_1} = 1$. In period 2 and

- $s = s_1$, strategies are $q_2^{s_1} = 1$, $r_2^{s_1} = 0$;
- $s = s_2$, player P's strategy is $q_2^{s_2} = 1$.

Using these strategies players' classes are revealed not later than in period 2. The following proposition summarises the conditions on the initial beliefs for which the described strategies form a PMPBE. To ease the exposition, coefficients A_5, A_6, A_7 and A_8 are defined in Appendix F, with $A_5, A_7 > 0$.

Proposition 4 *Let the initial state be $s = s_1$, and suppose that the following conditions hold:*

$$\begin{cases} i. & A_5\alpha_1 + A_6\beta_1 \geqslant d^{s_1} - c^{s_1}, \\ ii. & A_7\alpha_1 + A_8\beta_1 \geqslant d^{s_1} - c^{s_1}, \\ iii. & \dfrac{\beta_1}{\alpha_1} \leqslant \min\limits_{s=s_1,s_2} \left\{ \dfrac{\delta_P^s \mathbf{p}^s (V_{\mathbf{c}}(\delta_P) - V_{\mathbf{sc}}(\delta_P)) - (b^s - a^s)}{d^s - c^s} \right\}, \\ iv. & \dfrac{\beta_1}{\alpha_1} \geqslant \dfrac{\delta_M \mathbf{p}^{s_1} (V_{\mathbf{d}}(\delta_M) - V_{\mathbf{sc}}(\delta_M)) - (b^{s_1} - a^{s_1})}{d^{s_1} - c^{s_1}}. \end{cases}$$

Then the following strategies are PMPBE:

$$P: \ (q_1^{s_1}, q_2^{s_1}, q_2^{s_2}) = (1,1,1), \ M: \ (r_1^{s_1}, r_2^{s_1}) = (1,0).$$

In the cooperation phase, a cooperative equilibrium occurs if two players are of class P, while a semi-cooperative equilibrium occurs if two players are of class M or one player is of class P and the other one is of class M.

The equilibrium described in Proposition 4 shows the emergence of faking patience in the first period, where a player of class M cooperates and, by doing so, does not reveal herself. On the other hand, in the second period the M-class player would deviate in state s_1, and by doing so she reveals her type and the learning phase ends afterwards, by playing the semi-cooperative equilibrium in the cooperative phase. Intuitively, the faking patience effect is also what delays entering in the cooperation phase. Suppose that a player of type M keeps playing $r_t^{s_1} = 1$ for all periods t until a change of state takes place. In this case the learning phase goes on until state s_2 occurs.

Unlike the example where the learning phase lasts one period, the conditions of Proposition 4 are harder to interpret. We may however take a closer look at coefficients of α_1 in conditions *i.* and *ii.*. They are unambiguously positive,[10] suggesting that an increase in α_1 increases the chance that the two conditions hold. In words, the higher the belief that the competitor is patient, the higher the change of faking patience, the less likely the reaching of full cooperation. By contrast, the coefficient of β_1 in *i.* and *ii.* is ambiguous, as well as those of α_1 and β_1 in *iii.* and *iv.*. To fix ideas, in Section 4.2.3 we verify this intuition through a numerical simulation.

4.2.2 Initial state $s = s_2$

In state s_2, a player of class M always chooses action D. We examine the strategy profile according to which a player of class P chooses action C, i.e., $q_1^{s_2} = 1$. Therefore, if at least one of two players chooses action C in period 1, the learning phase is over, and the cooperation phase starts from period 2. According to rule (7), if both players choose C, then in the cooperation phase players implement a cooperative strategy profile. By contrast, if one player chooses C and the other one chooses D, then players play a non-cooperative strategy profile, according to (7).

If both players choose action D, then the learning phase transmits to period 2, and players can be either of type M or I, according to the following beliefs:

$$\alpha_2 = 0, \quad \beta_2 = \frac{\beta_1}{1-\alpha_1}, \quad \gamma_2 = \frac{1-\alpha_1-\beta_1}{1-\alpha_1}.$$

[10] See Appendix F.

Notice that these classes of players keep playing D until state s_1 is realized because of the Markovian property of strategies. Consider then the case where state s_1 occurs in the second period. A player of class I keeps playing D. On the other hand, a player of class M may choose action C (strategy $r_2^{s_1} = 1$) or action D (strategy $r_2^{s_1} = 0$). If he uses strategy $r_2^{s_1} = 0$, the beliefs remain the same and the strategy $r_t^{s_1}$ will be equal 0 until infinity because of the Markovian property[11]. Thus we focus on the conditions for which strategy $r_2^{s_1} = 1$ is a part of PMPBE.

Proposition 5 *Let the game start with state $s = s_2$, and suppose*

$$\begin{cases} \beta_1 \geqslant \dfrac{(d^{s_1} - c^{s_1})(1-\alpha_1)}{d^{s_1} - c^{s_1} + a^{s_1} - b^{s_1} + \delta_M \mathbf{p}^{s_1}(V_{\mathbf{sc}}(\delta_M) - V_{\mathbf{n}}(\delta_M))}, \\ \beta_1 \leq \dfrac{\alpha_1[d^{s_2} - c^{s_2} + a^{s_2} - b^{s_2} + \delta_P \mathbf{p}^{s_2}(V_{\mathbf{c}}(\delta_P) - V_{\mathbf{n}}(\delta_P))] - (d^{s_2} - c^{s_2})}{\delta_P \mathbf{p}^{s_2}(V_{\mathbf{sc}}(\delta_P) - V_{\mathbf{n}}(\delta_P))}. \end{cases}$$

Then the following strategies are PMPBE:

- *Class P players: $q_1^{s_2} = 1$. The learning phase is over at $t = 1$ if at least one of the players is of class P.*

- *Class M players: $r_k^{s_2} = 0$ from $k = 1$ onwards until $s = s_2$; $r_t^{s_1} = 1$, once $s = s_1$ at time $t > 1$. The learning phase is over at t if at least one of the two players is of class M.*

4.2.3 Numerical example

Given the limited tractability of this framework, we highlight the features of the equilibria in Propositions 4 and 5 using a numerical simulation. Consider the game represented by matrices

$$s = s_1: \begin{array}{cc} & \begin{array}{cc} C & D \end{array} \\ \begin{array}{c} C \\ D \end{array} & \begin{pmatrix} (11.3, 11.3) & (3, 12) \\ (12, 3) & (5, 5) \end{pmatrix} \end{array}, \quad s = s_2: \begin{array}{cc} & \begin{array}{cc} C & D \end{array} \\ \begin{array}{c} C \\ D \end{array} & \begin{pmatrix} (11, 11) & (3, 16) \\ (16, 3) & (10, 10) \end{pmatrix} \end{array},$$

for which Assumption 1 is true. Let the probabilities to transit from state s_1 to state s_1 and state s_2 be $\pi_1 = 0.8$ and $1 - \pi_1 = 0.2$ respectively, and the probabilities to transit from state s_2 to state s_1 and state s_2 be $\pi_2 = 0.1$ and $1 - \pi_2 = 0.9$ respectively. Assumption 2 is

[11] Hence, this strategy of player M is a part of a PMBE with infinite learning phase.

also true: we obtain the discount factors $\delta_1^* \approx 0.1245$ and $\delta_2^* \approx 0.7089$. Let players M and P have the discount factors $\delta_M = 0.7$ and $\delta_P = 0.9$ respectively.

If the game starts from state $s = s_1$, conditions of Proposition 4 amounts to

$$\begin{cases} 32.2162\alpha_1 + 20.1154\beta_1 \geqslant 2, \\ 11.4673\alpha_1 + 6.70729\beta_1 \geqslant 2, \\ \dfrac{\beta_1}{\alpha_1} \leqslant \min\{2.08243, 0.223938\}, \\ \dfrac{\beta_1}{\alpha_1} \geqslant 0.145989, \end{cases}$$

then the strategy profile $q_1^{s_1} = 1$, $r_1^{s_1} = 1$ in period 1 and $q_2^{s_2} = 1$ in period 2 is PMPBE. The region of (α_1, β_1) for which the system is satisfied (yellow color) is depicted on Fig. 3, case (i), and it is where the "faking patience effect" does not allow to reach a cooperation or semi-cooperation. The region of existence exhibits a combination of high values of α_1 and low values of β_1: a higher belief that the other player is of class P and not M is indeed what gives the incentive of a player of class M to fake patience.

If the game starts from state $s = s_2$, conditions of Proposition 5 are equivalent to

$$\begin{cases} \dfrac{1 - \alpha_1 - \beta_1}{\beta_1} \leqslant 4.11765, \\ 23.8919\alpha_1 - 15.3243\beta_1 \geqslant 7. \end{cases}$$

The range of parameters for which the strategy profile $q_1^{s_2} = 1$ and $r_t^{s_1} = 1$ for any $t > 1$ in the learning phase given in Proposition 5 is PMPBE in the game starting with $s = s_2$ is depicted in Fig. 3, case (ii) (orange area).

(i) The game starts with state $s = s_1$. (ii) The game starts with state $s = s_2$.

Figure 3: Equilibrium region.

5 Concluding remarks

In this paper we have analysed cooperation in an infinitely repeated prisoners' dilemma where a player's discount factor is private information. We have shown that the presence of different states of the world drastically affects the strategic role of beliefs. When the learning phase is limited to one period, a player that shifts from cooperation to deviation according to the state of the world has an incentive in faking patience in the good state. Since this behaviour is expected and increases with the belief in patience, the latter loses its role in determining cooperation. In case when the length of the learning phase is endogenously determined, the faking patience effect may still emerge. Given the multiplicity of equilibria, in the latter case we limit the attention to equilibrium configuration with pure strategies and where the cooperation phase starts as soon as possible.

An interesting extension might investigate the implementation of different strategy concepts. In the present analysis, we have considered grim trigger strategies. This class of strategies is used in folk theorems to prove the existence SPNE with cooperative outcomes. We have also referred to these class of strategies as they seemed to be natural to be used in the presence of incomplete information on the other player's discount factor (Maor and Solan, 2015). Future research may analyse equilibria using another trigger strategies such as tit-for-tat strategies (Axelrod and Hamilton, 1981), in which at every current stage the player chooses an action that the competitor played at the previous stage. In this case though, the profile of these strategies is not subgame perfect. Alternatively, the trigger strategies with limited number of punishing periods can also be used to construct the punishment of a deviating player.

References

[1] Abreu, D., Pearce, D, and Stacchetti, E. 1990. Toward a Theory of Discounted Repeated Games with Imperfect Monitoring. *Econometrica* **58**: 1041-1063.

[2] Athey, S. and Bagwell, K., 2008. Collusion with persistent cost shocks. *Econometrica* **76**: 493-540.

[3] Axelrod, R, and Hamilton, W. D., 1981. The evolution of cooperation. *Science* **211**: 1390-1396.

[4] Bagwell, K. and Staiger, R. 1997. Collusion over the business cycle. *RAND Journal of Economics* **28**: 82-106.

[5] Fudenberg, D. and Tirole, J. 1991. Perfect Bayesian equilibrium and sequential equilibrium. *Journal of Economic Theory*, **53**(2): 236-260.

[6] Fudemberg, D., Levine, D.K., and Maskin, E. 1994. The Folk Theorem with Imperfect Public Information. *Econometrica* **62**: 997-1040.

[7] Fudenberg, D. and Yamamoto, Y. 2010. Repeated games where the payoffs and monitoring structure are unknown. *Econometrica* **78**: 1673-1710.

[8] Fudenberg, D. and Yamamoto, Y. 2011. Learning from private information in noisy repeated games. *Journal of Economic Theory* **146**: 1733-1769.

[9] Gosh, P., and Ray, D. 1996. Cooperation in community interaction without information flows. *Review of Economic Studies* **63**: 491-519.

[10] Haltiwanger, J. and Harrington, J. 1991. The impact of cyclical demand movements on collusive behaviour. *RAND Journal of Economics* **22**: 89-106.

[11] Harrington Jr. J. E., Zhao W. 2012. Signaling and tacit collusion in an infinitely repeated Prisoners' Dilemma, *Mathematical Social Sciences* **64**: 277-289.

[12] Kandori, M., and Obara, I. 2006. Efficiency in Repeated Games Revisited: The Role of Private Strategies. *Econometrica* **74**: 499-519.

[13] Mailath, G. J. and Samuelson, L. 2006. Repeated Games and Reputations: Long-Run Relationships. Oxford. Oxford University Press.

[14] Maor, C. and Solan, E. 2015, Cooperation under incomplete information on the discount factors. *International Journal of Game Theory* **44**: 321-346.

[15] Peski, M. 2008. Repeated games with incomplete information on one side. *Theoretical Economics* **3**: 29-84.

[16] Peski, M. 2014. Repeated games with incomplete information and discounting. *Theoretical Economics* **9**: 651-694.

[17] Rotemberg, J. and Saloner, G. 1986. A supergame-theoretic model of price wars during booms. *American Economic Review* **76**: 390-407.

[18] Watson, J. 1999. Starting small and renegotiation. *Journal of Economic Theory* **85**: 52-90.

[19] Watson, J. 2002. Starting small and committment. *Games and Economic Behavior* **38**: 176-199.

[20] Wiseman, T. 2005. A partial folk theorem for games with unknown payoff distributions. *Econometrica* **73**: 629-645.

Appendix

Appendix A

Strategy profiles

The behaviour strategy profile is given by

$$\sigma = (\sigma_\nu : \nu \in \{i, j\}) \,. \tag{10}$$

In (10), $\sigma_\nu = \{\sigma^s_{\nu,t}\}_{t=1}^{\infty}$, where $\sigma^s_{\nu,t+1} : H(t) \to \mathbf{X}$ is an action of player ν in time period $t+1$ and state $s \in \{s_1, s_2\}$. $H(t) = ((s(1), x(1)), \ldots, (s(t), x(t)))$ is a history of time period t, where $s(t)$ is the state in time period t and $x(t)$ is the action profile played in state $s(t)$.

Definition 3 *A non-cooperative strategy of player ν is denoted as $\sigma_{\mathbf{n},\nu} = \{\sigma^s_{\nu,t}\}_{t=1,\ldots,\infty}^{s=s_1,s_2}$ such $\sigma^s_{\nu,t+1}(H(t)) = D$ for any $s = s_1, s_2$, $t = 1, \ldots, \infty$ and any history $H(t)$.*

We call the profile $\sigma_{\mathbf{n}} = (\sigma_{\mathbf{n},\nu} : \nu \in \{i, j\})$ as *non-cooperative strategy profile.*

Definition 4 *A semi-cooperative strategy of player ν is denoted as $\sigma_{\mathbf{sc},\nu} = \{\sigma^s_{\nu,t}\}_{t=1,\ldots,\infty}^{s=s_1,s_2}$ such that*

$$\sigma^s_{\nu,t+1}(H(t)) = \begin{cases} C, & \text{if } s = s_1 \text{ and } H(t) = H_{\mathbf{sc}}(t), \\ D, & \text{otherwise,} \end{cases}$$

while $H_{\mathbf{sc}}(t)$ is a history of time period t containing only the elements $(s_1, (C, C))$ and $(s_2, (D, D))$.

We call the profile $\sigma_{\mathbf{sc}} = (\sigma_{\mathbf{sc},\nu} : \nu \in \{i, j\})$ as *semi-cooperative strategy profile*, according to which players choose action C in state s_1 and action D in state s_2 if the deviation from history $H_{\mathbf{sc}}(t)$ is not observed. Otherwise, the players switch to playing action D in any state forever.

Definition 5 *A cooperative strategy of player ν is denoted as $\sigma_{\mathbf{c},\nu} = \{\sigma^s_{\nu,t}\}_{t=1,\ldots,\infty}^{s=s_1,s_2}$ such that*

$$\sigma^s_{\nu,t+1}(H(t)) = \begin{cases} C, & \text{if } H(t) = H_{\mathbf{c}}(t), \\ D, & \text{otherwise,} \end{cases}$$

and $H_{\mathbf{c}}(t) = ((s(1), (C, C)), \ldots, (s(t), (C, C)))$ is a history at time period t according to which both players choose action C in all time periods before $t + 1$.

We call the profile $\sigma_{\mathbf{c}} = (\sigma_{\mathbf{c},\nu} : \nu \in \{i, j\})$ as *cooperative strategy profile*, which prescribes players to choose action C in time period $t+1$ if the history shows past cooperation (i.e., no deviations are observed in the previous time periods). If a player observes deviation from action profile (C, C), then she chooses action D forever.

Expected payoffs

In this section we derive the value of the expected payoffs. For convenience, define

$$\widetilde{\Pi}(\delta_\nu) \equiv \frac{1}{(1-\delta_\nu)(1-\delta_\nu(\pi^{s_1}-\pi^{s_2}))} \begin{pmatrix} 1-\delta_\nu(1-\pi^{s_2}) & \delta_\nu(1-\pi^{s_1}) \\ \delta_\nu \pi^{s_2} & 1-\delta_\nu \pi^{s_1} \end{pmatrix}.$$

We can easily calculate the player's payoff in any equilibria $\sigma_{\mathbf{n}}$, $\sigma_{\mathbf{c}}$ or $\sigma_{\mathbf{sc}}$:

1. The discounted payoff of player ν in equilibrium $\sigma_{\mathbf{n}}$ is

$$V_{\mathbf{n}}(\delta_\nu) = \begin{pmatrix} V_{\mathbf{n}}^{s_1}(\delta_\nu) \\ V_{\mathbf{n}}^{s_2}(\delta_\nu) \end{pmatrix} = \widetilde{\Pi}(\delta_\nu) \begin{pmatrix} d^{s_1} \\ d^{s_2} \end{pmatrix}. \tag{11}$$

2. The discounted payoff of player ν in equilibrium $\sigma_{\mathbf{c}}$

$$V_{\mathbf{c}}(\delta_\nu) = \begin{pmatrix} V_{\mathbf{c}}^{s_1}(\delta_\nu) \\ V_{\mathbf{c}}^{s_2}(\delta_\nu) \end{pmatrix} = \widetilde{\Pi}(\delta_\nu) \begin{pmatrix} a^{s_1} \\ a^{s_2} \end{pmatrix}. \tag{12}$$

3. The discounted payoff of player ν in equilibrium $\sigma_{\mathbf{sc}}$

$$V_{\mathbf{sc}}(\delta_\nu) = \begin{pmatrix} V_{\mathbf{sc}}^{s_1}(\delta_\nu) \\ V_{\mathbf{sc}}^{s_2}(\delta_\nu) \end{pmatrix} = \widetilde{\Pi}(\delta_\nu) \begin{pmatrix} a^{s_1} \\ d^{s_2} \end{pmatrix}. \tag{13}$$

We obtain these formulas by calculating the payoff of player ν according to the profile definitions. The discounted payoff of player ν in equilibrium $\sigma_{\mathbf{n}}$ is

$$V_{\mathbf{n}}(\delta_\nu) = \begin{pmatrix} V_{\mathbf{n}}^{s_1}(\delta_\nu) \\ V_{\mathbf{n}}^{s_2}(\delta_\nu) \end{pmatrix} = \begin{pmatrix} d_{s_1} + \delta_\nu \mathbf{p}^{s_1} V_{\mathbf{n}}(\delta_\nu) \\ d_{s_2} + \delta_\nu \mathbf{p}^{s_2} V_{\mathbf{n}}(\delta_\nu) \end{pmatrix},$$

or in vectorial form:

$$V_{\mathbf{n}}(\delta_\nu) = \begin{pmatrix} d_{s_1} \\ d_{s_2} \end{pmatrix} + \delta \Pi V_{\mathbf{n}}(\delta_\nu).$$

This equation gives:

$$V_{\mathbf{n}}(\delta_\nu) = (\mathbb{I} - \delta_\nu \Pi)^{-1} \begin{pmatrix} d_{s_1} \\ d_{s_2} \end{pmatrix},$$

where $\mathbb{I}$ is an identity matrix of size 2×2. We denote $(\mathbb{I} - \delta_\nu \Pi)^{-1}$ by $\widetilde{\Pi}(\delta_\nu)$ and obtain the result. The inverse matrix $(\mathbb{I} - \delta_\nu \Pi)^{-1}$ always exists for any $\delta_\nu \in (0, 1)$.

Second, we calculate the discounted payoff of player ν in equilibrium $\sigma_{\mathbf{c}}$ that is:

$$V_{\mathbf{c}}(\delta_\nu) = \begin{pmatrix} V_{\mathbf{c}}^{s_1}(\delta_\nu) \\ V_{\mathbf{c}}^{s_2}(\delta_\nu) \end{pmatrix} = \begin{pmatrix} a^{s_1} \\ a^{s_2} \end{pmatrix} + \delta_\nu \begin{pmatrix} \pi^{s_1} & 1 - \pi^{s_1} \\ \pi^{s_2} & 1 - \pi^{s_2} \end{pmatrix} \begin{pmatrix} V_{\mathbf{c}}^{s_1}(\delta_\nu) \\ V_{\mathbf{c}}^{s_2}(\delta_\nu) \end{pmatrix}.$$

Rewriting this equation in vectorial form, we obtain equation (12).

Third, we calculate the discounted payoff of player player ν in equilibrium $\sigma_{\mathbf{sc}}$ that is:

$$V_{\mathbf{sc}}(\delta_\nu) = \begin{pmatrix} V_{\mathbf{sc}}^{s_1}(\delta_\nu) \\ V_{\mathbf{sc}}^{s_2}(\delta_\nu) \end{pmatrix} = \begin{pmatrix} a^{s_1} \\ d^{s_2} \end{pmatrix} + \delta_\nu \begin{pmatrix} \pi^{s_1} & 1 - \pi^{s_1} \\ \pi^{s_2} & 1 - \pi^{s_2} \end{pmatrix} \begin{pmatrix} V_{\mathbf{sc}}^{s_1}(\delta_\nu) \\ V_{\mathbf{sc}}^{s_2}(\delta_\nu) \end{pmatrix}.$$

Rewriting this equation in a vectorial form, we obtain equation (13).

Notice that, in expressions (11), (12) and (13), matrix $\widetilde{\Pi}(\delta_\nu)$ is the same.

Proof of Lemma 1

This is easily derived by expected players' payoffs $V_{\mathbf{n}}(\delta_\nu)$, $V_{\mathbf{c}}(\delta_\nu)$ and $V_{\mathbf{sc}}(\delta_\nu)$ given above in the Appendix, together with $a^{s_2} > d^{s_2}$ and $a^{s_1} > d^{s_1}$.

Appendix B: Proof of Proposition 1

We prove that the strategy profiles considered are subgame perfect using principle of non-profitability of one-shot deviations. Consider first the non-cooperative strategy profile. The fact that the strategy profile $\sigma_{\mathbf{n}}$ is a SPNE for every $\delta_\nu \in (0, 1)$, $\nu \in \{i, j\}$, discount factor δ is true because this profile prescribes players to choose the Nash equilibrium strategies in any state and one-shot deviations are non-profitable for any player ν with discount factor $\delta_\nu \in (0, 1)$.

Consider next the semi-cooperative strategy profile. The player ν's payoff in this profile is given by Appendix A. We need to prove that, if

$$\delta_\nu \geqslant \frac{\Delta_1 - \pi^{s_1}(b^{s_1} - d^{s_1}) - (1 - \pi^{s_2})(b^{s_1} - a^{s_1})}{2(\pi^{s_2} - \pi^{s_1})(b^{s_1} - d^{s_1})},$$

where

$$\Delta_1 = \left[(\pi^{s_1}(b^{s_1} - d^{s_1}) + (1 - \pi^{s_2})(b^{s_1} - a^{s_1}))^2 + 4(b^{s_1} - a^{s_1})(\pi^{s_2} - \pi^{s_1})(b^{s_1} - d^{s_1})\right]^{0.5},$$

then $\sigma_{\mathbf{sc}}$ is SPNE. If a deviation is observed, in the next stage of the game the deviating player is punished by getting the Nash equilibrium payoff in any state (d^{s_1} in state s_1 and d^{s_2} in state s_2). Clearly, deviation in state s_2 is not profitable for any $\delta_\nu \in (0, 1)$. Consider state s_1. The strategy profile $\sigma_{\mathbf{sc}}$ is a subgame perfect Nash equilibrium if for any player $\nu \in \{i, j\}$ there is no gain from deviation in state s_1:

$$a^{s_1} + \delta_\nu \mathbf{p}^{s_1} V_{\mathbf{sc}}(\delta_\nu) \geqslant b^{s_1} + \delta_\nu \mathbf{p}^{s_1} V_{\mathbf{n}}(\delta_\nu).$$

Substituting the expressions of $V_{\mathbf{sc}}(\delta_\nu)$ and $V_{\mathbf{n}}(\delta_\nu)$ from Appendix and rearranging, we obtain:

$$f_1(\delta_\nu) = \delta_\nu^2(\pi^{s_2} - \pi^{s_1})(b^{s_1} - d^{s_1}) + \delta_\nu \left\{\pi^{s_1}(b^{s_1} - d^{s_1}) + (1 - \pi^{s_2})(b^{s_1} - a^{s_1})\right\} - (b^{s_1} - a^{s_1}) \geqslant 0.$$

Function $f_1(\delta_\nu)$ is a quadratic function satisfying the following conditions: i) $f_1(0) \leqslant 0$, and ii) $\lim_{\delta \to 1} f_1(\delta_\nu) > 0$. Therefore, there is a unique solution $\delta_\nu = \delta_1^*$ of the equation $f_1(\delta_\nu) = 0$ for $\delta_\nu \in (0, 1)$, where δ_1^* is given by (2). The solution of the inequality $f_1(\delta_\nu) \geqslant 0$ is $\delta_\nu \in [\delta_1^*, 1)$. Therefore, if for any player $\nu \in \{i, j\}$ discount factor $\delta_\nu \in [\delta_1^*, 1)$, the semi-cooperative strategy profile $\sigma_{\mathbf{sc}}$ is SPNE.

Finally, consider the cooperative strategy profile. The player ν's payoff is given by Lemma 1. We need to prove that, for any

$$\delta_\nu \geqslant \frac{\Delta_2 - \pi^{s_2}(a^{s_1} - d^{s_1}) - (1 - \pi^{s_2})(b^{s_2} - d^{s_2}) - \pi^{s_1}(b^{s_2} - a^{s_2})}{2(\pi^{s_2} - \pi^{s_1})(b^{s_2} - d^{s_2})},$$

where

$$\Delta_2 = \left[(\pi^{s_2}(a^{s_1} - d^{s_1}) + (1 - \pi^{s_2})(b^{s_2} - d^{s_2}) + \pi^{s_1}(b^{s_2} - a^{s_2}))^2 + 4(b^{s_2} - a^{s_2})(\pi^{s_2} - \pi^{s_1})(b^{s_2} - d^{s_2})\right]^{0.5},$$

the strategy profile σ_c is SPNE. If a deviation is observed, in the next stage of the game the deviating player is punished by getting the Nash payoff in any state (d^{s_1} in state s_1 and d^{s_2} in state s_2). The strategy profile $\sigma_{\mathbf{c}}$ is SPNE if, for any player $\nu \in \{i, j\}$, there is no gain from deviation in any state. Player ν does not deviate in state s_1 if

$$a^{s_1} + \delta \mathbf{p}^{s_1} V_{\mathbf{c}}(\delta_\nu) \geqslant b^{s_1} + \delta_\nu \mathbf{p}^{s_1} V_{\mathbf{n}}(\delta_\nu),$$

and in state s_2 if

$$a^{s_2} + \delta_\nu \mathbf{p}^{s_2} V_{\mathbf{c}}(\delta_\nu) \geqslant b^{s_2} + \delta_\nu \mathbf{p}^{s_2} V_{\mathbf{n}}(\delta_\nu).$$

Substituting the expressions of $V_{\mathbf{c}}(\delta_\nu)$ and $V_{\mathbf{n}}(\delta_\nu)$ from Appendix A and rearranging, we obtain:

$$\begin{aligned} f_3(\delta_\nu) = \delta_\nu^2 &(\pi^{s_2} - \pi^{s_1})(b^{s_1} - d^{s_1}) \\ &+ \delta_\nu \left\{\pi^{s_1}(b^{s_1} - d^{s_1}) + (1-\pi^{s_1})(a^{s_2} - d^{s_2}) + (1-\pi^{s_2})(b^{s_1} - a^{s_1})\right\} - (b^{s_1} - a^{s_1}) \geqslant 0, \end{aligned}$$

$$\begin{aligned} f_2(\delta_\nu) = \delta_\nu^2 &(\pi^{s_2} - \pi^{s_1})(b^{s_2} - d^{s_2}) \\ &+ \delta_\nu \left\{\pi^{s_2}(a^{s_1} - d^{s_1}) + (1-\pi^{s_2})(b^{s_2} - d^{s_2}) + \pi^{s_1}(b^{s_2} - a^{s_2}))\right\} - (b^{s_2} - a^{s_2}) \geqslant 0. \end{aligned}$$

Here $f_2(\delta_\nu)$ and $f_3(\delta_\nu)$ are quadratic functions satisfying the following conditions. Given Assumption 1, $0 \geqslant f_3(0) \geqslant f_2(0)$ and $\lim_{\delta\to 1} f_3(\delta_\nu) = \lim_{\delta \to 1} f_2(\delta_\nu) > 0$. Therefore, there is a unique solution $\delta_\nu = \delta_3^*$ of the equation $f_3(\delta_\nu) = 0$ for $\delta_\nu \in (0,1)$ and a unique solution $\delta_\nu = \delta_2^*$ of the equation $f_2(\delta_\nu) = 0$ for $\delta_\nu \in (0,1)$, where δ_2^* is given by (3). Taking into account that both $f_2(\delta_\nu)$ and $f_3(\delta_\nu)$ are non-decreasing functions, we may easily prove that the solution of the system of inequalities

$$\begin{cases} f_3(\delta_\nu) \geqslant 0, \\ f_2(\delta_\nu) \geqslant 0, \end{cases}$$

is $\delta_\nu \in [\delta_2^*, 1)$. Therefore, if for any player $\nu \in \{i, j\}$ discount factor $\delta_\nu \in [\delta_2^*, 1)$ the cooperative strategy profile $\sigma_{\mathbf{c}}$ is SPNE.

Appendix C: Proof of Proposition 2

Before proving the proposition, it is convenient to introduce the following strategy profile, as it may emerge in the case of deviation. We will next proceed with the proof.

Deviating strategy profile

Definition 6 *A "deviating strategy profile" is denoted as* $\sigma_{\mathbf{d}} = (\sigma_{\mathbf{d},i}, \sigma_{\mathbf{c},j})$*, where* $\sigma_{\mathbf{d},i} = \{\sigma^s_{i,t}\}^{s=s_1,s_2}_{t=1,\ldots,\infty}$ *such that*

$$\sigma^s_{i,t+1}(H(t)) = \begin{cases} C, & \text{if } s = s_1 \text{ and } H(t) = H_{\mathbf{c}}(t) \\ D, & \text{if } s = s_1 \text{ and } H(t) \neq H_{\mathbf{c}}(t) \\ D, & \text{if } s = s_2 \end{cases}.$$

In this profile, a player j plays strategy $\sigma_{\mathbf{c},j}$ given by Definition 5 while player i applies strategy $\sigma_{\mathbf{d},i}$. This profile may occur when player j has a belief that the competitor i will play cooperatively while she will in fact deviate in state $s = s_2$. In turn, when player j observes a deviation from the cooperative strategy profile, she reacts with D in all stages afterwards according to strategy $\sigma_{\mathbf{c},j}$.

Denote by $V_{\mathbf{d}}(\delta_i)$ an expected payoff of deviating player i in strategy profile $\sigma_{\mathbf{d}}$. We compute the expected payoff $V_{\mathbf{d},i}(\delta_i)$ of a deviating player i which is:

$$V_{\mathbf{d}}(\delta_i) = \begin{pmatrix} V^{s_1}_{\mathbf{d}}(\delta_i) \\ V^{s_2}_{\mathbf{d}}(\delta_i) \end{pmatrix},$$

where $V^s_{\mathbf{d}}(\delta_i)$ is the payoff of player i in the subgame starting from state s. If the subgame starts from state s_1, player i gets

$$V^{s_1}_{\mathbf{d}}(\delta_i) = a^{s_1} + \delta(\pi^{s_1} V^{s_1}_{\mathbf{d},i}(\delta_i) + (1 - \pi^{s_1}) V^{s_2}_{\mathbf{d},i}(\delta_i)).$$

If the subgame starts from state s_2, player i deviates and gets b^{s_2}. Then she will be punished by playing (D, D) in any state from the next stage until infinity. Her total payoff will be

$$V^{s_2}_{\mathbf{d}}(\delta_i) = b^{s_2} + \delta \mathbf{p}^{s_2} V_{\mathbf{n}}(\delta_i).$$

From these two equations we obtain

$$V_{\mathbf{d}}(\delta_i) = \begin{pmatrix} \frac{1}{1-\delta_i \pi^{s_1}} [a^{s_1} + \delta_i (1 - \pi^{s_1})(b^{s_2} + \delta_i \mathbf{p}^{s_2} V_{\mathbf{n}}(\delta_i))] \\ b^{s_2} + \delta \mathbf{p}^{s_2} V_{\mathbf{n}}(\delta_i) \end{pmatrix}.$$

Initial state $s = s_1$. Strategy profile $(q_1^{s_1}, r_1^{s_1}) = (1, 0)$

Begin from a player of class P. If she does not deviate from $(1, 0)$, she gets

$$\alpha_1 \left(a^{s_1} + \delta_P \mathbf{p}^{s_1} V_{\mathbf{c}}(\delta_P)\right) + (1 - \alpha_1) \left(c^{s_1} + \delta_P \mathbf{p}^{s_1} V_{\mathbf{n}}(\delta_P)\right). \tag{14}$$

If she deviates from profile $(1, 0)$ $(q_1^{s_1} = 0)$, she gets:

$$\alpha_1 (b^{s_1} + \delta_P \mathbf{p}^{s_1} V_{\mathbf{n}}(\delta_P)) + (1 - \alpha_1)(d^{s_1} + \delta_P \mathbf{p}^{s_1} V_{\mathbf{n}}(\delta_P)). \tag{15}$$

The deviation is not profitable if (14) is larger or equal to (15), taking into account $\delta_P \geqslant \delta_2^*$.

Now consider the player of class M. Her payoff in profile $(1, 0)$ is

$$\alpha_1 (b^{s_1} + \delta_M \mathbf{p}^{s_1} V_{\mathbf{n}}(\delta_M)) + (1 - \alpha_1)(d^{s_1} + \delta_M \mathbf{p}^{s_1} V_{\mathbf{n}}(\delta_M)). \tag{16}$$

If she deviates from profile $(1, 0)$ (playing $r_1^{s_1} = 1$), then she gets:

$$\alpha_1 (a^{s_1} + \delta_M \mathbf{p}^{s_1} V_{\mathbf{d}}(\delta_M)) + (1 - \alpha_1)(c^{s_1} + \delta_M \mathbf{p}^{s_1} V_{\mathbf{n}}(\delta_M)), \tag{17}$$

where $V_{\mathbf{d}}(\delta_M)$ is the payoff of class M player when she cooperates in $s = s_1$ and deviates in state $s = s_2$ (which is profitable to her according to her discount factor). The deviation is not profitable if (16) is larger or equal than (17), taking into account inequality $\delta_1^* \leqslant \delta_M \leqslant \delta_2^*$ from Proposition 1. The strategy profile $(1, 0)$ is a PMPBE when one of the following systems has a solution:

$$\begin{cases} \alpha_1 \geqslant \dfrac{d^{s_1} - c^{s_1}}{d^{s_1} - c^{s_1} + a^{s_1} - b^{s_1} + \delta_P \mathbf{p}^{s_1} (V_{\mathbf{c}}(\delta_P) - V_{\mathbf{n}}(\delta_P))} \\ d^{s_1} - c^{s_1} + a^{s_1} - b^{s_1} + \delta_M \mathbf{p}^{s_1} (V_{\mathbf{d}}(\delta_M) - V_{\mathbf{n}}(\delta_M)) \leqslant 0 \end{cases},$$

or

$$\begin{cases} \alpha_1 \geqslant \dfrac{d^{s_1} - c^{s_1}}{d^{s_1} - c^{s_1} + a^{s_1} - b^{s_1} + \delta_P \mathbf{p}^{s_1} (V_{\mathbf{c}}(\delta_P) - V_{\mathbf{n}}(\delta_P))} \\ d^{s_1} - c^{s_1} + a^{s_1} - b^{s_1} + \delta_M \mathbf{p}^{s_1} (V_{\mathbf{d}}(\delta_M) - V_{\mathbf{n}}(\delta_M)) > 0 \\ \alpha_1 \leqslant \dfrac{d^{s_1} - c^{s_1}}{d^{s_1} - c^{s_1} + a^{s_1} - b^{s_1} + \delta_M \mathbf{p}^{s_1} (V_{\mathbf{d}}(\delta_M) - V_{\mathbf{n}}(\delta_M))} \end{cases}.$$

Now we need to verify the sign of expression:

$$d^{s_1} - c^{s_1} + a^{s_1} - b^{s_1} + \delta_M \mathbf{p}^{s_1} (V_{\mathbf{d}}(\delta_M) - V_{\mathbf{n}}(\delta_M)). \tag{18}$$

First, consider the difference $\delta_M \mathbf{p}^{s_1}(V_{\mathbf{d}}(\delta_M) - V_{\mathbf{n}}(\delta_M))$. Given Lemma 1, we obtain that

$$\delta_M \mathbf{p}^{s_1} V_{\mathbf{d}}(\delta_M) > \delta_M \mathbf{p}^{s_1} V_{\mathbf{sc}}(\delta_M)$$

and equivalently

$$\delta_M \mathbf{p}^{s_1}(V_{\mathbf{d}}(\delta_M) - V_{\mathbf{n}}(\delta_M)) > \delta_M \mathbf{p}^{s_1}(V_{\mathbf{sc}}(\delta_M) - V_{\mathbf{n}}(\delta_M)) \tag{19}$$

Taking into account that

$$\delta_M \mathbf{p}^{s_1}(V_{\mathbf{sc}}(\delta_M) - V_{\mathbf{n}}(\delta_M)) \geqslant b^{s_1} - a^{s_1},$$

we obtain

$$d^{s_1} - c^{s_1} + a^{s_1} - b^{s_1} + \delta_M \mathbf{p}^{s_1}(V_{\mathbf{d}}(\delta_M) - V_{\mathbf{n}}(\delta_M)) \geqslant d^{s_1} - c^{s_1}.$$

Therefore, the expression (18) is positive.

Simplifying the systems and considering $\delta_P \geqslant \delta_2^*$, we obtain the condition:

$$\alpha_1 \in [A_1, A_2], \tag{20}$$

where

$$A_1 \equiv \frac{d^{s_1} - c^{s_1}}{d^{s_1} - c^{s_1} + a^{s_1} - b^{s_1} + \delta_P \mathbf{p}^{s_1}(V_{\mathbf{c}}(\delta_P) - V_{\mathbf{n}}(\delta_P))},$$

and

$$A_2 \equiv \frac{d^{s_1} - c^{s_1}}{d^{s_1} - c^{s_1} + a^{s_1} - b^{s_1} + \delta_M \mathbf{p}^{s_1}(V_{\mathbf{d}}(\delta_M) - V_{\mathbf{n}}(\delta_M))}.$$

Initial state $s = s_1$. Strategy profile $(q_1^{s_1}, r_1^{s_1}) = (1, 1)$

Again, we begin from a player of class P. If she does not deviate from strategy $(1, 1)$, she gets:

$$(\alpha_1 + \beta_1)(a^{s_1} + \delta_P \mathbf{p}^{s_1} V_{\mathbf{sc}}(\delta_P)) + (1 - \alpha_1 - \beta_1)(c^{s_1} + \delta_P \mathbf{p}^{s_1} V_{\mathbf{n}}(\delta_P)). \tag{21}$$

If she deviates from $(1, 1)$ $(q_1 = 0)$, she gets:

$$(\alpha_1 + \beta_1)(b^{s_1} + \delta_P \mathbf{p}^{s_1} V_{\mathbf{n}}(\delta_P)) + (1 - \alpha_1 - \beta_1)(d^{s_1} + \delta_P \mathbf{p}^{s_1} V_{\mathbf{n}}(\delta_P)). \tag{22}$$

The deviation is not profitable if (21) is larger than or equal to (22), taking into account $\delta_P \geqslant \delta_2^*$ from Proposition 1.

Consider next a player of class M. Her payoff in profile $(1,1)$ is

$$(\alpha_1 + \beta_1)(a^{s_1} + \delta_M \mathbf{p}^{s_1} V_{\mathbf{sc}}(\delta_M)) + (1 - \alpha_1 - \beta_1)(c^{s_1} + \delta_M \mathbf{p}^{s_1} V_{\mathbf{n}}(\delta_M)). \tag{23}$$

If she deviates from profile $(1,1)$ $(r_1^{s_1} = 0)$ she gets:

$$(\alpha_1 + \beta_1)(b^{s_1} + \delta_M \mathbf{p}^{s_1} V_{\mathbf{n}}(\delta_M)) + (1 - \alpha_1 - \beta_1)(d^{s_1} + \delta_M \mathbf{p}^{s_1} V_{\mathbf{n}}(\delta_M)). \tag{24}$$

The deviation is not profitable if payoff (23) is larger than or equal to (24), taking into account $\delta_1^* \leqslant \delta_M \leqslant \delta_2^*$.

Thus, the strategy profile $(1,1)$ is a PMPBE if the following system has a solution:

$$\begin{cases} (\alpha_1 + \beta_1)\left[a^{s_1} - b^{s_1} - c^{s_1} + d^{s_1} + \delta_P \mathbf{p}^{s_1}(V_{\mathbf{sc}}(\delta_P) - V_{\mathbf{n}}(\delta_P))\right] \geqslant d^{s_1} - c^{s_1}, \\ (\alpha_1 + \beta_1)\left[a^{s_1} - b^{s_1} - c^{s_1} + d^{s_1} + \delta_M \mathbf{p}^{s_1}(V_{\mathbf{sc}}(\delta_M) - V_{\mathbf{n}}(\delta_M))\right] \geqslant d^{s_1} - c^{s_1}. \end{cases}$$

Since $\delta_M < \delta_P$, the system is equivalent to the following inequality:

$$\alpha_1 + \beta_1 \geqslant \frac{d^{s_1} - c^{s_1}}{d^{s_1} - c^{s_1} + a^{s_1} - b^{s_1} + \delta_M \mathbf{p}^{s_1}(V_{\mathbf{sc}}(\delta_M) - V_{\mathbf{n}}(\delta_M))}. \tag{25}$$

Initial state $s = s_2$. Strategy profile $(q_1^{s_2}) = (1)$

If the game starts in state s_2, the strategy profile $(q_1^{s_2}) = (1)$ is a PMPBE if the following inequality holds:

$$\alpha_1\left[a^{s_2} - b^{s_2} - c^{s_2} + d^{s_2} + \delta_P \mathbf{p}^{s_2}(V_{\mathbf{c}}(\delta_P) - V_{\mathbf{n}}(\delta_P))\right] \geqslant d^{s_2} - c^{s_2}.$$

Since $\delta_P \geqslant \delta_2^*$, then $\delta_P \mathbf{p}^{s_2}(V_{\mathbf{c}}(\delta_P) - V_{\mathbf{n}}(\delta_P)) \geqslant b^{s_2} - a^{s_2}$, so that:

$$\alpha_1 \geqslant \frac{d^{s_2} - c^{s_2}}{a^{s_2} - b^{s_2} - c^{s_2} + d^{s_2} + \delta_P \mathbf{p}^{s_2}(V_{\mathbf{c}}(\delta_P) - V_{\mathbf{n}}(\delta_P))}. \tag{26}$$

Appendix D: Proof of Corollary 1

The payoff of a P player in profile $(1,0)$ is

$$\alpha_1(a^{s_1}+\delta_P\mathbf{p}^{s_1}V_{\mathbf{c}}(\delta_P))+(1-\alpha_1)(c^{s_1}+\delta_P\mathbf{p}^{s_1}V_{\mathbf{n}}(\delta_P)),$$

and in profile $(1,1)$ is

$$(\alpha_1+\beta_1)(a^{s_1}+\delta_P\mathbf{p}^{s_1}V_{\mathbf{sc}}(\delta_P))+(1-\alpha_1-\beta_1)(c^{s_1}+\delta_P\mathbf{p}^{s_1}V_{\mathbf{n}}(\delta_P)).$$

The payoff of a P player in profile $(1,1)$ is not less than his payoff in profile $(1,0)$ if

$$\beta_1(c^{s_1}-a^{s_1}-\delta_P\mathbf{p}^{s_1}(V_{\mathbf{sc}}(\delta_P)-V_{\mathbf{n}}(\delta_P)))+\alpha_1\delta_P\mathbf{p}^{s_1}(V_{\mathbf{c}}(\delta_P)-V_{\mathbf{sc}}(\delta_P))\leqslant 0,$$

or

$$\frac{\beta_1}{\alpha_1}\geqslant\frac{\delta_P\mathbf{p}^{s_1}(V_{\mathbf{c}}(\delta_P)-V_{\mathbf{sc}}(\delta_P))}{a^{s_1}-c^{s_1}+\delta_P\mathbf{p}^{s_1}(V_{\mathbf{sc}}(\delta_P)-V_{\mathbf{n}}(\delta_P))}.$$

The payoff of an M player in profile $(1,0)$ is

$$\alpha_1(b^{s_1}+\delta_M\mathbf{p}^{s_1}V_{\mathbf{n}}(\delta_M))+(1-\alpha_1)(d^{s_1}+\delta_M\mathbf{p}^{s_1}V_{\mathbf{n}}(\delta_M))$$

and in profile $(1,1)$ is

$$(\alpha_1+\beta_1)(a^{s_1}+\delta_M\mathbf{p}^{s_1}V_{\mathbf{sc}}(\delta_M))+(1-\alpha_1-\beta_1)(c^{s_1}+\delta_M\mathbf{p}^{s_1}V_{\mathbf{n}}(\delta_M)).$$

The payoff of an M player in profile $(1,1)$ is not less than his payoff in profile $(1,0)$ if

$$\begin{aligned}\alpha_1(b^{s_1}-a^{s_1}+c^{s_1}-d^{s_1}+\delta_M\mathbf{p}^{s_1}(V_{\mathbf{n}}(\delta_M)-V_{\mathbf{sc}}(\delta_M)))&\\ +\beta_1(c^{s_1}-a^{s_1}+\delta_M\mathbf{p}^{s_1}(V_{\mathbf{n}}(\delta_M)-V_{\mathbf{sc}}(\delta_M)))+d^{s_1}-c^{s_1}&\leqslant 0.\end{aligned}$$

or

$$(\alpha_1+\beta_1)(d^{s_1}-b^{s_1}+a^{s_1}-c^{s_1}+\delta_M\mathbf{p}^{s_1}(V_{\mathbf{sc}}(\delta_M)-V_{\mathbf{n}}(\delta_M))\geqslant\beta_1(d^{s_1}-b^{s_1})+(d^{s_1}-c^{s_1}). \quad (27)$$

Taking into account that $\alpha_1+\beta_1\geqslant A_3$, we may state that

$$(\alpha_1+\beta_1)(d^{s_1}-b^{s_1}+a^{s_1}-c^{s_1}+\delta_M\mathbf{p}^{s_1}(V_{\mathbf{sc}}(\delta_M)-V_{\mathbf{n}}(\delta_M))\geqslant d^{s_1}-c^{s_1}.$$

The latter inequality guarantees that (27) is satisfied because $d^{s_1} - b^{s_1} < 0$.

Now consider the initial state $s = s_2$. The payoff of class P player in equilibrium $(q_1^{s_2}) = 1$ is

$$\alpha_1(a^{s_2} + \delta_P \mathbf{p}^{s_2} V_{\mathbf{c}}(\delta_P)) + (1-\alpha_1)(c^{s_2} + \delta_P \mathbf{p}^{s_2} V_{\mathbf{n}}(\delta_P)).$$

The payoff of class P player in equilibrium $(q_1^{s_2}) = q^*$ given by $q^* = A_4/\alpha_1$ (see Proposition 2) is

$$\begin{aligned}\alpha_1 qq(a^{s_2} - b^{s_2} + d^{s_2} - c^{s_2} + \delta_P \mathbf{p}^{s_2}(V_{\mathbf{c}}(\delta_P) - V_{\mathbf{n}}(\delta_P)) + \alpha_1 q(b^{s_2} - d^{s_2}) + q(c^{s_2} - d^{s_2}) \\ + d^{s_2} + \delta_P \mathbf{p}^{s_2} V_{\mathbf{n}}(\delta_P) = d^{s_2} + \alpha_1 q(b^{s_2} - d^{s_2}) + \delta_P \mathbf{p}^{s_2} V_{\mathbf{n}}(\delta_P).\end{aligned}$$

The payoff of class P player in profile $(q_1^{s_2}) = 1$ is not less than his payoff in profile $(q_1^{s_2}) = q$ if

$$\begin{aligned}&\alpha_1(a^{s_2} - c^{s_2} + \delta_P \mathbf{p}^{s_2}(V_{\mathbf{c}}(\delta_P) - V_{\mathbf{n}}(\delta_P))) \\ &\qquad \geqslant (d^{s_2} - c^{s_2}) \left[1 + \frac{b^{s_2} - d^{s_2}}{d^{s_2} - c^{s_2} + a^{s_2} - b^{s_2} + \delta_P \mathbf{p}^{s_2}(V_{\mathbf{c}}(\delta_P) - V_{\mathbf{n}}(\delta_P))}\right],\end{aligned}$$

which is always true for any $\alpha_1 \geqslant A_4$.

The payoff of player M in profile $(q_1^{s_2}) = 1$ is

$$\alpha_1(b^{s_2} + \delta_M \mathbf{p}^{s_2} V_{\mathbf{n}}(\delta_M)) + (1-\alpha_1)(d^{s_2} + \delta_M \mathbf{p}^{s_2} V_{\mathbf{n}}(\delta_M)),$$

and in profile $(q_1^{s_2}) = q$ is

$$\alpha_1(q(b^{s_2} + \delta_M \mathbf{p}^{s_2} V_{\mathbf{n}}(\delta_M)) + (1-q)(d^{s_2} + \delta_M \mathbf{p}^{s_2} V_{\mathbf{n}}(\delta_M))) + (1-\alpha_1)(d^{s_2} + \delta_M \mathbf{p}^{s_2} V_{\mathbf{n}}(\delta_M)).$$

The payoff of class M player in profile $(q_1^{s_2}) = 1$ is not less than his payoff in profile $(q_1^{s_2}) = q$ if

$$\alpha_1(b^{s_2} - qb^{s_2} - (1-q)d^{s_2}) \geqslant 0,$$

which is always true because $b^{s_2} > d^{s_2}$.

Appendix E: Proof of Proposition 3

Consider the payoffs of the players of classes P and M as functions of parameter α_1. By Proposition 2, there are three equilibria:

1. Equilibrium $(q_1^{s_1}, r_1^{s_1}) = (1, 0)$: the payoff of the player of class P is

$$\alpha_1(a^{s_1} + \delta_P \mathbf{p}^{s_1} V_{\mathbf{c}}(\delta_P)) + (1-\alpha_1)(c^{s_1} + \delta_P \mathbf{p}^{s_1} V_{\mathbf{n}}(\delta_P)).$$

   It is a linear function of α_1 with coefficient $a^{s_1} - c^{s_1} + \delta_P \mathbf{p}^{s_1}(V_{\mathbf{c}}(\delta_P) - V_{\mathbf{n}}(\delta_P))$ which is positive because $a^{s_1} > c^{s_1}$ and $V_{\mathbf{c}}(\delta) > V_{\mathbf{n}}(\delta)$ for any $\delta \in (0, 1)$.

   The payoff of the player of class M is

$$\alpha_1(b^{s_1} + \delta_M \mathbf{p}^{s_1} V_{\mathbf{n}}(\delta_M)) + (1-\alpha_1)(d^{s_1} + \delta_M \mathbf{p}^{s_1} V_{\mathbf{n}}(\delta_M)).$$

   It is also a linear function of α_1 with coefficient $b^{s_1} - d^{s_1}$ which is positive for any $\delta \in (0, 1)$.

2. Equilibrium $(q_1^{s_1}, r_1^{s_1}) = (1, 1)$: we begin with the player of class P. Her payoff is

$$(\alpha_1 + \beta_1)(a^{s_1} + \delta_P \mathbf{p}^{s_1} V_{\mathbf{sc}}(\delta_P)) + (1-\alpha_1-\beta_1)(c^{s_1} + \delta_P \mathbf{p}^{s_1} V_{\mathbf{n}}(\delta_P)).$$

   It is a a linear function of α_1 with coefficient $a^{s_1} - c^{s_1} + \delta_P \mathbf{p}^{s_1}(V_{\mathbf{sc}}(\delta_P) - V_{\mathbf{n}}(\delta_P))$ which is positive because $a^{s_1} > c^{s_1}$ and $V_{\mathbf{sc}}(\delta) > V_{\mathbf{n}}(\delta)$ for any $\delta \in (0, 1)$.

   Then, the payoff of the player of class M is

$$(\alpha_1 + \beta_1)(a^{s_1} + \delta_M \mathbf{p}^{s_1} V_{\mathbf{sc}}(\delta_M)) + (1-\alpha_1-\beta_1)(c^{s_1} + \delta_M \mathbf{p}^{s_1} V_{\mathbf{n}}(\delta_M)).$$

   This is a linear function of α_1 with coefficient $a^{s_1} - c^{s_1} + \delta_M \mathbf{p}^{s_1}(V_{\mathbf{sc}}(\delta_M) - V_{\mathbf{n}}(\delta_M))$ which is positive because $a^{s_1} > c^{s_1}$ and $V_{\mathbf{sc}}(\delta) > V_{\mathbf{n}}(\delta)$ for any $\delta \in (0, 1)$.

   The derivatives of the payoffs of the P and M players with respect to β_1 equal the corresponding derivatives subject to α_1. Therefore, the payoffs are also increasing functions of β_1.

3. Equilibrium $(q_1) = (1)$ in initial state $s = s_2$: the payoff of the player of class P is

$$\alpha_1(a^{s_2} + \delta_P \mathbf{p}^{s_2} V_{\mathbf{c}}(\delta_P)) + (1-\alpha_1)(c^{s_2} + \delta_P \mathbf{p}^{s_2} V_{\mathbf{n}}(\delta_P))$$

   It is a linear function of α_1 with coefficient $a^{s_2} - c^{s_2} + \delta_P \mathbf{p}^{s_2}(V_{\mathbf{c}}(\delta_P) - V_{\mathbf{n}}(\delta_P))$ which is positive because $a^{s_2} > c^{s_2}$ and $V_{\mathbf{c}}(\delta) > V_{\mathbf{n}}(\delta)$ for any $\delta \in (0, 1)$.

Appendix F: Proof of Proposition 4

Period 1, state s_1. If player P follows the described strategy $q_1^{s_1} = 1$, his payoff will be

$$\begin{aligned}
&\alpha_1 \left[a^{s_1} + \delta_P \mathbf{p}^{s_1} V_{\mathbf{c}}(\delta_P)\right] \\
&+ \beta_1 \left[a^{s_1} + \delta_P \pi^{s_1}(c^{s_1} + \delta_P \mathbf{p}^{s_1} V_{\mathbf{sc}}(\delta_P)) + \delta_P(1-\pi^{s_1})(c^{s_2} + \delta_P \mathbf{p}^{s_2} V_{\mathbf{sc}}(\delta_P))\right] \\
&+ (1-\alpha_1-\beta_1)\left[c^{s_1} + \delta_P \mathbf{p}^{s_1} V_{\mathbf{n}}(\delta_P)\right].
\end{aligned}$$

If he deviates to strategy $q_1^{s_1} = 0$, his class will be identified as I and his payoff will be

$$\alpha_1 \left[b^{s_1} + \delta_P \mathbf{p}^{s_1} V_{\mathbf{n}}(\delta_P)\right] + \beta_1 \left[b^{s_1} + \delta_P \mathbf{p}^{s_1} V_{\mathbf{n}}(\delta_P)\right] + (1-\alpha_1-\beta_1)\left[d^{s_1} + \delta_P \mathbf{p}^{s_1} V_{\mathbf{n}}(\delta_P)\right].$$

Remembering that $\mathbf{p}^s = (\pi^s, 1-\pi^s)$, the deviation of player P in period 1 is not profitable if

$$\begin{aligned}
&\alpha_1 \big[a^{s_1} - b^{s_1} + d^{s_1} - c^{s_1} + \delta_P \mathbf{p}^{s_1}(V_{\mathbf{c}}(\delta_P) - V_{\mathbf{n}}(\delta_P))\big] \\
&+ \beta_1 \big[a^{s_1} - b^{s_1} + d^{s_1} - c^{s_1} + \delta_P \pi^{s_1}(c^{s_1} - a^{s_1}) \\
&+ \delta_P(1-\pi^{s_1})(c^{s_2} - d^{s_2}) + \delta_P \mathbf{p}^{s_1}(V_{\mathbf{sc}}(\delta_P) - V_{\mathbf{n}}(\delta_P))\big] \\
&\geqslant d^{s_1} - c^{s_1}.
\end{aligned}$$

We call

$$\begin{aligned}
A_5 &\equiv \big[a^{s_1} - b^{s_1} + d^{s_1} - c^{s_1} + \delta_P \mathbf{p}^{s_1}(V_{\mathbf{c}}(\delta_P) - V_{\mathbf{n}}(\delta_P))\big], \\
A_6 &\equiv \big[a^{s_1} - b^{s_1} + d^{s_1} - c^{s_1} + \delta_P \pi^{s_1}(c^{s_1} - a^{s_1}) + \delta_P(1-\pi^{s_1})(c^{s_2} - d^{s_2}) \\
&\quad + \delta_P \mathbf{p}^{s_1}(V_{\mathbf{sc}}(\delta_P) - V_{\mathbf{n}}(\delta_P))\big].
\end{aligned}$$

Notice that $A_5 > 0$, given that $d^{s_1} - c^{s_1} > 0$, and that Proposition 1 combined with Lemma 1 implies

$$\delta_P > \frac{a^{s_1} - b^{s_1}}{\mathbf{p}^{s_1}(V_{\mathbf{c}}(\delta_P) - V_{\mathbf{n}}(\delta_P))}.$$

If player M follows the described strategy $r_1^{s_1}=1$, his payoff will be

$$\begin{aligned}
&\alpha_1\left[a^{s_1}+\delta_M\pi^{s_1}(b^{s_1}+\delta_M\mathbf{p}^{s_1}V_{\mathbf{sc}}(\delta_M))+\delta_M(1-\pi^{s_1})(b^{s_2}+\delta_M\mathbf{p}^{s_2}V_{\mathbf{sc}}(\delta_M))\right]\\
&+\beta_1\left[a^{s_1}+\delta_M\pi^{s_1}(d^{s_1}+\delta_M\mathbf{p}^{s_1}V_{\mathbf{sc}}(\delta_P))+\delta_M(1-\pi^{s_1})(d^{s_2}+\delta_M\mathbf{p}^{s_2}V_{\mathbf{sc}}(\delta_M))\right]\\
&+(1-\alpha_1-\beta_1)\left[c^{s_1}+\delta_M\mathbf{p}^{s_1}V_{\mathbf{n}}(\delta_M)\right].
\end{aligned}$$

If he deviates to strategy $r_1^{s_1}=0$, his class will be identified as I and his payoff will be

$$\alpha_1\left[b^{s_1}+\delta_M\mathbf{p}^{s_1}V_{\mathbf{n}}(\delta_M)\right]+\beta_1\left[b^{s_1}+\delta_M\mathbf{p}^{s_1}V_{\mathbf{n}}(\delta_M)\right]+(1-\alpha_1-\beta_1)\left[d^{s_1}+\delta_M\mathbf{p}^{s_1}V_{\mathbf{n}}(\delta_M)\right].$$

The deviation of player M in period 1 is not profitable if

$$\begin{aligned}
&\alpha_1\big[a^{s_1}-b^{s_1}+d^{s_1}-c^{s_1}+\delta_M\mathbf{p}^{s_1}(V_{\mathbf{sc}}(\delta_M)-V_{\mathbf{n}}(\delta_M))\\
&+\delta_M\pi^{s_1}(b^{s_1}-a^{s_1})+\delta_M(1-\pi^{s_1})(b^{s_2}-d^{s_2})\big]\\
&+\beta_1\big[a^{s_1}-b^{s_1}+d^{s_1}-c^{s_1}+\delta_M\mathbf{p}^{s_1}(V_{\mathbf{sc}}(\delta_M)-V_{\mathbf{n}}(\delta_M))\\
&+\delta_M\pi^{s_1}(d^{s_1}-a^{s_1})\big]\geqslant d^{s_1}-c^{s_1}.
\end{aligned}$$

We call

$$\begin{aligned}
A_7 &\equiv \big[a^{s_1}-b^{s_1}+d^{s_1}-c^{s_1}+\delta_M\mathbf{p}^{s_1}(V_{\mathbf{sc}}(\delta_M)-V_{\mathbf{n}}(\delta_M))+\delta_M\pi^{s_1}(b^{s_1}-a^{s_1})\\
&\quad+\delta_M(1-\pi^{s_1})(b^{s_2}-d^{s_2})\big],\\
A_8 &\equiv \big[a^{s_1}-b^{s_1}+d^{s_1}-c^{s_1}+\delta_M\mathbf{p}^{s_1}(V_{\mathbf{sc}}(\delta_M)-V_{\mathbf{n}}(\delta_M))+\delta_M\pi^{s_1}(d^{s_1}-a^{s_1})\big].
\end{aligned}$$

Notice that $A_7>0$, given that $b^s>a^s>d^s>c^s$, and that Proposition 1 combined with Lemma 1 implies

$$\delta_M>\frac{a^{s_1}-b^{s_1}}{\mathbf{p}^{s_1}(V_{\mathbf{sc}}(\delta_P)-V_{\mathbf{n}}(\delta_P))}.$$

Period 2. State s_1. If in period 1 the players' classes are not revealed, i. e. only action C was observed, then the learning phase continues and the updated beliefs are

$$\alpha_2=\frac{\alpha_1}{\alpha_1+\beta_1},\qquad \beta_2=\frac{\beta_1}{\alpha_1+\beta_1}.$$

If player P uses strategy $q_2^{s_1} = 1$, his payoff will be

$$\alpha_2 \left[a^{s_1} + \delta_P \mathbf{p}^{s_1} V_{\mathbf{c}}(\delta_P)\right] + \beta_2 \left[c^{s_1} + \delta_P \mathbf{p}^{s_1} V_{\mathbf{sc}}(\delta_P)\right].$$

If he deviates to strategy $q_2^{s_1} = 0$, his class will be identified as M and his payoff will be

$$\alpha_2 \left[b^{s_1} + \delta_P \mathbf{p}^{s_1} V_{\mathbf{sc}}(\delta_P)\right] + \beta_2 \left[d^{s_1} + \delta_P \mathbf{p}^{s_1} V_{\mathbf{sc}}(\delta_P)\right].$$

The deviation of player P in period 2, state s_1, is not profitable if

$$\alpha_2 \left[a^{s_1} - b^{s_1} + \delta_P \mathbf{p}^{s_1} (V_{\mathbf{c}}(\delta_P) - V_{\mathbf{sc}}(\delta_P))\right] + \beta_2 \left[c^{s_1} - d^{s_1}\right] \geqslant 0,$$

taking into account the expressions of α_2 and β_2, we obtain condition

$$\frac{\beta_1}{\alpha_1} \leqslant \frac{\delta_P \mathbf{p}^{s_1} (V_{\mathbf{c}}(\delta_P) - V_{\mathbf{sc}}(\delta_P)) - (b^{s_1} - a^{s_1})}{d^{s_1} - c^{s_1}}.$$

If player M uses strategy $r_2^{s_1} = 0$, his payoff will be

$$\alpha_2 \left[b^{s_1} + \delta_M \mathbf{p}^{s_1} V_{\mathbf{sc}}(\delta_M)\right] + \beta_2 \left[d^{s_1} + \delta_M \mathbf{p}^{s_1} V_{\mathbf{sc}}(\delta_M)\right].$$

If he deviates to strategy $r_2^{s_1} = 1$, his class will be identified as P and his payoff will be

$$\alpha_2 \left[a^{s_1} + \delta_M \mathbf{p}^{s_1} V_{\mathbf{d}}(\delta_M)\right] + \beta_2 \left[c^{s_1} + \delta_M \mathbf{p}^{s_1} V_{\mathbf{sc}}(\delta_M)\right].$$

The deviation of player M in period 2, state s_1, is not profitable if

$$\alpha_2 \left[b^{s_1} - a^{s_1} + \delta_M \mathbf{p}^{s_1} (V_{\mathbf{sc}}(\delta_M) - V_{\mathbf{d}}(\delta_M))\right] + \beta_2 \left[d^{s_1} - c^{s_1}\right] \geqslant 0,$$

taking into account the expressions of α_2 and β_2, we obtain condition

$$\frac{\beta_1}{\alpha_1} \geqslant \frac{\delta_M \mathbf{p}^{s_1} (V_{\mathbf{d}}(\delta_M) - V_{\mathbf{sc}}(\delta_M)) - (b^{s_1} - a^{s_1})}{d^{s_1} - c^{s_1}}.$$

Period 2. State s_2. If in period 2 player P uses strategy $q_2^{s_2} = 1$, his payoff will be

$$\alpha_2 \left[a^{s_2} + \delta_P \mathbf{p}^{s_2} V_{\mathbf{c}}(\delta_P)\right] + \beta_2 \left[c^{s_2} + \delta_P \mathbf{p}^{s_2} V_{\mathbf{sc}}(\delta_P)\right].$$

If he deviates to strategy $q_2^{s_2} = 0$, his class will be identified as M and his payoff will be

$$\alpha_2 \left[b^{s_2} + \delta_P \mathbf{p}^{s_2} V_{\mathbf{sc}}(\delta_P)\right] + \beta_2 \left[d^{s_2} + \delta_P p6s_2 V_{\mathbf{sc}}(\delta_P)\right].$$

The deviation of player P in period 2, state s_2, is not profitable if

$$\alpha_2 \left[a^{s_2} - b^{s_2} + \delta_P \mathbf{p}^{s_2}(V_{\mathbf{c}}(\delta_P) - V_{\mathbf{sc}}(\delta_P))\right] + \beta_2 \left[c^{s_2} - d^{s_2}\right] \geqslant 0.$$

Taking into account the expressions of α_2 and β_2, we obtain condition

$$\beta_1 \leqslant \frac{\delta_P \mathbf{p}^{s_2}(V_{\mathbf{c}}(\delta_P) - V_{\mathbf{sc}}(\delta_P)) - (b^{s_2} - a^{s_2})}{d^{s_2} - c^{s_2}} \alpha_1.$$

Combining all conditions in the system we prove the proposition.

Appendix G: Proof of Proposition 5

The deviation to strategy $r_2^{s_1} = 0$ is not profitable when

$$\begin{aligned} &\beta_2\big(a^{s_1} + \delta_M \mathbf{p}^{s_1} V_{\mathbf{sc}}(\delta_M)\big) + \gamma_2\big(c^{s_1} + \delta_M \mathbf{p}^{s_1} V_{\mathbf{n}}(\delta_M)\big) \\ &\geqslant \beta_2\big(b^{s_1} + \delta_M \mathbf{p}^{s_1} V_{\mathbf{n}}(\delta_M)\big) + \gamma_2\big(d^{s_1} + \delta_M \mathbf{p}^{s_1} V_{\mathbf{n}}(\delta_M)\big). \end{aligned}$$

Taking into account that $\beta_2 = \frac{\beta_1}{1-\alpha_1}$ and $\gamma_2 = \frac{1-\alpha_1-\beta_1}{1-\alpha_1}$, we obtain:

$$\beta_1 \geqslant \frac{(d^{s_1} - c^{s_1})(1 - \alpha_1)}{d^{s_1} - c^{s_1} + a^{s_1} - b^{s_1} + \delta_M \mathbf{p}^{s_1}(V_{\mathbf{sc}}(\delta_M) - V_{\mathbf{n}}(\delta_M))}.$$

The fact that an M-class player adopts C in period 2 and state s_1 affects in turn the choice of a competitor of class P: we now examine the condition under which $q_1^{s_2} = 1$ is a part of PMPBE if the M player chooses $r_2^{s_1} = 1$ in period 2 and state s_1. The deviation of player P to strategy $q_1^{s_2} = 0$ is not profitable if

$$\begin{aligned} &\alpha_1 \left[a^{s_2} - b^{s_2} + \delta_P \mathbf{p}^{s_2}(V_{\mathbf{c}}(\delta_P) - V_{\mathbf{n}}(\delta_P))\right] \\ &+ \beta_1[c^{s_2} - d^{s_2} + \delta_P \mathbf{p}^{s_2}(V_{\mathbf{n}}(\delta_P) - V_{\mathbf{sc}}(\delta_P))] + (1 - \alpha_1 - \beta_1)(c^{s_2} - d^{s_2}) \geqslant 0, \end{aligned}$$

which is equivalent to inequality

$$\beta_1 \leq \frac{\alpha_1[d^{s_2} - c^{s_2} + a^{s_2} - b^{s_2} + \delta_P \mathbf{p}^{s_2}(V_{\mathbf{c}}(\delta_P) - V_{\mathbf{n}}(\delta_P))] - (d^{s_2} - c^{s_2})}{\delta_P \mathbf{p}^{s_2}(V_{\mathbf{sc}}(\delta_P) - V_{\mathbf{n}}(\delta_P))}.$$

5.1 Appendix E: Non-cooperative outcomes

In this section we show the non-cooperative results. In this case, the non-cooperative strategy profile $\sigma_{\mathbf{n}}$ is formed by rule (7) in the cooperation phase of the game. We classify the equilibria according to which the non-cooperative strategy profile is played in the cooperation phase regardless of the players' classes. Like in the main text, we consider first the case where the learning phase lasts one period: the results are summarised in the following proposition.

Proposition 6 *Suppose $T = 1$. Then the strategy profiles $(q_1^{s_1}, r_1^{s_1}) = (0, 0)$ and $(q_1^{s_2}) = (0)$ are PMPBE for the game with the initial state $s = s_1$ and $s = s_2$ respectively.*

Proof. Consider the initial state s_1 and the strategy profile $(q_1^{s_1}; r_1^{s_1}) = (0, 0)$. A player P obtains the following payoff if she does not deviate from $(0, 0)$:

$$d^{s_1} + \delta_P \mathbf{p}^{s_1} V_{\mathbf{n}}(\delta_P). \tag{28}$$

If she deviates from profile $(0, 0)$ $(q_1^{s_1} = 1)$, she gets:

$$c^{s_1} + \delta_P \mathbf{p}^{s_1} V_{\mathbf{n}}(\delta_P). \tag{29}$$

Note that (28) is always greater than or equal to (29), since $d^{s_1} \geqslant c^{s_1}$. A deviation of a class M cannot be profitable either.

Consider the initial state s_2 and the strategy profile $(q_1^{s_2}) = (0)$. A player P obtains the following payoff if she does not deviate from (0):

$$d^{s_2} + \delta_P \mathbf{p}^{s_2} V_{\mathbf{n}}(\delta_P). \tag{30}$$

If she deviates from profile (0) $(q_1^{s_2} = 1)$, she gets:

$$c^{s_2} + \delta_P \mathbf{p}^{s_2} V_{\mathbf{n}}(\delta_P). \tag{31}$$

Note that (30) is always greater than or equal to (31), since $d^{s_2} \geqslant c^{s_2}$. Therefore, the strategy profile $(q_1^{s_2}, r_1^{s_2}) = (0, 0)$ when the game starts from s_1 and $(q_1^{s_2}) = (0)$ when the game starts from s_2 are PMPBE. ■

We now turn to the case where the learning phase is endogenously determined. Let the initial state be s_2. Consider the strategy $q_1^{s_2} = 0$ of player P in period 1 in state $s = s_2$. In this case, all players use action D and, after this period, the beliefs are not updated: $\alpha_2 = \alpha_1$, $\beta_2 = \beta_1$, $\gamma_2 = \gamma_1$. If in any further periods only state $s = s_2$ is realized, then the strategy of player of type P is $q_t^{s_2} = 0$ because of the Markovian property of the strategy.

The beliefs can be changed only if state $s = s_1$ is realized in the game. Let state s_1 be realized in period $t > 1$. If in this state players use strategies $q_1^{s_1} = 0$ and $r_1^{s_1} = 0$, then the beliefs do not change and again $\alpha_t = \alpha_1$, $\beta_t = \beta_1$. Therefore, using the Markovian property, we get by induction $q_k^{s_1} = r_k^{s_1} = 0$ for any k. These strategies determine a subgame perfect equilibrium with infinite learning phase when players always adopt action D in any state. The existence of a similar PMPBE can be proved when the game starts from state s_1 and players use actions D in this state and then in the firstly appeared state s_2 they also use actions D. The ongoing discussion can be summarised as follows.

Proposition 7 *For any initial probabilities $\alpha_1 > 0$, $\beta_1 > 0$ such that $\alpha_1 + \beta_1 < 1$, there always exists PMPBE in which the players' strategies for both initial states s_1 and s_2 are as follows: $q_t^{s_1} = r_t^{s_1} = 0$ and $q_t^{s_2} = 0$, $t = 1, 2, \ldots$ (players of all classes choose action D in any state forever). In this case the learning phase lasts forever.*