Repeated Coordination with Private Learning*

Pathikrit Basu[†] Kalyan Chatterjee[‡] Tetsuya Hoshino[§] Omer Tamuz[¶]

July 31, 2020

Abstract

We study a repeated game with payoff externalities and observable actions where two players receive information over time about an underlying payoff-relevant state, and strategically coordinate their actions. Players learn about the true state from private signals, as well as the actions of others. They commonly learn the true state (Cripps et al., 2008), but do not coordinate in every equilibrium. We show that there exist stable equilibria in which players can overcome unfavorable signal realizations and eventually coordinate on the correct action, for any discount factor. For high discount factors, we show that in addition players can also achieve efficient payoffs.

Keywords: repeated games, coordination, learning.

1 Introduction

Static coordination games have long been used as models of various important phenomena, such as currency attack and bank runs (Diamond and Dybvig, 1983). The subsequent literature relaxed the common knowledge assumption of payoffs by adding private information to the model, and this led to unexpected findings of a unique inefficient equilibrium.

One of the most important papers in this area is the electronic mail game of Rubinstein (1989). This game has the salient features of many of the other contributions to the literature and is arguably the simplest setting in which the key arguments can be made. His game has two pure-strategy Nash equilibria if the state of nature is "high," but one of these strategies is strictly dominant for each player if the state is "low."

*We thank Nageeb Ali, Sylvain Chassang, Drew Fudenberg, George Mailath, Manuel Mueller-Frank, Roger Myerson, Philipp Strack, Adam Wierman, and Yuichi Yamamoto for comments. We are grateful to the editor, Xavier Vives, and two anonymous referees for valuable suggestions. This work was supported by a grant from the Simons Foundation (#419427, Omer Tamuz).

[†]California Institute of Technology. E-mail - pathkrtb@caltech.edu

[‡]Pennsylvania State University. E-mail - kchatterjee@psu.edu

[§]ITAM. E-mail - tetsuya.hoshino@itam.mx

[¶]California Institute of Technology. E-mail - omertamuz@gmail.com

We study a repeated game with stage-game payoff structure resembling that of the electronic mail game, though no explicit communication is allowed (as in the models of bank runs and currency attacks) except through observable actions. Like other dynamic coordination games with learning (e.g., Angeletos et al., 2007), we have the state of nature determined at the beginning of the game, players having access to private signals informative about the state at each period and observing actions of other players in our model. Our game is infinitely repeated, and players have dynamic learning about the true state through both one's own private signals and others' actions. The learning takes place in a sequential equilibrium of the game, so it is endogenous and depends on the strategies used by players. This is a difference from the common learning paper of Cripps et al. (2008), which has private signals but no actions and therefore no concept of equilibrium learning. Our aim is to construct an equilibrium such that players commonly learn endogenously and also adjust their actions given this learning to achieve efficient coordination. This approach to showing efficiency differs from the stochastic game paper of Sugaya and Yamamoto (2020), which assumes the full-dimensionality of payoffs at *every* state of nature—referred to as statewise full-dimensionality. We do not have this assumption.

To be more specific about our model, we consider a repeated coordination game with observable actions and with stage payoffs that depend on an unknown, unchanging state of nature. The state is either high (H) or low (L). In each period, players choose between two actions: invest (I) or not invest (N). Not investing is a safe action that always yields stage payoff 0, whereas investment involves a cost. Investment yields a profit only if the state is high and the other player also invests.[1]

In each period, each player observes a private signal regarding the state, and chooses whether to invest or not. These signals are conditionally independent, and are observed regardless of the actions taken by the players.[2] Thus, for each player the uncertainty regarding the state eventually vanishes, through the aggregation of these signals.

Our aim is to find a sequential equilibrium that achieves efficient coordination and common learning. Earlier papers, discussed in Section 1.1, have either concentrated on learning without actions or used the notion of ex-post equilibrium because of the intractability of tracking Bayesian updating and beliefs based on private signals and actions together. We show that the players can coordinate on the correct action while learning. We construct equilibria with the stronger property of *action efficiency*, in which, from some time period on, the players choose the correct action with probability 1. It thus follows that these equilibria are *stable* (in the sense of Ellison (1994), who developed the concept in repeated games): when the state is high, players can recover from events of low signal realizations that cause coordination on the right action (for the high

[1]The stage payoffs can be treated as unobservable or as an expectation of a known distribution of payoffs that depends on the actions chosen. We discount future payoffs but the discount factor δ can also be taken as a probability of continuation with the accumulated stage payoffs banked and delivered to the players at the end of the game.

[2]Equivalently, when players both invest they receive independent random payoffs that depend on the state, but even when they do not invest, they observe the (random) payoff they would have gotten from investing. One could alternatively consider a setting in which players receive signals from distributions that also depend on their actions. We believe that our results hold as long as these signals are informative for any action.

state) to stop. Action efficiency is achievable for any discount factor. We further show that these equilibria achieve near optimal *payoff efficiency* when the discount factor is high. We show the latter by considering paths of payoffs directly.

Toward these results, we develop a partial construction of an equilibrium. The construction has prior analogues in other areas of economics, but we believe the combination of various elements of the proof is new in this paper.

Our construction uses the notion of "cool-off strategy." Cool-off strategies constitute a subset of all strategies and consist essentially of an initial cool-off phase, where players do not invest irrespective of their private signals—thus, the actions are not informative—and accumulate private signals; followed by an investment phase, where a player attempts to signal to the other player that his or her private signals suggest investing, which is a stage-game equilibrium only if the state is high. If any player gets sufficiently negative private information after this, she reverts to not investing and another cool-off phase starts. We show that there is an equilibrium in this restricted strategy space (in which each player uses a cool-off strategy) that achieves common learning and efficient action coordination. We then remove the restriction and show that the strategies constructed remain best responses to each other in the unrestricted strategy space.

The fixed point argument gives us a Nash equilibrium in the restricted strategy space. We, however, show that the best response to a strategy that satisfies the restriction also satisfies the restriction, so that, in fact, we have a Nash equilibrium in the entire strategy space. Moreover, we show that this Nash equilibrium is also outcome equivalent to a sequential equilibrium. Though a similar technique has been used elsewhere, this approach in this setting might be considered an innovation of this paper.

Our construction depends on some assumptions about the distributions of the private signals, which we shall discuss in detail later. One assumption is that informative signals are available no matter what action a player chooses. This rules out one-armed bandits, where there is a safe, uninformative arm. However, we can accommodate different (but commonly known) signal distributions that depend on the actions being played, so a "safe arm" could give less information than the "risky arm," using the terminology of one-armed bandits. We elaborate on the equilibrium in the next two paragraphs.

In the one-shot version of our game, when signals are strong enough, players can coordinate on the correct action with high probability; this follows from known results on common q-beliefs in static Bayesian games (Monderer and Samet, 1989). In our repeated setting, since signals accumulate over time, private information does indeed become strong, for example if players play action N for a long time, collect information, play action I if the information is favorable and then play action N forever if one (or both) of them receives a bad signal and reverts to playing action N forever. However, players are forward looking, and they can do better than playing action N forever after the first string of bad signals. Moreover, a player's beliefs are influenced both by his own private signals and the actions of the other player. Thus reasoning directly about players' beliefs is difficult, and one needs to revert to more abstract arguments, with the flavor of social learning, in order to analyze the dynamics. Indeed, the question that we ask regarding eventual coordination on the correct action closely corresponds to key questions regarding eventual learning addressed in the social learning literature (e.g., Smith and Sørensen, 2000).

We shall now consider briefly some relevant literature. Section 2 describes the model. Section 3, on equilibrium analysis, is the heart of the paper, with the formal definition of cool-off strategies and the formal arguments informally mentioned in the introduction. This also includes the main theorems of the paper, on action efficiency and payoff efficiency. Section 4 discusses specific aspects of the model in greater detail, and Section 5 concludes. Proofs are relegated to the Appendix unless otherwise mentioned.

1.1 Related literature

There is work related to this paper in three different parts of the economics literature.

A very similar setting of a repeated coordination game with private signals regarding a state of nature is offered as a motivating example by Cripps et al. (2008), who develop the notion of *common learning* to tackle this class of repeated games. The key assumption of this kind of game is that an action can be strictly dominant in some state of nature. This assumption poses a fundamental difficulty in the context of repeated games: conditional on such a state that a particular action is strictly dominant, the set of feasible and individually rational payoffs has dimension that is strictly less than the number of players. This implies that the standard full-dimensionality assumption fails to hold at that state, and consequently that we cannot use the standard techniques used in the literature of repeated games. We believe that the techniques we develop to solve our game should be easily applicable to theirs, as well as a wide spectrum of similar games.

Static coordination games with private information have long been used to model many real-world problems—for instance, currency attacks (Morris and Shin, 1998), bank runs (Goldstein and Pauzner, 2005), and political regime changes (Edmond, 2013). This literature has dynamic extensions. The dynamic game of regime change by Angeletos et al. (2007), for example, has a similar set-up in terms of private signals and publicly observable (in aggregate) actions, in a repeated game with a continuum of agents. While their game ends once the agents coordinate on a regime change, our game is infinitely repeated, but the underlying forces are similar—dynamic learning about the true state through both one's own private signals and other players' actions. Chassang (2010) is also related, but differs from our setting in that it has a state of nature that changes over time, and an action that is irreversible. Other examples include one of adoption of technology (Vives, 2007). This example supposes that each agent must decide whether or not to adopt a new technology, where action I corresponds to adoption and action N to no adoption. Motivated by Baliga and Sjöström (2004), we can also consider an example of two allies, each choosing whether to spend on domestic development, or on joint defense against a common adversary. The state of nature is the adversary's type, which ca be either aggressive or not aggressive. If the adversary is not aggressive, investing in domestic development is optimal for each player, regardless of the other player's action. But if the adversary is aggressive, coordinating on joint defense is optimal. Each ally obtains a sequence of private signals about the adversary's intentions.

The recent literature on *folk theorems* in repeated settings is possibly the closest, in its statement of the problem addressed, to our paper. The papers we consider in some detail below are Yamamoto (2014) and Sugaya and Yamamoto (2020). Most of the other existing work on learning in repeated games assumes that players observe public signals about the state, and studies equilibrium strategies in which players ignore private signals and learn only from public signals (Wiseman, 2005, 2012; Fudenberg and Yamamoto, 2010, 2011).[3]

Yamamoto (2014) and Sugaya and Yamamoto (2020) are the closest papers to ours. However, these papers are significantly different along the following dimensions. First, they focus on ex-post equilibria, which are defined as sequential equilibria in the infinitely repeated game in which the state θ is common knowledge for each θ. Since an ex-post equilibrium is a stronger notion, their folk theorem (based on ex-post equilibria) implies the folk theorem based on sequential equilibria. Their results do not apply to our setting, since they make assumptions that our model does not satisfy. Second, for their folk theorem, they assume that the set of feasible and individually rational payoffs has the dimension equal to the number of players for *every* state. As argued above, this assumption fails to hold in our environment, and therefore we cannot use techniques used in repeated games. Third, the folk theorem of Sugaya and Yamamoto (2020) fails to hold when there are two players and private signals are conditionally independent (See their Appendix D). Since our work considers such cases, it is complementary to theirs also in this sense.

2 Model

We consider an infinitely repeated game of incomplete information. There is an unknown payoff-relevant binary state of nature $\theta \in \Theta = \{H, L\}$. Time is discrete and the horizon is infinite, i.e., $t = 1, 2, 3....$ There are two players $\mathcal{N} = \{1, 2\}$. Each player i, at every period t, chooses an action $a_i \in A_i = \{I, N\}$. Action I can be interpreted as an action to *invest* and action N as *not invest*. We assume that the state of nature is determined at period $t = 1$ and stays fixed for the rest of the game. Further, both players hold a common prior $(p_0, 1 - p_0) \in \Delta(\Theta)$, where $p_0 \in (0, 1)$ is the initial probability that the state is $\theta = H$. The payoff to player i, in every period, from any action profile $a \in A = A_1 \times A_2$ depends on the state and is defined by a per-period payoff function $u_i(a, \theta)$. The following table shows the payoffs from actions associated with states $\theta = H$ and $\theta = L$ respectively:

	I	N
I	$1-c, 1-c$	$-c, 0$
N	$0, -c$	$0, 0$

Table 1: $\theta = H$

	I	N
I	$-c, -c$	$-c, 0$
N	$0, -c$	$0, 0$

Table 2: $\theta = L$

[3]The result of Wiseman (2012) holds even in the case where there is no public signal, but instead players' private signals are very closely correlated, if there are at least three players and cheap talk is allowed.

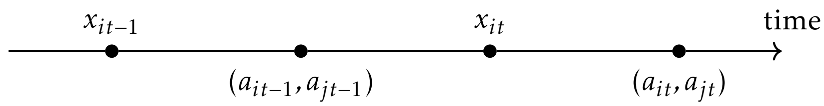

Figure 1: Timeline

Here $c \in (0,1)$ is a cost of investing, and so investment is beneficial if and only if the other player chooses to invest and the state is H. In this case both players obtain a payoff equal to $1-c$. In the state H, there are two stage-game equilibria in pure strategies, (I,I) and (N,N), where the former is Pareto dominant. In the state L, however, action N is strictly dominant, thereby the unique stage-game equilibrium being (N,N). We assume that stage payoffs are not observed by the players.

Private Signals. At every period t, each player i receives a *private signal* $x_{it} \in X$ about the state, where X is a finite signal space.[4] The signal distribution under the state θ is given by $f_\theta \in \Delta(X)$. We assume that conditional on the state θ, private signals are *independent* across players and across time. We also assume that $f_H \neq f_L$. This allows both players to individually learn the state of the nature asymptotically via their own private signals. Finally, we assume full support: $f_\theta(x) > 0$ for all $\theta \in \Theta$ and for all $x \in X$. Hence, no private signal fully reveals the state of nature.

Histories and Strategies. At every period t, each player i receives a private signal x_{it} and then chooses action a_{it}. Under our assumption of perfect monitoring of actions, player i's *private history* before taking an action a_{it} at period t is hence $h_{it} = (a^{t-1}, x_i^t)$, where we denote by $a^{t-1} = (a_1, a_2, \ldots, a_{t-1}) \in A^{t-1}$ the past action profiles and by $x_i^t = (x_{i1}, x_{i2}, \ldots, x_{it}) \in X^t$ player i's private signals. Let $H_{it} = A^{t-1} \times X^t$ be the set of all period-t private histories for player i. Note that action history a^{t-1} is publicly known. Figure 1 depicts the sequence of events. A *pure strategy* for player i is a function $s_i : \bigcup_t H_{it} \to A_i$. Let S_i denote the set of all player i's pure strategies. We equip this set with the product topology. Let $\Sigma_i = \Delta(S_i)$ denote the set of mixed strategies, which are the Borel probability measures on S_i.

Payoffs and Equilibria. Both players evaluate infinite streams of payoffs via discounted sum where the discount factor is $\delta \in (0,1)$. Hence, in the infinitely repeated game, if the players choose a path of actions $(a_t)_t \in A^\infty$, then player i receives a long-run payoff $\sum_{t \geq 1} (1-\delta)\delta^{t-1} u_i(a_t, \theta)$ when the state is θ. Hence, under a (mixed) strategy profile σ, the long-run expected payoff to the player is

$$\mathbb{E}_\sigma \Big[\sum_{t \geq 1} (1-\delta)\delta^{t-1} u_i(a_t, \theta) \Big],$$

[4] The assumption that the signal space is finite is not crucial for establishing our results. The key feature of the signal distributions that we exploit is that they generate bounded likelihood ratios.

where the expectation operator $\mathbb{E}_\sigma$ is defined with respect to the probability measure $\mathbb{P}_\sigma$ induced by the strategy profile σ on all possible paths of the repeated interaction $\Omega = \Theta \times (X^2)^\infty \times A^\infty$.

Our equilibrium concept is sequential equilibrium. We show that in our setting, for every Nash equilibrium there exists an outcome-equivalent sequential equilibrium (Appendix B). As usual, a strategy profile is a Nash equilibrium if no player has a profitable deviation.

3 Equilibrium Analysis

In this section, we study equilibria of the infinitely repeated game. Our goal is to establish the existence of equilibria that are efficient. We shall consider two notions of efficiency, namely *action* efficiency and *payoff* efficiency. Action efficiency requires that player actions converge almost surely to the correct actions, i.e., both players eventually invest (resp. not invest) under the state $\theta = H$ (resp. $\theta = L$). Payoff efficiency requires that players obtain payoffs close to efficient payoffs, i.e., in the state $\theta = H$ (resp. $\theta = L$), ex-ante long run payoffs of both players are close to $1-c$ (resp. close to 0). The main result is that action efficiency can always be obtained in equilibrium and payoff efficient equilibria exist for discount factor δ close to one.

One-period Example

To informally illustrate the nature of incentives that arise in our coordination framework, we first consider the scenario when there is only a single period of interaction. Suppose that players follow a threshold rule to coordinate on actions. That is, for each player i, there exists a threshold belief $\bar{p}_i$ such that player i invests (resp. does not invest) if his belief about the state $\theta = H$ is above (resp. below) the threshold. For simplicity, let us further assume that no posterior belief after the one period signal equals the threshold. Now, if player i's belief p_i is above his threshold and he chooses to invest, then he obtains a payoff equal to

$$p_i \mathbb{P}[p_j > \bar{p}_j \mid \theta = H] - c,$$

where $\mathbb{P}[p_j > \bar{p}_j \mid \theta = H]$ is the probability that player j invests conditional on the state $\theta = H$. Since player i obtains payoff of 0 from not investing, he has an incentive to invest only if this probability is high enough. Moreover, even if it were the case that the other player invests for sure in the state H, player i would still need his belief p_i to be above c in order to find investment optimal. Hence, when player j invests with high probability in the state H, the threshold of player i would be close to c. Likewise, if player j invests with low probability in the state H, the threshold would be close to one. There may hence arise multiple equilibria. There could be an equilibrium with low thresholds where both players invest with high probability in the state H. There could also be an equilibrium with high thresholds where this probability is low or even zero.[5]

[5]This multiplicity arises from the assumption that investment cannot be strictly dominant in either state, unlike global games (Carlsson and van Damme, 1993).

Note the roles played by both a player's belief about the state, and by his belief about his opponent's beliefs. If player i believes that the state is $\theta = H$ with probability 1, his payoff from investing is $\mathbb{P}[p_j > \bar{p}_j \mid \theta = H] - c$. Thus, he invests only if he believes that player j invests with probability at least c. On the other hand, if he believes that player j chooses to invest with probability 1, his payoff from investing is $p_i - c$. Thus, he invests only if he believes that the state is $\theta = H$ with probability at least c. This demonstrates that not only do players need high enough beliefs about the state H in order to invest, but also need to believe with high probability that their opponents hold high beliefs as well. Both players invest only if they acquire "approximate common knowledge" of the state being $\theta = H$ (Monderer and Samet, 1989). This naturally raises the question of whether an analogue of this insight obtains in the repeated game we consider here.

Now, imagine that the first period signals were informationally very precise. This would allow a very high probability of investment in the state H in the low threshold equilibrium. This suggests that if such precise signals were the result of accumulating a series of less precise ones, say received over time, the probability of coordinating on the right actions would be very high. In the infinitely repeated game, such a situation can arise, for example, when players choose to not invest and individually accumulate their private signals, followed by an attempt to coordinate with a low threshold.

In the dynamic context, because players are forward looking, such a threshold rule would also depend on how players behave in the future. For example, if non-investment triggers both players to not invest forever, the thresholds for investment would be low. This is because non-investment would yield a payoff of 0 whereas investment would yield expected payoff for the current period and a non-negative future continuation payoff (because a player can always guarantee himself a payoff of 0 by never investing on his own).[6] However, having non-investment trigger players not investing forever is not compatible with action efficiency, which requires the players to invest eventually in the high state. Thus action efficient equilibria must somehow incentivize players to revert back to investing.

3.1 Cool-off Strategies

We will focus on a particular class of strategies which we call *cool-off strategies*, in which play involves two kinds of phases: *cool-off phases* and *investment phases*. In a cool-off phase, both players accumulate private signals while playing action profile (N, N) for a certain period of time. After the cool-off phase, they enter an investment phase in which each player chooses action I with positive probability. In an investment phase, if the players' play action profile (I, I) then they remain in the investment phase, otherwise they restart a new cool-off phase.

Cool-off strategies are based on the notion of a *cool-off scheme*, which is defined as a partition $(\mathcal{C}, \mathcal{I})$ of the set of all action histories: the set of cool-off histories $\mathcal{C}$ and the set of investment histories $\mathcal{I}$. This classification is carried out by means of a *cool-off function* $T : \mathbb{N} \to \mathbb{N}$ which

[6] In the model of Angeletos et al. (2007), once sufficiently many players invest (or attack a status quo), they receive payoffs and the game ends. Hence, the issue of "repeated" coordination studied here does not arise in their work.

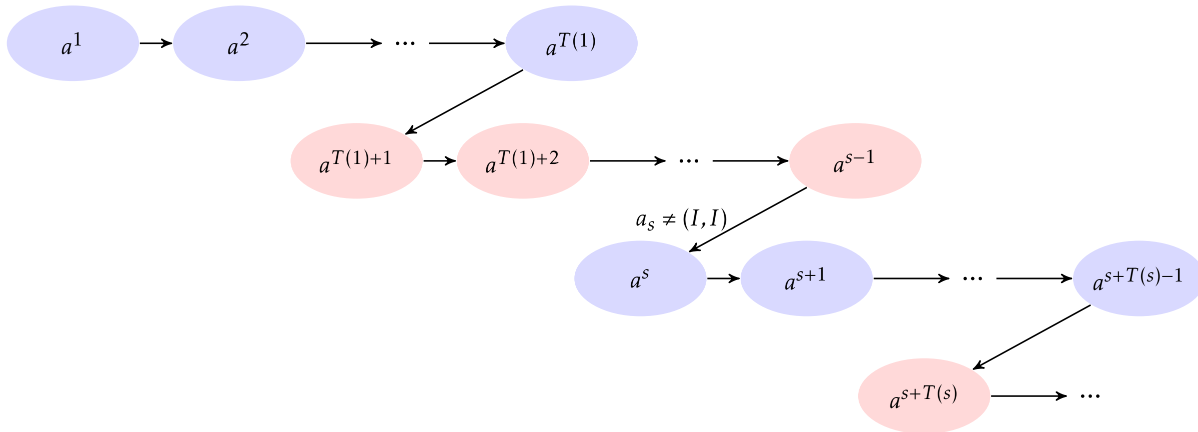

Figure 2: Cool-off Schemes

labels each action history a^t as being part of a cool-off phase or not. Formally, this is done as follows.

Definition 1. *A cool-off scheme $(\mathcal{C},\mathcal{I})$ induced by a cool-off function $T:\mathbb{N}\to\mathbb{N}$ is recursively defined as follows:*

1. *The empty action history is in $\mathcal{C}$.*
2. *For an action history a^t such that $t \geq 1$:*
   (a) *Suppose that $a^{t-1} \in \mathcal{I}$. Then, $a^t \in \mathcal{I}$ if $a_t = (I,I)$ and $a^t \in \mathcal{C}$ otherwise.*
   (b) *Suppose that $a^{t-1} \in \mathcal{C}$ and that there exists a subhistory of a^{t-1} in $\mathcal{I}$. Let a^{s-1} be the longest such subhistory. Then, $a^t \in \mathcal{C}$ if $t \leq s+T(s)-1$ and $a^t \in \mathcal{I}$ if $t > s+T(s)-1$.*
   (c) *Suppose that $a^{t-1} \in \mathcal{C}$ and that there does not exist a subhistory of a^{t-1} in $\mathcal{I}$. Then, $a^t \in \mathcal{C}$ if $t \leq T(1)$ and $a^t \in \mathcal{I}$ if $t > T(1)$.*

Each series of action histories in the sets $\mathcal{C}$ and $\mathcal{I}$ is called a cool-off phase *and an* investment phase *respectively.*

Figure 2 depicts a cool-off scheme. Initially, players are in a cool-off phase for $T(1)$ periods. At period $T(1)+1$, they are in the investment phase. If both players choose to invest, then they stay in the investment phase. If at some point during the investment phase, say at period s, at least one player chooses to not invest, then another cool-off phase begins at this action history a^s for $T(s)$ periods. Another investment phase starts at period $s+T(s)$.[7]

As is clear from Definition 1, the cool-off scheme $(\mathcal{C},\mathcal{I})$ induced by any cool-off function T satisfies both $\mathcal{C}\cap\mathcal{I}=\varnothing$ and $\mathcal{C}\cup\mathcal{I}=\bigcup_{t\geq 1}A^t$. Further, whether an action history a^t is in a cool-off phase or in an investment phase depends only on the cool-off function T but not on strategies.

[7] Players may never invest in an investment phase. For example, if action $a_{T(1)+1} \neq (I,I)$ then this investment phase ends immediately, and a new cool-off phase starts at period $T(1)+2$.

Definition 2. *Fix a cool-off function T. We define the set $S_i(T)$ as the set of all player i's pure strategies $s_i \in S_i$ with the following properties:*

1. *During a cool-off phase player i chooses action N for any realization of their private signals. That is, if $a^{t-1} \in \mathcal{C}$, then $s_i(a^{t-1}, x_i^t) = N$ for any $x_i^t \in X^t$.*

2. *In any action history a^{t-1} in which any player chose action I during a cool-off phase at some time $\tau < t$, player i chooses action N, i.e., $s_i(a^{t-1}, x_i^t) = N$.*

Let $\Sigma_i(T)$ be the set of mixed strategies whose support is contained in the set $S_i(T)$. Let $\Sigma(T) = \Sigma_1(T) \times \Sigma_2(T)$ be the set of such mixed strategy profiles.

The following claim is immediate.

Claim 1. *Fix a cool-off function T. Then, the set $\Sigma_i(T)$ is a closed and convex subset of the set Σ_i.*

As stated above, the investment phase corresponds to action histories in which players would invest with positive probability. We have not required that this be the case in Definition 2. For example, the trivial equilibrium, in which both players never invest, belongs to the class $\Sigma_i(T)$. Ideally, we would like strategies in the set $\Sigma(T)$ to be such that *whenever* the players are in an investment phase, there is a large probability of investment. Hence, we focus on a smaller class of strategies, which we introduce after the following definitions.

A private history (a^{t-1}, x_i^t) is said to be *consistent* with a pure strategy s_i if for all $\tau < t$ it holds that $s_i(a^{\tau-1}, x_i^\tau) = a_i^\tau$. A private history (a^{t-1}, x_i^t) is said to be consistent with a mixed strategy σ_i if it is consistent with some pure strategy s_i in the support of σ_i. Here, by support we mean the intersection of all closed subsets of S_i which have probability one under σ_i; this is equivalent to requiring that for each time t the restriction of s_i to time periods 1 to t is chosen with positive probability. Given a private history (a^{t-1}, x_i^t) that is consistent with σ_i, we can define the (in a sense) conditional probability of taking action I, which we denote by $\sigma_i(a^{t-1}, x_i^t)(I)$, as

$$\sigma_i(a^{t-1}, x_i^t)(I) = \frac{\sigma_i(s_i \text{ is consistent with } (a^{t-1}, x_i^t) \text{ and } s_i(a^{t-1}, x_i^t) = I)}{\sigma_i(s_i \text{ is consistent with } (a^{t-1}, x_i^t))}.$$

Note that this is simply the probability of taking action I at a private history (a^{t-1}, x_i^t) under the behavioral strategy that corresponds to σ_i (e.g., Osborne and Rubinstein, 1994). We say that an action history a^{t-1} is consistent with σ_i if (a^{t-1}, x_i^t) is consistent with σ_i for some x_i^t that occurs with positive probability. Since private signals are conditionally independent, we can define the conditional probability of investing at a consistent a^{t-1} in the state $\theta = H$ as

$$\sigma_i(a^{t-1})(I \mid H) = \frac{\sum_{x_i^t \in X^t} f^H(x_i^t)\left[\prod_{\tau=1}^{t-1} \sigma_i(x_i^\tau, a^{\tau-1})(a_{i\tau})\left(\sigma_i(x_i^t, a^{t-1})(I)\right)\right]}{\sum_{x_i^t \in X^t} f^H(x_i^t)\left[\prod_{\tau=1}^{t-1} \sigma_i(x_i^\tau, a^{\tau-1})(a_{i\tau})\right]}.$$

This is well-defined because a^{t-1} is consistent with σ_i. The important property that we will need, is that for any strategy profile $\sigma = (\sigma_i, \sigma_j)$ in which a given action history a^{t-1} is reached with positive probability, it holds that

$$\mathbb{P}_\sigma \left[a_{it} = I \mid \theta = H, a^{t-1} \right] = \sigma_i(a^{t-1})(I \mid H). \tag{1}$$

Hence, the conditional probability $\mathbb{P}_\sigma \left[a_{it} = I \mid \theta = H, a^{t-1} \right]$ is independent of σ_j. We show this formally in Appendix A (see Claim 3).

Definition 3. *Fix a cool-off function T and a constant $\varepsilon > 0$. We define the set $\Sigma_i(T, \varepsilon) \subset \Sigma_i(T)$ as the set of all player i's* cool-off strategies σ_i *with the following properties:*

1. *$\sigma_i \in \Sigma_i(T)$.*
2. *In an investment phase, player i invests with probability at least $1 - \varepsilon$ conditional on the state $\theta = H$. That is, for any action history $a^{t-1} \in \mathcal{I}$ which is consistent with σ_i,*

$$\sigma_i(a^{t-1})(I \mid H) \geq 1 - \varepsilon. \tag{2}$$

Let $\mathcal{S}(T, \varepsilon) = \Sigma_1(T, \varepsilon) \times \Sigma_2(T, \varepsilon)$ be the set of the cool-off strategy profiles.

As we noted above in (1), the second condition is equivalent to

$$\mathbb{P}_\sigma \left[a_{it} = I \mid \theta = H, a^{t-1} \right] \geq 1 - \varepsilon.$$

The following claim is again straightforward to prove.

Claim 2. *Fix a cool-off function T and a constant $\varepsilon > 0$. Then, the set $\Sigma_i(T, \varepsilon)$ is a closed and convex subset of the set Σ_i.*

3.2 Cool-off Equilibria as Fixed Points

We have introduced the key class of strategies. In what follows, we will find an equilibrium in this class. To do this, we make an appropriate choice of T and ε, and restrict the players to choose strategies from $\Sigma_i(T, \varepsilon)$. Since this set is closed and convex (Claim 2), we can appeal to a standard fixed point argument to show the existence of an equilibrium. We next show that this equilibrium in the restricted strategies is, in fact, still an equilibrium when the restriction is removed. Finally, we show that in this equilibrium, players will eventually coordinate on the right action. Moreover, when the players are very patient, the equilibrium achieves close to efficient payoffs.

Lemma 1 (Continuation Values)**.** *Let σ_j be any mixed strategy of player j. Fix an action history a^{t-1}. Then the continuation value of player i at the action history a^{t-1} is a non-decreasing convex function of his belief in the state $\theta = H$.*

Lemma 2 (Threshold Rule)**.** *For each $c,\delta \in (0,1)$, let $T_0 \in \mathbb{N}$ be large enough and $\varepsilon > 0$ small enough that*

$$\delta^{T_0} < (1-\delta)(1-c-\varepsilon). \tag{3}$$

Let a cool-off function T be such that $T(t) \geq T_0$ for all $t \geq 2$. Fix player j's strategy $\sigma_j \in \Sigma_j(T,\varepsilon)$ and an action history $a^{t-1} \in \mathcal{I}$. Then, for every best response of player i, there exists a threshold π such that at action history a^{t-1} player i invests (resp. does not invest) if her belief in the state $\theta = H$ is above (resp. below) the threshold π.

Lemma 2 implies that for an appropriate choice of T and ε, it holds in every equilibrium of the game restricted to cool-off strategies $\mathcal{S}(T,\varepsilon)$ that both players play threshold strategies: in each history each player has a threshold such that he invests above it and does not below it. Thus, when a player observes another player investing, he updates upwards (in the sense of stochastic dominance) his belief regarding the other player's signal. This idea is formally stated in the following corollary.

We shall denote by p_{it} the private belief of player i at period t before choosing action a_{it}. The belief p_{it} is computed based on player i's private history $h_{it} = (a^{t-1}, x_i^t)$. Hence, $p_{it} = \mathbb{P}[\theta = H \mid h_{it}]$. Finally, for convenience, we shall sometimes write the expectation operator $\mathbb{E}_\sigma$ and probability function $\mathbb{P}_\sigma$ without the subscript σ simply as $\mathbb{E}$ and $\mathbb{P}$.

Corollary 1 (Monotonicity of Expected Beliefs)**.** *For each $c,\delta \in (0,1)$, let $T_0 \in \mathbb{N}$ be large enough and $\varepsilon > 0$ small enough so as to satisfy condition (3) of Lemma 2. Let T be a cool-off function such that $T(t) \geq T_0$ for all $t \geq 2$. Then, in every equilibrium of the game restricted to $\mathcal{S}(T,\varepsilon)$ it holds that if a^{s-1} extends a^{t-1} with both players investing in periods $t, t+1, \ldots, s-1$, then*

$$\mathbb{E}\Big[p_{is} \mid \theta = H, a^{s-1}\Big] \geq \mathbb{E}\Big[p_{it} \mid \theta = H, a^{t-1}\Big].$$

The next two propositions are the key ingredient in showing that an equilibrium in our restricted game is also an equilibrium of the unrestricted game.

Proposition 1. *For each $c,\delta \in (0,1)$, take $T_0 \in \mathbb{N}$ large enough and $\varepsilon > 0$ small enough such that*

$$\frac{c}{(1-\varepsilon)-(1-c)\delta^{T_0}/(1-\delta)} < c(1+2\varepsilon) < 1-\varepsilon. \tag{4}$$

Let a cool-off function T be such that $T(t) \geq T_0$ for all $t \in \mathbb{N}$. Fix player j's strategy $\sigma_j \in \Sigma_j(T,\varepsilon)$ and an action history $a^{t-1} \in \mathcal{I}$. Then, in every best response of player i, if

$$\mathbb{E}\Big[p_{it} \mid \theta = H, a^{t-1}\Big] > 1-\varepsilon^2,$$

then player i invests with probability at least $1-\varepsilon$ in the state $\theta = H$. Moreover, player i invests if his private belief is above $c(1+2\varepsilon)$.

Proof. Denote by p the supremum of the beliefs in which player i does not invest at the action history a^{t-1}. Given the belief p, the payoff from investing is at least $(1-\delta)(-c+p(1-\varepsilon))$, while the payoff from not investing is at most $p\delta^{T_0}(1-c)$. Thus, $(1-\delta)(-c+p(1-\varepsilon)) \le p\delta^{T_0}(1-c)$. Rearranging the terms, we have

$$p \le \frac{c}{(1-\varepsilon)-(1-c)\delta^{T_0}/(1-\delta)},$$

which implies that $p < c(1+2\varepsilon)$ by condition (4). Hence, player i invests if his private belief p_{it} is above $c(1+2\varepsilon)$.

It follows from Markov's inequality that

$$\mathbb{P}(p_{it} \le p \mid \theta = H, a^{t-1}) = \mathbb{P}(1-p_{it} \ge 1-p \mid \theta = H, a^{t-1}) \le \frac{\mathbb{E}[1-p_{it} \mid \theta = H, a^{t-1}]}{1-p}.$$

Since $p < c(1+2\varepsilon)$ and $\mathbb{E}[p_{it} \mid \theta = H, a^{t-1}] > 1-\varepsilon^2$,

$$\frac{\mathbb{E}[1-p_{it} \mid \theta = H, a^{t-1}]}{1-p} < \frac{1-(1-\varepsilon^2)}{1-c(1+2\varepsilon)} < \varepsilon,$$

where the latter inequality follows from condition (4). Hence, $\mathbb{P}(p_{it} \le p \mid \theta = H, a^{t-1}) \le \varepsilon$. That is, player i invests with probability at least $1-\varepsilon$. ■

Proposition 2. *For each $c, \delta \in (0,1)$, let $T_0 \in \mathbb{N}$ and $\varepsilon > 0$ be such that condition (3) in Lemma 2 is satisfied. Then, there is a cool-off function T with $T(t) \ge T_0$ for all $t \ge 2$ such that, in every equilibrium strategy profile $\sigma \in \mathcal{S}(T,\varepsilon)$, it holds in every action history $a^{t-1} \in \mathcal{I}$ that for each $i = 1,2$,*

$$\mathbb{E}_\sigma\Big[p_{it} \mid \theta = H, a^{t-1}\Big] > 1-\varepsilon^2.$$

Proof. Fix $T_0 \in \mathbb{N}$ and $\varepsilon > 0$. Since private signals are bounded, for each $s \ge 1$, there exists $\underline{p}_s > 0$ such that for each $i = 1,2$,

$$\mathbb{P}\Big[p_{is} > \underline{p}_s\Big] = 1.$$

irrespective of the strategies of the players. For example, $\underline{p}_s$ could be defined as the belief that a player would have if he observed both players received the worst signal repeatedly up to period s, where the worst signal is defined as the one which has the lowest likelihood ratio. Note that $\underline{p}_s$ only depends on the period s and not on the action history or the private history of any player.

We now construct the cool-off function T. For each $s \ge 1$, define $T(s) \ge T_0$ large enough to be such that if a cool-off phase were to begin at period s and a player's belief were $\underline{p}_s$, then when the cool-off phase ends at time period $s+T(s)-1$, the expectation of a player's private belief conditional on $\theta = H$ would be above $1-\varepsilon^2$.

For this cool-off function T, from Corollary 1 it follows that for each equilibrium strategy profile $\sigma \in \mathcal{S}(T,\varepsilon)$ and action history $a^{t-1} \in \mathcal{I}$, we have that

$$\mathbb{E}_\sigma\left[p_{it} \mid \theta = H, a^{t-1}\right] > 1 - \varepsilon^2,$$

which completes the proof. ■

Now we prove the existence of equilibria in cool-off strategies—called cool-off equilibria. Applying a fixed-point theorem to the game restricted to cool-off strategies $\mathcal{S}(T,\varepsilon)$, we find an equilibrium $\sigma^* \in \mathcal{S}(T,\varepsilon)$. We then show that this equilibrium remains an equilibrium of the unrestricted game, using Propositions 1 and 2.

Proposition 3. *For each $c,\delta \in (0,1)$, let $T_0 \in \mathbb{N}$ be large enough and $\varepsilon > 0$ small enough so as to satisfy condition (3) in Lemma 2 and condition (4) in Proposition 1. Then, there exists a cool-off function T with $T(t) \geq T_0$ for all $t \geq 2$ and there exists an equilibrium $\sigma^* \in \mathcal{S}(T,\varepsilon)$.*

Finally, we discuss the evolution of belief processes. We show that under any—even non-equilibrium—strategy profile σ, a player's belief process is a submartingale (resp. supermartingale) conditional on the state $\theta = H$ (resp. $\theta = L$). Here is an intuition for this result. We consider the case of the state $\theta = H$, as the case of the state $\theta = L$ is analogous. Since player i observes "high" signals x_{it+1} with high probability, her belief increases in expectation. She also observes player j's action a_{jt} to learn about his signals. As player j's signals are likely to be "high" in the state $\theta = H$, player i is likely to observe player j take an action that corresponds to his "high" signals, so that player i's belief should increase in expectation.

Proposition 4. *Given any strategy profile σ, player i's belief process $\{p_{it}\}$ is a submartingale (resp. supermartingale) conditional on the state $\theta = H$ (resp. $\theta = L$). That is,*

$$\mathbb{E}_\sigma[p_{it+1} \mid \theta = H, p_{it}] > p_{it} > \mathbb{E}_\sigma[p_{it+1} \mid \theta = L, p_{it}].$$

From this proposition, it follows that both players individually learn a true state θ in the long run. This result is immediate from the Martingale Convergence Theorem.[8]

Corollary 2. *Given any strategy profile σ, both players individually learn a true state θ almost surely:*

$$\mathbb{P}_\sigma\left(\lim_{t\to\infty} p_{it}(\omega) = 1 \quad \forall i = 1,2 \mid \theta = H\right) = \mathbb{P}_\sigma\left(\lim_{t\to\infty} p_{it}(\omega) = 0 \quad \forall i = 1,2 \mid \theta = L\right) = 1.$$

3.3 Efficiency

Theorem 1 (Action Efficiency)**.** *For each $c,\delta \in (0,1)$, there exist a cool-off function T, a constant $\varepsilon > 0$ such that for any equilibrium $\sigma^* \in \mathcal{S}(T,\varepsilon)$, players eventually choose the right actions. That is, the equilibrium action profile converges to (I,I) in the state $\theta = H$ and to (N,N) in the state $\theta = L$ almost surely.*

[8] For completeness, a self-contained proof of this corollary is included in Appendix A.

Proof. The underlying probability space for defining the events of interest will be the set of all outcomes $\Omega = \Theta \times A^\infty \times (X^2)^\infty$. Let $\sigma^* \in \mathcal{S}(T,\varepsilon)$ be any equilibrium. Define the events

$$A_I^\infty = \left\{\omega \in \Omega : \lim_{t\to\infty} a_t(\omega) = (I,I)\right\}, \quad A_N^\infty = \left\{\omega \in \Omega : \lim_{t\to\infty} a_t(\omega) = (N,N)\right\}.$$

We wish to show that $\mathbb{P}(A_I^\infty \mid \theta = H) = \mathbb{P}(A_N^\infty \mid \theta = L) = 1$. Consider the latter equality. It would suffice to show that there exists a $\underline{p} \in (0,1)$ such that no player invests when his private belief is below $\underline{p}$. By Corollary 2, in the state $\theta = L$, players' private beliefs almost surely converge to 0 and hence will eventually fall below $\underline{p}$, leading both players to not invest from some point on. We derive such a $\underline{p}$.

When a player has belief p, the payoff from investing is at most $(1-\delta)(p-c) + \delta(p(1-c))$ and the payoff from not investing is at least 0. Hence, if a player finds it optimal to invest at p, then

$$p \geq \frac{c - c\delta}{1 - c\delta}.$$

Hence, we can define $\underline{p} = \frac{c-c\delta}{1-c\delta}$. It now follows that $\mathbb{P}(A_N^\infty \mid \theta = L) = 1$.

We next show $\mathbb{P}(A_I^\infty \mid \theta = H) = 1$. Consider the event $B^\infty = \{\omega \in \Omega : \lim_{t\to\infty} p_{it}(\omega) = 1 \text{ for both } i = 1,2\}$. From Corollary 2, we know that $\mathbb{P}(B^\infty \mid \theta = H) = 1$. Hence, it suffices to show that $B^\infty \subseteq A_I^\infty$.

Let $\omega \in B^\infty$. It suffices to show that on the outcome path induced by ω, there must only be finitely many cool-off phases. Suppose not. Then, there exist infinitely many cool-off phases in ω. Recall from Proposition 1 that a player invests with probability one in an investment phase when his private belief is above $\bar{p} = c(1+2\varepsilon)$. Now, since $\omega \in B^\infty$, there exists a $\bar{T}$ such that $p_{it}(\omega) > \bar{p}$ for all $t \geq \bar{T}$ for both $i = 1,2$. Now, let t' be the starting period of the first cool-off phase after $\bar{T}$. Then, $T(t') + t' \geq \bar{T}$. Hence, when this cool-off period ends at time $s = T(t') + t' - 1$, we enter an investment phase. Since this investment phase at $s+1$ starts beyond period T', the private beliefs of both players would be above $\bar{p}$ and both players would invest in time period $s+1$. Next period, players would continue to be in the investment phase and again invest. Proceeding this way, players would keep on investing which means the investment phase would last forever. This contradicts the fact that there are infinitely many cool-off phases. Hence, there exist only finitely many cool-off phases in ω, which means that players invest forever after some point of time. Hence, $\omega \in A_I^\infty$. ■

Theorem 2 (Payoff Efficiency)**.** *Fix any $c \in (0,1)$. For every $\Delta > 0$, there exists $\bar{\delta} \in (0,1)$ such that for each $\delta \geq \bar{\delta}$, there exist a cool-off function T, a constant $\varepsilon > 0$, such that in any equilibrium $\sigma^* \in \mathcal{S}(T,\varepsilon)$, both players obtain a payoff of at least $1 - c - \Delta$ in the state $\theta = H$ and a payoff of at least $-\Delta$ in the state $\theta = L$.*

Proof. Fix $c, \Delta \in (0,1)$. Let $\varepsilon > 0$ be small enough so that $c < c(1+2\varepsilon)(1-\varepsilon) < (1-\varepsilon)^2$ and moreover

$$(1-\varepsilon)^2\left(\left(1 - \frac{2\varepsilon}{1-\bar{p}}\right)(1-c) + \frac{2\varepsilon}{1-\bar{p}}(-c)\right) > 1 - c - \Delta. \tag{5}$$

and

$$(1-\varepsilon)\left(\frac{-\Delta}{2}\right)-\varepsilon c>-\Delta. \tag{6}$$

where we recall from Proposition 1 that $\bar{p}=c(1+2\varepsilon)$. In what follows, we will show that players achieve a payoff equal to the LHS of (5) in the state $\theta=H$ and a payoff equal to the LHS of (6) in the state $\theta=L$. Let $T_1\in\mathbb{N}$ be such that for any strategy profile which implements a cool-off scheme T with $T(1)=T_1$, it holds that (i) $\mathbb{E}[p_{iT_1}\mid\theta=H]>1-\varepsilon^2$ for all $i=1,2$, (ii) $\mathbb{P}[p_{iT_1}>1-\varepsilon$ for each $i=1,2\mid\theta=H]>1-\varepsilon$, and (iii) $\mathbb{P}[p_{iT_1}<\frac{\Delta}{2(1-c)+\Delta}$ for each $i=1,2\mid\theta=L]>1-\varepsilon$. Now let $\bar{\delta}\in(0,1)$ be such that $\bar{\delta}^{T_1}>1-\varepsilon$. Let $\delta>\bar{\delta}$. Suppose that T_0 is such that condition (3) in Lemma 2 and condition (4) in Proposition 1 are satisfied for T_0 and ε. From Proposition 3, it follows that there exists a cool-off function with $T(1)=T_1$ and $T(t)\geq T_0$ for all $t\geq 2$ and there exists an equilibrium of the repeated game restricted to strategies in $\mathcal{S}(T,\varepsilon)$. Let σ^* be one such equilibrium.

We shall show that in this equilibrium, each player obtains a long-run expected payoff of at least $1-c-\Delta$ in the state $\theta=H$ and $-\Delta$ in the state $\theta=L$. We first establish the former. Note that we have $1-\varepsilon>c(1+2\varepsilon)$. Hence, if the private beliefs of both players are above $1-\varepsilon$ at the end of the first cool-off phase with length T_1, and stay above $1-\varepsilon$ forever, both players would invest forever. We shall show that the probability with which this happens is at least $1-\frac{2\varepsilon}{1-\bar{p}}$. We argue as follows. Suppose that private beliefs for players 1 and 2 are $p_1,p_2\geq 1-\varepsilon$ at time T_1. Hence, from period T_1 onwards, the probability that both players invest forever is at least

$$\mathbb{P}\Big[p_{it}>\bar{p}\text{ for each }i=1,2\text{ and }t\geq T_1\mid p_{iT_1}=p_i,p_{jT_1}=p_j,\theta=H\Big]. \tag{7}$$

Note that conditional on $p_{iT_1}=p_i,p_{jT_1}=p_j$ and $\theta=H$, the belief process p_{it} is a bounded submartingale. Now define the stopping time

$$T_i(\omega)=\min\{t\mid p_{it}(\omega)\leq\bar{p}\}.$$

Let p'_{it} be the process p_{it} stopped at time T_i. It follows that p'_{it} is a bounded submartingale as well. From the Martingale Convergence Theorem, it converges almost surely to a limit $p'_{i,\infty}$. Hence, the conditional probability in (7) is at least

$$\mathbb{P}\Big[p'_{i,\infty}>\bar{p}\text{ for each }i=1,2\mid p_{iT_1}=p_i,p_{jT_1}=p_j,\theta=H\Big].$$

From the Optional Stopping Theorem, we have that

$$\mathbb{E}\Big[p'_{i,\infty}\mid p_{iT_1}=p_i,p_{jT_1}=p_j,\theta=H\Big]\geq p_i.$$

Since $1-p'_{i,\infty}$ is non-negative and $p_i \geq 1-\varepsilon$, from Markov's inequality it follows that for each i,

$$\mathbb{P}\Big[p'_{i,\infty} \leq \bar{p} \mid p_{iT_1} = p_i, p_{jT_1} = p_j, \theta = H\Big] \leq \frac{\varepsilon}{1-\bar{p}}.$$

From the above we obtain that the probability that both players will invest from period T_1 onwards is at least $1-\frac{2\varepsilon}{1-\bar{p}}$. Hence, the long-run expected payoff to each player in state $\theta = H$ will be at least equal to the LHS of (5) which is greater than $1-c-\Delta$.

Finally, we show that under σ^*, the payoff in the state L is at least $-\Delta$. To see this, we first make the following observation. Suppose at the end of the first T_1 cool-off periods, a player's private belief is $p \leq \frac{\Delta}{2(1-c)+\Delta}$. Now from Lemma 1, we know that a player's optimal continuation value $v(\cdot)$ is a non-decreasing convex function of his private belief p. We show that if the optimal continuation strategy of the player yields payoffs (v_H, v_L) at p, then $v_L \geq -\frac{\Delta}{2}$. Since (v_H, v_L) is the optimal continuation value at p, it follows that the vector (v_H, v_L) supports the convex function $v(\cdot)$ at p. Hence,

$$v(p) = pv_H + (1-p)v_L.$$

Note that $v(p) \geq 0$ since a player can always guarantee himself a payoff of zero forever and $v_H \leq 1-c$ since the best payoff a player can get in state H is $1-c$. These, together with the fact that $p \leq \frac{\Delta}{2(1-c)+\Delta}$, imply $v_L \geq -\frac{\Delta}{2}$. Since $\mathbb{P}[p_{iT_1} < \frac{\Delta}{2(1-c)+\Delta}$ for each $i = 1,2 \mid \theta = L] > 1-\varepsilon$, the ex-ante payoff in the state L to each player is at least equal to the LHS of (6), which is greater than $-\Delta$. ■

4 Discussion

4.1 Ex-post Equilibria

Most work in repeated games of complete information assumes that the set of feasible and individually rational payoffs is full-dimensional. That is, the dimension of the set equals the number of players. Analogously, existing work in repeated games of incomplete information assumes so-called statewise full-dimensionality (e.g., Wiseman, 2005, 2012; Yamamoto, 2014; Sugaya and Yamamoto, 2020). However, our payoff structure does not satisfy statewise full-dimensionality. In addition, Yamamoto (2014) and Sugaya and Yamamoto (2020) focus on a special class of sequential equilibria, called ex-post equilibria. Yet, we will see that any ex-post equilibrium is inefficient in our model, and this inefficiency is due to the lack of the statewise full-dimensionality. We discuss each of these two differences in turn.

Statewise Full-dimensionality. For each θ, we define the set of feasible set of payoffs by

$$W(\theta) = \mathrm{co}\Big(\Big\{v \in \mathbb{R}^2 : \exists\, a \in A \quad \forall\, i \in \mathcal{N} \quad u_i(a,\theta) = v_i\Big\}\Big).$$

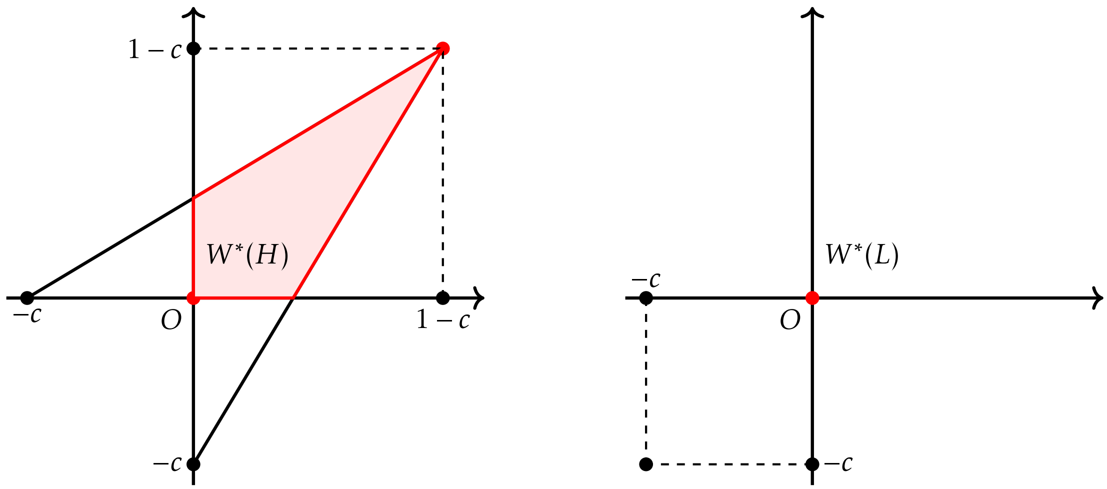

Figure 3: the Lack of Statewise Full-dimensionality

Further, we define player i's minimax payoff in state θ by

$$\underline{w}_i(\theta) = \min_{\alpha_j \in \Delta A_j} \max_{a_i \in A_i} u_i(a_i, \alpha_j, \theta)$$

and the set of all feasible and individually rational payoffs in state θ by

$$W^*(\theta) = \left\{ w \in W(\theta) : \forall\, i \in \mathcal{N} \quad w_i \geq \underline{w}_i(\theta) \right\}.$$

Statewise full-dimensionality is satisfied if the dimension of the set $W^*(\theta)$ equals the number of players for *every* θ.[9] The statewise full-dimensionality is assumed in the existing studies of repeated games of incomplete information, but not in our study. As in Figure 3, the set $W^*(H)$ is of dimension 2, but the set $W^*(L) = \{(0,0)\}$ is not. Intuitively speaking, this lack of statewise full-dimensionality means that players' incentives, thus behavior, are quite different from state to state. This could make it difficult to design an appropriate punishment scheme that supports cooperation.

Note that statewise full-dimensionality is lost at a dominance region (i.e., at the state $\theta = L$). The existence of dominance regions is a key assumption in the literature of equilibrium robustness to (higher-order) belief perturbation, including the electronic mail game and global games.

Ex-post Equilibria. Two closely related papers to ours, Yamamoto (2014) and Sugaya and Yamamoto (2020), focus on a special class of sequential equilibria, called *ex-post equilibria*. An ex-post equilibrium is defined as a sequential equilibrium in the infinitely repeated game in which the strategies are such that they would be optimal even if the state θ were common knowledge. In an ex-post equilibrium, player i's continuation play after history h_{it} is optimal regardless of the

[9] Without uncertainty of payoff-relevant states, the statewise full-dimensionality is equivalent to the full-dimensionality in the literature of repeated games of complete information.

state θ for each h_{it} (including off-path histories).

No ex-post equilibrium, however, can approximate efficiency in our model. More precisely, in every ex-post equilibrium, each player chooses not to invest at every private history. To see this, note that in every ex-post equilibrium, a player's strategy would be optimal if the state L were common knowledge. Since in the state L, a player's best response to any strategy of the opponent is to not invest at any private history, the player never invests even in the state H. Thus, neither player ever invests in any ex-post equilibrium.

Ex-post equilibria may yield trivial play if the payoff structure does not satisfy statewise full-dimensionality. Further, they may not exist if a stage game is like a global game. To see this, suppose, for example, that there is another state in which irrespective of the opponent's action, a player gains payoff $1-c$ if he invests and payoff of 0 if he does not. Then, players always invest if that state were common knowledge, but they would never invest if the state L were common knowledge; therefore, there would not exist ex-post equilibria.

4.2 Common Learning

A coordination game which is very similar to ours is presented as one of the motivations for the introduction of the concept of common learning by Cripps et al. (2008). We now explain this concept, and show that, in our game, common learning is attained in every strategy profile.

Recall $\Omega = \Theta \times A^\infty \times (X^2)^\infty$ is the set of all possible outcomes. Any strategy profile $\sigma = (\sigma_i, \sigma_j)$ induces a probability measure $\mathbb{P}$ on Ω which is endowed with the σ-algebra $\mathcal{F}$ generated by cylinder sets. For each $F \in \mathcal{F}$ and $q \in (0,1)$, define the event $B^q_{it}(F) = \{\omega \in \Omega : \mathbb{P}(F \mid a^{t-1}, x^t_i) \geq q\}$. Now, define the following events:

1. $B^q_t(F) = B^q_{it}(F) \cap B^q_{jt}(F)$. That is, this is the event that both players assign probability at least q to the event F on the basis of their private information.

2. $[B^q_t]^n(F) = B^q_t(B^q_t(B^q_t ... B^q_t(F)))$. That is, this is the event that both players assign probability at least q that both players assign probability at least q that ... both players assign probability at least q to the event F on the basis of their private information.

3. $C^q_t(F) = \bigcap_{n\geq 1}[B^q_t]^n(F)$.

We say that players have *common q-belief* in the event F at $\omega \in \Omega$, at time t, if $\omega \in C^q_t(F)$. We say that players *commonly learn* $\theta \in \Theta$ if for all $q \in (0,1)$, there exists a T such that for all $t \geq T$,

$$\mathbb{P}\left[C^q_t(\theta) \mid \theta\right] > q.$$

Proposition 5. *Let σ be any strategy profile which induces the probability measure $\mathbb{P}$. Then, under $\mathbb{P}$, players commonly learn the state θ for each $\theta \in \Theta$.*

5 Conclusion

This paper has attempted to answer the question of whether rational Bayesian players in a coordination game, each of whom is receiving private signals about the unknown state of nature, can both learn to play the action appropriate to the true state and do so in a manner such that the loss in payoffs becomes negligible as the discount factor goes to one. The answer, in a simple setup, is affirmative, even though the particular coordination game we consider does not have properties such as statewise full-dimensionality.

Future research can be considered along several dimensions. We comment here on some of them. First, we believe our methods would extend to games with the same features as the stage game in this paper, except the number of players would be an arbitrary finite number. We could modify the cool-off strategies appropriately, so if any player plays action N in an investment phase a new cool-off phase begins. Second, though we need some informative signals to be obtained by a player who is playing action N it is not necessary that the signals be identically distributed to the ones obtained by someone playing action I so long as the property of bounded likelihood ratios is satisfied for these distributions. Our method will not give the same efficiency result if playing action N leads to no information being obtained by the player. More generally, our techniques should apply to a large class of similar games that involve state dependent repeated coordination with private signals.

Our analysis assumes that action profile (N,N) is a stage-game Nash equilibrium in both states. If the state is high, there is an opportunity cost of playing this equilibrium rather than the efficient one. If this were not a stage-game Nash equilibrium, we would need a construction to ensure the sequential rationality of players playing a particular action profile in a cool-off phase, which might be problematic in the absence of statewise full dimensionality to bring punishment into play.

Finally, another interesting follow-up to this work would be to see if a similar set of results can be obtained for boundedly rational players who do not use full Bayesian updating but are constrained by memory restrictions. We have no conjectures about possible results for this problem.

A Appendix

A.1 Strategies

Claim 3. *(Equivalence of conditions) Let $\sigma_i \in \mathcal{S}(T,\varepsilon)$ and let σ_j be any strategy of player j. Now, suppose a^{t-1} is reached with positive probability under $\sigma = (\sigma_i, \sigma_j)$. Then,*

$$\mathbb{P}_\sigma\left[a_{it} = I \mid \theta = H, a^{t-1}\right] = \sigma_i(a^{t-1})(I \mid H). \tag{8}$$

Hence, the conditional probability $\mathbb{P}_\sigma\left[a_{it} = I \mid \theta = H, a^{t-1}\right]$ is independent of σ_j.

Proof. Firstly note that since a^{t-1} is reached with positive probability under σ, it is the case that

a^{t-1} is consistent with σ_i. Hence, the conditional probability in (8) is well-defined. Note that:

$$\mathbb{P}_\sigma\left[a_{it} = I \mid \theta = H, a^{t-1}\right] = \frac{\mathbb{P}_\sigma\left[a_{it} = I, \theta = H, a^{t-1}\right]}{\mathbb{P}_\sigma\left[\theta = H, a^{t-1}\right]}.$$

The numerator is equal to

$$\begin{aligned}
&\sum_{x_i^t \in X^t} \sum_{x_j^t \in X^t} f^H(x_i^t) f^H(x_j^t) \Big[\prod_{\tau=1}^{t-1} \sigma_i(x_i^\tau, a^{\tau-1})(a_{i\tau}) \big(\sigma_i(x_i^t, a^{t-1})(I)\big)\Big]\Big[\prod_{\tau=1}^{t-1} \sigma_j(x_j^\tau, a^{\tau-1})(a_{j\tau})\Big] \\
&= \Big[\sum_{x_i^t \in X^t} f^H(x_i^t) \prod_{\tau=1}^{t-1} \sigma_i(x_i^\tau, a^{\tau-1})(a_{i\tau}) \big(\sigma_i(x_i^t, a^{t-1})(I)\big)\Big]\Big[\sum_{x_j^t \in X^t} f^H(x_j^t) \prod_{\tau=1}^{t-1} \sigma_j(x_j^\tau, a^{\tau-1})(a_{j\tau})\Big].
\end{aligned}$$

The equality follows from the fact that signals are conditionally independent across players, which means that the probability that players receive signals (x_i^t, x_j^t) is equal to $f^H(x_i^t) f^H(x_j^t)$. By the same token, the denominator is equal to

$$\begin{aligned}
&\sum_{x_i^t \in X^t} \sum_{x_j^t \in X^t} f^H(x_i^t) f^H(x_j^t) \Big[\prod_{\tau=1}^{t-1} \sigma_i(x_i^\tau, a^{\tau-1})(a_{i\tau})\Big]\Big[\prod_{\tau=1}^{t-1} \sigma_j(x_j^\tau, a^{\tau-1})(a_{j\tau})\Big] \\
&= \Big[\sum_{x_i^t \in X^t} f^H(x_i^t) \prod_{\tau=1}^{t-1} \sigma_i(x_i^\tau, a^{\tau-1})(a_{i\tau})\Big]\Big[\sum_{x_j^t \in X^t} f^H(x_j^t) \prod_{\tau=1}^{t-1} \sigma_j(x_j^\tau, a^{\tau-1})(a_{j\tau})\Big].
\end{aligned}$$

Now, notice that since the denominator is strictly positive, it follows that the term

$$\sum_{x_j^t \in X^t} f^H(x_j^t) \prod_{\tau=1}^{t-1} \sigma_j(x_j^\tau, a^{\tau-1})(a_{j\tau})$$

in the denominator is also strictly positive. This terms also cancels out from the numerator and the denominator and we obtain

$$\mathbb{P}_\sigma\left[a_{it} = I \mid \theta = H, a^{t-1}\right] = \frac{\sum_{x_i^t \in X^t} f^H(x_i^t) \Big[\prod_{\tau=1}^{t-1} \sigma_i(x_i^\tau, a^{\tau-1})(a_{i\tau}) \big(\sigma_i(x_i^t, a^{t-1})(I)\big)\Big]}{\sum_{x_i^t \in X^t} f^H(x_i^t) \Big[\prod_{\tau=1}^{t-1} \sigma_i(x_i^\tau, a^{\tau-1})(a_{i\tau})\Big]}.$$

Hence, $\mathbb{P}_\sigma\left[a_{it} = I \mid \theta = H, a^{t-1}\right] = \sigma_i(a^{t-1})(I \mid H)$. ■

A.2 Two Lemmas

The following lemmas shall be useful in our analysis. The first lemma states that beliefs in an event—which are a martingale—form a submartingale when conditioned on the same event.

Lemma 3. *Let $\mathcal{F}_1 \subseteq \mathcal{F}_2$ be sigma-algebras in a finite probability space, let $p = \mathbb{P}[E \mid \mathcal{F}_1] \neq 0$ be the prior probability of an event E, and let $q = \mathbb{P}[E \mid \mathcal{F}_2]$ be its posterior probability. Then,*

$$\mathbb{E}[q \mid E, p] = \frac{\mathbb{E}[q^2 \mid p]}{\mathbb{E}[q \mid p]} \geq p,$$

where the inequality holds with equality if and only if $\mathbb{P}(q \neq p \mid p) = 0$.

Proof. Since the space is finite, we can write

$$\mathbb{E}[q \mid E, p] = \sum_{x \in [0,1]} x \cdot \mathbb{P}[q = x \mid E, p] = \sum_{x \in [0,1]} x \cdot \mathbb{P}[E \mid q = x, p] \frac{\mathbb{P}[q = x \mid p]}{\mathbb{P}[E \mid p]},$$

where we use Bayes' rule. Note that $\mathbb{P}[E \mid p] = p$ and that since $\mathcal{F}_1 \subseteq \mathcal{F}_2$, $\mathbb{P}[E \mid q = x, p] = x$. Hence, the RHS is equal to

$$\sum_{x \in [0,1]} x^2 \cdot \frac{\mathbb{P}[q = x \mid p]}{p} = \frac{\mathbb{E}[q^2 \mid p]}{p}.$$

Again since $\mathcal{F}_1 \subseteq \mathcal{F}_2$, $p = \mathbb{E}[q \mid p]$, and thus we have shown that

$$\mathbb{E}[q \mid E, p] = \frac{\mathbb{E}[q^2 \mid p]}{\mathbb{E}[q \mid p]}. \tag{9}$$

Since $\mathbb{E}[(q-p)^2 \mid p] \geq 0$ and since $\mathbb{E}[q \mid p] = p$, we have (9) ≥ 0. Finally, (9) $= 0$ if and only if $\mathbb{E}[(q-p)^2 \mid p] = 0$, i.e., $\mathbb{P}(q \neq p \mid p) = 0$. ■

The next lemma states that if a random variable taking values in the unit interval has a mean close to one, then it must assume values close to one with high probability.

Lemma 4. *Let X be a random variable which takes values in $[0,1]$. Suppose $\mathbb{E}[X] \geq 1 - \varepsilon^2$ where $\varepsilon \in (0,1)$. Then,*

$$\mathbb{P}[X > 1 - \varepsilon] \geq 1 - \varepsilon.$$

Proof. Define the random variable $Y = 1 - X$. It suffices to show that $\mathbb{P}[Y \geq \varepsilon] \leq \varepsilon$. Since $\mathbb{E}[Y] \leq \varepsilon^2$ and since Y is non-negative, by Markov's inequality, we obtain it. ■

A.3 Proofs

Proof of Claim 1. The conclusion follows by definition. If $\{\sigma_i^n\}_n \subseteq \Sigma_i(T)$ and $\sigma_i^n \to \sigma_i^*$ in the weak-* topology. Since the support of each σ_i^n is contained in $S_i(T)$, so is the support of σ_i^*. It also follows that any convex combination $\sigma_i^\lambda = \lambda\sigma_i + (1-\lambda)\sigma_i'$ of two mixed strategies $\sigma_i, \sigma_i' \in \Sigma_i(T)$ also belongs to $\Sigma_i(T)$. ■

Proof of Claim 2. We first argue that $\Sigma_i(T,\varepsilon)$ is closed. Suppose $\{\sigma_i^n\} \subseteq \Sigma_i(T,\varepsilon)$ and suppose $\sigma_i^n \to \sigma_i^*$ in the weak-* topology. Firstly, note that from Claim 1, it follows that $\sigma_i^* \in \Sigma_i(T)$. Now suppose a^{t-1} in an investment phase is consistent with σ_i^*. Then, the denominator of $\sigma_i^*(a^{t-1})(I \mid H)$, as defined, will be strictly positive. This means that for large n, the denominator of $\sigma_i^n(a^{t-1})(I \mid H)$ will be strictly positive also. This means that a^{t-1} is consistent with σ_i^n for large n. It follows from the definition that $\sigma_i^n(a^{t-1})(I \mid H) \to \sigma_i^*(a^{t-1})(I \mid H)$ which implies that $\sigma_i^*(a^{t-1})(I \mid H) \geq 1-\varepsilon$.

We now show convexity. Let $\sigma_i, \sigma_i' \in \Sigma_i(T,\varepsilon)$ be two mixed strategies, and let $\lambda \in (0,1)$. Now, consider the mixed strategy defined as $\sigma_i^\lambda = \lambda\sigma_i + (1-\lambda)\sigma_i'$. Then, we show $\sigma_i^\lambda \in \Sigma_i(T,\varepsilon)$.

Clearly $\sigma_i^\lambda \in \Sigma_i(T)$ since $\Sigma_i(T)$ is convex from Claim 1. Let $\bar{\sigma}_j$ be the strategy of player j that plays each action with probability 1/2 at each private history. Denote as $\mathbb{P}^\lambda$, $\mathbb{P}$ and $\mathbb{P}'$, the probability measures induced by the strategy profiles $(\sigma_i^\lambda, \sigma_j), (\sigma_i, \sigma_j)$ and (σ_i', σ_j). Note that $\mathbb{P}^\lambda = \lambda\mathbb{P} + (1-\lambda)\mathbb{P}'$.

We wish to show that σ_i^λ also satisfies condition (2). To this end, let a^{t-1} be an action history in an investment phase that is consistent with σ_i^λ. It follows that a^{t-1} is reached with positive probability under σ_i^λ. Then, from equation (8), it suffices to show that

$$\mathbb{P}^\lambda\left[a_{it} = I \mid \theta = H, a^{t-1}\right] \geq 1-\varepsilon.$$

Now, note by hypothesis, since $\sigma_i, \sigma_i' \in \Sigma_i(T,\varepsilon)$, we already have that $\mathbb{P}\left[a_{it} = I, \theta = H, a^{t-1}\right] \geq (1-\varepsilon)\mathbb{P}\left[\theta = H, a^{t-1}\right]$ and $\mathbb{P}'\left[a_{it} = I, \theta = H, a^{t-1}\right] \geq (1-\varepsilon)\mathbb{P}'\left[\theta = H, a^{t-1}\right]$. These together imply that

$$\frac{\lambda\mathbb{P}\left[a_{it} = I, \theta = H, a^{t-1}\right] + (1-\lambda)\mathbb{P}'\left[a_{it} = I, \theta = H, a^{t-1}\right]}{\lambda\mathbb{P}\left[\theta = H, a^{t-1}\right] + (1-\lambda)\mathbb{P}'\left[\theta = H, a^{t-1}\right]} \geq 1-\varepsilon,$$

which is equivalent to

$$\mathbb{P}^\lambda\left[a_{it} = I \mid \theta = H, a^{t-1}\right] = \frac{\mathbb{P}^\lambda\left[a_{it} = I, \theta = H, a^{t-1}\right]}{\mathbb{P}^\lambda\left[\theta = H, a^{t-1}\right]} \geq 1-\varepsilon.$$

■

Proof of Lemma 1. Denote by p player i's belief at the action history a^{t-1}. For each possible continuation strategy s of player i (which we think of as depending only on his future signals) let v_H and v_L denote the expected payoff in the high and low state, respectively, so that the expected payoff for the continuation strategy s is given by $pv_H + (1-p)v_L$. In any best response of player i, his expected payoff is given by

$$v(p) = \max_s \left\{pv_H + (1-p)v_L\right\},$$

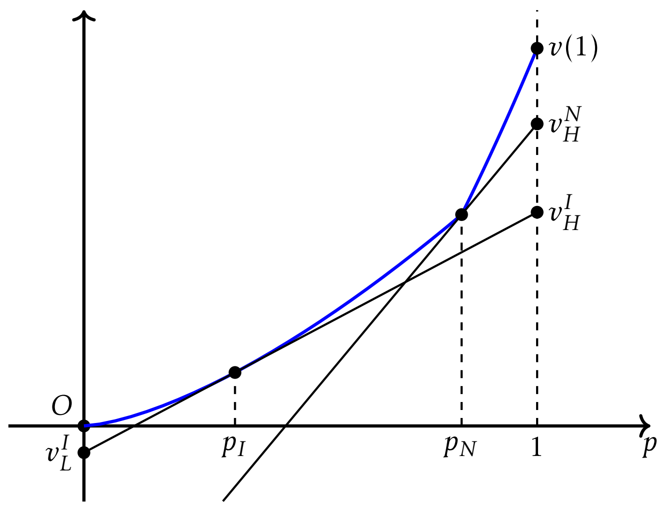

Figure 4: the continuation value $v(p)$

where the maximization is over all continuation strategies for player i. Since it is the maximum of convex functions, it is convex. To see that it is increasing, note that continuation value $v_L \leq 0$ for every continuation strategy s and that $v_L = v_H = 0$ for some continuation strategy s—specifically in the strategy in which the player never invests. Hence the derivative of continuation payoff $v(p)$ is at least 0 at $p = 0$, and thus continuation payoff $v(p)$, as a function of belief p, has a non-negative subderivative in the entire interval $[0, 1]$. ■

Proof of Lemma 2. Denote by p player i's belief at the action history a^{t-1}. Let s_I^* be the strategy in which player i invests at a^{t-1} and then never again. The conditional payoffs for this strategy are

$$v_H^* \geq (1-\delta)(1-c-\varepsilon), \quad v_L^* = -(1-\delta)c.$$

where the former holds true because the opponent invests with probability at least $1-\varepsilon$ in the state $\theta = H$.

Let s_N be any strategy in which player i does not invest at a^{t-1}. Then, because of the cool-off scheme, the payoffs for this strategy are

$$v_H^N \leq (1-c)\delta^{T_0}. \tag{10}$$

Let s_I be any strategy in which player i invests at a^{t-1}. Then

$$v_L^I \leq -(1-\delta)c. \tag{11}$$

Assume by contradiction that it is optimal to choose the strategy s_I at belief p_I and to choose the strategy s_N at belief $p_N > p_I$. Recall that the continuation payoff $v(p)$, defined in the proof of Lemma 1, is supported by the line $pv_L^I + (1-p)v_H^I$ at the belief p_I and by $pv_L^N + (1-p)v_H^N$ at p_N.

Since it is non-decreasing and convex (Lemma 1), it follows that $v_H^I \leq v_H^N$. Inequality (10) implies that $v_H^I \leq \delta^{T_0}$. By condition (3), we have that $v_H^I < v_H^*$. By inequality (11) we have that $v_L^I \leq v_L^*$, and thus the continuation strategy s_I^* strictly dominates s_I for any belief $p \neq 0$. Hence, the continuation strategy s_I cannot be optimal in such a belief. Moreover, it cannot be optimal at the belief $p = 0$ either, since it is better to not invest. We have thus reached a contradiction. ■

Proof of Corollary 1. It suffices to establish the corollary in the case where $s = t + 1$. That is, $a^{s-1} = a^t = (a^{t-1}, (I, I))$. Now, we have,

$$\begin{aligned}\mathbb{E}\Big[p_{it+1} \mid \theta = H, a^{t-1}, a_{it} = I, a_{jt} = I\Big] &\geq \mathbb{E}\Big[p_{it+1} \mid \theta = H, a^{t-1}, a_{jt} = I\Big] \\ &\geq \mathbb{E}\Big[p_{it+1} \mid \theta = H, a^{t-1}\Big],\end{aligned}$$

where the first inequality follows from the fact that player i uses a threshold rule at a^{t-1} (Lemma 2), hence conditioning on $a_{it} = I$ increases the expectation of players i's private belief. The second inequality follows since player j uses a threshold rule, hence conditional on $a_{jt} = I$, player i learns that player j's belief is above a threshold and hence, his beliefs move upwards. Now, by the tower property of expectations,

$$\mathbb{E}\Big[p_{it+1} \mid \theta = H, a^{t-1}\Big] = \mathbb{E}\Big[\mathbb{E}\Big[p_{it+1} \mid \theta = H, a^{t-1}, p_{it}\Big] \mid \theta = H, a^{t-1}\Big].$$

From Lemma 3, and since $p_{it} = \mathbb{P}[\theta = H \mid a^{t-1}, x_i^t]$, it follows that $\mathbb{E}[p_{it+1} \mid \theta = H, a^{t-1}, p_{it}] \geq p_{it}$ almost surely, and so

$$\mathbb{E}\Big[p_{it+1} \mid \theta = H, a^{t-1}\Big] \geq \mathbb{E}\Big[p_{it} \mid \theta = H, a^{t-1}\Big].$$

■

Proof of Proposition 3. The strategy space Σ_i is compact. By Claim 2, the set $\Sigma_i(T, \varepsilon)$ is closed, thus compact, and convex. Payoffs are continuous, which follows from the fact that they are discounted exponentially and that the topology on Σ_i is the weak-* topology. Hence, by Glicksberg's Theorem, the game in which the players are restricted to $\mathcal{S}(T, \varepsilon)$ has a Nash equilibrium $\sigma^* = (\sigma_i^*, \sigma_j^*) \in \mathcal{S}(T, \varepsilon)$. It follows from Propositions 1 and 2 that this fixed point is, in fact, a Nash equilibrium of the unrestricted game, since any best response to any strategy in $\Sigma_j(T, \varepsilon)$ would be in $\Sigma_i(T, \varepsilon)$. ■

Proof of Proposition 4. Note that $\mathbb{P}(p_{it+1} \neq p_{it} \mid p_{it}) \neq 0$. From Lemma 3, it follows that $\mathbb{E}[p_{it+1} \mid \theta = H, p_{it}] > p_{it}$. The case of the state $\theta = L$ is analogous. ■

Proof of Corollary 2. Define probability measures $\mathbb{P}^H := \mathbb{P}(\cdot \mid \theta = H)$ and $\mathbb{P}^L := \mathbb{P}(\cdot \mid \theta = H)$ on the paths of play, in the states H and L respectively, under the given strategy. Also, define the probability measure $\mathbb{Q} = p_0 \mathbb{P}^H + (1 - p_0)\mathbb{P}^L$, where p_0 is the prior. Note that both $\mathbb{P}^H$ and $\mathbb{P}^L$

are absolutely continuous with respect to $\mathbb{Q}$. Now, let $\mathcal{F}_{it}$ be the filtration generated by player i's private information (x_i^t, a^{t-1}). From Blackwell-Dubins Merging Theorem, it follows that

$$\|\mathbb{Q}(\cdot \mid \mathcal{F}_{it}) - \mathbb{P}^H(\cdot \mid \mathcal{F}_{it})\|_{TV} \to 0, \quad \text{a.s. } \mathbb{P}^H.$$

Now, define the event

$$E = \left\{\omega \in \Omega \mid \lim_{t\to\infty} \frac{|\{\tau \le t : x_{i\tau} = x\}|}{t} = f_H(x) \text{ for all } x \in X\right\}.$$

Since private signals are i.i.d., by the strong law of large numbers, in the states $\theta = H, L$, we have $\mathbb{P}^H(E \mid \mathcal{F}_{it}) = 1$ and $\mathbb{P}^L(E \mid \mathcal{F}_{it}) = 0$. Further, note that

$$|\mathbb{Q}(E \mid \mathcal{F}_{it}) - \mathbb{P}^H(E \mid \mathcal{F}_{it})| = 1 - p_{it}.$$

By the definition of the total variational distance, we obtain that $\|\mathbb{Q}(\cdot \mid \mathcal{F}_{it}) - \mathbb{P}^H(\cdot \mid \mathcal{F}_{it})\|_{TV} \ge 1 - p_{it}$. Hence $p_{it} \to 1$ almost surely in the state $\theta = H$. The argument is the same for both states and players. ■

Proof of Proposition 5. We prove the claim in the case of the state $\theta = H$. The proof for the case of the state $\theta = L$ is analogous. We proceed in two steps.

Step 1: Fix a strategy profile σ, and suppose that there exists an on-path action history $\bar{a}^{t-1} \in A^{t-1}$ such that

$$\mathbb{E}[p_{it} \mid \theta = H, \bar{a}^{t-1}] \ge 1 - \varepsilon^2$$

for each $i = 1, 2$. From Lemma 4, it follows that

$$\mathbb{P}[p_{it} > 1 - \varepsilon \mid \theta = H, \bar{a}^{t-1}] \ge 1 - \varepsilon. \tag{12}$$

We will show that

$$\mathbb{P}\Big[C_t^{(1-\varepsilon)^2}(H) \mid \theta = H, \bar{a}^{t-1}\Big] \ge (1 - \varepsilon)^2. \tag{13}$$

Note that beliefs p_{1t} and p_{2t} are independent conditional on the state $\theta = H$ and action history $\bar{a}^{t-1}$, since private signals are conditionally independent. Now, consider the event

$$F_t = \Big\{\omega \in \Omega : a^{t-1}(\omega) = \bar{a}^{t-1},\ p_{it}(\omega) \ge 1 - \varepsilon \text{ for each } i = 1, 2\Big\}.$$

Note that from inequality (12), we have $\mathbb{P}[F_t \mid \theta = H, \bar{a}^{t-1}] \ge (1 - \varepsilon)^2$. Hence, it suffices to show that $F_t \subseteq C_t^{(1-\varepsilon)^2}(H)$. By the result of Monderer and Samet (1989), it suffices to show that the event F_t is $(1-\varepsilon)^2$-evident (i.e., $F_t \subseteq B_t^{(1-\varepsilon)^2}(F_t)$) and that $F_t \subseteq B_t^{(1-\varepsilon)^2}(H)$. The latter follows since $F_t \subseteq B_t^{1-\varepsilon}(H) \subseteq B_t^{(1-\varepsilon)^2}(H)$. Thus, we show the former. At any $\omega \in F_t$, in period t player i's belief about the event F_t is

$$\mathbb{P}\Big[F_t \mid \bar{a}^{t-1}, x_i^t\Big] \geq p_{it}(\omega)\mathbb{P}\Big[F_t \mid \theta = H, \bar{a}^{t-1}, x_i^t\Big] \geq (1-\varepsilon)^2.$$

This implies that $\omega \in B_t^{(1-\varepsilon)^2}(F_t)$. That is, $F_t \subseteq B_t^{(1-\varepsilon)^2}(F_t)$. Hence, inequality (13) follows. Moreover, since $C_t^{(1-\varepsilon)^2}(H) \subseteq C_t^{(1-\varepsilon)^2(1-\varepsilon^2)}(H)$, we have

$$\mathbb{P}\Big[C_t^{(1-\varepsilon)^2(1-\varepsilon^2)}(H) \mid \theta = H, \bar{a}^{t-1}\Big] \geq (1-\varepsilon)^2.$$

Step 2: Fix any $\varepsilon \in (0,1)$. By Corollary 2, there exists some $\bar{T} \in \mathbb{N}$ such that for each $i = 1,2$ and $t \geq \bar{T}$,

$$\mathbb{E}\Big[p_{it} \mid \theta = H\Big] \geq 1 - \left(\frac{\varepsilon^2}{2}\right)^2.$$

By the tower property of conditional expectations,

$$\mathbb{E}[p_{it} \mid \theta = H] = \mathbb{E}\Big[\mathbb{E}\Big[p_{it} \mid \theta = H, a^{t-1}\Big] \mid \theta = H\Big].$$

From Lemma 4, we have that for each $i = 1,2$,

$$\mathbb{P}\Big[\mathbb{E}\Big[p_{it} \mid \theta = H, a^{t-1}\Big] \geq 1 - \frac{\varepsilon^2}{2} \mid \theta = H\Big] \geq 1 - \frac{\varepsilon^2}{2}.$$

Then,

$$\mathbb{P}\Big[\mathbb{E}\Big[p_{it} \mid \theta = H, a^{t-1}\Big] \geq 1 - \varepsilon^2 \text{ for each } i = 1,2 \mid \theta = H\Big] \geq 1 - \varepsilon^2.$$

Hence, for each $t \geq \bar{T}$,

$$\mathbb{P}\Big[C_t^{(1-\varepsilon)^2(1-\varepsilon^2)}(H) \mid \theta = H\Big] \geq (1-\varepsilon)^2(1-\varepsilon^2),$$

which establishes the desired result. ■

B Sequential Equilibrium

A *behavioral strategy* for player i is defined as a map from private histories to probabilities over actions i.e., $\sigma_i : \bigcup_t X^t \times A^{t-1} \to \Delta(\{I,N\})$. For a given strategy profile $\sigma = (\sigma_i, \sigma_j)$, we define *continuation values* as $V_i^\theta(\sigma \mid h_t)$ for each $h_t = (a^{t-1}, x_i^t, x_j^t) = (h_{it}, h_{jt})$. This value represents the long-run expected payoff player i would obtain at history h_t, if players play according to the strategies σ and the state is θ. At period t, depending on the private history, player i has beliefs about the state of nature and private signals of his opponent $(\theta, x^t) \in \Theta \times X^t$. Player i's *belief function* is given by $\mu_i : \bigcup_t H_{it} \to \Delta(\Theta \times X^t)$. An *assessment* is a pair (σ, μ) consisting of a behavioural strategy profile $\sigma = (\sigma_i)_i$ and belief functions for each player $\mu = (\mu_i)_i$. We use sequential equilibrium as our solution concept, which we define in the current setting as follows:

Definition 4. *An assessment* (σ^*, μ^*) *is a* sequential equilibrium *if:*

1. (Sequential Rationality) *The strategy profile* σ^* *is sequentially rational for each player given beliefs* μ^**. This means, for each* $h_{it} = (a^{t-1}, x_i^t) \in H_{it}$ *and* $\sigma_i \in \Sigma_i$,

$$\sum_{(\theta, x_j^t)} \mu_i^*\big(h_{it}\big)\big(\theta, x_j^t\big) V_i^\theta\big(\sigma^* \mid a^{t-1}, x_i^t, x_j^t\big) \geq \sum_{(\theta, x_j^t)} \mu_i^*\big(h_{it}\big)\big(\theta, x_j^t\big) V_i^\theta\big(\sigma_i, \sigma_{-i}^* \mid a^{t-1}, x_i^t, x_j^t\big). \tag{14}$$

2. (Consistency) *There exists a sequence of profiles of completely mixed strategies (that assigns positive probability to all actions at all private histories)* $\{\sigma^n\}_{n\geq 1}$ *such that the belief functions* $\{\mu^n\}_{n\geq 1}$ *uniquely induced by them satisfy*

$$\lim_{n\to\infty} (\mu_i^n(h_{it}), \sigma_i^n(h_{it})) = (\mu_i^*(h_{it}), \sigma_i^*(h_{it})) \tag{15}$$

for all $h_{it} \in H_{it}$*, and convergence is with respect to the product topology on* $\Delta(\Theta \times X^t) \times A_i$.

The definition of sequential equilibrium is borrowed from Kreps and Wilson (1982) and is adapted to our setting. While their definition corresponds to finite extensive-form games, the extension to the current setting is justified since there are only countably many information sets for each player and moreover, each information set is finite.

The following lemma reduces the problem of finding a sequential equilibrium to that of finding a Nash equilibrium.

Lemma 5. *For every mixed strategy Nash equilibrium* σ^* *in our game there exists an outcome equivalent sequential equilibrium* (σ^{**}, μ^*).

Proof. [10] Let σ^* be a mixed-strategy Nash equilibrium. From Kuhn's theorem, there exists an outcome equivalent behavioral strategy profile β^*. It follows that β^* is also a Nash equilibrium.

Player i's private history $h_{it} = (a^{t-1}, x_i^t)$ is on-path if there are private signals x_j^{t-1} such that $\beta_i^*(a^{s-1}, x_i^s)$ assigns positive probability to a_{is} and $\beta_j^*(a^{s-1}, x_j^s)$ assigns positive probability to a_{js} for each $s = 1, 2, \ldots, t-1$. We classify the set of his off-path histories into two. First, player i's (off-path) history h_{it} follows his own unobservable deviation if $\beta_i^*(a^{s-1}, x_i^s)$ assigns probability zero to a_{is} for some $s = 1, 2, \ldots, t-1$ but there are private signals $(\tilde{x}_i^{t-1}, x_j^{t-1})$ such that $\beta_i^*(a^{s'-1}, \tilde{x}_i^{s'})$ assigns positive probability to $a_{is'}$ and $\beta_j^*(a^{s'-1}, x_j^{s'})$ assigns positive probability to $a_{js'}$ for each $s' = 1, 2, \ldots, t-1$. Second, player i's (off-path) history h_{it} follows either player's observable deviation if there are some $s = 1, 2, \ldots, t-1$ and some player k's action a_{ks} to which $\beta_k^*(a^{s-1}, x_k^s)$ assigns probability zero for all $x_k^s \in X^s$.[11]

[10] Similar arguments to some parts of this proof are in Sekiguchi (1997).

[11] For example, players using cool-off strategies are supposed to choose action N in a cool-off phase, irrespective of private signals. Hence, choosing action I in a cool-off phase is an observable deviation.

Now we define another behavioral strategy profile σ^{**}. Let $\sigma_i^{**}(h_{it}) = \beta_i^*(h_{it})$ for any on-path h_{it}, while let $\sigma_i^{**}(h_{it}) = N$ for any off-path h_{it} that follows either player's observable deviation. Then, it suffices to consider an off-path h_{it} that follows player i's own unobservable deviation. Given such a h_{it}, consider player i's belief about player j's continuation strategy, derived from β_j^*. There exists an optimal continuation strategy for player i at h_{it}, since the set of his continuation strategies are compact and the payoff function is continuous with respect to the product topology. Choose such a continuation strategy, and let $\sigma_i^{**}(h_{it})$ be the (mixed) action specified under that continuation strategy at h_{it}. By construction, σ_i^{**} is a best response to β_j^*.

Now we construct the belief system μ^*. The beliefs at any on-path histories or off-path history that follows an unobservable deviation are uniquely determined by Bayes' rule, while the beliefs at any off-path history that follows an observable deviation are set as follows. When a player makes an observable deviation by taking an action that is not supported by any realization of her private signals, players update their beliefs as if they observed the other action. Note that $\sigma^{**} = (\sigma_i^{**}, \sigma_j^{**})$ is sequentially rational given μ^*.

To see that μ^* satisfies consistency, for each $i = 1, 2$ and $n > 1$, define the behavioral strategy σ_i^n as follows:

$$\sigma_i^n(a^{t-1}, x_i^t) = \begin{cases} 1 - \frac{1}{n} & \text{if } \sigma_i^{**}(a^{t-1}, x_i^t) = 1 \\ \frac{1}{n} & \text{if } \sigma_i^{**}(a^{t-1}, x_i^t) = 0 \\ \sigma_i^{**}(a^{t-1}, x_i^t) & \text{o.w.} \end{cases}$$

Note that σ_i^n chooses each action at each private history with positive probability. In particular, σ^n differs from σ^{**} in that whenever under σ^{**} player i chooses an action with probability one, under σ^n player i chooses that same action instead with probability $1 - \frac{1}{n}$. So we can see $\lim_{n\to\infty} \sigma_i^n = \sigma_i^{**}$. (Note that this change does not depend on the player's private history. Hence, when an observable deviation occurs under σ^{**}, the beliefs of the players are unchanged by assumption and are close to the beliefs generated by σ^n as n tends to infinity.) ■

References

George-Marios Angeletos, Christian Hellwig, and Alessandro Pavan. Dynamic global games of regime change: Learning, multiplicity, and the timing of attacks. *Econometrica*, 75(3):711–756, 2007.

Sandeep Baliga and Tomas Sjöström. Arms races and negotiations. *Review of Economic Studies*, 71 (2):351–369, 2004.

Hans Carlsson and Eric van Damme. Global games and equilibrium selection. *Econometrica*, (5): 989–1018, 1993.

Sylvain Chassang. Fear of miscoordination and the robustness of cooperation in dynamic global games with exit. *Econometrica*, 78(3):973–1006, 2010.

Martin W Cripps, Jeffrey C Ely, George J Mailath, and Larry Samuelson. Common learning. *Econometrica*, 76(4):909–933, 2008.

Douglas W Diamond and Philip H Dybvig. Bank runs, deposit insurance, and liquidity. *Journal of Political Economy*, 91(3):401–419, 1983.

Chris Edmond. Information manipulation, coordination, and regime change. *Review of Economic Studies*, 80(4):1422–1458, 2013.

Glenn Ellison. Cooperation in the prisoner's dilemma with anonymous random matching. *Review of Economic Studies*, 61(3):567–588, 1994.

Drew Fudenberg and Yuichi Yamamoto. Repeated games where the payoffs and monitoring structure are unknown. *Econometrica*, 78(5):1673–1710, 2010.

Drew Fudenberg and Yuichi Yamamoto. Learning from private information in noisy repeated games. *Journal of Economic Theory*, 146(5):1733–1769, 2011.

Itay Goldstein and Ady Pauzner. Demand–deposit contracts and the probability of bank runs. *The Journal of Finance*, 60(3):1293–1327, 2005.

David M Kreps and Robert Wilson. Sequential equilibria. *Econometrica*, (4):863–894, 1982.

Dov Monderer and Dov Samet. Approximating common knowledge with common beliefs. *Games and Economic Behavior*, 1(2):170–190, 1989.

Stephen Morris and Hyun Song Shin. Unique equilibrium in a model of self-fulfilling currency attacks. *American Economic Review*, 88(3):587–597, 1998.

Martin J Osborne and Ariel Rubinstein. *A course in game theory*. MIT press, 1994.

Ariel Rubinstein. The electronic mail game: Strategic behavior under "almost common knowledge". *American Economic Review*, 79(3):385–391, 1989.

Tadashi Sekiguchi. Efficiency in repeated prisoner's dilemma with private monitoring. *Journal of Economic Theory*, 76(2):345–361, 1997.

Lones Smith and Peter Sørensen. Pathological outcomes of observational learning. *Econometrica*, 68(2):371–398, 2000.

Takuo Sugaya and Yuichi Yamamoto. Common learning and cooperation in repeated games. *Theoretical Economics*, 15(3):1175–1219, 2020.

Xavier Vives. Games of strategic complementarities: an application to bayesian games. *Spanish Economic Review*, 9(4):237–247, 2007.

Thomas Wiseman. A partial folk theorem for games with unknown payoff distributions. *Econometrica*, 73(2):629–645, 2005.

Thomas Wiseman. A partial folk theorem for games with private learning. *Theoretical Economics*, 7(2):217–239, 2012.

Yuichi Yamamoto. Individual learning and cooperation in noisy repeated games. *Review of Economic Studies*, 81(1):473–500, 2014.