Unprecedented*

Satoshi Fukuda† Yuichiro Kamada‡

November 7, 2022

Abstract

This paper considers a dynamic game in which each player can take a new action only if either she privately learns it or the opponent takes it. The new action profile is a Nash equilibrium, and is Pareto dominated by the default action profile. Under the assumptions that taking the new action is an irreversible choice and moves are asynchronous, we show that, when probability of private learning is low and players are patient, there is a unique perfect Bayesian equilibrium. In the unique equilibrium, the new action is never taken, i.e., the new action remains unprecedented. This is the case even though, after many periods, it is almost common knowledge among the players that they have learned the new action.

JEL Classification: C73; D80

Keywords: precedents; asynchronous moves; unique equilibrium; almost common knowledge; repeated game

*We would like to thank Yu Awaya, Mehmet Ekmekci, Jeff Ely, Francesco Fabbri, Akihiko Matsui, Stephen Morris and Burkhard Schipper for their comments. We also thank the seminar participants at UC Berkeley lunch seminar and Drew Fudenberg's 65th birthday conference. We thank Francesco Bilotta for his excellent research assistantship.

†Department of Decision Sciences and IGIER, Bocconi University, Via Roentgen 1, Milan 20136, Italy, e-mail: satoshi.fukuda@unibocconi.it

‡Haas School of Business, University of California Berkeley, 2220 Piedmont Avenue, Berkeley, CA 94720-1900, USA, e-mail: y.cam.24@gmail.com

1 Introduction

There are many situations where an action can be taken only when there is a precedent. Absence of a precedent is often the public sector's favorite excuse to avoid doing something new, and this "precedentism" is a target of criticism not only by mass media and the general public but also by academics (Merton, 1940, 1968).

If, however, an action cannot be taken without a precedent, then that action can never be taken. In real life, however, some people may accidentally find out a way to take such an action or is granted a right to take the action for some exogenous reason. For example, a firm may invent a promotion strategy to effectively attract customers, and once this firm employs such a strategy, it may be used by any other firms.[1] Technological innovations by firms or development of war weapons by countries are also an example of an action that can be taken after seeing a precedent.[2] Even when a city hall has been declining the request to provide data to researchers because there is no precedent, a researcher may accidentally get a connection to the city mayor which enables her to get a permission to obtain data from the city hall. Once she uses this right and obtains data, it becomes a precedent and other researchers can also obtain data. In sports, once an athlete acquires a new technique and performs it in a competition, other athletes will start mimicking that technique. The "Tomoa Skip" in the speed category of sport climbing and the "Fosbury Flop" in the high jump are now used by almost all athletes in the respective competitions.[3]

In this paper, we are interested in the incentives of players when there is an action that needs a precedent for it to be played. Specifically, we consider the situation in which a 2×2 stage game is repeatedly played by two players. In the stage game, one action is the "old action" that the players can play from the initial period and the other action is the "new action" that a player can only take when either she has accidentally learned it privately or the opponent has chosen it in the past. We consider a particular class of stage games in which the old action profile Pareto dominates the new action profile (as in the case of technological innovations that are harmful for the climate or development of war weapons) while the latter is a Nash equilibrium. This

[1] For instance, since American Airlines has launched the world's first mileage-based frequent-flyer program (AAdvantage) in 1981, more than 130 airlines issue miles as of 2005 (The Economist, 2005).

[2] See, for instance, Pandya (2019).

[3] There are many other examples of performance breakthroughs in sports that have become new standards. See Veenendaal (2018).

class includes coordination games as well as prisoner's dilemma.

Under the assumptions to be discussed shortly, we show that the model has a unique perfect Bayesian equilibrium when the probability of private learning is small and the players are sufficiently patient. In the unique equilibrium, players keep playing the old action, even if both of them have privately learned the new action. Hence, *the new action will remain unprecedented*.

There are two assumptions in obtaining this conclusion: *irreversibility* and *asynchronicity*. Irreversibility means that once a player chooses a new action, she cannot switch back to the old action. Asynchronicity means that the two players alternate in choosing actions.[4] We show that if one of these assumptions fails, there are multiple equilibria.

Why will the new action remain unprecedented? To understand the intuition, suppose that the probability of private learning is small and players use pure strategies. If a player happens to privately learn the new action in the first period, she may not want to play it right away because it is unlikely that the opponent has already learned it, and thus keeping the new action a "secret" is a good strategy. This argument may seem to break down when the time passes and the probability that each player has learned the new action has become large. Suppose that player i is supposed to play the new action at (large) time t. If she followed this strategy, then she might obtain the instantaneous gain, but from the next period on the two players would be playing the new action forever. If, however, she deviates and plays the old action, that would substantially change the opponent j's belief about the likelihood of player i knowing the new action, incentivizing j to play the old action as in the first period. Hence, player i has an incentive to deviate and play the old action. The actual proof is more contrived as we consider mixed strategies as well.

The irreversibility and asynchronicity play key roles in the preceding argument. We used irreversibility when claiming that once a player takes the new action, the players will be playing the new action forever. Without irreversibility, the game after the new action is played is the standard repeated game, so the folk theorem

[4] We do not claim that these are the most natural assumptions in all applications, but they may be natural in some settings. For an example of irreversibility, a firm may have a reputation concern and may not be able to switch back to providing no promotion once a promotion strategy is launched. Another way to think of irreversibility as a natural assumption is to view the problem as that of an optimal stopping problem. For an example of asynchronicity, the decision times of competing firms are typically not coordinated with each other.

generically holds in this continuation game. Players can condition their continuation play on who has played the new action first, so as to incentivize players to play the new action as soon as they learn it. This results in the existence of an equilibrium in which players play the new action as soon as they can. Also, we used asynchronicity when discussing the consequence of a deviation to play the old action. If moves are synchronous, the opponent may be playing the new action at the same time as the player takes the new action. When it is likely that the opponent has learned the new action and hence plays it with high probability, it is risky to take the old action. This implies the existence of an equilibrium in which the new action is played on the equilibrium path.[5]

The insight that a sharper prediction can be obtained when moves are asynchronous appears in the literature, e.g., Maskin and Tirole (1987, 1988a,b) and Lagunoff and Matsui (1997, 2001). The key factor in these papers is that a player can commit to an action when the opponent chooses an action. In our model, however, it is rather that a player can take a certain action to influence the opponent's belief update while the opponent is not choosing an action.

In our unique equilibrium, the new action remains unprecedented even when it is very close to common knowledge that the players have learned the new action. This may look at odds with the result of Monderer and Samet (1989) that shows certain continuity of the equilibrium set with respect to the degree of common beliefs. Section 4.3 discusses how to reconcile this seeming contradiction, where we argue that in our model, there is an action that can change what is almost common knowledge, which contributes to the difference.

We assume that a new action is initially unavailable to the players. One interpretation of this assumption is that it is a feasibility constraint. Another interpretation is that the player is initially unaware of the new action, and she learns it through private learning or a precedent. Defining strategies and equilibria in dynamic games with unawareness is a tricky task because a player's strategy cannot condition on what she is unaware of. In our model, however, the interpretation based on unawareness is still valid as long as there are only two actions (as in our main model). This is because a player's strategy is only relevant when the player has choices between

[5]For some parameter range, we use a public randomization device to obtain the existence of such an equilibrium. Our uniqueness result under asynchronicity, which is our main theorem, holds even if a public randomization device exists.

multiple actions, and this happens only when the player has already learned the new action.[6] There is a burgeoning literature, such as Feinberg (2004), Chung and Fortnow (2016), Jéhiel and Newman (2019), Heifetz et al. (2021), and Schipper (2021), which studies dynamic games that involve unawareness about actions. In particular, Feinberg (2004) considers a finitely repeated prisoner's dilemma in which one player may be unaware of the "defect" action, while she can play it if the other player—who knows both actions—plays it. He showed that some cooperation will be achieved in equilibrium. His result uses the finite horizon of the game and reminiscent of the reputation argument where the probability of unawareness is fixed. We, in contrast, consider infinite horizon and the probability of unawareness converges to zero due to continual private learning.

The paper is structured as follows. The rest of this section discusses the related literature. Section 2 formulates our model and Section 3 provides our main uniqueness result. Section 4 discusses several topics. First, Sections 4.1 and 4.2 discuss the models that violate irreversibility or asynchronicity and show that multiple equilibria exist. Section 4.3 discusses the relationship of our result with the one in Monderer and Samet (1989). Section 4.4 extends our model to the case with more than two actions, and Section 4.5 considers the case when players may not be able to learn the new action privately. Section 4.6 discusses the cases of low discount factors. Section 5 provides concluding remarks. The proofs are relegated to Appendix A. The Online Supplementary Appendix contains additional results and discussions.

2 Model

There are two players, 1 and 2. Time is discrete and the horizon is infinite: $t = 1, 2, \ldots$.

Player 1 chooses her action $a_{1,t}$ from set $A_{1,t} \subseteq \{O, N\}$ at every period t that is odd.[7] Player 2 chooses her action $a_{2,t}$ from set $A_{2,t} \subseteq \{O, N\}$ at every period t that is even. At each period in which player i moves, before i makes a choice of her action,

[6] This argument breaks down if there are three or more actions as in Section 4.4. Whether player i's strategy (defined in that section) is a best response when i has a choice between two actions depends on what the opponent j would do after j has learned the third action which i is currently unaware of.

[7] The notation O stands for an "old" or "original" action, and N stands for a "new" action.

she privately learns action N with probability $p_i \in [0,1]$, independently across players and periods, if she has not learned it yet.

To define the players' action sets at every period, we define player i's private history $h_{i,t}$ at time t as the following form:

$$h_{i,t} = \big((a^\tau)_{\tau \in \{1,\ldots,t-1\}}, s\big) \in (\{O,N\} \times \{O,N\})^{t-1} \times \{1,\ldots,t+1\},$$

where the interpretation is that each a^τ is the action profile taken at time period τ in the past, and s is the period at which player i has learned action N by chance if she has learned it by chance before or at time t, and $s = t+1$ otherwise. If we can write $h_{i,t}$ in the above form, we write $s(h_{i,t}) := s$. We say that a private history is *feasible* if it can happen with positive probability under the above-specified rule in which the action sets evolve.

Player i's action set $A_{i,t}$ is recursively defined according to the following rule. Initially, only action O is available to i. Action N becomes available to i if either i privately learns it or the opponent plays it in the past. Once player i chooses N, N becomes the only choice for i.

Formally, the set of actions available to the moving player i at period t can be written as:

$$A_{i,t} = A(h_{i,t}) = \begin{cases} \{O\} & \text{if } s(h_{i,t}) = t+1 \text{ and } a_{-i,\tau} = O \text{ for all } \tau \leq t-1 \\ \{N\} & \text{if } a_{i,\tau} = N \text{ for some } \tau \leq t-2 \\ \{O,N\} & \text{otherwise} \end{cases}.$$

Let $H_{i,t}$ be the set of all feasible private histories of player i at time t. Let $H_i := \bigcup_{t \in \mathbb{N}} H_{i,t}$.

Player i's strategy is a mapping $\sigma_i : H_i \to \Delta(\{O,N\})$ such that $\sigma_i(h_{i,t})(A(h_{i,t})) = 1$ for every $h_{i,t} \in H_i$. Let Σ_i be the set of all strategies of player i, and let $\Sigma = \Sigma_1 \times \Sigma_2$.

Player i's stage-game payoff at time t under action profile (a_1^t, a_2^t), which we denote by $u_i(a_1^t, a_2^t)$, is given by the payoff matrix as in Table 1, where $\ell_i > 0$ and $x_i \in \mathbb{R}$ for each $i = 1, 2$.

Thus, we consider the class of (stage) games with an inefficient Nash equilibrium. While the action profile (N,N) is a Nash equilibrium, the action profile (O,O) Pareto-dominates (N,N). When $x_i > 1$ for each $i = 1, 2$, the stage game is the prisoner's

	O	N
O	$1, 1$	$-\ell_1, x_2$
N	$x_1, -\ell_2$	$0, 0$

Table 1: The payoff matrix of the stage game

dilemma. When $x_i < 1$ for each $i = 1, 2$, the stage game is a coordination game (in this case, the action profile (O, O) is also a stage-game Nash equilibrium).[8] In other words, the class of (stage) games we consider contains the prisoner's dilemma and coordination games.

The discount factor is $\delta_i \in (0, 1)$ for each i. Given the action sequence $(a^t)_{t=1}^{\infty}$, player i's payoff in the supergame is $(1 - \delta_i) \sum_{t=1}^{\infty} \delta_i^{t-1} u_i(a^t)$. We sometimes use the unnormalized payoff in the supergame, and it is $\sum_{t=1}^{\infty} \delta_i^{t-1} u_i(a^t)$.

Given i's private history $h_{i,t} \in H_i$ and strategy profile $\sigma \in \Sigma$, we can define player i's continuation payoff $\pi_i(\sigma | h_{i,t})$.

Our solution concept is a perfect Bayesian equilibrium (PBE).

3 Main Result: Unique Equilibrium

Let σ^{G} be the grim trigger strategy profile, that is, each player chooses action O if and only if no one has chosen N in the past.

Theorem 1. *There exist $\overline{p} \in (0, 1)$ and $\overline{\delta} \in (0, 1)$ such that if $p_i < \overline{p}$ and $\delta_i > \overline{\delta}$ for all $i = 1, 2$, then the grim trigger strategy profile σ^{G} is a unique PBE.*

That is, in the unique equilibrium, action N keeps being unprecedented. This result is of particular interest because on the path of equilibrium play, after many periods, it is "almost common knowledge" among the players that they have learned action N. When $x_i > 1$ for each $i = 1, 2$, although σ^N, which we define to be a strategy profile in which each player plays N whenever she learns action N, is an equilibrium in the game in which the "almost common knowledge" is replaced with the exact common knowledge, it is not an equilibrium in our model.[9] We will discuss this issue in more depth in Section 4.3 after explaining the proof.

[8] When $x_i \in (0, 1)$, the stage game is a stag hunt game.

[9] In our context, common knowledge is equivalent to common certainty (common belief with probability one).

To explain the proof of this theorem, we introduce some notation. Define $\pi_i(\delta_i)$ as

$$\pi_i(\delta_i) := p_{-i}[1 + \delta_i(-\ell_i) + \delta_i^2 \cdot 0] + (1 - p_{-i})[1 + \delta_i \cdot 1 + \delta_i^2 x_i].$$

To interpret this, suppose that the players play σ^N and, at period t, the moving player i learns action N while $-i$ has not. Then, $\pi_i(\delta_i)$ is her unnormalized payoff when she makes a one-period deviation to play O at t (and follows the original strategy σ_i^N later on).

For each i, for a given $p_{-i} \in [0, 1]$, we define

$$\begin{aligned}
\overline{\delta}_i(p_{-i}) &:= \min\{\delta_i' \in [0,1] \mid x_i < \pi_i(\delta_i) \text{ for all } \delta_i \in (\delta_i', 1)\}, \text{ and} \\
\underline{\delta}_i(p_{-i}) &:= \max\{\delta_i' \in [0,1] \mid x_i < \pi_i(\delta_i) \text{ for all } \delta_i \in [0, \delta_i')\},
\end{aligned}$$

where we sometimes omit the dependence on p_{-i}. The interpretation of these variables is the following. Suppose that players play σ^N and, at period t, the moving player i learns action N while $-i$ has not. The left-hand side in the defining inequality of $\overline{\delta}_i(p_{-i})$, x_i, is i's unnormalized payoff when i follows this strategy. The right-hand side, $\pi_i(\delta_i)$, is her unnormalized payoff when she makes a one-period deviation to play O at t (and follows the original strategy σ_i^N later on). Thus, the discount factor $\overline{\delta}_i(p_{-i})$ is the minimum one above which such one-period deviation is always profitable for player i. Likewise, the discount factor $\underline{\delta}_i(p_{-i})$ is the maximum one below which such one-period deviation is always profitable for player i.

We also define

$$\hat{\delta}_i := \min\{\delta_i \in [0,1] \mid (1 - \delta_i)x_i \leq 1\}.$$

To interpret this, consider the strategy profile σ^{G}. Suppose that at period t, the moving player i learns action N while $-i$ has not. The left-hand side in the defining inequality of $\hat{\delta}_i$, $(1 - \delta_i)x_i$, is i's payoff when i makes a one-period deviation to play N at t (and, by the assumption on the evolution of the action sets, play N forever). The right-hand side, 1, is her payoff when she follows strategy σ_i^{G}. The discount factor $\hat{\delta}_i$ is the minimum one for which such one-period deviation is not profitable for player i. Note that if $x_i > 1$ then $\hat{\delta}_i = \frac{x_i - 1}{x_i}$ and $\hat{\delta}_i = 0$ otherwise.

Now, we are ready to give a proof of Theorem 1.

Proof of Theorem 1. We prove the theorem through a series of lemmas, whose formal proofs we provide in the Appendix. Here we provide the formal statements of the lemmas as well as their intuitions.

First, the standard argument shows that the grim trigger strategy profile σ^{G} is an equilibrium when the players are sufficiently patient.

Lemma 1. *Suppose $\delta_i \geq \hat{\delta}_i$ for each $i = 1, 2$. Then, for any $p_1, p_2 \in [0, 1]$, σ^{G} is a PBE.*

Next, we show that σ^{G} is the only possibility for an equilibrium under which action N is not observed on the path of play.

Lemma 2. *For any $\delta_1, \delta_2 \in (0, 1)$ and $p_1, p_2 \in [0, 1]$, fix any PBE σ in which, on the path of play, there is ex ante probability zero that some player plays N. Then, $\sigma = \sigma^{\mathrm{G}}$.*

This lemma is straightforward as well because i's playing N is a unique best response once $-i$ chooses N. This follows from $-\ell_i < 0$ and the assumption that $-i$ will not be able to change his action from N.

Now, notice that at time t, the moving player i needs to make a choice from multiple actions only at private histories of the form $h_{i,t}$ such that $h_{i,t}$ involves no observation of action N and she has already learned action N (i.e., $s(h_{i,t}) \leq t$). Let the set of such histories be $\overline{H}_{i,t}$.

Fix $p_1, p_2 \in [0, 1]$ and $\delta_1, \delta_2 \in (0, 1)$. Suppose for a contradiction that there is a (potentially mixed) PBE other than σ^{G}, and fix an arbitrary one of them, σ^*. Let t_1 be the first period at which some player's equilibrium strategy σ_i^* assigns positive probability to action N after some history:

$$t_1 := \min\{t \in \mathbb{N} \mid \sigma_i^*(h_{i,t})(N) > 0 \text{ for some } i \in \{1, 2\} \text{ and some } h_{i,t} \in \overline{H}_{i,t}\},$$

with a convention that $t_1 := \infty$ if there is no player i, time t and her private history $h_{i,t} \in \overline{H}_{i,t}$ such that $\sigma_i^*(h_{i,t})(N) > 0$. Since we have assumed $\sigma^* \neq \sigma^{\mathrm{G}}$, Lemma 2 implies $t_1 < \infty$. Now we recursively define t_k. First, if $t_{k-1} < \infty$, let t_k be the k-th period at which some player i's strategy σ_i^* assigns positive probability to action N after some history: for each $k \in \mathbb{N}$ with $k \geq 2$,

$$t_k := \min\{t \in \mathbb{N} \mid t > t_{k-1} \text{ and } \sigma_i^*(h_{i,t})(N) > 0 \text{ for some } i \in \{1, 2\} \text{ and some } h_{i,t} \in \overline{H}_{i,t}\}$$

with the same convention as before that $t_k := \infty$ if there is no player i, time t with

$t > t_{k-1}$ and her private history $h_{i,t} \in \overline{H}_{i,t}$ such that $\sigma_i^*(h_{i,t})(N) > 0$. If $t_{k-1} = \infty$, then we let $t_k = \infty$.

We show that t_k's are consecutive if they are finite:

Lemma 3. *If $\delta_i > \hat{\delta}_i$, then for all $k \in \mathbb{N}$ such that i moves at t_k, $t_k < \infty$ implies $t_{k+1} = t_k + 1$.*

We also show that, for any time t and any histories in $\overline{H}_{i,t}$, the moving player i has the same continuation payoff.

Lemma 4. *For any i, t, and $h_{i,t}, h'_{i,t} \in \overline{H}_{i,t}$, we have $\pi_i(\sigma^*|h_{i,t}) = \pi_i(\sigma^*|h'_{i,t})$.*

Hence, at and after time t_1, the moving player i's continuation payoff given any $h_{i,t_1} \in \overline{H}_{i,t}$ must be equal to the payoff from her playing action N, which is $(1-\delta_i)x_i$.

Lemma 5. *We have $\overline{\delta}_i \geq \hat{\delta}_i$ for each $i = 1, 2$.*

We next show that at no t_k, the moving player i plays a pure action (i.e., she takes action N with probability 1).

Lemma 6. *Suppose $\delta_i > \hat{\delta}_i$ and $\delta_{-i} \in (0, \underline{\delta}_{-i}) \cup (\overline{\delta}_{-i}, 1)$. Then, for any $k \in \mathbb{N}$ such that i moves at t_k, $\sigma_i^*(h_{i,t_k})(N) < 1$ holds for some $h_{i,t_k} \in \overline{H}_{t_k}$.*

The intuition behind this result is as follows. If i assigned probability 0 to action O at some t_k when she has privately learned N, then conditional on i's playing O, $-i$ at the next period become sure that i has not learned action N yet. This would make it the unique best response for $-i$ to play O in order to make action N a "secret," instead of playing N which teaches player i the existence of action N. This is a contradiction to our earlier conclusion in Lemma 3 which would imply that N must be a best response for $-i$ at period $t_k + 1$. Condition $\delta_i > \hat{\delta}_i$ ensures that one-period deviation by i is not worthwhile at t_k, and condition $\delta_{-i} \in (0, \underline{\delta}_{-i}) \cup (\overline{\delta}_{-i}, 1)$ ensures that $-i$ is either (i) impatient enough so that the impact of the one-period loss from playing N is too large ($0 < \underline{\delta}_{-i}$ holds only when $x_{-i} < 1$) or (ii) patient enough so that sticking to O is strictly better than playing N to obtain a one-period gain in the payoff.

For any time t at which i moves, for any private history $h_{i,t} \in \overline{H}_{i,t}$, let $\mu(h_{i,t})$ be the posterior probability that i at $h_{i,t}$ assigns to the event that $-i$ knows action N at the next period that $-i$ plays (that is, after the learning stage in the next period). Notice that $\mu(h_{i,t}) = \mu(h'_{i,t})$ must hold for any $h_{i,t}, h'_{i,t} \in \overline{H}_{i,t}$ because private learning happens independently across players and $\mu(\cdot)$ is derived from Bayes rule because, under σ^*, the event under which player $-i$ chooses action O at every period in the past is assigned a strictly positive probability (since it is possible that $-i$ has never privately learned action N at any past periods). So, simply denote $\mu_t := \mu(h_{i,t})$ for any $h_{i,t} \in \overline{H}_{i,t}$.

Lemma 7. *Suppose $p_j \in [0,1)$ for each $j = 1,2$ and suppose $\delta_i > \hat{\delta}_i$ and $\delta_{-i} \in (0, \underline{\delta}_{-i}) \cup (\overline{\delta}_{-i}, 1)$. Then, there exist $j = 1,2$ and $\alpha_j^* > 1$ such that, for any $t \geq t_1$ such that j moves, we have $1 - \mu_{t+2} = \alpha_j^*(1 - \mu_t)$.*

The intuition behind this lemma is as follows. Take any $t \geq t_1$ and let the moving player be i. First, Lemma 3 shows that N is a best response for i at some history in $\overline{H}_{i,t}$, where we use Lemma 5 to show the condition necessary for Lemma 3, i.e., $\delta_{-i} > \hat{\delta}_{-i}$, is satisfied. Also, Lemma 6 implies that O is a best response for i at some history in $\overline{H}_{i,t}$. Hence, Lemma 4 implies that both O and N are a best response for i at every history in $\overline{H}_{i,t}$.

Now, the fact that i is indifferent at every history in $\overline{H}_{i,t}$ at every t_k has implications on the evolution of the belief μ_t. To see this, first note that since i is indifferent at t_k, her payoff from playing O at t_k is $(1-\delta_i)x_i$, which is the payoff she gets when she plays N. Suppose i plays O at t_k. If $-i$ plays N at $t_k + 1$, then i receives a payoff of $(1-\delta_i)(-\ell_i)$ from period $t_k + 1$ onward. The probability of this event is $\mu_t r_{t+1}$, where r_{t+1} is the probability that $-i$ plays N at period $t_k + 1$, conditional on his having privately learned N at $t_k + 1$. If $-i$ plays O at $t_k + 1$, then i receives a payoff of $(1-\delta_i)x_i$ from period $t_k + 2$ onward by the indifference at $t_k + 2$. The probability of this event is $1 - \mu_t r_{t+1}$. Summarizing, by the indifference at period t_k, we have that $(1-\delta_i)x_i$ is equal to a convex combination of some constants, where the weighs are given by $\mu_t r_{t+1}$ and $1 - \mu_t r_{t+1}$. This implies that $\mu_t r_{t+1}$ is a constant across all t. Note that $\mu_t r_{t+1} < 1$ must hold if $p_{-i} < 1$.

Finally, when $p_{-i} < 1$, the Bayes rule implies

$$1 - \mu_{t+2} = \frac{(1-p_{-i})(1-\mu_t)}{1-\mu_t r_{t+1}},$$

and since $\mu_t r_{t+1}$ is a constant, we have $1 - \mu_{t+2} = \alpha_i(1-\mu_t)$, where $\alpha_i := \frac{1-p_{-i}}{1-\mu_t r_{t+1}}$ is a constant. With some more work we can show $\alpha_j > 1$ for some $j = 1,2$.

Now, suppose that there is $i = 1,2$ such that $\delta_i > \hat{\delta}_i$ and $\delta_{-i} > \overline{\delta}_{-i}(p_i)$. Since

Lemma 7 implies there is $j = 1, 2$ such that $1-\mu_{t+2} = \alpha_j^*(1-\mu_t)$ with $\alpha_j^* > 1$ whenever j moves at t, there exists $\bar{t}$ such that $\mu_{\bar{t}} < 0$. But this cannot hold because μ_t must be a probability. This is a contradiction, and thus there cannot exist a (potentially mixed) PBE other than σ^{G}.

Lemma 8. *There exists $\overline{p} > 0$ such that $\overline{\delta}_i(p_{-i}) < 1$ for all $i = 1, 2$ and $p_{-i} \in (0, \overline{p})$.*

This lemma is straightforward because of the continuity of the function π_i that defines $\overline{\delta}_i(p_{-i})$: we have $\pi_i(\delta_i) = 2 + x_i > x_i$ when $\delta_i = 1$ and $p_{-i} = 0$, and π_i is continuous in δ_i and p_{-i} at these values.

Now, take $\overline{p}$ as in Lemma 8 and $\overline{\delta} = \max(\overline{\delta}_1, \overline{\delta}_2)$. The proof of Theorem 1 is now complete. □

In the unique equilibrium, action N is never played. For any small $\varepsilon > 0$, there exists $t \in \mathbb{N}$ such that, at t, it is common belief among the players that they have learned action N with probability greater than $1 - \varepsilon$. On the one hand, if $\varepsilon = 0$, then there is an equilibrium in which N is played. On the other hand, if $\varepsilon > 0$, then N cannot be played in the unique equilibrium (under the condition of Theorem 1).

4 Discussion

4.1 Reversible Case

Our main model assumes irreversibility of actions: Once player i chooses N, she will never be able to take action O again. This section argues that some type of irreversibility assumption is necessary for our main result. To do this, consider the model that is the same as our main model except that the action set of player i evolves in a different manner[10]: player i initially knows O only, and privately learns N with probability p_i each period she moves if she has not learned it yet. Once she privately learns N or $-i$ chooses N, i's action set is $\{O, N\}$. That is, the action set includes N under the same condition as in our main model, while O is always in the action set. Call this model the *reversible model*. Notice that for any $p_1, p_2 \in [0, 1]$, σ^{G} is a PBE in this model if $\delta_1, \delta_2 < 1$ are sufficiently high. A stage game is a *pure coordination game* if $x_i = -\ell_{-i}$ for each $i = 1, 2$. We can show the following:

[10]We note that the same argument as the one we make in this section can be made for the case when the moves are synchronous, as in the next section (Section 4.2).

Proposition 1. *Suppose that the stage game is not a pure coordination game. For any $p_1, p_2 \in (0, 1]$, there exists $\delta' < 1$ such that if $\delta_i \in (\delta', 1)$ for each $i = 1, 2$, then the reversible model has a PBE in which action N is played with probability 1.*

To show this, we construct an equilibrium strategy profile. In this equilibrium, the first player who privately learns N plays N. Let this first player be j. In the continuation game after j has played N, an equilibrium $\sigma^{(j)}$ will be played. The strategy profiles $\sigma^{(1)}$ and $\sigma^{(2)}$ are set up so that the implied equilibrium payoff profile $(u_1^{(1)}, u_2^{(1)})$ and $(u_1^{(2)}, u_2^{(2)})$ satisfy $u_1^{(1)} > u_1^{(2)} > 0$ and $u_2^{(2)} > u_2^{(1)} > 0$. Since the stage game is not a pure coordination game, such equilibria can be easily constructed when the discount factors are high by a standard argument of folk theorems on asynchronous moves (e.g., Dutta, 1995; Yoon, 2001). This concludes the proof of the theorem.[11]

The key is that the players' flexibility to switch back and forth between the two actions—such a back-and-forth happens in equilibria $\sigma^{(1)}$ and $\sigma^{(2)}$—makes it possible, in the continuation game after some player j has chosen N, to set up a "reward" for j and a punishment for $-j$, and this incentivizes players to choose N as soon as possible.

4.2 Synchronous Case

Our main model assumes asynchronous moves. To clarify the role of asynchronicity, this section considers a model with synchronous moves, and shows that there are multiple equilibria when the players are patient.

Formally, suppose that at each period $t = 1, 2, \ldots$, the two players simultaneously make a choice. As in the main model, each player i initially knows only O, but she privately learns N with probability p_i each period. Player i can choose action N if either she has privately learned it or the opponent has taken it. Once i takes action N, that will be the only choice for her thereafter.[12] Suppose also that at each period, a public randomization device is available.[13] Specifically, at the beginning of each period, a real number q_t is drawn (independently over time) according to the uniform

[11] If $p_i > 0 = p_{-i}$, then the conclusion of Proposition 1 holds if and only if $x_i \geq 1$. We set $u_i^{(i)} \geq 1$ in the proof.

[12] We note that the same argument as the one to be made in this section can be made for the case when the action set evolves as in the previous subsection (Section 4.1).

[13] The public randomization device can be anything, such as a sunspot, that the players can commonly observe but does not directly affect the payoffs. Our main theorem (Theorem 1) holds even when a public randomization device exists, which strengthens our uniqueness result.

distribution over $[0,1]$, and both players observe its realization. Call this model the *synchronous model*.

The set of private histories $H_{i,t}$, strategies, and PBE are defined analogously to Section 2. Consider the following strategy $\sigma_i^{(T,\overline{q})}$: At any $h_{i,t} \in H_{i,t}$, i plays N if and only if (i) some player has already taken N, or (ii) i has privately learned N, $t = n \cdot T$ for some $n = 1, 2, \ldots$, and $q_t \leq \overline{q}$.

Proposition 2. *For any $p_1, p_2 \in (0,1]$ and $\overline{q} \in [0,1)$, there exist $\delta' \in (0,1)$ and $T' < \infty$ such that if $\delta_i \in (\delta', 1)$ for each $i = 1, 2$ and $T > T'$, then $\sigma^{(T,\overline{q})}$ is a PBE in the synchronous model.*

The intuition behind this result hinges on synchronous moves. For any p_{-i}, if we take T sufficiently high, then the posterior probability that the opponent has learned N is sufficiently high at period T. Hence, when $q_t \leq \overline{q}$, i has an incentive to play N because she attaches probability close to 1 to the event that the opponent plays N as well. The opponent is in the same situation and hence plays N. This argument would break down if moves are asynchronous because $-i$ would not play when i plays, and thus i could take O to manipulate $-i$'s belief in the next period, without receiving the payoff of $-\ell_i$.

When $x_i > 1$ for some $i = 1, 2$, the public randomization device was used to incentivize players to play O at period $T - 1$: Given $\overline{q} < 1$, there is a positive probability that players continue playing O for additional T periods which is long, and thus players are incentivized to take O instead of receiving an instantaneous gain of x_i.

We note that public randomization is not necessary for multiple equilibria if $x_i \leq 1$ for each i. Moreover, it is not necessary even if $x_i > 1$ for some $i = 1, 2$ for a certain parameter range such that there is a unique PBE under the main model of asynchronous moves. These points are investigated in the Online Supplementary Appendix.

Given that T' in Proposition 2 can be taken uniformly across δ, the payoff from $\sigma^{(T,\overline{q})}$ can be shown to approach 0 as $\delta \to 1$ for any $\overline{q} \in (0,1)$.

Proposition 3. *For any $p_1, p_2 \in (0,1]$ and $\overline{q} \in (0,1)$, there is $T < \infty$ such that (i) $\sigma^{(T,\overline{q})}$ is a PBE of the synchronous model and (ii) the expected payoff from $\sigma^{(T,\overline{q})}$ converges to 0 as $\delta_i \to 1$ for each $i = 1, 2$.*

4.3 Almost Common Knowledge of the Unprecedented Action

The unique equilibrium σ^{G} has the feature that, for any fixed (p_1, p_2) and $\varepsilon > 0$, there is $T < \infty$ such that at period T, it is common $(1-\varepsilon)$-belief among players that they have privately learned N. Note that if $x_i > 0$ for each $i = 1, 2$, and if common $(1-\varepsilon)$-belief was replaced with common knowledge, then σ^N would also be an equilibrium in the continuation game starting at period T. One may wonder how this seeming discontinuity can be reconciled with the result of Monderer and Samet (1989), which provides a certain continuity property of the equilibrium set with respect to the degree of common beliefs.

To understand this, recall that Monderer and Samet (1989) consider a game Γ in which each player obtains some information about which of the multiple possible games $(\Gamma_1, \ldots, \Gamma_m)$ is the true game, where those games share a common action set for each player. They show that for any $\varepsilon > 0$, if it is common p-belief in Γ that players are playing game Γ_1 with sufficiently high p, then for any Nash equilibrium s of Γ_1, "players play as in s when the true game is in fact Γ_1" is part of an ε-Nash equilibrium of the original game Γ where best response is required conditional on the information each player has at T.[14] In our context, we could let Γ be the continuation game starting at period T, and imagine there are four possible games, $\Gamma_1, \ldots, \Gamma_4$: Both players have learned N in Γ_1, only player 1 has learned it in Γ_2, only 2 has learned it in Γ_3, and no one has learned it in Γ_4. Each player knows whether she herself has learned N or not. For any $\varepsilon > 0$, we can take T large enough so that it is common p-belief with high p that the game is Γ_1.[15] By the result of Monderer and Samet (1989), since σ^N is a Nash equilibrium of Γ_1, there must exist an ε-Nash equilibrium of the continuation game Γ in which each player plays N as soon as possible at Γ_1.

First, we note that Monderer and Samet (1989)'s result only implies ε-equilibrium, as opposed to the exact equilibrium. Thus, the fact that σ^N is not an exact equilibrium in Γ is not a contradiction *per se*. We now discuss why σ^N is only an ε-equilibrium of Γ, and the nature of a deviation that a player would like to make.

[14] This version of the statement corresponds to their result that requires "ex post" ε equilibrium.

[15] Precisely speaking, there are more states because a player can learn N at one of many periods in the past, but this variation is not essential to our discussion here. For this reason, this remark (and only this remark) assumes that players' actions do not condition on the period in which they learned N.

To see where the approximation appears in the entire game Γ, note first that the moving player i at T is taking a best response given that the opponent plays σ_{-i}^N. This is because, if the opponent will take N with high probability in the next period, i wants to take N at the current period. A suboptimal action is taken, however, at period $T+1$ after i has played O at T. Although σ^N dictates that $-i$ plays N even under such a history, playing N is suboptimal *because $-i$ has updated his belief* to assign only a small probability to the event that i would have learned N at the next period. Since the event that i plays O at T happens with a very small probability under σ^N (it happens only when i has not learned N yet), what $-i$ plans to do after such an event affects $-i$'s expected payoff only minimally, and thus $-i$ is still taking an ε-best response under σ^N.[16] Conditional on the event that i plays O at T, however, σ_{-i}^N is not only ε away from a best response at $T+1$ but is substantially suboptimal.

The key is that *there is an action that can change what is almost common knowledge*: the event that players have learned N by T was almost common knowledge at T but, once player i takes O at T, it is no longer almost common knowledge. Monderer and Samet (1989)'s model, in contrast, does not allow for such an action because their model is static.

It is worth noting the two results that Fudenberg and Tirole (1991) prove. First, they present a corollary of Monderer and Samet (1989), which says that if there is a strict Nash equilibrium in each Γ_k, then there is an exact (not ε-) equilibrium of the entire game Γ that plays s when the true game is in fact Γ_1 (which is common p-belief). In our context, the hypothesis of this corollary does not apply because σ^N is not a strict equilibrium in Γ_1: Player $-i$ assigns probability 0 to i's playing O at T, and hence he is indifferent among any actions under such a history at $T+1$. Indeed, as we have explained, the equilibrium we construct is not exact because $-i$ was playing suboptimally under such a history.

Second, Fudenberg and Tirole (1991) prove that there is an (exact) Nash equilibrium of Γ that is close to playing a Nash equilibrium at each game Γ_k. Their proof uses the theorem by Wu and Jiang (1962) which assumes finite strategic-form games (as Fudenberg and Tirole (1991) has clarified in their statement). Since our strategy space is infinite, the theorem does not apply.

[16]Note that we are still talking about the notion of *ex post* equilibrium here in the sense defined in Monderer and Samet (1989) because the game has already started.

4.4 The Case with More Than Two Actions

In the main analysis, we assumed that each player has two actions in the stage game. In this section, we consider the case that allows for more than two actions.

Formally, each player i's action set is $\{a^1, \ldots, a^K\}$ (note that an action set is the same across players because a player may learn from the opponent). We suppose the following:

1. For each $k \in \{1, \ldots, K-1\}$, an action profile (a^k, a^k) Pareto dominates (a^{k+1}, a^{k+1}).

2. $u_i(a^l, a^k) < u_i(a^k, a^k)$ for $l < k$.

Note that, by (2), the action profile (a^K, a^K) is a Nash equilibrium of the stage game.

The initial action set is $\{a^1\}$ for each $i = 1, 2$. Once player i chooses a^k then she cannot take actions in $\{a^1, \ldots, a^{k-1}\}$. For each $k, k' = 1, \ldots, K$ with $k < k'$, there is $p^i_{k,k'} \in [0, 1]$. In each period player i moves, when her current action set is $\{a^l, \ldots, a^k\}$ (where $1 \leq l \leq k \leq K$) with $k < k'$, (i) i privately learns $a_{k'}$ with probability $p^i_{k,k'}$ and (ii) she learns $a_{k'}$ if her opponent's latest action is $a^{k'}$. Under any one of these events, her action set becomes $\{a^l, \ldots, a^{k'}\}$. In other words, learning $a^{k'}$ makes all actions a^m with $k < m < k'$ available as well. We call this model the *ordered-action model.*

The set of private histories $H_{i,t}$, strategies, and PBE are defined analogously to Section 2. We extend the notion of grim-trigger strategy profile σ^{G} as follows: at a history $h_{i,t} \in H_{i,t}$ at which each player j's latest action is a^{k_j}, player i plays $\sigma^{\mathrm{G}}_i(h_{i,t}) = a^l$ where $l = \max\{k_1, k_2\}$. Note that, in the two-action case, this reduces to the grim-trigger strategy profile σ^{G} that we defined in Section 3.

Theorem 2. *Consider the ordered-action model. There exist $\overline{p} \in (0, 1)$ and $\overline{\delta} \in (0, 1)$ such that for any $(p^i_{k,k'})_{i,k,k'}$ with $p^i_{k,k'} < \overline{p}$ for each k, k' such that $k < k'$ and each $i = 1, 2$ and $\delta_i > \overline{\delta}$ for each $i = 1, 2$, σ^* is a unique PBE.*

Proof Sketch of Theorem 2. The proof of this theorem can be done by induction. To do this, consider the following claim:

Claim k. *There exists $\hat{\delta}^{(k)} \in (0, 1)$ such that for all $\delta_1, \delta_2 > \hat{\delta}^{(k)}$ and all $k_1, k_2 \geq k$, at a history $h_{i,t} \in H_{i,t}$ at which each player j's latest action is a^{k_j}, player i plays a^l where $l = \max(k_1, k_2)$ under any PBE of the ordered-action model.*

	A	B	C
A	$1, 1$	$-2, 2$	$-2, 2$
B	$2, -2$	$-1, -1$	$0, 3$
C	$2, -2$	$3, 0$	$-2, -2$

Table 2: The payoff matrix of the stage game in Example 1

In Appendix A.2.3, we show that for any $k = 1, \cdots, K-1$, Claim k holds if Claim l holds for every $l = k+1, \cdots, K$.

By induction, this shows that Claim k holds for every k, which proves the theorem.

□

To see that the assumptions on payoffs and the evolution of action sets are essential, we provide two examples in which the assumptions are violated and the conclusion of Theorem 2 does not hold. In the first example, the assumption on payoffs is violated.

Example 1. Consider the symmetric normal-form game as depicted in Table 2.

Note that this game does not satisfy our conditions on the payoffs: If it were to satisfy the conditions, then, since $u_i(A, A) > u_i(B, B) > u_i(C, C)$ for each i, we must have $(a^1, a^2, a^3) = (A, B, C)$. However, we have $u_i(C, B) \geq u_i(C, C)$.

Suppose, in contrast, that the evolution of action sets satisfies the assumption specified in this subsection, where the actions are ordered as $(a^1, a^2, a^3) = (A, B, C)$. We assume $p_{A,B} \in (0, 1)$ and $p_{A,C} = p_{B,C} \in (0, 1)$.

However, the following strategy profile does not play A forever, as the payoff structure violates the assumptions in this subsection.

- At a history at which player i has taken C, i plays C (this is her only choice).
- At a history at which player i has not taken C:
  - if the opponent has taken C, then player i takes B.
  - if the opponent has not taken C either, then:
    - if action C is available to player i, then i takes C.
    - if action C is not available to player i, then:
      - if the opponent has played B, then player i takes B.
      - if the opponent has not played B, then player i takes A.

	A	B	C
A	$2, 2$	$-2, 3$	$-3, 4$
B	$3, -2$	$1, 1$	$-1, 2$
C	$4, -3$	$2, -1$	$0, 0$

Table 3: The payoff matrix of the stage game in Example 2

In this strategy profile, the players play C as soon as it becomes available. The reason why this constitutes an equilibrium is that in the small game that excludes action A, there are two Pareto-unranked Nash equilibria: (B, C) and (C, B), where each player prefers the one in which she plays C. Given that the opponent $-i$ will play C as soon as possible, it is i's best response to play C as soon as possible. This sort of construction depends on the multiple Pareto-unranked equilibria in the small game, and our assumption on the payoff functions excludes such a possibility. We show in Appendix A.2.3 that the above strategy profile is a PBE. □

The second pertains to the assumption on the evolution of action sets.

Example 2. Consider the symmetric normal-form game as depicted in Table 3.

Note that this game satisfies the conditions on the payoffs, where the actions are ordered as $(a^1, a^2, a^3) = (A, B, C)$.

Suppose, however, that the evolution of action sets does not satisfy our condition but is given as follows.

1. Each player's action set is $\{A\}$ at the beginning.
2. With probability p, each player's action set $\{A\}$ becomes $\{A, B, C\}$.
3. Once i plays B or C, her action set becomes $\{B, C\}$ forever.
4. If i has been always playing A and $-i$ has played B or C in the past, then i's action set is $\{A, B, C\}$.

The following strategy profile, in which the players play A forever on the path of play, is a PBE when $\delta \geq \frac{1}{2}$: (i) play A as long as only A has been taken; otherwise, play C. See Appendix A.2.3 for the proof of this claim.

We argue that the following strategy profile σ^* does not play A forever and is a PBE. For each i, let $\sigma^{(i)}$ be a PBE strategy profile and denote by $u_j^{(i)}$ the payoff this strategy profile yields to player j. We suppose that $u_1^{(1)} > u_1^{(2)}$ and $u_2^{(2)} > u_2^{(1)}$ (the standard argument shows that we can construct $(\sigma^{(i)})_{i=1,2}$ satisfying this condition). Now we define σ^* as follows:

1. Player i plays C as soon as she privately learns it.

2. If player i takes C for the first time at time t, then players play $\sigma^{(i)}$ forever.

We show in Appendix A.2.3 that σ^* is a PBE when the players are sufficiently patient. The reason why this constitutes an equilibrium is analogous to the analysis of the reversible model in Section 4.1 because the "subgame" after each player chooses an action other than A is that of a reversible model. Recall that the reversible model has multiple equilibria due to the flexibility of action switches. Our assumption on the evolution of action sets in this section shuts down such flexibility of action switches. □

4.5 Possibility of No Private Learning of a Precedent

In our main analysis in Section 3, each player privately learns action N with positive probability p in each period (if she has not learned it yet). In reality, it may be the case that, with certain probability, no player may be able to learn the existence of action N at all. For another instance, with certain probability, the only way in which a player learns action N is through the observation of the opponent's action.

This subsection shows that our main result (Theorem 1) is robust to such possibilities. Formally, we consider the following two variants of our main model. In the first model, there are two states, which realize with positive probability before the start of the game. In the first state, the action N does not exist. That is, no player privately learns action N. In the second state, the action N exists and the players learn N as in our main model. That is, whether there is a possibility of private learning is common across players. Call this model the *common-possibility model.* In this model, the characterization of the players' incentives is the same as our main model. This is because, in order to consider a player's incentive to take action N, we only need to consider histories at which the player has learned N and both players have been taking O. Conditional on such histories, the player knows that action N exists, and our analysis reduces to the main one in Section 3. Thus, our main result (Theorem 1) carries over to this model. In fact, our full analysis in Section 4.6 also carries over to this extension.

In the second model, for each player, there exists a probability-q event (which occurs independently across the players) under which the player cannot privately learn action N on her own (i.e., she learns N only when the opponent plays it). That is, whether there is a possibility of learning is independently determined. Call this model the *independent-possibility model*. We argue that, in this second model, there exist $\overline{\delta}_1, \overline{\delta}_2 < 1$ such that if $\delta_i > \overline{\delta}_i$ for each $i = 1, 2$, then for any $p_1, p_2 \in [0, 1]$, the unique PBE is σ^{G}. In order to consider player i's incentive to take action N, it is sufficient to consider histories at which player i has learned N and both players have been taking O. If player i takes N at one of such histories, then her continuation payoff at that history is $(1 - \delta_i)x_i$. In contrast, if she obeys strategy σ_i^{G}, then her continuation payoff is at least

$$q \cdot 1 + (1 - q)\left((1 - \delta_i) \cdot 1 + \delta_i(1 - \delta_i)(-\ell_i)\right).$$

While the latter converges to $q > 0$ as δ_i tends to 1, the former converges to 0 as δ_i tends to 1. Thus, there exists $\overline{\delta}_i \in (0, 1)$ such that if $\delta_i > \overline{\delta}_i$ then the deviation of playing N is not profitable.

We summarize the findings of this section as follows.

Proposition 4. *In both the common-possibility model and the independent-possibility model, the conclusion of Theorem 1 holds.*[17]

4.6 2×2 Full Analysis

Our main theorem pertains to the case when p_i is small and δ_i is high. For 2×2 games, we provide an equilibrium characterization for any p_i and δ_i. The following lemma provides basic properties of best responses.

Lemma 9. *Fix $p_1, p_2 \in [0, 1]$. For each $i = 1, 2$, the following hold.*

1. *The strategy σ_i^N is a unique best response against σ_{-i}^N if and only if $x_i > \pi_i(\delta_i)$.*
2. *The strategy σ_i^{G} is a unique best response against σ_{-i}^N if and only if $x_i < \pi_i(\delta_i)$.*
3. *If $\delta_i \in (0, \hat{\delta}_i)$, then player i's strategy must be σ_i^N in any PBE.*

[17] Furthermore, it will be straightforward to see that the conclusion of Theorem 3 in Section 4.6 holds as well in the common-possibility model.

<table>
<tr><td></td><td></td><td>$\hat{\delta}_2 > \delta_2$</td><td colspan="2">$\hat{\delta}_2 < \delta_2$</td></tr>
<tr><td></td><td></td><td colspan="2">$x_2 > \pi_2(\delta_2)$</td><td>$x_2 < \pi_2(\delta_2)$</td></tr>
<tr><td rowspan="2">$\delta_1 > \hat{\delta}_1$</td><td>$x_1 < \pi_1(\delta_1)$</td><td>Unique $(\sigma_1^{\mathrm{G}}, \sigma_2^N)$ (Part (2))</td><td colspan="2" rowspan="2">Unique σ^{G} (Part (1))</td></tr>
<tr><td rowspan="2">$x_1 > \pi_1(\delta_1)$</td><td rowspan="2">Unique σ^N (Part (4))</td></tr>
<tr><td>$\delta_1 < \hat{\delta}_1$</td><td>$\sigma^{\mathrm{G}}$ and σ^N (Part (3))</td><td>Unique $(\sigma_1^N, \sigma_2^{\mathrm{G}})$ (Part (2))</td></tr>
</table>

Figure 1: Illustration of Theorem 3

Using this lemma, we obtain the following theorem.

Theorem 3. *Fix $p_1, p_2 \in [0, 1]$. Let $\delta_1, \delta_2 \in (0, 1)$.*

1. *If $\delta_i > \hat{\delta}_i$ and $x_{-i} < \pi_{-i}(\delta_{-i})$, then σ^{G} is a unique PBE.*

2. *If $\delta_i < \hat{\delta}_i$ and $x_{-i} < \pi_{-i}(\delta_{-i})$, then $(\sigma_i^N, \sigma_{-i}^{\mathrm{G}})$ is a unique PBE.*

3. *If $\delta_j > \hat{\delta}_j$ and $x_j > \pi_j(\delta_j)$ for each $j = 1, 2$, then both σ^{G} and σ^N are a PBE.*

4. *If $\delta_i < \hat{\delta}_i$ and $x_{-i} > \pi_{-i}(\delta_{-i})$, then σ^N is a unique PBE.*

Figure 1 illustrates Theorem 3. The Online Supplementary Appendix uses Theorem 3 to provide an alternative characterization of PBE as a function of (δ_1, δ_2) for almost all profiles (x_1, x_2). Here, we provide graphical illustrations of the characterization. Figure 2 illustrates PBE for different values of (δ_1, δ_2). We note that, if p_{-j} is small, then $\overline{\delta}_j(p_{-j}) < 1$ in the left and right panels and $\underline{\delta}_j(p_{-j}) = 1$ in the central panel and thus indeed σ^{G} is a unique PBE when the discount factors are high as shown in Theorem 1. For the symmetric cases in which $p := p_1 = p_2$, $\delta := \delta_1 = \delta_2$, $x := x_1 = x_2$, and $l := \ell_1 = \ell_2$, Figure 3 illustrates the characterization for different values of (p, δ). These figures, in particular, demonstrate that there may exist multiple equilibria when private learning is likely or the players are impatient. The Online Supplementary Appendix provides additional graphical illustrations to cover the cases of asymmetric parameter values.

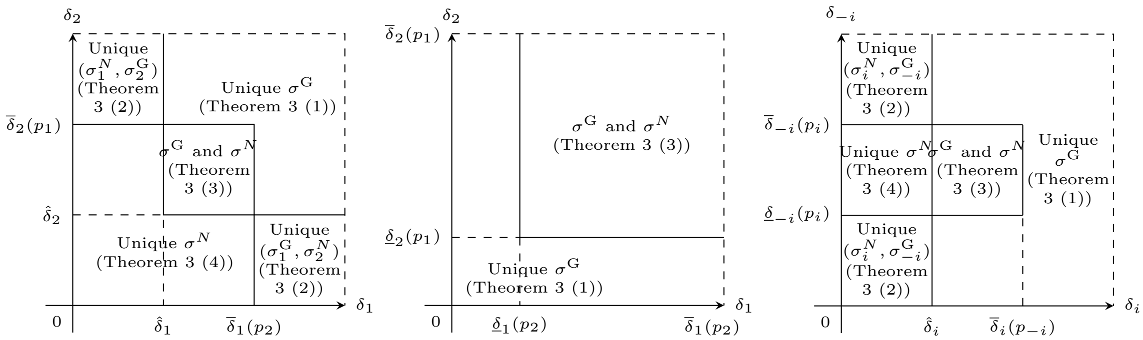

Figure 2: The alternative characterization of PBE. *Left*: The case with $x_i > 1$ for each $i = 1, 2$. *Center*: The case with $x_i < 1$ for each $i = 1, 2$. The graph depicts the case in which $\underline{\delta}_i(p_{-i}) \leq \overline{\delta}_i(p_{-i})$ (otherwise, $\underline{\delta}_i(p_{-i}) = 1$ and $\overline{\delta}_i(p_{-i}) = 0$, and thus σ^{G} is a unique PBE in the entire region. This happens when p_{-i} is low.) for each $i = 1, 2$. In this case, $\overline{\delta}_i(p_{-i}) = 1$ for each i. *Right*: The case with $x_i > 1 > x_{-i}$. The graph depicts the case in which $\underline{\delta}_{-i}(p_i) \leq \overline{\delta}_{-i}(p_i)$ (otherwise, $\underline{\delta}_{-i}(p_i) = 1$ and $\overline{\delta}_{-i}(p_i) = 0$, and thus only σ_{-i}^{G} is played in the entire region).

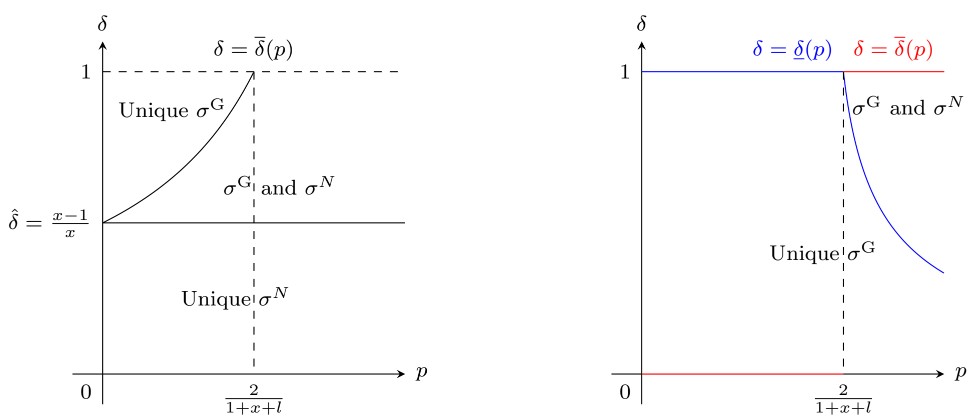

Figure 3: The alternative characterization of PBE for the symmetric cases. *Left*: The case with $x > 1$. *Right*: The case with $x < 1$. We let $\overline{\delta}(p) = \overline{\delta}_1(p) = \overline{\delta}_2(p)$, $\underline{\delta}(p) = \underline{\delta}_1(p) = \underline{\delta}_2(p)$, and $\hat{\delta} = \hat{\delta}_1 = \hat{\delta}_2$.

5 Conclusion

This paper studied a dynamic game in which players are not able to take certain actions without a precedent, i.e., unless some other player has already taken them. For the 2×2 stage game with an inefficient Nash equilibrium in which an action that corresponds to the inefficient Nash equilibrium corresponds to a precedent (e.g., action N in the prisoner's dilemma), we showed that, in the perfect Bayesian equilibrium which we show is unique, each player takes the initial default action as long as the opponent has only taken it, when the players are sufficiently patient and each player's probability of privately learning the new action is low. We also characterized perfect Bayesian equilibria for all discount factors and probabilities. We have also formulated the ordered-action model in which players may have more than two actions, and extended the main result to this model: if the players are sufficiently patient and probabilities of privately learning a new action are low, then in the unique perfect Bayesian equilibrium each player takes the initial default action as long as the opponent has only taken it.

Our model is simple and provides various avenues for future research. As the first paper to consider the effect of precedents in a dynamic interaction, however, those extensions of interest were beyond its scope. For example, it may be interesting to consider extensions of the no-private learning models as in Section 4.5 to the environment with a larger action set. In such extensions, the probability of no private learning for an action depends on the action itself or the current action set (recall the probability $p^i_{k,k'}$ in Section 4.4). This probability would affect players' incentives to take a new action, and would complicate the analysis.

References

Chung, K.-S. and L. Fortnow (2016): "Loopholes," *Economic Journal*, 126, 1774–1797.

Dutta, P. K. (1995): "A Folk Theorem for Stochastic Games," *Journal of Economic Theory*, 66, 1–32.

Feinberg, Y. (2004): "Subjective Reasoning—Games with Unawareness," Working paper.

Fudenberg, D. and J. Tirole (1991): *Game Theory*, MIT Press.

Heifetz, A., M. Meier, and B. C. Schipper (2021): "Prudent Rationalizability in Generalized Extensive-form Games with Unawareness," *B.E. Journal of Theoretical Economics*.

Jéhiel, P. and A. F. Newman (2019): "Loopholes and the Evolution of Contract Form," Working paper.

Lagunoff, R. and A. Matsui (1997): "Asynchronous Choice in Repeated Coordination Games," *Econometrica*, 65, 1467–1477.

——— (2001): "Are "Anti-Folk Theorems" in Repeated Games Nongeneric?" *Review of Economic Design*, 6, 397–412.

Maskin, E. and J. Tirole (1987): "A Theory of Dynamic Oligopoly, III: Cournot Competition," *European Economic Review*, 56, 549–569.

——— (1988a): "A Theory of Dynamic Oligopoly, I: Overview and Quantity Competition with Large Fixed Costs," *Econometrica*, 56, 571–599.

——— (1988b): "A Theory of Dynamic Oligopoly, II: Price Competition, Kinked Demand Curves, and Edgeworth Cycles," *Econometrica*, 31, 947–968.

Merton, R. K. (1940): "Bureaucratic Structure and Personality," *Social Forces*, 18, 560–568.

——— (1968): *Social Theory and Social Structure*, Free Press, enlarged ed.

Monderer, D. and D. Samet (1989): "Approximating common knowledge with common beliefs," *Games and Economic Behavior*, 1, 170–190.

Pandya, J. (2019): "The Dual-Use Dilemma of Artificial Intelligence," *Forbes*, January 7.

Schipper, B. C. (2021): "Discovery and Equilibrium in Games with Unawareness," *Journal of Economic Theory*, 198, 105365.

The Economist (2005): "Funny Money," December 20.

Veenendaal, A.-K. (2018): *Toward a Better Understanding of Rule-Breaking Market Behavior: Insights from Performance Breakthroughs in Sports*, Springer.

Wu, W.-T. and J.-H. Jiang (1962): "Essential Equilibrium Points of n-person Non-cooperative Games," *Scientia Sinica*, 11, 1307–1322.

Yoon, K. (2001): "A Folk Theorem for Asynchronously Repeated Games," *Econometrica*, 69, 191–200.

A Proofs

A.1 Proofs for Section 3

Proof of Lemma 1. Under σ^{G}, suppose that player i moves at time t and has not played N.

Suppose first that her opponent $-i$ has already played N. Then, if i chooses O at time t, her payoff is at most $(1-\delta_i)\cdot(-\ell_i)+\delta_i\cdot 0<0$. If i instead chooses N at time t, her payoff is 0. Hence, it is a best response to choose N at time t.

Next, suppose that $-i$ has not already played N. Then, if i chooses O at time t, her payoff is 1. If i instead chooses N at time t, her payoff is $(1-\delta_i)\cdot x_i+\delta_i\cdot 0=(1-\delta_i)x_i$. This is no greater than 1 if $(1-\delta_i)x_i\leq 1$, or $\hat{\delta}_i\leq\delta_i$. Hence, it is a best response to choose O at time t.

This completes the proof. □

Proof of Lemma 2. Fix σ that satisfies the condition stated in the lemma. Then, suppose that i moves at time t given some history in which $-i$ has already played action N. Then, the same proof as in the proof of Lemma 1 shows that it is the unique best response to choose action N at time t. Hence, action N is chosen under σ whenever the opponent has chosen action N in the past. Since σ coincides with σ^{G} in histories where the opponent has not chosen action N, this completes the proof. □

Proof of Lemma 3. Fix $\delta_i>\hat{\delta}_i$. Suppose to the contrary that there exists k such that $t_k+1<t_{k+1}$, where i is the moving player at time t_k. Fix an arbitrary private history $h_{i,t_k}\in\overline{H}_{i,t_k}$ at time t_k. Player i's payoff from playing action N at h_{i,t_k} is

$$(1-\delta_i)x_i+\delta_i\cdot 0=(1-\delta_i)x_i.$$

On the other hand, if she plays action O at h_{i,t_k} and then plays action N at time $t_k + 2$, then her payoff is

$$(1-\delta_i)\cdot 1 + \delta_i((1-\delta_i)\cdot 1 + \delta_i(1-\delta_i)x_i) = (1-\delta_i^2)\cdot 1 + \delta_i^2(1-\delta_i)x_i$$

because player $-i$ will not play action N at time t_k+1 by our supposition. Note that

$$(1-\delta_i)x_i < (1-\delta_i^2)\cdot 1 + \delta_i^2(1-\delta_i)x_i$$

because $\delta_i > \hat{\delta}_i$ implies $(1-\delta_i)x_i < 1$. Hence, player i assigns probability 0 to action N at h_{i,t_k}. Since h_{i,t_k} was an arbitrary history in $\overline{H}_{i,t_k}$, this contradicts the definition of t_k. Thus, we have shown that $t_{k+1} = t_k + 1$ for every $k \in \mathbb{N}$ such that i moves at t_k. □

Proof of Lemma 4. We first provide two definitions. First, write $h_{i,t} \sqsubseteq h'_{i,t'}$ if $t \le t'$, $s(h_{i,t}) = s(h_{i,t'}) \le t$, and the action profiles in $h_{i,t}$ appear as the first t elements of $h_{i,t'}$.

Second, for $h_{i,t}$ and $h'_{i,t'}$ such that $t \le t'$ and $s(h_{i,t}) \le t$, write $h_{i,t} \circ h'_{i,t'}$ for a history such that $s(h_{i,t} \circ h'_{i,t'}) = s(h_{i,t})$, the first t elements of $h_{i,t} \circ h'_{i,t'}$ are the action profiles in $h_{i,t}$ and the last $t'-t$ action profiles of $h_{i,t} \circ h'_{i,t'}$ are the last $t'-t$ action profiles in $h'_{i,t'}$.

With these definitions in hand, take $h_{i,t}$ and $h'_{i,t}$ satisfying the stated conditions. Suppose that $\pi_i(\sigma^*|h_{i,t}) > \pi_i(\sigma^*|h'_{i,t})$. Then, consider a strategy σ'_i such that, for $t' \ge t$,

$$\sigma'_i(\tilde{h}_{i,t'}) = \begin{cases} \sigma^*_i(h_{i,t} \circ \tilde{h}_{i,t'}) & \text{if } h'_{i,t} \sqsubseteq \tilde{h}_{i,t'} \\ \sigma^*_i(\tilde{h}_{i,t'}) & \text{otherwise} \end{cases}.$$

Then, at history $h'_{i,t}$, player i has a profitable deviation to play σ'_i. This contradicts our assumption that σ^* is a PBE (so σ^*_i must be a best response to σ^*_{-i}). The case when $\pi_i(\sigma^*|h_{i,t}) < \pi_i(\sigma^*|h'_{i,t})$ is symmetric. Thus, we must have $\pi_i(\sigma^*|h_{i,t}) = \pi_i(\sigma^*|h'_{i,t})$. □

Proof of Lemma 5. Fix $i = 1, 2$. First, suppose $x_i \le 1$. Since $\hat{\delta}_i = 0$, we have $\overline{\delta}_i \ge \hat{\delta}_i$. Next, consider the case in which $x_i > 1$. In this case, $\hat{\delta}_i = \frac{x_i - 1}{x_i} < 1$ is the positive solution of

$$x_i = 1 + \delta_i \cdot 1 + \delta_i^2 x_i. \tag{1}$$

Notice that the left-hand side is constant in δ_i, while the right-hand side is 1 at $\delta_i = 0$ and is increasing in $\delta_i \geq 0$.

If $\overline{\delta}_i = 1$, then we are done. Suppose $\overline{\delta}_i < 1$. Recall

$$\pi_i(\delta_i) = p_{-i}[1 + \delta_i(-\ell_i) + \delta_i^2 \cdot 0] + (1 - p_{-i})[1 + \delta_i \cdot 1 + \delta_i^2 x_i].$$

If $p_{-i} = 1$, then $\pi_i(0) = 1$ and $\pi_i(\delta_i)$ is decreasing in δ_i. Thus, by the definition of $\overline{\delta}_i$, we have $\overline{\delta}_i = 1$. Hence, we have $p_{-i} < 1$. Then, $\pi_i(\delta_i)$ is quadratic in δ_i with a positive coefficient, so it is convex in δ_i and diverges to infinity as δ_i goes to infinity. Thus, by the definition of $\overline{\delta}_i$, $\overline{\delta}_i$ is the unique positive solution of $x_i = \pi_i(\delta_i)$.

Finally, the right-hand side of equation (1) is no less than $\pi_i(\delta_i)$ because $-\ell_i < 1$ and $0 < x_i$.

Combining these facts, we obtain $\overline{\delta}_i \geq \hat{\delta}_i$ when $x_i > 1$. In sum, we obtain $\overline{\delta}_i \geq \hat{\delta}_i$ for all x_i. □

Proof of Lemma 6. Fix $\delta_i > \hat{\delta}_i$ and $\delta_{-i} \in (0, \underline{\delta}_{-i}) \cup (\overline{\delta}_{-i}, 1)$. Suppose, to the contrary, that there exists $k \in \mathbb{N}$ such that i moves at t_k and assigns probability one to action N under every history in $\overline{H}_{i,t_k}$. At $t_k + 1$, consider any history $h_{-i,t_k+1} \in \overline{H}_{-i,t_k+1}$. If the moving player $-i$ takes action N, her continuation payoff is $(1 - \delta_{-i})x_{-i}$. If, instead, player $-i$ takes action O at h_{-i,t_k+1}, then her continuation payoff is at least[18]

$$(1 - \delta_{-i}) \cdot 1 + \delta_{-i} \left(p_i((1 - \delta_{-i})(-\ell_{-i}) + \delta_{-i} \cdot 0) + (1 - p_i)((1 - \delta_{-i}) \cdot 1 + \delta_{-i}(1 - \delta_{-i})x_{-i})\right).$$

Then, we have the latter being strictly greater than the former if $\delta_{-i} \in (0, \underline{\delta}_{-i}) \cup (\overline{\delta}_{-i}, 1)$ by the definition of $\underline{\delta}_{-i}$ and $\overline{\delta}_{-i}$. In fact, once both expressions are divided by $(1 - \delta_{-i})$, we have

$$x_{-i} < \pi_{-i}(\delta_{-i}) = p_i[1 + \delta_{-i}(-\ell_{-i}) + \delta_{-i}^2 \cdot 0] + (1 - p_i)[1 + \delta_{-i} \cdot 1 + \delta_{-i}^2 x_{-i}].$$

The former (x_{-i}) is the left-hand side of the inequality defining $\underline{\delta}_{-i}$ and $\overline{\delta}_{-i}$ and the latter is the right-hand side. By the definition of $\underline{\delta}_{-i}$ and $\overline{\delta}_{-i}$, at h_{-i,t_k+1}, player $-i$ has a strict incentive to take action O. Hence, at h_{-i,t_k+1}, player $-i$ assigns probability 0

[18] If player i takes action N at $t_k + 2$ (after learning action N), player $-i$'s continuation payoff from time $t_k + 2$ on is $(1 - \delta_{-i})(-\ell_{-i}) + \delta_{-i} \cdot 0 = (1 - \delta_{-i})(-\ell_{-i})$. If instead player i takes action O (even after learning action O) then player $-i$ can secure a continuation payoff of $(1 - \delta_{-i})1 + \delta_{-i} \cdot 0 = 1 - \delta_{-i}$. Since $(1 - \delta_{-i})(-\ell_{-i}) \leq 1 - \delta_{-i}$, player $-i$ can secure a payoff of $(1 - \delta_{-i})(-\ell_{-i}) + \delta_{-i} \cdot 0$.

to action N. Since the choice of h_{-i,t_k+1} was arbitrary from $\overline{H}_{-i,t_k+1}$, this implies that $t_{k+1} > t_k + 1$. But this contradicts Lemma 3 whose conclusion (i.e., $t_{k+1} = t_k + 1$) must hold as $\delta_i > \hat{\delta}_i$. Therefore, we conclude that there exists no $k \in \mathbb{N}$ such that the moving player i at time t_k assigns probability one to action N under some history in $\overline{H}_{i,t_k}$. □

Proof of Lemma 7. We start with showing that any $t \geq t_1$ satisfies $t = t_k$ for some $k \in \mathbb{N}$. To see this, first, $\delta_i > \hat{\delta}_i$ by assumption. Second, if $\underline{\delta}_{-i} > 0$, then $x_{-i} \leq 1$, because if $x_{-i} > 1$ then $\pi_i(0) = 1 < x_i$ and thus $\underline{\delta}_{-i} = 0$. This implies that $\hat{\delta}_{-i} = 0$. Third, Lemma 5 implies $\overline{\delta}_{-i} \geq \hat{\delta}_{-i}$. Hence we have $\delta_{-i} \geq \hat{\delta}_{-i}$, which enables us to invoke Lemma 3. The lemma implies that any $t \geq t_1$ satisfies $t = t_k$ for some $k \in \mathbb{N}$.

Now, for each t_k such that i moves, by definition, there is a history $h'_{i,t_k} \in \overline{H}_{i,t_k}$ such that N is i's best response at t_k. Also, Lemma 6 implies that there is a history $h''_{i,t_k} \in \overline{H}_{i,t_k}$ such that O is i's best response at t_k. Thus, Lemma 4 implies that at any history $h_{i,t_k} \in \overline{H}_{i,t_k}$, both O and N are best responses, and thus in particular they induce the same continuation payoff. Then, since we have shown that any $t \geq t_1$ satisfies $t = t_k$ for some $k \in \mathbb{N}$, it follows that, at any $t \geq t_1$, for any history $h_{i,t} \in \overline{H}_{i,t}$ of the moving player i, actions O and N induce the same continuation payoff.

Letting the conditional probability assigned to action N at time $t+1$ be r_{t+1} (conditional on $h_{i,t}$),[19] i's indifference condition at time t implies

$$\begin{aligned}(1-\delta_i)x_i = (1-\delta_i)\cdot 1 + \delta_i[&\mu_t r_{t+1}((1-\delta_i)(-\ell_i) + \delta_i \cdot 0) \\ &+ (1-\mu_t r_{t+1})((1-\delta_i)\cdot 1 + \delta_i(1-\delta_i)x_i)]. \qquad (2)\end{aligned}$$

Note that the left-hand side does not depend on t. The right-hand side depends on t only through $\mu_t r_{t+1}$. Thus, $\mu_t r_{t+1}$ does not depend on t.

Now, notice that, conditional on history $h_{i,t}$ and i playing O at time t, the probability that the history is in $\overline{H}_{i,t+2}$ at time $t+2$ is $1 - \mu_t r_{t+1}$ because $\mu_t r_{t+1}$ is the probability that $-i$ will take action N at time $t+1$. Also, under the same conditioning, the probability that $-i$ does not know action N at time $t+3$ is $(1-p_{-i})(1-\mu_t)$

[19] That is,

$$r_{t+1} = \sum_{h_{-i,t+1} \in \overline{H}_{-i,t+1}} \text{Prob}(h_{-i,t+1}|h_{i,t})\sigma^*_{-i}(h_{-i,t+1})(N).$$

We note that $\text{Prob}(h_{-i,t+1}|h_{i,t}) = \text{Prob}(h_{-i,t+1}|h'_{i,t})$ for any $h_{i,t}, h'_{i,t} \in \overline{H}_{i,t}$ and $h_{-i,t+1} \in \overline{H}_{-i,t+1}$ for a reason analogous to the one presented in the first paragraph of the current proof, and thus r_{t+1} does not depend on $h_{i,t}$ as long as $h_{i,t} \in \overline{H}_{i,t}$.

because $-i$ does not know N with probability $1-\mu_t$ at $t+1$ (when he chooses an action), and he does not learn action N at time $t+3$ with probability $1-p_{-i}$. Hence, by the Bayes rule, we have

$$1-\mu_{t+2} = \frac{(1-p_{-i})(1-\mu_t)}{1-\mu_t r_{t+1}} = \alpha_i(1-\mu_t),$$

where $\alpha_i := \frac{1-p_{-i}}{1-\mu_t r_{t+1}}$ is a constant, which is well defined because $p_{-i} < 1$ implies $\mu_t < 1$, and does not depend on t because $\mu_t r_{t+1}$ does not depend on t.

Now, define

$$f_i(z) := (1-\delta_i)\cdot 1 + \delta_i \left(z((1-\delta_i)(-\ell_i) + \delta_i \cdot 0) + (1-z)((1-\delta_i)\cdot 1 + \delta_i(1-\delta_i)x_i) \right).$$

Note that f_i is strictly decreasing in z because $(1-\delta_i)(-\ell_i) + \delta_i \cdot 0 < 0 < (1-\delta_i)\cdot 1 + \delta_i(1-\delta_i)x_i$. Also, equation (2) shows that $f_i(\mu_t r_{t+1}) = (1-\delta_i)x_i$.

Case 1. Suppose that $\delta_i \in (\hat{\delta}_i, \underline{\delta}_i) \cup (\overline{\delta}_i, 1)$. The proof of Lemma 6 shows that $f_i(p_{-i}) > (1-\delta_i)x_i$ in this case. Therefore, $\mu_t r_{t+1} > p_{-i}$. Hence, $\alpha_i = \frac{1-p_{-i}}{1-\mu_t r_{t+1}} > 1$. Letting $j = i$ and $\alpha_i^* = \alpha_i$ completes the proof for this case.

Case 2. Suppose that $\delta_i \notin (\hat{\delta}_i, \underline{\delta}_i) \cup (\overline{\delta}_i, 1)$. The proof of Lemma 6 shows that $f_i(p_{-i}) < (1-\delta_i)x_i$ in this case. Therefore, $\mu_t r_{t+1} < p_{-i}$. Hence, $\alpha_i = \frac{1-p_{-i}}{1-\mu_t r_{t+1}} < 1$. This implies that $\mu_t < \mu_{t+2}$ for all $t \geq t_1$ such that i moves. Since $\mu_t r_{t+1}$ is independent of t and r_{t+1} is a probability, we have $r_{t+1} \in (0,1)$ for all $t \geq t_1$ such that i moves. This implies that there are histories $h'_{-i,t+1}, h''_{-i,t+1} \in \overline{H}_{-i,t+1}$ such that O and N are $-i$'s best responses, respectively. Thus, Lemma 4 implies that at any history $h_{-i,t+1} \in \overline{H}_{-i,t+1}$, both O and N are best responses, and thus in particular they induce the same continuation payoff. Then we can follow the same argument from the second paragraph of the current proof (i.e., the proof of Lemma 7) to Case 1, with the roles of players i and $-i$ switched, that $\mu_{t+1} r_{t+2} > p_i$ and thus $\alpha_{-i} = \frac{1-p_i}{1-\mu_{t+1} r_{t+2}} > 1$. Letting $j = -i$ and $\alpha_{-i}^* = \alpha_{-i}$ completes the proof for this case as well.

□

Proof of Lemma 8. Recall

$$\pi_i(\delta_i, p_{-i}) = p_{-i}[1 + \delta_i(-\ell_i) + \delta_i^2 \cdot 0] + (1-p_{-i})[1 + \delta_i \cdot 1 + \delta_i^2 x_i],$$

where we wrote p_{-i} as an argument of π_i to make the dependence clear. Then,

$$\pi_i(1,0) = 2 + x_i > x_i.$$

The conclusion of the lemma follows because π_i is a continuous function of (δ_i, p_{-i}). □

A.2 Proofs for Section 4

A.2.1 Proofs for Section 4.1

Proof of Proposition 1. The proof is provided in the main text. □

A.2.2 Proofs for Section 4.2

Proof of Proposition 2. If $\overline{q} = 0$, then $\sigma^{(T,\overline{q})}$ is the grim-trigger strategy σ^{G} for any T. Thus, similarly to Lemma 1, there exists $\delta' \in (0,1)$ such that $\sigma^{(T,\overline{q})}$ is a PBE if $\delta_i \in (\delta', 1)$ for each $i = 1, 2$. Henceforth, assume $\overline{q} \in (0,1)$.

Let $\overline{H}_{i,t}$ be the set of i's private histories $h_{i,t}$ at period t such that she has privately learned but has not taken N.

First, let $H_i^{1-\overline{q}}$ be the set of histories of the form $h_{i,t} \in \overline{H}_{i,t}$ such that $t = n \cdot T$ for some $n = 1, 2, \dots$ and $q_t > \overline{q}$. Let the continuation payoff under $\sigma^{(T,\overline{q})}$ at i's history $h_{i,t}$ be $V_i(h_{i,t})$. Then, we have

$$V_i(h_{i,t}) \leq (1-\delta_i^T) \cdot 1 + \delta_i^T \left(\overline{q}(1-\delta_i)\max\{x_i, 0\} + (1-\overline{q}) \sup_{h'_{i,t} \in H_i^{1-\overline{q}}} V_i(h'_{i,t}) \right).$$

Since this inequality must hold for every $h_{i,t} \in H_i^{1-\overline{q}}$, we have

$$\sup_{h'_{i,t} \in H_i^{1-\overline{q}}} V_i(h'_{i,t}) \leq (1-\delta_i^T) \cdot 1 + \delta_i^T \left(\overline{q}(1-\delta_i)\max\{x_i, 0\} + (1-\overline{q}) \sup_{h'_{i,t} \in H_i^{1-\overline{q}}} V_i(h'_{i,t}) \right).$$

This implies that

$$\sup_{h'_{i,t} \in H_i^{1-\overline{q}}} V_i(h'_{i,t}) \leq \frac{(1-\delta_i^T) + \delta_i^T \overline{q}(1-\delta_i)\max\{x_i, 0\}}{1-\delta_i^T(1-\overline{q})}.$$

Note that

$$\frac{(1-\delta_i^T)+\delta_i^T\overline{q}(1-\delta_i)\max\{x_i,0\}}{1-\delta_i^T(1-\overline{q})} \le \frac{1-\delta_i}{\overline{q}}\left(\frac{(1-\delta_i^T)}{1-\delta_i}+\delta_i^T\overline{q}\max\{x_i,0\}\right)$$
$$\le \frac{1-\delta_i}{\overline{q}}\left(T+\delta_i^T\overline{q}\max\{x_i,0\}\right),$$

where the first inequality follows from $1-\delta_i^T(1-\overline{q}) \ge \overline{q}$ and the second from $\frac{1-\delta_i^T}{1-\delta_i} \le T$. Thus, we have

$$\sup_{h'_{i,t}\in H_i^{1-\overline{q}}} V_i(h'_{i,t}) \le \frac{1-\delta_i}{\overline{q}}\left(T+\delta_i^T\overline{q}\max\{x_i,0\}\right).$$

Second, consider any history $h_{i,t}\in \overline{H}_{i,t}$ such that $t=n\cdot T$ for some $n=1,2,\dots$ and $q_t \le \overline{q}$. If i plays N, her continuation payoff is at least $(1-p_{-i})^T(1-\delta_i)\min\{x_i,0\}$. If instead she plays O, her continuation payoff is at most

$$(1-(1-p_{-i})^T)(1-\delta_i)(-\ell_i)$$
$$+(1-p_{-i})^T\left((1-\delta_i^T)\cdot 1+\delta_i^T\left(\overline{q}(1-\delta_i)\max\{x_i,0\}+(1-\overline{q})\sup_{h'_{i,t}\in H_i^{1-\overline{q}}} V_i(h'_{i,t})\right)\right),$$

which is no greater than

$$(1-(1-p_{-i})^T)(1-\delta_i)(-\ell_i)$$
$$+(1-p_{-i})^T(1-\delta_i)\left(T+\left(\overline{q}\max\{x_i,0\}+\frac{1-\overline{q}}{\overline{q}}\left(T+\overline{q}\max\{x_i,0\}\right)\right)\right).$$

This is equal to:

$$(1-(1-p_{-i})^T)(1-\delta_i)(-\ell_i)+\frac{(1-p_{-i})^T}{\overline{q}}(1-\delta_i)\left(T+\overline{q}\max\{x_i,0\}\right).$$

Thus, the continuation payoff from playing N is no less than the one from playing O if

$$(1-p_{-i})^T\min\{x_i,0\} \ge (1-(1-p_{-i})^T)(-\ell_i)+\frac{(1-p_{-i})^T}{\overline{q}}\left(T+\overline{q}\max\{x_i,0\}\right). \quad (3)$$

Since $(1-p_{-i})^T T \to 0$ as $T\to\infty$, it follows that as $T\to\infty$, the left-hand side converges to 0 while the right-hand side converges to $-\ell_i$, which is strictly less than 0. Hence, there exists $T' < \infty$ such that for each $i = 1, 2$ and for all $T > T'$, N is a best response at any $h_{i,t} \in \overline{H}_{i,t}$ such that $t = n \cdot T$ for some $n = 1, 2, \ldots$ and $q_t \leq \overline{q}$. Note that $T' < \infty$ can be taken uniformly for δ because (3) is independent of δ.

Third, consider any history $h_{i,t} \in \overline{H}_{i,t}$ such that $t \neq n \cdot T$ for any $n = 1, 2, \ldots$ or $q_t > \overline{q}$. Let $s \in \{1, \ldots, T\}$ be such that the current period is $t = nT - s$ for some $n = 1, 2, \ldots$.

If i plays N, her continuation payoff is $(1 - \delta_i)x_i$. If instead she plays O, her continuation payoff is at least

$$(1-\delta_i^s)\cdot 1+\delta_i^s \left(\overline{q}\cdot(1-\delta_i)\min\{x_i,0\}+(1-\overline{q})((1-\delta_i^T)\cdot 1+\delta_i^T(1-\delta_i)\min\{x_i,0\})\right),$$

which is equal to

$$(1-\delta_i^s)+\delta_i^s(1-\overline{q})(1-\delta_i^T)+\delta_i^s(1-\delta_i)\min\{x_i,0\}(\overline{q}+(1-\overline{q})\delta_i^T).$$

The continuation payoff from playing O is no less than the one from playing N if

$$(1-\delta_i)x_i \leq (1-\delta_i^s)+\delta_i^s(1-\overline{q})(1-\delta_i^T)+\delta_i^s(1-\delta_i)\min\{x_i,0\}(\overline{q}+(1-\overline{q})\delta_i^T),$$

which is equivalent to:

$$x_i \leq \frac{1-\delta_i^s}{1-\delta_i}+\delta_i^s(1-\overline{q})\frac{1-\delta_i^T}{1-\delta_i}+\delta_i^s\min\{x_i,0\}(\overline{q}+(1-\overline{q})\delta_i^T). \tag{4}$$

Note that (4) holds if

$$x_i+|x_i| \leq \frac{1-\delta_i^s}{1-\delta_i}+\delta_i^s(1-\overline{q})\frac{1-\delta_i^T}{1-\delta_i}.$$

Since the right-hand side is a convex combination of $\frac{1}{1-\delta_i}$ and $(1-\overline{q})\frac{1-\delta_i^T}{1-\delta_i}$ and since $\frac{1}{1-\delta_i} \geq (1-\overline{q})\frac{1-\delta_i^T}{1-\delta_i}$, the above inequality holds if

$$x_i+|x_i| \leq (1-\overline{q})\frac{1-\delta_i^T}{1-\delta_i}. \tag{5}$$

Now, we show that there exist $\delta' \in (0, 1)$ and $T' < \infty$ such that the above inequality holds at $(\delta_i, T) = (\delta', T')$ for each $i = 1, 2$. In fact, for any $T' < \infty$, the right-hand side of (5) converges to $(1-\overline{q})T'$ as $\delta_i \to 1$. Thus, noting that $\overline{q}<1$ and taking T' that satisfies $\frac{x_i+|x_i|}{1-\overline{q}} < T'$ for each $i=1,2$, we can choose $\delta'<1$ satisfying (5) for each $i=1,2$. Since the right-hand side of (5) is non-decreasing in δ_i and T, it follows that if $\delta_i \in (\delta',1)$ for each $i=1,2$ and $T>T'$, each i's continuation payoff from playing O is no less than the one from playing N. □

Proof of Proposition 3. Fix any $\overline{q}\in(0,1)$. We have shown in the proof of Proposition 2 that there exist $\delta'\in(0,1)$ and $T'<\infty$ such that $\sigma^{(T,\overline{q})}$ is a PBE of the synchronous model if $\delta_i\in(\delta',1)$ for each $i=1,2$ and $T>T'$.

Let H_i^1 be the set of histories at period $t=n\cdot T+1$ for some $n=0,1,\ldots,$ where the realization of private learning or the public randomization device has not occurred yet.[20] For any $h_{i,t}\in H_i^1$, denoting by $V_i(h_{i,t})$ player i's continuation payoff at $h_{i,t}$ under $\sigma^{(T,\overline{q})}$, we have:

$$V_i(h_{i,t}) \le (1-\delta_i^{T-1})\cdot 1 + \delta_i^{T-1}\Bigg(\overline{q}(1-(1-p_1)^T(1-p_2)^T)(1-\delta_i)\max\{x_i,0\} + (\overline{q}(1-p_1)^T(1-p_2)^T+1-\overline{q})((1-\delta_i)+\delta_i \sup_{h'_{i,t}\in H_i^1} V_i(h'_{i,t}))\Bigg).$$

Since this inequality must hold for every $h_{i,t}\in H_i^1$, we have

$$\sup_{h'_{i,t}\in H_i^1} V_i(h'_{i,t}) \le (1-\delta_i^{T-1})\cdot 1 + \delta_i^{T-1}\Bigg(\overline{q}(1-(1-p_1)^T(1-p_2)^T)(1-\delta_i)\max\{x_i,0\} + (\overline{q}(1-p_1)^T(1-p_2)^T+1-\overline{q})((1-\delta_i)+\delta_i \sup_{h'_{i,t}\in H_i^1} V_i(h'_{i,t}))\Bigg).$$

This is equivalent to:

$$(1-\delta_i^{T-1}(\overline{q}(1-p_1)^T(1-p_2)^T+1-\overline{q})\delta_i)\sup_{h'_{i,t}\in H_i^1} V_i(h'_{i,t}) \le (1-\delta_i^{T-1})\cdot 1 + \delta^{T-1}\left(\overline{q}(1-(1-p_1)^T(1-p_2)^T)(1-\delta_i)\max\{x_i,0\} + (\overline{q}(1-p_1)^T(1-p_2)^T+1-\overline{q})(1-\delta_i)\right). \tag{6}$$

[20]Strictly speaking, these are not "histories" as defined in Section 2, and thus this involves an abuse of terminology. Nonetheless, the payoff is still well defined.

For any fixed $T < \infty$, notice that the left-hand side of (6) goes to

$$(1 - (\overline{q}(1-p_1)^T(1-p_2)^T + 1 - \overline{q})) \sup_{h'_{i,t} \in H_i^1} V_i(h_{i,t})$$

as $\delta_i \to 1$. Notice that $1 - (\overline{q}(1-p_1)^T(1-p_2)^T + 1 - \overline{q}) > 0$ because $\overline{q} > 0$ and $p_1, p_2 > 0$. On the other hand, the right-hand side of (6) goes to 0 as $\delta_i \to 1$. This implies that

$$\sup_{h'_{i,t} \in H_i^1} V_i(h'_{i,t}) \leq 0.$$

Finally, notice that we must have $V_i(h_{i,t}) \geq 0$ for any $h_{i,t}$ because the limit payoff of σ_i^N as $\delta_i \to 1$ is no less than 0 for any strategy of the opponent. Hence, for any $h_{i,t} \in H_i^1$, i's expected payoff under $\sigma^{(T,\overline{q})}$ at that history converges to 0 as $\delta_i \to 1$. By setting $n = 0$, we conclude that i's expected payoff of the game under $\sigma^{(T,\overline{q})}$ converges to 0 as $\delta_i \to 1$. □

A.2.3 Proofs for Section 4.4

Proof of Theorem 2. Fix $k \leq K-1$. Suppose, as an induction hypothesis, that Claim l holds for every $l = k+1, \cdots, K$. Then, it suffices to show that Claim k holds. For this purpose, assume $\delta > \hat{\delta}^{(k+1)}$ and fix a PBE σ. Fix i and t. The proof consists of three steps.

For the first step, consider a history $h_{i,t} \in H_{i,t}$ at which player i's current latest action is a^k and player $-i$'s current latest action is $a^{k'}$, where $k' > k$. The induction hypothesis implies the following.

1. If i plays a^l such that $l > k'$, then her payoff is

   $$(1-\delta_i)u_i(a^l, a^{k'}) + \delta_i u_i(a^l, a^l).$$

   Due to condition (1), there is $\hat{\delta}_i < 1$ such that for all $\delta_i \in (\hat{\delta}_i, 1)$, this is strictly less than $u_i(a^{k'}, a^{k'})$.

2. If i plays a^l such that $l = k'$, then her payoff is

   $$u_i(a^{k'}, a^{k'}).$$

3. If i plays a^l such that $k < l < k'$, then her payoff is at most

$$(1-\delta_i^2)u_i(a^l, a^{k'}) + \delta_i^2 u_i(a^{k'}, a^{k'}).$$

Due to condition (2), this is strictly less than $u_i(a^{k'}, a^{k'})$.

4. If i plays a^l such that $l = k$, then her payoff is a convex combination of $u_i(a^k, a^{k'})$ and at most $u_i(a^{k'}, a^{k'})$ when $\delta_i \in (\hat{\delta}_i, 1)$ from cases 1–3 above, with a strictly positive weight on the former payoff.

Overall, we find that it is i's unique best response to play $a^{k'}$ if $k' > k$, provided $\delta_i > \max(\hat{\delta}^{(k+1)}, \hat{\delta}_i) < 1$.

For the second step, consider a history $h_{i,t} \in H_{i,t}$ at which player i's current latest action is $a^{k'}$ and player $-i$'s current latest action is a^k, where $k' > k$. Then, if i plays a^l such that $l \geq k'$, we know from the first step that $-i$ plays a^l at $t+1$, and by the induction hypothesis, the action profile (a^l, a^l) will be played thereafter at every period. Hence, the payoff is

$$(1-\delta_i)u_i(a^l, a^k) + \delta_i u_i(a^l, a^l).$$

Due to condition (1), there is $\hat{\delta}_i' < 1$ such that for all $\delta \in (0, \hat{\delta}_i')$, this is maximized when $a^l = a^k$. Hence, we find that it is i's unique best response to play $a^{k'}$ if $k' > k$, provided $\delta > \max(\hat{\delta}^{(k+1)}, \hat{\delta}_{-i}, \hat{\delta}_i') < 1$.

For the third step, consider a history $h_{i,t} \in H_{i,t}$ at which player i's current latest action is a^k and player $-i$'s current latest action is also a^k.

We first show that for all time $t' \geq t$, for all history $h_{j,t'}$ where j is the moving player at t' such that the current action profile is (a^k, a^k), j does not play a^l with $k+1 < l$. To see this, note that our first step implies that the payoff from playing $a^{l'}$ with $k < l'$ is

$$(1-\delta_j)u_j(a^{l'}, a^k) + \delta_j u_j(a^{l'}, a^{l'}).$$

Due to condition (1), there is $\hat{\delta}_j'' < 1$ such that for all $\delta_j \in (0, \hat{\delta}_j'')$, this is maximized when $l' = k+1$ among $l' \in \{k+1, \ldots, K\}$. Since action a^{k+1} is available to a player when action $a^{l'}$ with $k+1 < l'$ is available to that player, this implies that when the current action profile is (a^k, a^k), the moving player will never take any action a^l with $k+1 < l$.

Note that the first step, the second step, and the induction hypothesis pin down the players' action under σ when the current action profile is (a^k, a^{k+1}),(a^{k+1}, a^k), and (a^{k+1}, a^{k+1}), respectively, and they imply that the chosen action is a^{k+1} with probability 1 in either case. This implies that at every time $t' \geq t$ such that the current action profile is (a^k, a^k), the moving player's optimization problem is the same as that of general 2×2 game. Thus, by Theorem 1, the proof is complete. □

Proof for Example 1. We show that the given strategy profile is a PBE. First, at a history at which player i has taken C, the only available action to her is C. Thus, below we consider a history at which player i has not taken C.

Second, suppose that player $-i$ has taken C. Since player $-i$ will keep taking C in the future, it is player i's best response to take B. For this reason, below we assume that no player has taken C.

Third, suppose that action C is available to player i. If her opponent $-i$ has taken B (note that this implies that player $-i$ has taken B in the last period), then it is player i's best response to follow the prescribed strategy to obtain a maximum continuation payoff of 3. Thus, assume that player $-i$ has taken only A. Now, if player i takes C, then her continuation payoff is $V_C := (1-\delta) \cdot 2 + \delta \cdot 3$. If player i takes A (this means that both players have been taking A), then her continuation payoff is

$$V_A = (1-\delta) \cdot 1 + \delta \left(\tilde{p} \left((1-\delta)(-2) + \delta \cdot 0 \right) + (1-\tilde{p}) \left((1-\delta) \cdot 1 + \delta \cdot V_C \right) \right),$$

where $\tilde{p}$ is the probability that player $-i$ takes C in the next period. Since V_A is a convex combination of V_C and a term which is less than 2 and since $V_C \geq 2$, we have $V_C \geq V_A$, and thus the deviation is not profitable.

If player i takes B instead, then, letting $\tilde{q}$ be the probability that player $-i$ takes C in the next period, her continuation payoff is

$$V_B := (1-\delta)2 + \delta \left(\tilde{q} \cdot 0 + (1-\tilde{q})((1-\delta)(-1) + \delta \cdot 3) \right).$$

We have $V_B \leq V_C$ because V_B is a convex combination of 2 and a term which is less than 3, V_C is a convex combination of 2 and 3, and the weights on the convex combinations are the same.

Fourth, assume that action C is not available to player i and that the opponent has played B. If player i follows the prescribed strategy, then her continuation payoff W_B satisfies:

$$W_B = (1-\delta)(-1) + \delta\left(q\cdot 0 + (1-q)(q\cdot 3 + (1-q)W_B)\right).$$

If A is available to player i and she takes A, then her continuation payoff is

$$(1-\delta)(-2) + \delta\left(q((1-\delta)(-2) + \delta\cdot 0) + (1-q)W_B\right).$$

To see that such a deviation is not profitable, it is enough to show that $q\cdot 3 + (1-q)W_B \geq W_B$, that is, $3 \geq W_B$. However, this inequality follows because player i's maximum possible continuation payoff is 3.

Fifth, suppose that action C is not available to player i and player $-i$ has been taking A. If the only available action to player i is A, then she takes A. Suppose action B is available to player i. Let $\tilde{r}$ be the probability that player $-i$ takes C in the next period. Since $p_{A,C} = p_{B,C} =: p$, we have $\tilde{r} = p$.[21] If player i follows the prescribed strategy, then her continuation payoff X_A satisfies

$$X_A = (1-\delta)\cdot 1 + \delta\left(\tilde{r}((1-\delta)(-2) + \delta\cdot 0) + (1-\tilde{r})((1-\delta)1 + \delta\left(p\cdot V_C + (1-p)X_A\right))\right).$$

By continuity, the limit of X_A as $\delta \to 1$, which we denote by X_A^*, satisfies

$$X_A^* = (1-\tilde{r})(3p + (1-p)X_A^*).$$

Thus,

$$X_A^* = \frac{(1-\tilde{r})3p}{1-(1-\tilde{r})(1-p)} = \frac{(1-p)3p}{1-(1-p)^2}.$$

If player i instead takes B, then her continuation payoff is

$$X_B = (1-\delta)\cdot 2 + \delta\left(p\cdot 0 + (1-p)\left((1-\delta)(-1) + \delta Y_B\right)\right),$$

where

$$Y_B = p\cdot 2 + (1-p)\left((1-\delta)(-1) + \delta\left(p\cdot 0 + (1-p)Y_B\right)\right).$$

[21] In fact, for the same reason, we also have $\tilde{p} = \tilde{q} = p$.

By continuity, the limit of Y_B as $\delta \to 1$, which we denote by Y_B^*, satisfies

$$Y_B^* = p \cdot 2 + (1-p)^2 Y_B^*.$$

By continuity, the limit of X_B as $\delta \to 1$, which we denote by X_B^*, satisfies

$$X_B^* = (1-p) Y_B^*.$$

We have $X_A^* > X_B^*$ if and only if

$$\frac{(1-p)3p}{1-(1-p)^2} > \frac{(1-p)2p}{1-(1-p)^2},$$

which holds because $p \in (0,1)$. Thus, there exists $\overline{\delta} \in (0,1)$ such that player i's deviation is not profitable for all $\delta \in (\overline{\delta}, 1)$. □

Proof for Example 2. For the first strategy profile, we show that this constitutes a PBE if $\delta \geq \frac{1}{2}$. We consider player i's incentives. If player i can choose her action from $\{A, B, C\}$ at h^t because the probability p event hits to her (i.e., the opponent has been taking A), then she receives a payoff of 2 as long as she follow the strategy. If she deviates by taking C, then she receives a payoff of $(1-\delta)4$ (if she deviates by taking B, her payoff is bounded by $(1-\delta)3 < (1-\delta)4$). Thus, she plays A if

$$(1-\delta) \cdot 4 \leq 2, \text{ that is, } \delta \geq \frac{1}{2}.$$

If player i can choose her action from $\{A, B, C\}$ at h^t because the opponent has taken B or C in the past, then it is player i's best response to always choose C.

If player i's available action set is $\{B, C\}$ at h^t because player i has taken B in the past, then it is player i's best response to always choose C.

Second, we show that the second strategy profile σ^* is a PBE when the players are sufficiently patient. Player i plays C when she privately learns it, if

$$(1-\delta) \cdot 4 + \delta u_i^{(i)} \geq (1-\delta)2 + \delta\left(p\left\{(1-\delta)(-3) + \delta u_i^{(-i)}\right\} + (1-p)\left\{(1-\delta)4 + \delta u_i^{(i)}\right\}\right),$$

or,

$$(1-\delta(1-p))\left((1-\delta)\cdot 4+\delta u_i^{(i)}\right)\geq(1-\delta)2+\delta p\left\{(1-\delta)(-3)+\delta u_i^{(-i)}\right\}.$$

When $\delta = 1$, the above inequality holds as it reduces to $u_i^{(i)} > u_i^{(-i)}$ (assuming $p \neq 0$). Thus, there exists $\overline{\delta} \in (0,1)$ such that the above inequality holds when $\delta \in (\overline{\delta}, 1)$. □

A.2.4 Proofs for Section 4.6

Proof of Lemma 9.

1. Fix $i = 1, 2$. First, notice that σ_i^N designates a unique best response at any history such that the opponent has already chosen N in the past.

   So, consider a history in $\overline{H}_{i,t}$. If she takes action N, her continuation payoff is $(1-\delta_i)x_i$. If, instead, she takes action O, then her continuation payoff is

   $$(1-\delta_i)\cdot 1+\delta_i\left(p_{-i}((1-\delta_i)(-\ell_i)+\delta_i\cdot 0)+(1-p_{-i})((1-\delta_i)\cdot 1+\delta_i(1-\delta_i)x_i)\right),$$

   which is $(1-\delta_i)\pi_i(\delta_i)$. Then, we have the former being strictly greater than the latter if and only if $x_i > \pi_i(\delta_i)$.[22] Thus, σ_i^N is a unique best response against σ_{-i}^N whenever $x_i > \pi_i(\delta_i)$.

2. Fix $i = 1, 2$. First, notice that σ_i^{G} designates a unique best response at any history such that the opponent has already chosen N in the past.

   So, consider a history in $\overline{H}_{i,t}$. If she takes action N, her continuation payoff is $(1-\delta_i)x_i$. If, instead, she takes action O, then her continuation payoff is

   $$(1-\delta_i)\cdot 1+\delta_i\left(p_{-i}((1-\delta_i)(-\ell_i)+\delta_i\cdot 0)+(1-p_{-i})((1-\delta_i)\cdot 1+\delta_i(1-\delta_i)x_i)\right),$$

   which is $(1-\delta_i)\pi_i(\delta_i)$. Then, we have the latter being strictly greater than the former if and only if $x_i < \pi_i(\delta_i)$.[23] Thus, σ_i^{G} is a unique best response against σ_{-i}^N whenever $x_i < \pi_i(\delta_i)$.

3. Fix any PBE. Note first that, if player i chooses N at time t, then $-i$ chooses N at time $t+1$. Consider a history $h_{i,t} \in \overline{H}_{i,t}$ at time t in which player

[22] Note that this calculation is essentially the same as in the proof of Lemma 6.
[23] Note that this calculation is essentially the same as in the proof of Lemma 6.

i moves. Notice that if i plays N at t, her payoff is $(1-\delta_i)\cdot x_i + \delta_i \cdot 0 = (1-\delta_i)x_i$. If she plays O at t but $-i$ plays N at time $t+1$, then i's payoff is $(1-\delta_i)\cdot 1 + \delta_i[(1-\delta_i)\cdot(-\ell_i) + \delta_i \cdot 0] = (1-\delta_i)(1-\delta_i\ell_i)$. If no one plays N at any time, i's continuation payoff is 1. Hence, i's continuation payoff at $h_{i,t}$ is in the following set:

$$\text{co}\left(\{\delta_i^{2n}(1-\delta_i)x_i | n = 0,1,\dots\} \cup \{\delta_i^{2n+1}(1-\delta_i)(1-\delta_i\ell_i) | n = 0,1,\dots\} \cup \{1\}\right),$$

where co(S) denotes the set of convex combinations of the set S. If $\delta_i < \frac{x_i-1}{x_i}$, the maximum element in this set is $(1-\delta_i)x_i$ because $(1-\delta_i)x_i > 1$ and $(1-\delta_i)(1-\delta_i\ell_i) < 1$. Hence, any action that induces the continuation payoff of $(1-\delta_i)x_i$ is a best response at $h_{i,t}$. Hence, choosing action N is a best response at $h_{i,t}$. Since there is an action that induces the payoff of $(1-\delta_i)x_i$ at $h_{i,t}$, no other action is a best response. Hence, at $h_{i,t}$, choosing action N is a unique best response and thus, i chooses N at $h_{i,t}$ under the PBE we fixed.

Since this was true for any choice of $h_{i,t}$ from $\overline{H}_{i,t}$, this shows that if $\delta_i < \frac{x_i-1}{x_i}$, in any PBE, i chooses action N at any history in $\overline{H}_{i,t}$. This completes the proof.

□

Proof of Theorem 3.

1. This part follows from Lemmas 1 to 7.
2. By Lemma 1 (3), in any PBE, player i's equilibrium strategy is σ_i^N. Then, by Lemma 9 (2), player $-i$'s unique best response is σ_{-i}^{G}. Thus, the result obtains.
3. Lemma 1 shows that σ^{G} is a PBE. Also, by Lemma 9, σ^N is a PBE.
4. By Lemma 1 (3), in any PBE, player i's equilibrium strategy is σ_i^N. Then, by Lemma 9, player $-i$'s unique best response is σ_{-i}^N. Thus, the result obtains.

□