Penn Institute for Economic Research
Department of Economics
University of Pennsylvania
3718 Locust Walk
Philadelphia, PA 19104-6297
pier@econ.upenn.edu
http://economics.sas.upenn.edu/pier

PIER Working Paper 17-011

We Can Cooperate Even When the Monitoring Structure Will Never Be Known

by

Yuichi Yamamoto

http://ssrn.com/abstract=2969936

We Can Cooperate Even When the Monitoring Structure Will Never Be Known*

Yuichi Yamamoto†

First Draft: March 29, 2014
This Version: April 8, 2017

Abstract

This paper considers infinite-horizon stochastic games with hidden states and hidden actions. The state changes over time, players observe only a noisy public signal about the state each period, and actions are private information. In this model, uncertainty about the monitoring structure does not disappear. We show how to construct an approximately efficient equilibrium in a repeated Cournot game. Then we extend it to a general case and obtain the folk theorem using ex-post equilibria under a mild condition.

Journal of Economic Literature Classification Numbers: C72, C73.

Keywords: stochastic game, hidden state, public monitoring, pseudo-ergodic strategy, folk theorem, ex-post equilibrium.

*The author thanks George Mailath and Takuo Sugaya for helpful conversations, and seminar participants at various places.

†Department of Economics, University of Pennsylvania. Email: yyam@sas.upenn.edu

Contents

1 Introduction

The theory of repeated games provides a framework to study the role of long-term relationships in facilitating cooperation. Past work has shown that reciprocation can lead to more cooperative equilibrium outcomes even if there is *imperfect public monitoring* so that each period, players observe only a noisy public signal about the actions played (Abreu, Pearce, and Stacchetti (1990) and Fudenberg, Levine, and Maskin (1994, hereafter FLM)). This work has covered a range of applications, from oligopoly pricing (e.g. Green and Porter (1984) and Athey and Bagwell (2001)), repeated partnerships (Radner, Myerson, and Maskin (1986)), and relational contracts (Levin (2003)). A common assumption in the literature is that players know the monitoring structure, i.e., they know the distribution of public signals as a function of the actions played. Fudenberg and Yamamoto (2010) relax this assumption and consider players who initially do not know the monitoring structure. However, in their model, the true monitoring structure is fixed over time, and thus players can learn it from observed signals in the long run. In particular, their analysis relies on the fact that patient players care only about payoffs in a distant future in which uncertainty about the monitoring structure vanishes.

Assuming (asymptotically) perfect knowledge of the monitoring structure is restrictive. To address this concern, this paper considers a model in which uncertainty about the monitoring structure never disappears. Specifically, we consider a model with *unknown, perpetually changing monitoring structure*. In this model, players may obtain some information about the current monitoring structure from the signal today, *after* they choose actions. But then in the next period, the monitoring structure will stochastically change, so players will continue to face new uncertainty.

Changing monitoring structures naturally arise when the underlying economic conditions change over time. One example is a repeated Cournot model with hidden correlated demand shocks. Suppose that the state of the economy ω, which influences the distribution of the market price today, is hidden information and positively correlated over time. As in Green and Porter (1984), the realized price is regarded as a noisy public signal about the current actions (quantities), because higher quantities induce low prices more likely. In this model, the signal (price) distributions are different for different periods, since the state ω is not i.i.d.. Also the true signal distribution is unknown each period, as the state ω is not observable. Hence the monitoring structure is unknown and changing. Another example is a repeated principal-agent problem. If the agent's productivity is unobservable, and is changing due to experience, then the true distribution of the output is unknown and changing. This paper shows that long-run relationships facilitate cooperation even in such situations. In particular, we show that the folk theorem obtains using public-strategy equilibria.

Formally, we consider a new class of stochastic games in which a hidden, changing state influences the monitoring structure (the signal distribution) in the current stage game. Actions can influence the state transition. Each player's stage-game payoff depends on her own action and the public signal, so the payoff does not contain more information than the signal. In this setup, the hidden state indirectly influences the stage-game payoffs through the distribution of the public signal. For example, in a repeated oligopoly, firms have higher expected payoffs at a state in which high prices are more likely. So uncertainty about the probability of high prices leads to uncertainty about the expected payoffs of the stage game.

Since we assume that actions are private information, even if players have an initial common prior about the state, their posterior beliefs can potentially diverge in later periods. For example, if a player chooses a mixed action, the realized action is her private information, and she updates her posterior given this information. Similarly, if a player deviates from an equilibrium strategy, she will update her posterior given her deviation, while the opponents will update without knowing it. A common technique in the literature is to allow cheap-talk communications to resolve conflicting information (e.g., Kandori and Matsushima (1998)), but it does not seem to easily apply to our setup.[1]

Diverging posterior beliefs can cause a miscoordination problem. Suppose

[1] Kandori and Matsushima (1998) consider repeated games with private monitoring and communication, in which there is no payoff-relevant state and each player reports private signals about the opponents' actions. They focus on "public equilibria" in which the play depends only on public reports. Then a player's continuation payoff is a function of the past public reports, which allows them to use recursive tools to characterize the equilibrium payoff set. In contrast, in this paper, each player i has a private belief μ_i about the payoff-relevant state, and this private belief directly influences her continuation payoffs. That is, the continuation payoff depends on the true belief μ_i and is not a function of public histories. Hörner, Takahashi, and Vieille (2015) argue that the equilibrium analysis becomes significantly harder in such a case. They show that it is still possible to provide truthful incentives if some assumptions are satisfied (e.g., independent private values); but unfortunately, these assumptions do not hold in our setup.

that there are two players who want to reward each other (i.e., they each want to give high payoffs to the other). If they have the same belief μ about the state, this can be done by playing a welfare-maximizing strategy profile, say $s^{\text{eff}}(\mu)$. Note that this profile depends on the belief μ, because the stage-game payoffs depend on the hidden state ω. On the other hand, when players have different beliefs and these beliefs are private information, it is less clear what they should do. Indeed, if each player simply chooses the welfare-maximizing strategy corresponding to her own belief, the resulting strategy profile may not maximize the welfare due to miscoordination. A similar problem arises when we consider a player who wants to punish the opponent. If the opponent's belief is private information, it is unclear how to punish the opponent, as the effective punishment depends on the opponent's belief in general.

To overcome this problem, this paper introduces the idea of "pseudo-ergodic strategies." In general, given a strategy profile in the infinite-horizon game, different initial priors induce different payoff streams, which result in different average payoffs. Pseudo-ergodic strategies are a special class of strategy profiles in which all initial priors yield approximately the same average payoffs. Such a property may sound demanding, but it turns out that in our setting, for an arbitrarily fixed belief $\tilde{\mu}$, the corresponding welfare-maximizing strategy profile $s^{\text{eff}}(\tilde{\mu})$ is a pseudo-ergodic strategy which approximates the welfare-maximizing payoff regardless of the true belief μ. A rough idea is that when players play this strategy profile $s^{\text{eff}}(\tilde{\mu})$, the resulting payoff stream in the infinite-horizon game has a flavor of ergodicity in the sense that the initial belief does not influence the continuation payoff after a long time. This indeed implies that the strategy profile $s^{\text{eff}}(\tilde{\mu})$ yields almost the same payoff for all initial beliefs, as patient players care only about payoffs in a distant future. See Section 4.1 for more details.

So if players want to reward each other, they may ignore their private beliefs and simply play the above profile $s^{\text{eff}}(\tilde{\mu})$. That is, they may form a "dummy public belief" $\tilde{\mu}$ and play the corresponding strategy. This approximates the efficient payoff regardless of their initial private beliefs. Similarly, if a player plays the minimax strategy for some dummy belief $\tilde{\mu}$, it approximates the minimax payoff regardless of the true belief μ. In this way, players can reward or punish their opponent without fine-tuning the strategy depending on their private beliefs.

The next question is whether we can actually construct an equilibrium by assembling these pseudo-ergodic strategies: We need to find an effective punishment mechanism when players have diverging beliefs about the true monitoring structure. To solve this problem, we consider a punishment mechanism in which a deviation today will lower continuation payoffs regardless of the current hidden state ω. Under this mechanism, *ex-post incentive compatibility* is satisfied in that any deviation today is prevented regardless of the current hidden state ω; hence this mechanism works even if there is uncertainty about the state ω.

Of course, ex-post incentive compatibility is more demanding than Bayesian incentive compatibility, and in general, the set of ex-post equilibrium payoffs is smaller than the set of sequential equilibrium payoffs. However, in our environment, it is possible for ex-post equilibria to approximate the Pareto-efficient frontier. Indeed, our main result is the folk theorem: We show that any feasible and individually rational payoff can be approximated by a public ex-post equilibrium, if players are patient and if *cross-state individual full rank* and *cross-state pairwise full rank* hold. The cross-state full-rank condition is an extension of individual full rank and pairwise full rank of FLM, and requires that a public signal can statistically distinguish the current state and the chosen action profile.

Fudenberg and Yamamoto (2010) also consider ex-post equilibria when players face uncertainty about the monitoring structure. However, there are important differences between their work and this paper. As noted, in Fudenberg and Yamamoto (2010), players can learn the true monitoring structure in the long run. Then players' incentive problems can be decomposed state by state; this is because it is possible to influence players' incentives in some state ω without affecting incentives in other states, by changing players' continuation play in a distant future in which players have learned the true state ω. This property helps to provide ex-post incentives.[2]

On the other hand, in our model, the state today is never revealed to players, and thus the above idea does not apply. Accordingly, incentive problems for different states are entangled in a non-trivial way, and providing ex-post incentive compatibility becomes quite delicate. In particular, the "state-specific punishment" of Fudenberg and Yamamoto (2010) do not work effectively in our environment.

[2]Their analysis is more complicated than the discussion here, because ex-post incentives must be provided *each period*. They develop a useful recursive method and show that it is indeed possible to provide such incentives.

More detailed discussions will be given in Section 3.7.

The contributions of this paper are two-fold. First, we provide a general idea on how to construct an equilibrium in a new environment, at least for high discount factors δ. In particular, we illustrate how dynamic incentives can be effectively and simply provided via public pseudo-ergodic strategies.

Second, we show that ex-post equilibrium can approximate efficient outcomes, even if the state changes over time so that state learning is impossible. As will be explained in the next subsection, "utility transfer across players" of FLM cannot provide appropriate incentives in our environment, and we construct a new punishment mechanism which works out even when there is a hidden changing state.

1.1 Overview of the Argument

To understand the critical steps in our proof, it is useful to review the ideas of FLM, who prove the folk theorem for repeated games with public monitoring. Their main finding is that when players are patient, any ball W in the interior of the feasible and individually rational payoff set (see Figure 1) is *self-generating*. That is, each payoff v in the ball W is achievable by (some action profile today and) continuation payoffs in the ball W itself. As shown by Abreu, Pearce, and Stacchetti (1990), such a ball W is attained by public equilibria, and hence the folk theorem indeed follows.

How do they prove that the ball W is self-generating? As a first step, they show that each payoff v on the boundary of the ball W can be achieved using continuation payoffs w on a *translate of the tangent line*. For example, take the target payoff v as in Figure 1. (As we will soon see, this is the most difficult case in our proof.) FLM show that this payoff v is achievable by the action profile a^X which yields the payoff X in the figure, and by some continuation payoffs on the horizontal line L. Here, the continuation payoffs take different values for different signals, so that player 1's deviation today is deterred. Also, since player 2's continuation payoff is constant on the line L and the action profile a^X achieves the best payoff X for player 2, she has no incentive to deviate either. So appropriate incentives are indeed provided by the continuation payoffs on the line L. Without loss of generality, we can assume that the variation in continuation payoffs (the distance between w and w'' in the figure) is of order $O(1-\delta)$; such continuation payoffs can indeed deter player 1's deviation today, because her gain by deviating is of order $O(1-\delta)$. Also, the length D in the figure, which measures the distance from the payoff v to the line L, is of order $O(1-\delta)$. This is so because v must be exactly achieved as the weighted average of today's payoff X and the expected continuation payoff on the line L, where the weight on today's payoff is $1-\delta$.

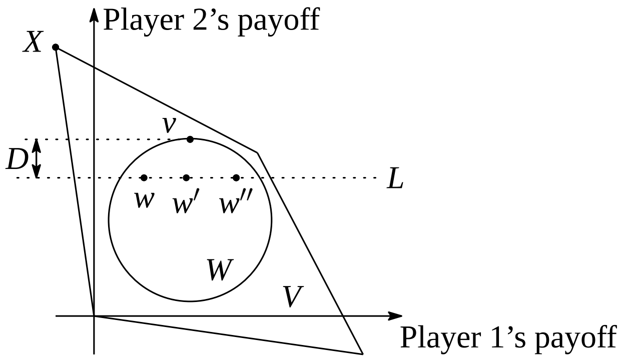

Figure 1: Continuation payoffs w, w', and w'' are on the line L.

Then as a second step, FLM show that if continuation payoffs w move only on the line L (and if the variation is of order $O(1-\delta)$), they stay in the interior of the ball W. The proof idea is illustrated in the left panel of Figure 2; as one can see, the distance to the boundary of the ball W is of order $O(\sqrt{1-\delta})$, which is much larger than the variation in the continuation payoffs w, and thus w never goes to the outside of the ball W. This result implies that the continuation payoffs constructed in the first step are in the ball W, so any boundary point v of the ball W is achievable by continuation payoffs in W. They also show that the same result holds even if v is an interior point of W, so in sum, any payoff v in the ball W is achievable by continuation payoffs in W itself. Hence W is indeed self-generating.

To summarize, the key technique of FLM is to decompose the target payoff v into two parts: The one-shot action profile a^X and the continuation payoffs on the line L (which are always in the ball W). Our proof extends this technique to the case in which the monitoring structure is unknown and changing. Since there is a hidden changing state in our model, new complications arise, and we need to modify the proof accordingly. Specifically, we make the following changes:

- We replace the action profile a^X above with a *pseudo-ergodic block strategy*, which approximates the payoff X regardless of players' private beliefs.

- We allow the continuation payoffs to move *vertically*, so they are *not* on the line L. (But they are still in the ball W, so the ball is self-generating. See Figure 2.)

In what follows, we will explain why we need such changes, and how they work.

To begin with, note that the definition of the feasible payoff set in our environment is different from the one in the standard repeated game; since the stage game payoffs are influenced by a hidden, changing state ω, they are not "feasible payoffs" in the infinite-horizon game. Instead, given the initial prior μ about the state and the discount factor δ, we define the feasible payoff set $V^{\mu}(\delta)$ as the set of all possible payoff vectors in the infinite-horizon game. Let V^{μ} denote the limit of the feasible payoff set as $\delta \to 1$; intuitively, this is the feasible payoff set when players are patient. In the special case in which the state is observable and follows a Markov process, this limit feasible payoff set V^{μ} does not depend on the initial prior μ, because the state eventually converges to the stationary distribution regardless of the initial state. Our Proposition 2 shows that under a mild condition, the same result holds even for our general model in which the state is unobservable and influenced by actions. So we denote this limit feasible payoff set by V, as in Figure 1.

Since the feasible payoff set is quite different from the stage-game payoffs, each extreme point of the feasible payoff set V may not be attained by any one-shot action profile in our model. For example, in order to decompose the payoff v in Figure 1, FLM use the action profile a^X, which yields the best payoff X for player 2 within the feasible payoff set. In our model, such an action profile may not exist, as the payoff X is a payoff in the infinite-horizon game, rather than a

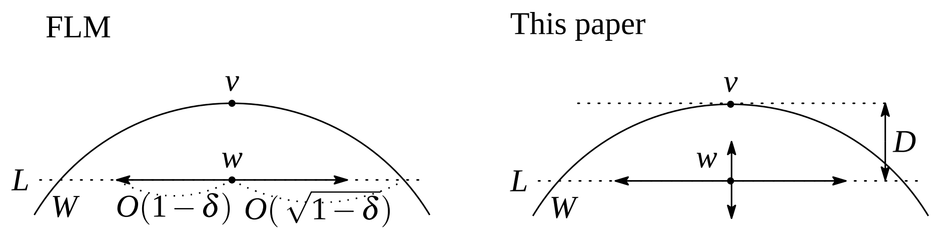

Figure 2: Vertical move of w must be less than $D = O(1-\delta)$. This is more restrictive than the bound on the horizontal move, which is of order $O(\sqrt{1-\delta})$.

stage-game payoff. To fix this problem, we regard the infinite horizon as a series of *blocks*, and treat each block as a "big" stage game. The point is that when the block is sufficiently long, each extreme payoff of the feasible payoff set V is approximated by the average payoff in the block (i.e., the payoff in the "big" stage game). That is, the difference between the stage-game payoffs and the feasible payoff set disappears, if we regard a long block as a big stage game.

In particular, we use a pseudo-ergodic strategy in each block, so that players' private beliefs about the state have almost no impact on the block payoff. For example, instead of the action profile a^X in FLM, we use a pseudo-ergodic block strategy whose block payoff approximates the payoff X *regardless of players' beliefs*. To see how to find such a pseudo-ergodic strategy, pick a dummy belief $\tilde{\mu}$ arbitrarily, and let s^X be the block strategy which would maximize player 2's block payoff if players' initial common prior was $\tilde{\mu}$. In general, this strategy s^X needs not maximize player 2's payoff when the true belief μ differs from $\tilde{\mu}$. However, our Proposition 3 shows that it *approximates* the best payoff X regardless of the true belief μ. This s^X is the pseudo-ergodic strategy we use.[3]

This explains why we need the change stated in the first bullet point: By replacing one-shot action profiles in FLM with pseudo-ergodic block strategies, we can approximate each extreme point of the feasible payoff set, regardless of players' beliefs. However, this is not the only change we must make: As noted in the second bullet point, we consider continuation payoffs which are not on a translate of the tangent line. We make this change because we need to construct a public equilibrium in the presence of the hidden changing state, which requires continuation payoffs to satisfy a more demanding condition than in the standard repeated game. Moving continuation payoffs only on a translate of the tangent line is too restrictive to satisfy this new condition.

To be more specific, take an arbitrary ball W in the feasible payoff set V. Our goal is to show that this ball W is achieved by public equilibria. For this, it is sufficient to show that the ball W is self-generating; but the definition of self-generation here is slightly different from the one in FLM, due to the hidden

[3]The idea of the block itself is not new: Dutta (1995) uses the same technique in stochastic games with observable states. The novelty here is to use a pseudo-ergodic strategy, which allows players to approximate a desired payoff even though their block strategy cannot depend on the hidden state. In Dutta (1995), the state is observable, and thus players can use different block strategies for different initial states.

changing state. To illustrate the difference, take the payoff v as in Figure 1. For the ball W to be self-generating in our sense, we need to find continuation payoffs w in the ball W such that regardless of players' beliefs μ, (i) the payoff v is exactly achieved as the sum of the block payoff by the pseudo-ergodic strategy s^X (which approximates the payoff X) and the continuation payoff w, and (ii) any deviation from the strategy s^X during the block is not profitable. This condition is more demanding than that in FLM, because the choice of w must be independent of players' initial belief μ, that is, our continuation payoffs must work for all beliefs μ. Note in particular that the condition (ii) above requires s^X to be an ex-post equilibrium, in that playing s^X is optimal for each player even if the initial state ω is revealed.[4] To satisfy this condition, we consider continuation payoffs which move vertically, as in the right panel of Figure 2. Allowing vertical move is useful for two reasons:

(a) The block strategy s^X approximates the payoff X, but does not exactly achieve it. In particular, different initial beliefs μ yield (slightly) different block payoffs to player 2. This payoff difference must be offset by continuation payoffs, as we want the same payoff v to be achieved for all beliefs. So player 2's continuation payoff cannot be constant, and thus w must move vertically.

(b) When player 2's belief differs from the dummy belief $\tilde{\mu}$, the block strategy s^X does not maximize her block payoff, so she can earn a positive profit by deviating from s^X. We need to punish such a deviation via a variation in continuation payoffs. That is, we need to burn player 2's continuation value (relative to the line L) after some signals.[5]

These issues (a) and (b) could be easily handled if we could choose continuation payoffs in an arbitrarily way, but unfortunately, there is a constraint; we must choose the continuation payoffs from the ball W. (Otherwise, the ball W is not

[4] But this condition is weaker than "perfect" public ex-post equilibria of Fudenberg and Yamamoto (2010), which requires that *in each period t*, the continuation strategy is a Nash equilibrium even if the state ω^t in that period is revealed. See Section 3.7 for more detailed discussions.

[5] The problem (b) here is relevant only when the tangent at the point v is a coordinate vector. Indeed, when we consider v whose tangent is not a coordinate vector (this is the case of "regular directions" in FLM), we can incentivize both players by moving continuation payoffs on a translate of the tangent line. See Section 3.5 for more details.

self-generating.) This in particular implies that the vertical move of the continuation payoffs cannot be greater than the length D in Figure 2. It turns out that this constraint is quite restrictive, and makes our problem substantially different from the one in FLM, in the following sense. Recall that FLM consider the horizontal move only, in which case the distance to the boundary of the ball W is of order $O(\sqrt{1-\delta})$. This constraint is "loose" in that continuation payoffs are always in the ball W as long as the variation is of order $O(1-\delta)$. In contrast, the bound D on the vertical move is of order $O(1-\delta)$. Hence the continuation payoff may go to the outside of the ball, even if the variation toward the vertical direction is of order $O(1-\delta)$.

So in sum, we need to solve the above problems (a) and (b), subject to the constraint that the vertical move is sufficiently "small." Note that this constraint is deeply related to the inefficiency result of Radner, Myerson, and Maskin (1986), who show that the use of huge value burning causes huge inefficiency. In order to construct an approximately efficient equilibrium, we must avoid such inefficiency, so we need to minimize the amount of value burning.

It is relatively easy to show that small value burning is indeed enough to solve the problem (a). Since the block strategy s^X yields almost the same block payoff for all beliefs μ, only a small perturbation of the continuation payoffs is enough to offset this payoff difference.

The problem (b) is more delicate. Since the block strategy s^X approximates the best payoff X for player 2, her *ex-ante* expected gain by deviating from s^X is small; that is, deviating from s^X cannot improve the block payoff by much, if we evaluate payoffs by taking expectations over the future states and the future histories. However, this property needs not imply that small value burning is enough to solve the problem (b). To see why, suppose that we are now in period $t > 1$ of the block and the history within the block so far is h^{t-1}. If we want to deter player 2's deviation in the current period t via small value burning, we have to show that her gain by such a deviation is small *conditional on the current history* h^{t-1}. Obviously, this condition needs not be satisfied even if the ex-ante gain (which takes the expectation over h^{t-1}) is small.

So in order to solve the problem (b) with small value burning, we need to carefully evaluate player 2's gain when she deviates in later periods of the block. To do so, it is useful to examine how her posterior belief about the state evolves over time. Since the state is changing, the belief evolution in our model is complex, and keeping track of it over a long block is computationally demanding. Nonetheless, we find that under a mild condition, the *belief convergence theorem* holds, so that the impact of the current belief on the posterior in a distant future is almost negligible. In other words, after a long history, all initial beliefs μ induce almost the same posterior.

An important consequence of the belief convergence theorem is that even if the true belief μ is quite different from the dummy belief $\tilde{\mu}$ in period one, after a long time, they induce asymptotically the same posteriors, μ^t and $\tilde{\mu}^t$. (See Figure 3.) This result is useful to obtain an effective bound on player 2's gain by deviating in a later period t of the block: Recall that the block strategy s^X maximizes player 2's payoff given the dummy initial belief $\tilde{\mu}$. So after every history h^{t-1}, deviating from s^X in the continuation game is not profitable if player 2's true posterior μ^t equals the dummy posterior $\tilde{\mu}^t$. Of course, these posteriors μ^t and $\tilde{\mu}^t$ need not be equal, if the initial belief μ differs from the dummy belief $\tilde{\mu}$; but as noted above, the belief convergence theorem ensures that the posteriors μ^t and $\tilde{\mu}^t$ are *asymptotically* the same for large t, even if the initial belief μ is quite different from $\tilde{\mu}$. Hence, player 2's gain by deviating in a later period t is small, and converges to zero as t increases. This property (in particular the fact that the gain converges to zero) is useful to find an effective bound on the amount of value burning which deters player 2's deviation in *all periods* of the block. See Section 3.6 for more details.

1.2 Literature Review

The framework of stochastic games was proposed by Shapley (1953). Dutta (1995) proved the folk theorem for the case of observable actions, and Fudenberg and Yamamoto (2011b) and Hörner, Sugaya, Takahashi, and Vieille (2011) extend it to games with public monitoring. All these papers assume that the state of the world is publicly observable at the beginning of each period. Yamamoto (2016) considers hidden states, but assumes that actions are observable. Accordingly, the belief is always common across players, and the model reduces to the stochastic game in which players' belief is a common state variable. In this paper, players' beliefs are private information and there is no common state variable.

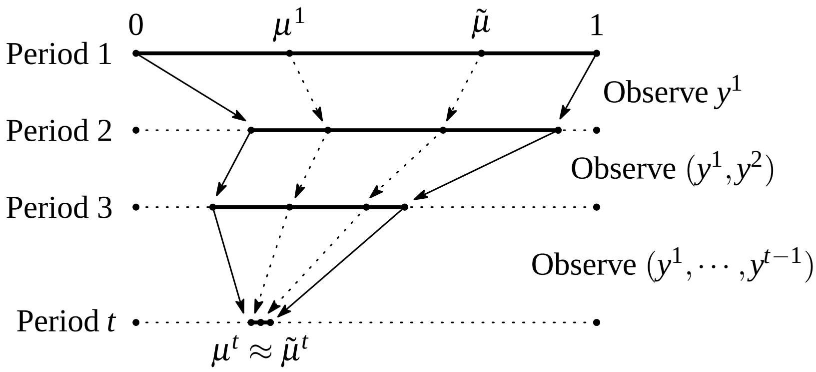

Figure 3: Belief evolution when there are only two states. The whole belief space is $[0,1]$. Each thick line is the set of all possible posteriors given the past history. It shrinks over time, so eventually all initial priors induce the same posterior.

Athey and Bagwell (2008), Escobar and Toikka (2013), and Hörner, Takahashi, and Vieille (2015) consider repeated Bayesian games in which the state changes as time goes and players have private information about the state each period. They assume hat the state of the world is a collection of players' private information, so if players report their information truthfully, the state is perfectly revealed before they choose actions.[6] In contrast, in this paper, the state is not perfectly revealed.

Wiseman (2005), Fudenberg and Yamamoto (2010), Fudenberg and Yamamoto (2011a), and Wiseman (2012) study repeated games with unknown states. They assume that the state does not change over time, so that players can (almost) perfectly learn the true state by aggregating all the past public signals. In our model, the state changes as time goes and players never learn it perfectly.

Ex-post equilibria have been recently used in various dynamic models, such as Hörner and Lovo (2009), Fudenberg and Yamamoto (2010), Fudenberg and Yamamoto (2011a), Hörner, Lovo, and Tomala (2011), and Yamamoto (2014). They consider the case in which the state is fixed at the beginning. Again this paper differs from their work, because we consider changing states.[7]

[6]Sections 4 and 5 of Hörner, Takahashi, and Vieille (2015) consider equilibria in which some players do not reveal information, but their analysis relies on the independent private value assumption.

[7]There are many papers that discuss ex-post equilibria in undiscounted repeated games; see Koren (1992) and Shalev (1994), for example.

2 Setup

2.1 Stochastic Games with Hidden States

Let $I = \{1, \cdots, N\}$ be the set of players. At the beginning of the game, Nature chooses the state of the world ω^1 from a finite set Ω. The state may change as time passes, and the state in period $t = 1, 2, \cdots$ is denoted by $\omega^t \in \Omega$. The state ω^t is not observable to players, so they have an initial common prior $\mu \in \triangle\Omega$ about ω^1.

In each period t, players move simultaneously, with player $i \in I$ choosing an action a_i from a finite set A_i. Let $A \equiv \times_{i \in I} A_i$ be the set of action profiles $a = (a_i)_{i \in I}$. Actions are not observable, and instead players observe a public signal y from a finite set Y. Then players go to the next period $t+1$, with a new (hidden) state ω^{t+1}. The distribution of y and ω^{t+1} depends on the current state ω^t and the current action profile $a \in A$; let $\pi^\omega(y, \tilde{\omega}|a)$ denote the probability that players observe a signal y and the next state becomes $\omega^{t+1} = \tilde{\omega}$, given $\omega^t = \omega$ and a. In this setup, a public signal y can be informative about the current state ω *and* the next state $\tilde{\omega}$. This is so because the distribution of y may depend on ω, and y may be correlated with $\tilde{\omega}$. Let $\pi_Y^\omega(y|a)$ denote the marginal probability of y.

Player i's payoff in period t is a function of her current action a_i and the current public signal y, and is denoted by $u_i(a_i, y)$. Her expected stage-game payoff conditional on the current state ω and the current action profile a is $g_i^\omega(a) = \sum_{y \in Y} \pi_Y^\omega(y|a) u_i(a_i, y)$. Here the hidden state ω influences a player's expected payoff through the distribution of y. Let $g^\omega(a) = (g_i^\omega(a))_{i \in I}$ be the vector of expected payoffs. Let $\overline{g}_i = \max_{\omega, a} |2g_i^\omega(a)|$, and let $\overline{g} = \sum_{i \in I} \overline{g}_i$. Also let $\overline{\pi}$ be the minimum of $\pi^\omega(y, \tilde{\omega}|a)$ over all $(\omega, \tilde{\omega}, a, y)$ such that $\pi^\omega(y, \tilde{\omega}|a) > 0$.

Our formulation encompasses the following examples:

- *Stochastic games with observable states*. Let $Y = \Omega \times \Omega \times Y_A$, and suppose that $\pi^\omega(y, \tilde{\omega}|a) = 0$ for $y = (y_1, y_2, y_A)$ such that $y_1 \neq \omega$ or $y_2 \neq \tilde{\omega}$. That is, the first component of the signal y reveals the current state, and the second component reveals the next state. The third component is a noisy signal about actions. Since the signal in the previous period perfectly reveals the current state, players know the state ω^t *before* they choose an action profile a^t. Also, the stage-game payoff $u_i(a_i, y)$ directly depends on the current state through the first component y_1 of the signal. This is exactly the standard stochastic games studied in the literature.

- *Delayed observation.* Let $Y = \Omega \times Y_A$, and assume that $\pi_Y^\omega(y|a) = 1$ for each $y = (y_\Omega, y_A)$ such that $y_\Omega \neq \omega$. That is, the first component of the current signal reveals the current state. The second component is a noisy signal about actions. This is the case in which players observe the current state *after* they choose their actions.

In the infinite-horizon stochastic game, players have a common discount factor $\delta \in (0,1)$. Let $(\omega^\tau, a^\tau, y^\tau)$ be the state, the action profile, and the public signal in period τ. Player i's history up to period $t \geq 1$ is $h_i^t = (a_i^\tau, y^\tau)_{\tau=1}^t$. Let H_i^t denote the set of all h^t, and let $H_i^0 = \{\emptyset\}$. A public history up to period $t \geq 1$ is denoted by $h^t = (y^\tau)_{\tau=1}^t$. Let H^t denote the set of all h^t, let $H^0 = \{\emptyset\}$, and let $H = \bigcup_{t=0}^\infty H^t$ be the set of all public histories. A strategy for player i is a mapping $s_i : \bigcup_{t=0}^\infty H_i^t \to \triangle A_i$. Let S_i be the set of all strategies for player i, and let $S = \times_{i \in I} S_i$. For each strategy s_i, let $s_i|_{h_i^t}$ be the continuation strategy induced by s_i after history h_i^t.

A strategy s_i is *public* if it depends only on public information, i.e., $s_i(h_i^t) = s_i(\tilde{h}_i^t)$ for t, h_i^t, and $\tilde{h}_i^t$ such that $y^\tau = \tilde{y}^\tau$ for all τ. A strategy profile s is *public* if s_i is public for all i. For each public strategy s_i, let $s_i|_{h^t}$ be the continuation strategy induced by s_i after public history h^t. Similarly, $s|_{h^t}$ denotes the continuation strategy profile after public history h^t.

Let $v_i^\mu(\delta, s)$ denote player i's average payoff in the stochastic game when the initial prior is μ, the discount factor is δ, and players play the strategy profile s. Let $v^\mu(\delta, s) = (v_i^\mu(\delta, s))_{i \in I}$ be the payoff vector achieved by the strategy profile s, given μ and δ. We write $v_i^\omega(\delta, s)$ and $v^\omega(\delta, s)$ instead of $v_i^\mu(\delta, s)$ and $v^\mu(\delta, s)$, when the initial prior μ puts probability one on the state ω. As Yamamoto (2016) shows, for each initial prior μ, discount factor δ, and public strategy s_{-i}, player i's best reply s_i exists. A strategy profile s is a *Nash equilibrium* for an initial prior μ if $v_i^\mu(\delta, s) \geq v_i^\mu(\delta, \tilde{s}_i, s_{-i})$ for all i and $\tilde{s}_i$. Also, a strategy profile is an *ex-post equilibrium* if it is a Nash equilibrium for all μ. As Sekiguchi (1997) shows, under the full support assumption (which will be stated in the next subsection), the difference between Nash and sequential equilibria is not essential. Indeed, given an initial prior μ, if a payoff v is achieved by some ex-post equilibrium s, there is a sequential equilibrium $\tilde{s}$ which achieves the same payoff v.

In what follows, we assume that the function π has a full support:

Definition 1. The *full support assumption* holds if $\pi^{\omega}(y,\tilde{\omega}|a) > 0$ for all ω, $\tilde{\omega}$, a, and y.

The full support assumption requires that regardless of the current state ω and the current action profile a, any signal y can be observed and any state $\tilde{\omega}$ can realize tomorrow. Under this assumption, any public history h^t can happen with positive probability. Also, since any state can happen with positive probability, in any period $t > 1$, a player's posterior belief about the state is always interior, i.e., she assigns at least probability $\overline{\pi}$ on any state ω.

The full support assumption is imposed only for the sake of exposition. In Appendix D, we show that our result remains valid even if the full support assumption is replaced with a weaker condition. In particular, we show that the folk theorem holds in the examples presented above.

2.2 Belief Convergence Theorem

As noted in the introduction, since actions are private information, players' beliefs can possibly diverge in our model. In this subsection, we present the *belief convergence theorem*, which shows that the current belief has only a negligible impact on the posterior belief after a sufficiently long history h^t. This result implies that if players play pure strategies and do not deviate, their private beliefs will eventually merge.

Formally, given a pure public strategy s_i, a public strategy s_{-i}, and an initial prior μ, let $\mu_i(h^t|\mu,s) \in \triangle\Omega$ denote player i's belief about the state ω^{t+1} in period $t+1$ after the public history h^t. That is, $\mu_i(h^t|\mu,s)$ is the posterior belief when no one deviates from the strategy profile s. This belief is well-defined under the full support assumption, because all public histories can appear with positive probability on the path. The following is the belief convergence theorem; the proof relies on weak ergodicity of inhomogeneous Markov matrices, see Appendix B.

Proposition 1. *Suppose that the full support assumption holds, and let* $\beta = 1 - \frac{\overline{\pi}}{|\Omega|} \in (0,1)$. *Then for each i, pure public strategy* s_i, *public strategy* s_{-i}, $t \geq 0$, h^t,

μ, and $\tilde{\mu}$, we have

$$\left|\mu_i(h^t|\mu,s)-\mu_i(h^t|\tilde{\mu},s)\right|\leq\beta^t.$$

To interpret this result, pick a strategy profile s as stated, and pick an arbitrary public history h^t. In general, given this history h^t, different initial priors μ and $\tilde{\mu}$ induce different posterior beliefs, $\mu_i(h^t|\mu,s)$ and $\mu_i(h^t|\tilde{\mu},s)$. However, the above proposition ensures that these two posterior beliefs get closer as t increases, at a rate at least geometric with parameter β. So the impact of player i's current belief on her posterior belief in a distant future is almost negligible, as shown in Figure 3 in Section 1.1. This ensures that even if the opponents do not know player i's current belief, after a long time, they will eventually obtain very precise information about player i's posterior belief.

The result above relies on the assumption that player i chooses a pure strategy s_i and does not deviate. Indeed, if s_i is a mixed strategy, player i's belief in period t crucially depends on her private information about what actions are actually chosen, and hence the opponents cannot obtain precise information about her posterior belief. Similarly, if player i deviates to other strategy $\tilde{s}_i\neq s_i$. then her posterior belief is $\mu_i(h^t|\mu,\tilde{s}_i,s_{-i})$, which can be quite different from the opponents' expectation $\mu_i(h^t|\mu,s)$. In other words, player i can always possess new private information about the true state by deviating from a prescribed strategy s_i.

The belief convergence theorem does not ensure that two posterior beliefs induced by different public history h^t and $\tilde{h}^t$ will merge. That is, different public histories may yield quite different beliefs even after a long time. In this sense, the belief evolution is path-dependent, and state learning never ends.

3 Example: Stochastic Cournot Competition

To illustrate the key ideas of the paper, in this section, we consider a Cournot example and show how to construct an approximately efficient equilibrium.

3.1 Model

There are two firms, and each firm i produces product i. We consider differentiated products, but these products are quite similar; so the prices of the two products are highly correlated.[8] (For example, think about the price of coffee beans from Brazil and the one from Kenya.) In each period, each firm i chooses quantity a_i. There are three possible values of a_i: $H = 20$ (high), $M = 10$ (middle), or $L = 0$ (low).

There is a persistent demand shock and an i.i.d. demand shock. The persistent demand shock is represented by a hidden state ω, which follows a Markov process. Specifically, the state is either a boom ($\omega = \omega_G$) or a slump ($\omega = \omega_B$), and after each period, the state stays at the current state with probability 0.8. Actions (quantities) do not influence the state evolution. Let $\mu \in (0,1)$ be the probability of ω_G in period one.

Due to the i.i.d. demand shock, the market price is stochastically distributed, conditional on the current economic condition ω and the quantity $a = (a_1, a_2)$. Let $y_i \in \{0,10,20,30,40,50\}$ denote the price of product i, and let $y = (y_1, y_2)$. For each state ω and each quantity a, let $\pi_Y^\omega(\cdot|a)$ denote the distribution of the price vector y over $Y = \{0,10,20,30,40,50\}^2$. We assume that both y_1 and y_2 are publicly observable. The precise specification of π_Y will be given in Appendix A, and here we list only the key properties of π_Y:

- $\pi_Y^\omega(y|a) > 0$ for each ω, a, and y, so the full support assumption holds.
- The distributions $\{(\pi_Y^\omega(y|a))_{y\in Y}\}_{(\omega,a)}$ are linearly independent. This implies that the firms can statistically distinguish (ω, a) through y.
- The expected price $E[y_i|\omega, a] = \sum_{y\in Y} \pi_Y^\omega(y|a) y_i$ conditional on the state $\omega = \omega_G$ is as in the left table below: For each cell, the first component represents the expectation of y_1, and the second is of y_2. Similarly, the expected price conditional on the state $\omega = \omega_B$ is as in the right table.

	L	M	H
L	42, 42	41, 40	23, 22
M	40, 41	23, 23	16, 15
H	22, 23	15, 16	13, 13

	L	M	H
L	36, 36	35, 34	17, 16
M	34, 35	17, 17	10, 9
H	16, 17	9, 10	7, 7

[8]Even when the firms produce homogeneous products, as in Green and Porter (1984), our folk theorem (Proposition 6) applies so that we can construct an approximately efficient equilibrium. However, the equilibrium strategy becomes a bit more complicated in that case. The reason is that when the products are homogeneous, symmetric action profiles never have pairwise full rank, and thus we need to perturb the optimal policy for the signal to be informative about the identity of the deviator. See pages 1020-1021 of FLM for more details.

Since the two firms produce similar products, the expected prices are (almost) determined by the total production a_1+a_2. For example, the three action profiles (H,L), (M,M), and (L,H) induce the same production level $a_1+a_2=20$, and hence result in similar expected prices, 22 or 23, at the good state ω_G (the left table). Also, as one can see from the right table, when the state changes to the bad state ω_B, the expected price drops by 6, compared to the case with the good state ω_G.

For simplicity, we assume that the marginal cost of production is $C=0$ for each firm. Hence firm i's actual profit is $a_i y_i$, and its expected payoff given ω and a is $a_i \cdot E[y_i|\omega,a]$. We normalize the payoff (subtract 200 and then divide by 10) and denote it by $g_i^{\omega}(a)$. That is, let $g_i^{\omega}=\frac{a_i \cdot E[y_i|\omega,a]-200}{10}$. These payoffs $g_i^{\omega}(a)$ are summarized as follows; the left table describes the payoffs for the good state ω^G, and the right table describes the ones for the bad state ω^B.

	L	M	H
L	$-20, -20$	$-20, 20$	$-20, 24$
M	$20, -20$	$3, 3$	$-4, 10$
H	$24, -20$	$10, -4$	$6, 6$

	L	M	H
L	$-20, -20$	$-20, 14$	$-20, 12$
M	$14, -20$	$-3, -3$	$-10, -2$
H	$12, -20$	$-2, -10$	$-6, -6$

As one can see, (H,H) is a Nash equilibrium of the stage game, regardless of ω. Also, "Always (H,H)" is a sequential equilibrium in the infinite-horizon game regardless of the initial prior μ, since the state transition does not depend on actions. The payoff of this equilibrium is $\frac{1}{2}(6,6)+\frac{1}{2}(-6,-6)=(0,0)$ in the limit as $\delta \to 1$, because the time average of the hidden state ω is $\frac{1}{2}$-$\frac{1}{2}$ in the long run.

In this game, the "efficient" action profile (i.e., the action profile which maximizes the total profit of the firms) is (H,H) at the state ω_G, but is (M,M) at the state ω_B. So in order to maximize the total profit, the firms should produce less than the Nash equilibrium quantity when they are pessimistic about the current state of the economy. The next subsection studies this issue in more details.

3.2 Feasible Payoff Set and Optimal Policies

Given the initial prior μ and the discount factor δ, different strategy profiles s yield different payoffs in the infinite-horizon game. The set of all such payoff vectors is the feasible payoff set in our environment. That is, the feasible payoff set given the initial prior μ and the discount factor δ is defined as

$$V^{\mu}(\delta) = \text{co}\{v^{\omega}(\delta, s) | s \in S\}.$$

The welfare-maximizing point in this set $V^{\mu}(\delta)$ can be computed by dynamic programming. To see this, note that the welfare-maximizing point must be achievable by a pure strategy, as mixed strategies can achieve only a convex combination of pure-strategy payoffs. When the firms use a pure strategy profile and do not deviate, they do not have private information, so the posterior belief is common after each period. Thus the maximal welfare must be achieved by a pure Markovian strategy profile in which the posterior belief μ^t is a common state variable. This implies that the maximal welfare must solve the following Bellman equation: Let $f(\mu)$ be the maximal welfare given the initial prior μ, and let $\tilde{\mu}(y|\mu,a)$ be the posterior belief in period two given that the initial prior is μ and the outcome in period one is (a,y). Then the function f must solve

$$f(\mu) = \max_{a \in A} \left[(1-\delta)(g_1^{\mu}(a) + g_2^{\mu}(a)) + \delta \sum_{y \in Y} \pi_Y^{\mu}(y|a) f(\tilde{\mu}(y|\mu,a)) \right] \qquad (1)$$

Intuitively, (1) asserts that the maximal welfare $f(\mu)$ is the sum of today's profit $g_1^{\mu}(a) + g_2^{\mu}(a)$ and the expectation of the future profits $f(\tilde{\mu}(y|\mu,a))$. The current action should maximize this sum, and hence we take the maximum with respect to a.

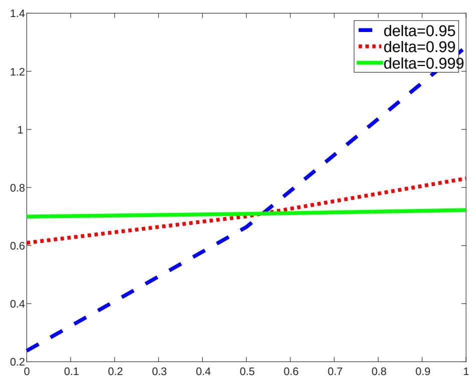

Figure 4: Value Functions for High δ
x-axis: belief μ. y-axis: payoffs.

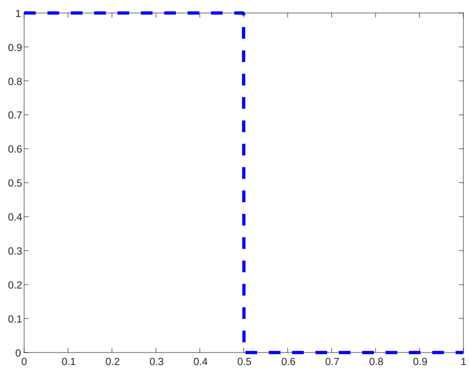

Figure 5: Optimal Policy
x-axis: belief μ. y-axis: actions.

For each discount factor $\delta \in (0,1)$, we can derive an approximate solution to (1) by value function iteration with a discretized belief space. Figure 4 illustrates how the value function f changes when the firms become more patient; it describes the value functions for $\delta = 0.95$, $\delta = 0.99$, and $\delta = 0.999$. As one can see, the value function becomes almost flat as the discount factor approaches one, that is, the firms' initial prior has almost no impact on the efficient payoff. For δ close to one, the value function (the maximal welfare $f(\mu)$) approximates 0.70 regardless of the initial prior μ.

The optimal policy for $\delta = 0.95$ is described in Figure 5, where 1 in the vertical axis means $a = (M,M)$, and 0 means $a = (H,H)$. It shows that the optimal policy is a simple cut-off strategy, which chooses (M,M) when the current belief μ is less than $\frac{1}{2}$, and (H,H) otherwise.[9] In what follows, let $s^{\text{eff}}(\delta,\mu)$ denote the optimal policy given δ and μ. That is, $s^{\text{eff}}(\delta,\mu)$ is the strategy for the infinite-horizon game which achieves the efficient payoff within the feasible payoff set, given the discount factor δ and the initial prior μ. Without loss of generality, we assume that the optimal policy $s^{\text{eff}}(\delta,\mu)$ is a pure public strategy profile.

Using a similar technique, we can compute other extreme points of the feasible payoff set $V^{\mu}(\delta)$. For example, the highest payoff for firm 1 within the feasible payoff set can be computed by solving

$$\tilde{f}(\mu) = \max_{a \in A}\left[(1-\delta)g_1^{\mu}(a) + \delta \sum_{y \in Y} \pi_Y^{\mu}(y|a)\tilde{f}(\tilde{\mu}(y|\mu,a))\right].$$

Again, we can derive an approximate solution using value function iteration. It turns out that when δ is close to one, the value function is almost flat and approximates 18.2 regardless of the initial prior μ. The optimal policy is again a cut-off strategy; it chooses (M,L) when $\mu \leq \frac{1}{3}$, and (H,L) when $\mu > \frac{2}{3}$, regardless of δ. Let $s^1(\delta,\mu)$ denote the optimal policy given δ and μ. That is, $s^1(\delta,\mu)$ is the strategy for the infinite-horizon game which maximizes firm 1's payoff. Similarly, let $s^2(\delta,\mu)$ denote the strategy which maximizes firm 2's payoff.

To summarize, when δ is close to one, the initial prior does not influence the maximal welfare or the highest payoff for each firm. More generally, our

[9] Note that this optimal policy is identical with the myopic policy, which maximizes the stage-game payoff each period. This follows from the fact that in this example, the distribution of the belief tomorrow does not depend on the current action profile a, as explained in Appendix A. Our equilibrium construction does not rely on this property, and is valid in more general environments.

Proposition 2 shows that the feasible payoff set does not depend on the initial prior μ, in the limit as δ goes to one. Let V denote this limit feasible payoff set, and let V^* be the set of all payoffs in V which Pareto-dominates the trivial equilibrium payoff, $(0,0)$. Figure 6 describes (a subset of) the limit feasible payoff set. Here the profit maximizing point is $(0.35, 0.35)$, because the value function f approximates 0.70 as δ goes to one. Also the down-right corner is $(18.2, -20)$ because the value function $\tilde{f}$ approximates 18.2, and the optimal policy $s^1(\delta, \mu)$ asks firm 2 to play L forever, which yields -20 each period. Figure 6 is only a subset of V, because we have not computed other extreme points of V. Figure 7 describes (a subset of) the feasible and individually rational payoff set V^*. In what follows, we will construct an equilibrium which approximates the efficient payoff vector $(0.35, 0.35)$.

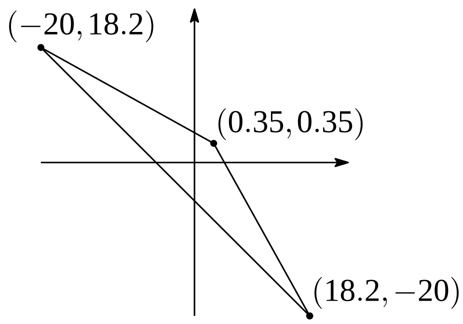

Figure 6: Subset of V

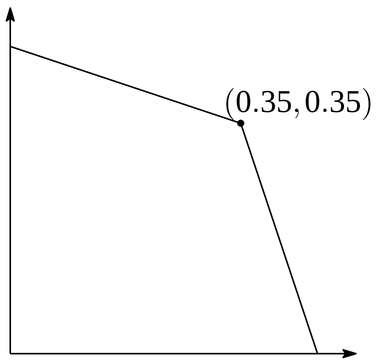

Figure 7: Subset of V^*

3.3 Pseudo-Ergodic Strategies

As explained, when the firms have a common belief μ, the efficient payoff vector $(0.35, 0.35)$ is (approximately) achieved if the firms coordinate and play the optimal policy $s^{\text{eff}}(\delta, \mu)$. However, in our model, the firms may not have a common belief, which causes a possible miscoordination problem in the following sense: Suppose that each firm i's current belief is μ_i where $\mu_1 \neq \mu_2$, and that these beliefs are private information. If each firm chooses the optimal policy corresponding to its own belief, then the resulting profile $(s_1^{\text{eff}}(\delta, \mu_1), s_2^{\text{eff}}(\delta, \mu_2))$ is quite different from the optimal policy.

To avoid this miscoordination problem, in our equilibrium, the firms play "pseudo-ergodic strategies" which do not depend on the private beliefs. For example, if the firms want to cooperate and approximate $(0.35, 0.35)$ from now on, they form a "dummy public belief" $\tilde{\mu} = \frac{1}{2}$ and play the corresponding optimal policy $s^{\text{eff}}(\delta, \frac{1}{2})$. By the definition, this strategy $s^{\text{eff}}(\delta, \frac{1}{2})$ is not optimal unless the firms' current beliefs are $\mu_1 = \mu_2 = \frac{1}{2}$. However, as Proposition 3 shows, it is approximately optimal for all initial priors μ, that is, it approximates the efficient payoff $(0.35, 0.35)$ regardless of the true belief. So if the firms want to cooperate, they may ignore their private beliefs and simply play $s^{\text{eff}}(\delta, \frac{1}{2})$, the optimal policy for the dummy public belief $\tilde{\mu} = \frac{1}{2}$.

Proposition 3 also shows that the same result holds for other optimal policies. For example, the optimal policy $s^1(\delta, \frac{1}{2})$, which achieves the best payoff for firm 1 given the dummy belief $\tilde{\mu} = \frac{1}{2}$, approximates the down-right corner $(18.2, -20)$ of the feasible payoff set regardless of the initial prior μ. So the firms may play it if they want to reward firm 1 (by giving 18.2) while punishing firm 2 (by giving -20). Similarly, the optimal policy $s^2(\delta, \frac{1}{2})$ for the dummy belief $\frac{1}{2}$ approximates $(-20, 18.2)$ regardless of the initial prior μ. This strategy can be used when the firms want to reward firm 2 while punishing firm 1.

Also, any constant action profile is a pseudo-ergodic strategy in that it achieves approximately the same payoff regardless of the initial belief. For example, if the firms always play (H,H), the payoff $(0,0)$ is achieved in the limit as $\delta \to 1$, regardless of the initial belief. As will be explained, the firms use this "Always (H,H)" when they want to punish each other.

3.4 Random Blocks and Self-Generation

Now we construct an equilibrium approximating the efficient payoff $(0.35, 0.35)$, by assembling the pseudo-ergodic strategies in the previous subsection. In what follows, we assume that public randomization z, which follows the uniform distribution on $[0,1]$, is available at the end of each period.

As in Yamamoto (2016), we regard the infinite horizon as a sequence of *random blocks*, the length of which is determined by public randomization. Specifically, at the end of each period t, the firms check the public randomization z^t. If $z^t \geq p$ for some fixed number $p \in [0,1]$, then the current random block terminates and the new block begins from the next period $t+1$. If $z^t < p$, then the current block does not terminate and the next period $t+1$ is included in the current block. So the probability that the current block terminates is $1-p$ each period, and the length of the block is geometrically distributed. We take p close to one, so the expected length of the block is long.

Due to the random termination probability $1-p$, each random block is payoff-equivalent to the infinite-horizon game with the discount factor $p\delta$. Indeed, given the initial prior μ and the strategy profile s, the unnormalized expected payoff in the first block is $\sum_{t=1}^{\infty}(p\delta)^{t-1}E[g^{\omega^t}(a^t)|\mu,s]$, where p^{t-1} is the probability of the t-th period of the block being actually played. This payoff can be rewritten as $\frac{v^{\mu}(p\delta,s)}{1-p\delta}$, which equals the unnormalized payoff for the infinite-horizon game with the discount factor $p\delta$.

In each random block, the firms choose one of the four pseudo-ergodic strategies: $s^{\text{eff}}(p\delta,\frac{1}{2})$, $s^1(p\delta,\frac{1}{2})$, $s^2(p\delta,\frac{1}{2})$, or "Always (H,H)." On the equilibrium path, the firms do not change the strategy in the middle of the block. That is, once a strategy is selected, they play it until the block ends. (Of course, they can deviate whenever they want.) Since the firms play pseudo-ergodic strategies, their block payoffs are not affected by their beliefs by much. For example, if the firms choose $s^{\text{eff}}(p\delta,\frac{1}{2})$ during the block, the block payoff approximates the efficient payoff $(0,35,0.35)$ regardless of their beliefs. Here the firms use the optimal policy for the discount factor $p\delta$ (rather than δ), because the "effective discount factor" for each random block is $p\delta$, as explained above.

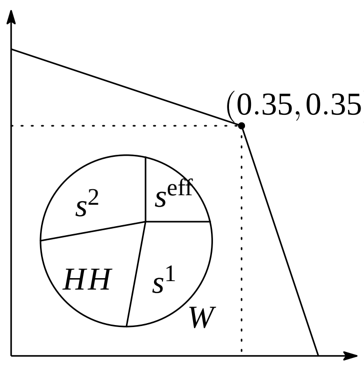

Figure 8: Block Strategies

Take a ball W in the interior of the feasible and individually rational payoff set V^*. For the sake of exposition, we assume that W is in the interior of the rectangle $[0,0.35]\times[0,0.35]$, as in Figure 8. Our goal is to show that this ball W is attainable by public equilibria, when the firms are patient. Thanks to the self-generation theorem (Proposition B1 in Appendix B), it is sufficient to show that the ball W is self-generating. That is, for each target payoff v in the ball W, there is a block strategy s and a continuation payoff function $w : H \to W$ such that

$$v_i = \frac{1-\delta}{1-p\delta} v_i^{\omega}(p\delta, s) + \sum_{t=1}^{\infty}(1-p)p^{t-1}\delta^t E[w_i(h^t)|\omega, s] \tag{2}$$

for all ω and i, and

$$v_i \geq \frac{1-\delta}{1-p\delta} v_i^{\omega}(p\delta, \tilde{s}_i, s_{-i}) + \sum_{t=1}^{\infty}(1-p)p^{t-1}\delta^t E[w_i(h^t)|\omega, \tilde{s}_i, s_{-i}] \tag{3}$$

for all ω, i, and $\tilde{s}_i$.

(2) is the promise-keeping condition, which implies that regardless of the initial state ω, the target payoff v is exactly achieved as the sum of the block payoff by the strategy s and the continuation payoff w chosen from the ball W. Indeed, the first term in the right-hand side is the block payoff by s, and the second term is the expected continuation payoff. More precisely, $\frac{v_i^{\omega}(p\delta, s^v)}{1-p\delta}$ in the first term is the unnormalized payoff during the block, and we multiply $1-\delta$ because we consider the average payoff with the discount factor δ. $(1-p)p^{t-1}$ in the second term denotes the probability of the current block terminating after period t, and $w_i(h^t)$ is the continuation payoff in that case.

(3) is the incentive compatibility condition, which ensures that any deviation from the block strategy s is not profitable, regardless of the initial state ω. This ensures that the block strategy s is an ex-post equilibrium in the random block, that is, s is a Nash equilibrium even if the initial state ω is revealed.

To show that the ball W is indeed self-generating, for each target payoff $v \in W$, we need to find a block strategy s and a continuation payoff function w which satisfy (2) and (3). We choose the strategy s as in Figure 8; here the ball W is divided into four parts, depending on the corresponding block strategy s. For example, if the target payoff v is in the top-right part, we let $s = s^{\text{eff}}$.

What remains is to show that for each $v \in W$, there are continuation payoffs w which indeed satisfy (2) and (3). We will work on this in the next two subsections.

3.5 Regular Directions

Choose v as in Figure 9. That is, let v be a boundary point of the ball W with the unit normal $\lambda = (\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}})$. Then take the block strategy s as in Figure 8; since v is in the top-right part, we let $s = s^{\text{eff}}$, whose block payoff approximates the efficient payoff $(0.35, 0.35)$ regardless of the initial prior μ. In what follows, we will explain that there is a continuation payoff function w satisfying (2) and (3).

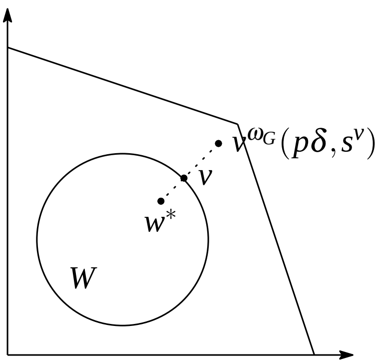

Figure 9: Choice of w^*

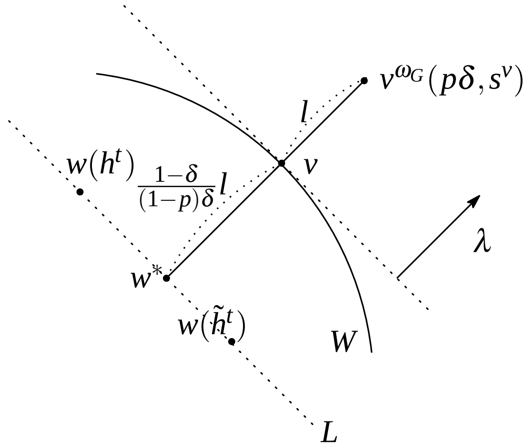

Figure 10: Zoomed Picture

As a first step, we construct w which satisfies the promise-keeping condition (2) for state ω_G. Suppose that the function w is constant, i.e., $w(h^t) = w^*$ for all h^t, where $w^* \in \boldsymbol{R}^N$ is a constant. Then since $\sum_{t=1}^{\infty}(1-p)p^{t-1}\delta^t = \frac{(1-p)\delta}{1-p\delta} = 1 - \frac{1-\delta}{1-p\delta}$, the promise-keeping condition (2) for state ω_G is rewritten as

$$v = \frac{1-\delta}{1-p\delta} v^{\omega_G}(p\delta, s) + \left(1 - \frac{1-\delta}{1-p\delta}\right) w^*. \tag{4}$$

Choose the constant w^* so that the above equality holds, that is, let

$$w^* = v - \frac{1-\delta}{(1-p)\delta}(v^{\omega_G}(p\delta, s) - v).$$

Intuitively, w^* is chosen so that the target payoff v is exactly achieved as the sum of the block payoff $v^{\omega_G}(p\delta, s^v)$ and of the continuation payoff w^*. (See Figure 9.) Pick p close to one, and then choose δ close to one. Then $\frac{1-\delta}{(1-p)\delta}$ is close to zero, and thus w^* is in the interior of the ball W, as in Figure 9.

By the definition, this constant function $w(h^t) = w^*$ satisfies the promise-keeping condition (2) for the good state ω_G. However, it does not satisfy the promise-keeping condition (2) for the bad state ω_B. (This is precisely the problem (a) discussed in Section 1.1; we need to achieve the same payoff v regardless of the initial state ω.) Also, it does not satisfy the incentive compatibility condition (3). So we need to modify the constant function in such a way that these constraints are satisfied.

To satisfy the promise-keeping condition (2) for the bad state ω_B, we only need to perturb the continuation payoff a bit. To see why, recall that the block strategy s^{eff} approximates the payoff $(0.35, 0.35)$ regardless of the initial state ω. This implies that the initial state influences the term $v_i^{\omega}(p\delta, s)$ in the right-hand side of (2) only by a small amount. To offset this payoff difference, we only need a small perturbation of the continuation payoff. For more details, see Lemma B16 in Appendix B.

To satisfy the incentive compatibility condition (3), we borrow the idea of "utility transfer across players" of FLM. Roughly, if the public history h^t during the current block indicates firm 1's deviation, then we change the continuation payoff from w^* to $w(h^t)$ in Figure 10. By doing so, we punish firm 1 by reducing the continuation payoff (relative to w^*), while we reward firm 2 by increasing the continuation payoff. Likewise, if the public history $\tilde{h}^t$ indicates firm 2's deviation, we use $w(\tilde{h}^t)$ in the figure as the continuation payoff. This punishment mechanism does not involve value burning, that is, regardless of the realization of h^t, the continuation payoff is always on the line L in Figure 10, so the sum of the firms' profits is constant. This property ensures that we can avoid inefficiency, even though punishment occurs on the equilibrium path due to imperfect monitoring of actions.

In the rest of this subsection, we formally show that the idea of utility transfers indeed works in our environment. Since this is an analogue of FLM, readers who are not interested in details may skip this part and go to Section 3.6.

To satisfy (3), we consider the function $w : H \to W$ with the form

$$w(h^t) = w^* + z^t(h^t) \tag{5}$$

for each t and h^t. That is, we add a perturbation term z^t to the constant payoff w^*; the superscript t on the perturbation term z represents the time at which the block terminates.

For each t, we will choose the perturbation term z^t carefully so that any deviation in period t is not profitable regardless of the past history h^{t-1} and the current state ω^t. Specifically, note that $z^t(h^t)$ can be written as $z^t(h^{t-1},y)$, where y is the public signal in period t. Then given t and h^{t-1}, we choose $(z^t(h^{t-1},y))_{y\in Y}$ such that

$$\delta(1-p)\sum_{y\in Y}\pi_Y^{\omega}(y|a_i,s_{-i}(h^{t-1}))z_i^t(h^{t-1},y)=\begin{cases}0 & \text{if } a_i=s_i(h^{t-1})\\ -\frac{1-\delta}{1-p\delta}\overline{g} & \text{otherwise}\end{cases} \tag{6}$$

for all ω, i, and a_i, and

$$z_1^t(h^{t-1},y)+z_2^t(h^{t-1},y)=0 \tag{7}$$

for all y. The first condition (6) ensures that any one-shot deviation in period t after the public history h^{t-1} is not profitable, regardless of the current state ω^t. To see this, suppose that we are now in period t of the block and that the past public history is h^{t-1}. If firm i deviates in the current period t, it influences the distribution of the public signal y today, and hence the expected value of the perturbation term $z_i^t(h^{t-1},y)$. The left-hand side of (6) measures how much this influences firm i's stochastic-game payoff, evaluated at period t; the term $1-p$ is the probability that the block terminates right after the current period t, in which case the continuation payoff is indeed w^*+z^t. (6) asserts that this effect is $\frac{1-\delta}{1-p\delta}\overline{g}$, which is large enough to deter firm i's one-shot deviation in period t. Indeed, any deviation cannot increase the block payoff by more than $\frac{1-\delta}{1-p\delta}\overline{g}$, so the gain by such a deviation is less than the loss.

The second condition (7) asserts that the modified continuation payoff $w(h^t)=w^*+z^t(h^t)$ is on the dotted line L in Figure 10, which is a translate of the tangent line. We can show that the above system of equations (6) and (7) indeed has a solution z^t. Also, (7) ensures that the resulting continuation payoffs $w(h^t)=w^*+z^t(h^t)$ are in the ball W. (The proof is very similar to that of FLM, and hence omitted.)

We have explained that the above perturbation term z^t ensures ex-post incentive compatibility in period t. That is, any deviation in period t is not profitable, even if the state ω^t is revealed. Since we choose such z^t for each t, ex-post incentive compatibility is satisfied each period, and thus the incentive compatibility condition (3) indeed holds. (Here, we can ignore the possibility that a deviation in period t influences the perturbation term $z^{\tilde{t}}$ for $\tilde{t}\neq t$. This is because $z^{\tilde{t}}$ satisfies (6) so that its expected value is zero as long as firm i does not deviate in period $\tilde{t}$.)

3.6 Coordinate Directions and Value Burning

In the previous subsection, we have explained that when v is chosen as in Figure 9, we can find continuation payoffs w satisfying (2) and (3). A similar argument shows that the same result holds for almost all payoffs v in the ball W. The only exceptions are the case in which v is a boundary point of W whose unit normal is a coordinate vector, $\lambda = (1,0)$ or $\lambda = (0,1)$. In this subsection, we will explain how to find w for such a case.

Since the game is symmetric, without loss of generality, we focus on the case in which the unit normal is $\lambda = (0,1)$. That is, we choose v as in Figure 1 in Section 1.1. Choose the block strategy s as in Figure 8, i.e., let $s = s^2$. This block strategy approximates $(-20, 18.2)$ regardless of the initial prior μ.

As in the previous subsection, take the constant continuation payoff $w(h^t) = w^*$ so that the promise-keeping condition (2) holds for the good state ω_G. This constant function does not satisfy the promise-keeping condition (2) for ω_B or the incentive compatibility condition (3), so we need to modify it. As explained in the previous subsection, the promise-keeping condition (2) for ω_B can be satisfied by perturbing the continuation payoff a bit.

How about the incentive compatibility condition (3)? In the previous subsection, we have shown that it can be satisfied by moving continuation payoffs on the line L in Figure 10, which is a translate of the tangent line. Unfortunately, this idea does not extend here: When firm 2's belief is $\mu \neq \frac{1}{2}$, the block strategy s^2 does not maximize firm 2's block payoff, and thus firm 2 can improve its block payoff by deviating from s^2. For (3) to hold, we need to deter such a deviation via a variation in continuation payoffs, so the continuation payoff w must move vertically. This means that w cannot be on the line L in Figure 1. Note that this is precisely the problem (b) discussed in Section 1.1.

In what follows, we show that we can indeed deter firm 2's deviation via vertical move of the continuation payoffs w, while keeping w in the ball W. (We focus only on firm 2's problem, because firm 1's incentive condition can be easily satisfied using the horizontal move.) Formally, we show that there is firm 2's continuation payoff w_2 which satisfies the incentive compatibility condition (3) and

$$\left|w_2(h^t)-w_2^*\right| \leq \frac{(1-\delta)p}{2(1-p\delta)}l \tag{8}$$

for all t and h^t, where $l = v_2^{\omega_G}(p\delta, s^v) - v_2$. Intuitively, (8) ensures that the vertical move of the continuation payoff, which is measured by $|w_2(h^t)-w_2^*|$, is small and less than a half of the length D in Figure 2. Indeed, the y-intercept of the line L in Figure 2 is w_2^*, so we have $D = v_2 - w_2^* = \frac{(1-\delta)p}{1-p\delta}l$. As shown in Lemma B8 in Appendix B, (8) ensures that the vertical move of w is small enough w never goes to the outside of the ball W.

The proof consists of two steps. In the first step, we show that firm 2's gain G^t by deviating in period t of the block is small for each t, and that even the *infinite sum* of the gains, $\sum_{t=1}^{\infty} G^t$, is small. Note that for this to be the case, we need to show that the gains G^t for later periods t converge to zero as t increases; indeed, if G^t is small but constant (e.g., $G^1 = G^2 = \cdots = \varepsilon$), the sum becomes infinitely large. In the second step, we construct continuation payoffs w such that the incentive condition (3) holds. The size of the vertical move of this continuation payoff w is proportional to the sum of the gains, $\sum_{t=1}^{\infty} G^t$; this is so because we need to deter firm 2's deviation in *all* periods in the block. From the first step, we know that the sum of the gains is small, and thus the vertical move of w is small. Hence this w indeed satisfies the desired inequality (8).

To begin, we define firm 2's gain G^t when it deviates in period t of the block. Consider a random block with firm 2's initial belief μ. Suppose that we are now in period t of the block, and the public history within the block is h^{t-1}. Suppose that firm 2 has not deviated within the block so far, so its posterior belief is $\mu_2(h^{t-1}|\mu, s^2)$. If firm 2 deviates in the continuation game, it can improve the block payoff by

$$G^t(\mu, h^{t-1}) = \max_{s_2 \in S_2} v^{\mu_2(h^{t-1}|\mu, s^2)}(p\delta, s_1^2|_{h^{t-1}}, s_2) - v^{\mu_2(h^{t-1}|\mu)}(p\delta, s^2|_{h^{t-1}}).$$

Here the first term of the right-hand side is the payoff in the continuation game when firm 2 deviates, while the second term is the one when firm 2 does not deviate. Note that we take the maximum over all s_2, so we allow firm 2 to deviate not only in period t, but in later periods; so G^t should be interpreted as the gain by firm 2 when "firm 2 follows the strategy s^2 until period $t-1$, but then deviates in period t and then choose a best reply after that." As one can see from the formula above, the initial belief μ influences the gain $G^t(\mu, h^{t-1})$ through the posterior belief $\mu_2(h^{t-1}|\mu, s^2)$. When the initial belief is $\mu = \frac{1}{2}$, we have $G^t(\mu, h^{t-1}) = 0$, because s^2 is the optimal policy for this belief and firm 2 has no reason to deviate.

Let $G^t = \max_{\mu \in \triangle\Omega} \max_{h^{t-1} \in H^{t-1}} G^t(\mu, h^{t-1})$ denote the maximal possible gain in the continuation game from period t on, and let $\beta < 1$ be as in Proposition 1. We claim that for any small $\varepsilon > 0$, if p and δ are large enough (i.e., if the expected block length is large and players are patient), we have

$$G^t < \beta^{t-1}\varepsilon \tag{9}$$

for each t. That is, the gain G^1 by deviating in period one is small, and the gain G^t by deviating in a later period t is even smaller and converges to zero at a rate at least geometric with parameter β. This inequality ensures that the sum of the gains is small; indeed, summing both sides over t, we obtain $\sum_{t=1}^{\infty} G^t = \frac{\varepsilon}{1-\beta}$. Since ε can be arbitrarily small, the infinite sum $\sum_{t=1}^{\infty} G^t$ is also small.

It is easy to see that (9) holds for $t = 1$. Indeed, since s^2 approximates the best payoff for firm 2 regardless of the initial belief μ, the gain by deviating from s^2 in period one must be small. The proof for $t > 1$ is more involved, but the idea is roughly as follows. The initial belief μ influences the gain $G^t(\mu, h^{t-1})$ through the posterior belief $\mu_2(h^{t-1}|\mu, s^2)$, but the belief convergence theorem implies that this posterior does not depend on the initial belief by much, after a long history. Thus all initial beliefs induce the same gain asymptotically, that is, for any initial beliefs μ and $\tilde{\mu}$, the difference in gains, $|G^t(\mu, h^{t-1}) - G^t(\tilde{\mu}, h^{t-1})|$, converges to zero as $t \to \infty$. Then, plugging $G^t(\tilde{\mu}, h^{t-1}) = 0$ for $\tilde{\mu} = \frac{1}{2}$, we can conclude that the gain $G^t(\mu, h^{t-1})$ converges to zero as $t \to \infty$, for all initial beliefs μ. We can also show that the rate of convergence is at least geometric, see Lemma B9 in Appendix B for the formal proof.

Now we proceed to the second step of the proof; we construct continuation payoffs w which deters firm 2's deviation from s^2 so that the incentive condition (3) holds. In what follows, we focus on firm 2's incentive only, so we omit the subscript 2 to simplify the notation. Consider firm 2's continuation continuation payoff w with the following form:

$$w(h^1) = w^* + z^1(h^1),$$
$$w(h^2) = w^* + z^1(h^1) + z^2(h^2),$$
$$w(h^3) = w^* + z^1(h^1) + z^2(h^2) + z^3(h^3),$$

and so on. In words, if the random block ends in period one, firm 2's continuation payoff from the next block is the constant term w^* plus a perturbation term $z^1(h^1)$, which will be defined later. If the block ends in period two, we have an additional perturbation term $z^2(h^2)$. In this way, we have more perturbation terms when the block terminates in a later period. For each t, we will carefully choose z^t so that a particular form of deviations is not profitable. Specifically, we choose z^t such that "follow s^2 until period $t-1$ ends, then deviate in period t, and then play a best reply after that" is not profitable regardless of the initial state ω^1. If we can find such z^t for each t, deviating to any block strategy $s_2 \neq s_2^2$ is not profitable regardless of the initial state, and thus (3) holds.

In order to find such z^t, note that $z^t(h^t)$ can be written as $z^t(h^{t-1}, y)$, where y is the signal in period t. Then for each t and h^{t-1}, we choose the perturbation terms $(z^t(h^{t-1}, y))_{y \in Y}$ such that

$$\sum_{\tilde{t}=0}^{\infty} \delta^{\tilde{t}+1} p^{\tilde{t}}(1-p) \sum_{y \in Y} \pi_Y^{\omega}(y|a_i, s_{-i}(h^{t-1})) z^t(h^{t-1}, y) = \begin{cases} 0 & \text{if } a_2 = s_2(h^{t-1}) \\ -\frac{1-\delta}{1-p\delta} G^t & \text{otherwise} \end{cases} \quad (10)$$

for each ω and a_2. To interpret this condition, suppose that we are now in period t of the block, and that the past public history is h^{t-1}. Suppose that no one has deviated from s^2 so far. If firm 2 deviates in the current period t, it influences the public signal y in period t and hence the expected value of $z^t(h^{t-1}, y)$. The left-hand side of (10) measures the expected discounted value of this change, evaluated at period t; here the term $p^{\tilde{t}}(1-p)$ is the probability that the block terminates at the end of period $t+\tilde{t}$, and we take the expectation with respect to the termination date $t+\tilde{t}$. (10) asserts that this effect is large enough that "follow s^2 until period $t-1$ ends, then deviate in period t, and then play a best reply thereafter" is not profitable regardless of the initial state ω^1. Indeed, the gain by such a deviation is at most $\frac{1-\delta}{1-p\delta} G^t(\mu, h^{t-1})$, which is less than the loss $\frac{1-\delta}{1-p\delta} G^t$.

What remains is to show that the above perturbation is small enough that (8) holds. Note that $\sum_{\tilde{t}=0}^{\infty} \delta^{\tilde{t}+1} p^{\tilde{t}}(1-p) = \frac{(1-p)\delta}{1-p\delta}$. Plugging this into (10) and dividing both sides by $\frac{(1-p)\delta}{1-p\delta}$, we get

$$\sum_{y \in Y} \pi_Y^{\omega}(y|a_i, s_{-i}(h^t)) z^t(h^{t-1}, y) = \begin{cases} 0 & \text{if } a_2 = s_2(h^{t-1}) \\ -\frac{1-\delta}{(1-p)\delta} G^t & \text{otherwise} \end{cases} .$$

Thus there is some $C > 0$ such that $|z^t(h^t)| \leq \frac{1-\delta}{(1-p)\delta} C G^t$ for each t and h^t. That is, the perturbation term z^t is proportional to G^t. Hence

$$|z^1(h^1)| + |z^2(h^2)| + \cdots \leq \frac{1-\delta}{(1-p)\delta} C G^*,$$

where $G^* = \sum_{t=1}^{\infty} G^t$ is the sum of the gains. This shows that the maximal size of the perturbation is proportional to the sum G^* of the gains. As explained, when p and δ are large enough, the sum G^* can be arbitrarily small, and in particular $CG^* < \frac{l}{2}$. This implies (8), as desired.

3.7 Comments and Remarks

3.7.1 State-Specific Punishment vs Uniform Punishment

Our promise-keeping condition (2) requires that the same payoff v be achieved regardless of the hidden state ω, using the same continuation payoff function w. Accordingly, in our equilibrium, after every block, each player's payoff in the continuation payoff is independent of the hidden state ω. This in particular implies that our punishment mechanism is *not* state-specific, in the following sense: Pick an equilibrium strategy profile, and suppose that after some history h^{t-1}, a player's payoff in the continuation payoff is low conditional on some state ω^t. Then in this continuation game, her payoff must be the same (low) value even if the state were $\tilde{\omega}^t \neq \omega^t$. So after this history h^{t-1}, she is punished not only at a particular state ω^t, but at all states uniformly.

In contrast, Fudenberg and Yamamoto (2010) consider ex-post equilibria with state-specific punishments. Specifically, they take a ball W from the *extended space* $\boldsymbol{R}^{|\Omega| \times N}$ (rather than $\boldsymbol{R}^N$), and choose the target payoff $v = (v_i^{\omega})_{(i,\omega)}$ and the continuation payoff $w = (w_i^{\omega})_{(i,\omega)}$ from this ball. In this framework, both v and w can directly depend on the state ω, so in their equilibrium, a player's payoff depends on ω after every history. In particular, since w is chosen from the extended space $\boldsymbol{R}^{|\Omega|\times N}$, it is possible to lower player i's continuation payoff w_i^{ω} at state ω, while not affecting her payoff $w_i^{\tilde{\omega}}$ at other states $\tilde{\omega} \neq \omega$. In this sense, their punishment is state-specific. Fudenberg and Yamamoto (2010) show that such state-specific punishments are indeed useful to provide ex-post incentives: They show that ex-post incentives can be provided by moving continuation payoffs only on a translate on the tangent hyperplane in the extended space $\boldsymbol{R}^{|\Omega|\times N}$, so value burning does not occur in their equilibrium.[10]

What will happen if we consider equilibria with state-specific punishments in our environment? To answer this question, take a ball W from the extended space $\boldsymbol{R}^{|\Omega|\times N}$ as in Fudenberg and Yamamoto (2010). The self-generation theorem still holds, that is, this ball W is supported by public equilibria if for each payoff $v \in W$, there is a block strategy s and a continuation payoff $w : H \to W$ such that

$$v_i^{\omega} = \frac{1-\delta}{1-p\delta} v_i^{\omega}(p\delta, s) + \sum_{t=1}^{\infty}(1-p)p^{t-1}\delta^t E[w_i^{\omega^{t+1}}(h^t)|\omega, s] \tag{11}$$

for all ω and i, and

$$v_i^{\omega} \geq \frac{1-\delta}{1-p\delta} v_i^{\omega}(p\delta, \tilde{s}_i, s_{-i}) + \sum_{t=1}^{\infty}(1-p)p^{t-1}\delta^t E[w_i^{\omega^{t+1}}(h^t)|\omega, \tilde{s}_i, s_{-i}], \tag{12}$$

for all ω, i, and $\tilde{s}_i$. The difference from (2) and (3) is that both v and w directly depend on the state ω, which allows us to use state-specific punishments. However, it turns out that when the state changes over time, the self-generation conditions (11) and (12) are intractable, which makes it difficult to characterize the equilibrium payoff set.

[10] To see how the things works, suppose that there are two players ($N=2$) and two states ($|\Omega| = 2$), so that the extended space is $\boldsymbol{R}^{|\Omega|\times N} = \boldsymbol{R}^4$. Take the unit ball $W = \{v \in \boldsymbol{R}^4 | |v| \leq 1\}$. Then take the payoff $v = (1,0,0,0)$ so that its unit normal is a coordinate vector. This payoff v yields the highest payoff to player 1 within the ball W (hence she is rewarded) at state ω_1, but not at state ω_2. So her continuation payoff $w_1^{\omega_2}$ at state ω_2 can be both higher and lower than the payoff v, and in this sense the choice of the continuation payoff is flexible. Indeed, if we move continuation payoffs w on a translate of the tangent hyperplane in the extended space $\boldsymbol{R}^4$, player 1's continuation payoff $w_1^{\omega_1}$ at state ω_1 must be constant, but her continuation payoff $w_1^{\omega_2}$ at state ω_2 can take arbitrary values. This helps to provide ex-post incentives at state ω_2. In contrast, in our setup, if we take the payoff v as in in Figure 1 in Section 1.1, it achieves the best payoff for player 2 within the ball W at *all* states simultaneously. Accordingly, if we move continuation payoffs w on the line L in the figure, player 2's continuation payoff must be constant at all states.

To illustrate this issue, note first that in the standard repeated game (in which there is no hidden state ω), adding a constant to continuation payoffs does not change players' incentives. Formally, suppose that an action profile α is *enforceable* by some continuation payoff function $w : Y \to \boldsymbol{R}^N$, in that no one has a profitable deviation from α. Then the profile α is still enforceable even if we add a constant $c \in \boldsymbol{R}^N$ to the continuation payoff, i.e., α is enforceable by the continuation payoff $\tilde{w}$ such that $\tilde{w}(y) = w(y) + c$ for all y. This property is used to derive various useful results in the literature, such as the linear programming characterization of Fudenberg and Levine (1994).

A similar property still holds in the setup of Fudenberg and Yamamoto (2010). That is, adding a constant to continuation payoffs does not change ex-post incentives, even if there is a hidden *fixed* state ω, and even if we consider the extended space $\boldsymbol{R}^{|\Omega| \times N}$. Suppose that α is *ex-post enforceable* by some continuation payoff $w : Y \to \boldsymbol{R}^{|\Omega| \times N}$, in that no one has a profitable deviation from α given any state ω. Then the profile α is still ex-post enforceable even if we consider a modified continuation payoff $\tilde{w}$ such that $\tilde{w}(y) = w(y) + c$, where $c \in \boldsymbol{R}^{|\Omega| \times N}$ is a constant. (The constant term c can specify different values c_i^ω for different states ω, as it is chosen from the extended space.) Using this property, Fudenberg and Yamamoto (2010) show that the linear programming technique of Fudenberg and Levine (1994) remains valid even in their setup.

In contrast, in our model, adding a constant to continuation payoffs *does* influence ex-post incentives. Indeed, when the state changes as time goes, the state today can be possibly different from the state tomorrow, and thus the continuation payoff w_i^ω for state ω appears not only in the incentive condition (12) for ω, but also in the ones for other states $\tilde{\omega} \neq \omega$. In other words, the incentive conditions (12) for (ω, i) and $(\tilde{\omega}, i)$ are *entangled* through the term w_i^ω. This implies that if we add a constant to the continuation payoff w_i^ω for state ω, it influences player i's incentives at *all states* in a complicated way. Due to this problem, Lemma 3 of Fudenberg and Yamamoto (2010) (the linear programming characterization) does not extend, and accordingly it is not clear how to compute the limit equilibrium payoff set in our model.

To avoid this problem, we focus on the payoff space $\boldsymbol{R}^N$, and consider equilibria in which the resulting payoffs are perfectly correlated across states. When we work on this space $\boldsymbol{R}^N$, adding a constant to continuation payoffs does not change ex-post incentives,[11] which allows us to borrow useful repeated-game techniques. Of course, focusing on the space $\boldsymbol{R}^N$ limits flexibility of the way we provide intertemporal incentives. In particular, we need to use value burning in order to provide ex-post incentives, as explained in Section 3.6. However, we find that this value burning does not cause significant inefficiency, and thus we can approximate efficient outcomes without using state-specific punishments.

3.7.2 Perfect Ex-Post IC vs Periodic Ex-Post IC

Our incentive compatibility condition (3) requires that s be a Nash equilibrium in the block even if the initial state ω is revealed, that is, ex-post incentive compatibility must hold in period one. However, it needs not imply ex-post incentive compatibility in any later period $t > 1$ of the block. Indeed, even if (3) holds, there may be a public history h^{t-1} such that the continuation strategy $s|_{h^{t-1}}$ is not a Nash equilibrium (i.e., deviating from $s|_{h^{t-1}}$ is profitable) if the state ω^t in period t is revealed. Accordingly, in our equilibrium, ex-post incentive compatibility holds only periodically; it holds only at the initial period of each block. This is weaker than "perfect ex-post incentive compatibility" of Fudenberg and Yamamoto (2010), which requires ex-post incentive compatibility each period.

More formally, our condition (3) is equivalent to requiring that in each period t of the block, the continuation strategy $s_i|_{h^{t-1}}$ is a best reply for player i if her posterior μ_i^t is chosen from the set $\{\mu_i(h^{t-1}|\mu,s)|\forall\mu\}$, which is represented by the thick line in Figure 3 in Section 1.1. In any period $t > 1$ of the block, this thick line is a strict subset of the whole belief space, and hence ex-post incentive compatibility does not hold in these periods.

Why should we be interested in the condition (3), rather than perfect ex-post incentive compatibility? There are two reasons. First, while our condition (3) is weaker than perfect ex-post incentive compatibility, it is not "too weak," and it still has a nice robustness property. Specifically, (3) ensures that as long as a player has not deviated during the current block, her best reply is not influenced by the belief μ^1 at the beginning of the block, and hence not influenced by what happened before the current block begins. In turns out that this property is sufficient for our

[11] This is formally stated as follows. Suppose that a payoff vector $v \in \boldsymbol{R}^N$ is enforced by some public strategy profile s and some continuation payoff w in the sense of Definition B1. Then $\tilde{v} = v + c$ can be enforced by $\tilde{w} = w + \frac{(1-p)\delta}{1-p\delta}c$, where $c \in \boldsymbol{R}^N$ is a constant.

purpose, in that we can construct a public equilibrium using a recursive method. See Proposition B1 in Appendix B.

Second, since our condition is weaker than perfect ex-post incentive compatibility, it can be satisfied with a smaller variation in continuation payoffs. This is crucial when we show that the vertical move of the continuation payoffs w in Section 3.6 is small. Indeed, if we require perfect ex-post incentive compatibility rather than (3), the vertical move of w must be larger, and w may go to the outside of the ball W.[12] In this sense, our condition is not "too strong."

4 Pseudo-Ergodic Strategies

4.1 Feasible Payoff Set and Pseudo-Ergodic Strategies

Now we consider the general model. As in the Cournot example, let $V^{\mu}(\delta) = \text{co}\{v^{\mu}(\delta,s)|s \in S\}$ be the set of feasible payoffs when the initial prior is μ and the discount factor is δ.

Let Λ be the set of all directions $\lambda \in \boldsymbol{R}^N$ with $|\lambda| = 1$. For each direction λ, we define the "score" as

$$\max_{v \in V^{\mu}(\delta)} \lambda \cdot v.$$

A standard argument shows that this maximization problem indeed has a solution. Intuitively, the score characterizes the extreme point of the feasible payoff set $V^{\mu}(\delta)$ toward the direction λ. For example, in a two-player game, the score for $\lambda = (1,0)$ equals the best possible payoff for player 1 within the feasible payoff set, and the score for $\lambda = (\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}})$ corresponds to the welfare-maximizing point.

[12] The perturbation term z^t defined in (10) ensures our incentive compatibility condition (3), but it does not satisfy perfect ex-post incentive compatibility. This is so because the term G^t is the maximal gain by deviating from s^2 conditional on that player 2's belief is chosen from the thick line in Figure 3. Of course, if we replace the term G^t in (10) with

$$\tilde{G}^t = \max_{\mu^t \in \triangle\Omega} \left(\max_{s_2 \in S_2} v^{\mu^t}(p\delta, s_1^2|_{h^{t-1}}, s_2) - v^{\mu^t}(p\delta, s^2|_{h^{t-1}}) \right),$$

then the resulting perturbation term $\tilde{z}^t$ ensures ex-post incentive compatibility in period t. But this gain $\tilde{G}^t$ needs not converge to zero even if we take $t \to \infty$, so its infinite sum $\sum_{t=1}^{\infty} \tilde{G}^t$ is infinitely large. This implies that the vertical move of the continuation payoff is quite large and does not satisfy (8). So the continuation payoff w goes to the outside of the ball W.

As explained in Section 3.2, each extreme point of the feasible payoff set is achievable by a pure public strategy profile, and thus the score can be computed using dynamic programming. Fix δ and λ, and let $f(\mu)$ be the score when the initial prior is μ. Then this score function f must solve

$$f(\mu) = \max_{a\in A}\left[(1-\delta)\lambda\cdot g^{\mu}(a) + \delta\sum_{y\in Y}\pi_Y^{\mu}(y|a)f(\tilde{\mu}(y|\mu,a))\right].$$

This Bellman equation is exactly the same as that for the perfect-monitoring case studied by Yamamoto (2016), and hence his invariance result remains true:

Proposition 2. *Suppose that the full support assumption holds. Then for each* $\varepsilon>0$*, there is* $\overline{\delta}\in(0,1)$ *such that for any* λ*,* $\delta\in(\overline{\delta},1)$*,* μ*, and* $\tilde{\mu}$*,*

$$\left|\max_{v\in V^{\mu}(\delta)}\lambda\cdot v - \max_{\tilde{v}\in V^{\tilde{\mu}}(\delta)}\lambda\cdot\tilde{v}\right|<\varepsilon.$$

This proposition asserts that when δ is sufficiently large, the scores $\max_{v\in V^{\mu}(\delta)}\lambda\cdot v$ are similar across all priors μ. This implies that the feasible payoff sets $V^{\mu}(\delta)$ are similar across all initial priors μ, when δ is close to one.

As Yamamoto (2016) shows, the score $\max_{v\in V^{\mu}(\delta)}\lambda\cdot v$ has a limit as $\delta\to 1$, so let V^{μ} be the set of all payoff vectors v such that $\lambda\cdot v\leq\lim_{\delta\to1}\max_{v\in V^{\mu}(\delta)}\lambda\cdot v$ for all λ. From Proposition 2, this set V^{μ} is independent of μ, so we denote it by V. Intuitively, this set V is the "limit feasible payoff set" in that the feasible payoff set $V^{\mu}(\delta)$ approximates V regardless of μ when δ is close to one.

The next proposition shows that there is a "pseudo-ergodic" strategy which achieves (approximately) the same payoff regardless of the initial prior. As explained in the Cournot example, such a strategy profile plays a crucial role in our equilibrium construction; it ensures that players can approximate an extreme point of the feasible payoff set even if they do not know the opponents' private beliefs.

Proposition 3. *Suppose that the full support assumption holds. Then for each* $\varepsilon>0$*, there is* $\overline{\delta}\in(0,1)$ *such that for each* λ*, for each* $\delta\in(\overline{\delta},1)$*, for each* μ*, for each* $\tilde{\mu}$*, for each pure public strategy profile* $s^{\tilde{\mu}}\in\arg\max_{s\in S}\lambda\cdot v^{\tilde{\mu}}(\delta,s)$*, for each* $t\geq 0$*, and for each* h^t*,*

$$\left|\max_{v\in V^{\mu}(\delta)}\lambda\cdot v-\lambda\cdot v^{\mu}(\delta,s^{\tilde{\mu}}|_{h^t})\right|<\varepsilon.$$

To interpret this proposition, pick a direction λ and a "dummy belief" $\tilde{\mu}$ arbitrarily. Consider the optimal policy $s^{\tilde{\mu}}$ for this dummy belief $\tilde{\mu}$. Then the proposition shows that this strategy profile $s^{\tilde{\mu}}$ approximates the score regardless of the initial prior μ (so it is a pseudo-ergodic strategy). The proposition also shows that any continuation strategy profile $s^{\tilde{\mu}}|_{h^t}$ has a similar property, that is, it approximates the score regardless of the initial prior μ.

To illustrate the intuition behind the above proposition, consider the Cournot example in Section 3, and the welfare-maximizing strategy $s^{\text{eff}}(\delta, \frac{1}{2})$ for the dummy belief $\tilde{\mu} = \frac{1}{2}$. The above proposition asserts that this strategy approximates the efficient payoff $(0.35, 0.35)$ even if the true belief is $\mu \neq \frac{1}{2}$. This result can be explained as follows. Pick T sufficiently large, and then take δ close to one. Since patient players do not care about payoffs in the first T periods, the average payoff in the overall game is approximated by the expected continuation payoff after period T. So it suffices to explain that this expected continuation payoff after period T approximates $(0.35, 0.35)$.

Suppose that the initial prior is $\mu \neq \frac{1}{2}$, and that players have played $s^{\text{eff}}(\delta, \frac{1}{2})$ for the first T periods. Let h^T be the realized history, and let μ^{T+1} be the posterior belief after this history h^T given the initial belief μ. Also, let $\tilde{\mu}^{T+1}$ be the "dummy posterior belief" induced by the same history h^T but given the dummy initial belief $\tilde{\mu} = \frac{1}{2}$. Since the optimal policy is Markov, the continuation strategy after this history h^T is $s^{\text{eff}}(\delta, \tilde{\mu}^{T+1})$, that is, the continuation strategy maximizes the social welfare if the posterior belief μ^{T+1} matches the dummy posterior $\tilde{\mu}^{T+1}$. Of course, these posteriors μ^{T+1} and $\tilde{\mu}^{T+1}$ are not exactly the same in general, but the belief convergence theorem ensures that they are approximately the same. So the continuation strategy $s^{\text{eff}}(\delta, \tilde{\mu}^{T+1})$ approximates the maximal social welfare under the true posterior μ^{T+1}, i.e., it approximates $(0.35, 0.35)$. Since the same result holds given any history h^T, the expected continuation payoff after period T approximates $(0.35, 0.35)$, as desired.

Remark 1. As discussed in Yamamoto (2016), ε in Proposition 2 can be replaced with $O(1-\delta)$. Using this result, we can show that ε in Proposition 3 can be also replaced with $O(1-\delta)$. See the proof of the proposition for more details.

4.2 Minimax Payoffs and Pseudo-Ergodic Strategies

Given the initial prior μ and the discount factor δ, player i's *minimax payoff* is defined to be

$$\underline{v}_i^{\mu}(\delta) = \inf_{s_{-i} \in S_{-i}^{\text{pub}}} \max_{s_i \in S_i} v_i^{\mu}(s) \tag{13}$$

where S_{-i}^{pub} is the set of all public strategies s_{-i}. In the Cournot example, the minimax strategy is "Always (H,H)" for all μ and δ.

A couple of remarks are in order. First, we assume that the opponents use a public strategy s_{-i} to punish player i, and this is a loss of generality. The general minimax payoff, which allows the opponents to take private strategies, can be possibly lower than our minimax value above. However, we need this restriction because we focus on public-strategy equilibria in this paper. Note that focusing on public strategies is still more general than Escobar and Toikka (2013) and Hörner, Takahashi, and Vieille (2015), who assume that the opponents play a constant pure action a_{-i} over time.

Second, the opponents' strategy s_{-i} is not necessarily Markovian. Since we assume that actions are not observable, the opponents cannot observe player i's deviation and hence cannot know her belief. Accordingly, player i's belief cannot be used as a common state variable, and a minimax strategy needs not be Markovian here.

Third, we take the infimum over s_{-i} instead of the minimum, so the solution to the above minimax problem may not exist. On the other hand, player i's best reply exists for any given s_{-i}, so we take the maximum over $s_i \in S_i$. This difference essentially comes from the fact that player i knows her own posterior belief after every history, while the opponents do not.

When actions are observable, Yamamoto (2016) shows that the minimax payoff is invariant to the initial prior in the limit as $\delta \to 1$. The following proposition extends this result to the imperfect-monitoring case. Our proof technique is quite different from that of Yamamoto (2016), because minimax strategies are not Markovian in our setup. The proof will be given in Appendix B.

Proposition 4. *Suppose that the full support assumption holds. Then for each $\varepsilon > 0$, there is $\overline{\delta} \in (0,1)$ such that $|\underline{v}_i^{\mu}(\delta) - \underline{v}_i^{\tilde{\mu}}(\delta)| < \varepsilon$ for each i, $\delta \in (\overline{\delta}, 1)$, μ, and $\tilde{\mu}$.*

In general, a minimax strategy for player i depends on the initial prior μ. However, the next proposition shows that there is a strategy which minimaxes player i regardless of the initial prior μ. This ensures that even if the opponents do not know player i's private belief μ, they can punish player i by playing such a strategy. The proof is more complicated than that of Proposition 3, because minimax strategies are not Markov. See Appendix B for the formal proof.

Proposition 5. *Suppose that the full support assumption holds. Then for each $\varepsilon > 0$, there is $\overline{\delta} \in (0,1)$ such that for any i, for any $\delta \in (\overline{\delta},1)$, and for any $\tilde{\mu}$, there is a public strategy $s_{-i}^{\tilde{\mu}}$ such that for each pure strategy $s_i^{\tilde{\mu}} \in \arg\max_{\tilde{s}_i \in S_i} v_i^{\tilde{\mu}}(\delta, \tilde{s}_i, s_{-i}^{\tilde{\mu}})$,*

$$\left| v_i^{\mu}(\delta, s^{\tilde{\mu}}|_{h^t}) - \underline{v}_i^{\mu}(\delta) \right| < \varepsilon \tag{14}$$

and

$$\max_{\tilde{s}_i \in S_i} v_i^{\mu}(\delta, \tilde{s}_i, s_{-i}^{\tilde{\mu}}|_{h^t}) - v_i^{\mu}(\delta, s^{\tilde{\mu}}|_{h^t}) < \varepsilon \tag{15}$$

for each μ, $t \geq 0$, and h^t.

The above proposition states only "there is a public strategy $s_{-i}^{\tilde{\mu}}$," but in the proof, we explain how to find such $s_{-i}^{\tilde{\mu}}$. There are two cases to be considered. First, if the minimax problem (13) has a solution, then let $s_{-i}^{\tilde{\mu}}$ be a minimax strategy given a dummy belief $\tilde{\mu}$. The first inequality (14) asserts that this strategy $s^{\tilde{\mu}}$ approximates the minimax payoff $\underline{v}_i^{\mu}(\delta)$ regardless of the true belief μ. The second inequality (15) ensures that player i's gain by deviating from this profile $s^{\tilde{\mu}}$ is almost negligible regardless of the true belief μ. The proposition also shows that any continuation strategy $s^{\tilde{\mu}}|_{h^t}$ of this profile satisfies the same properties.

Second, if the minimax problem (13) does not have a solution, then the minimax strategy does not exist. So instead, we let $s_{-i}^{\tilde{\mu}}$ be a strategy which approximates the minimax payoff given the belief $\tilde{\mu}$. It turns out that the same result holds even in this case, as long as we carefully choose the strategy $s_{-i}^{\tilde{\mu}}$. See the proof of the proposition for how to choose $s_{-i}^{\tilde{\mu}}$.

As in Yamamoto (2016), we can show that given any initial prior, the minimax payoff has a limit as $\delta \to 1$. Let $\underline{v}_i^{\mu}$ denote this limit, that is, $\underline{v}_i^{\mu} = \lim_{\delta \to 1} \underline{v}_i^{\mu}(\delta)$. Proposition 4 ensures that the limit minimax payoff $\underline{v}_i^{\mu}$ does not depend on μ, so we denote it by $\underline{v}_i$. Let V^* denote the set of feasible payoffs $v \in V$ such that $v_i \geq \underline{v}_i$ for each i. That is, V^* is the set of feasible payoffs which Pareto-dominate the minimax payoff.

5 The Folk Theorem

In this section, we show that the folk theorem holds under a mild condition. In ex-post equilibria, a player's deviation must be punished appropriately regardless of the current hidden state; this requires that actions and states should be statistically distinguished through a public signal. Specifically, we will assume *cross-state individual full-rank* and *cross-state pairwise full-rank* conditions, which strengthen individual full-rank and pairwise full-rank conditions of FLM and Fudenberg and Yamamoto (2010).

For each i and each mixed action profile α, let $\Pi_i(\alpha)$ be a matrix with rows $\pi_Y^\omega(a_i,\alpha_{-i}) = (\pi_Y^\omega(a_i,\alpha_{-i}))_{y\in Y}$ for all ω and a_i. In words, the matrix $\Pi_i(\alpha)$ is a collection of the marginal distributions of the public signal y induced by player i's unilateral deviation from α for all possible states ω. For each (i,j) with $i \neq j$ and for each α, let $\Pi_{ij}(\alpha)$ be a matrix constructed by stacking the two matrices $\Pi_i(\alpha)$ and $\Pi_j(\alpha)$. That is, $\Pi_{ij}(\alpha)$ is the collection of the marginal distributions of y induced by a unilateral deviation by i or j.

Definition 2. An action profile α has *cross-state individual full rank for i* if the matrix $\Pi_i(\alpha)$ has rank equal to $|\Omega| \times |A_i|$. An action profile α has *cross-state individual full rank* if it has cross-state individual full rank for all i.

Cross-state individual full rank requires that the hidden state ω and player i's action a_i can be statistically distinguished by a public signal y. This condition is stronger than individual full rank of FLM and Fudenberg and Yamamoto (2010), since we require the hidden state ω to be statistically distinguished.

Definition 3. An action profile α has *cross-state pairwise full rank for (i,j)* if the matrix $\Pi_{ij}(\alpha)$ has rank equal to $|\Omega| \times (|A_i| + |A_j| - 1)$. An action profile α has *cross-state pairwise full rank* if it has cross-state pairwise full rank for all pairs (i,j) with $i \neq j$.

Cross-state pairwise full rank says that if someone unilaterally deviates from α, then her identity (as well as the hidden state ω) can be revealed by a public signal y. Again, this condition is stronger than pairwise full rank of FLM and Fudenberg and Yamamoto (2010).

We impose the following assumptions. They are generically satisfied if there are many signals so that $|Y| \geq |\Omega| \times (|A_i| + |A_j| - 1)$ for all i and j,

Condition IFR. Every pure action profile has cross-state individual full rank.

Condition PFR. For each (i, j) with $i \neq j$, there is an action profile α that has cross-state pairwise full rank for (i, j).

Now we are ready to present the folk theorem. The proof extends the idea presented in Section 3 to the general case, and can be found in Appendix B.

Definition 4. A subset W of $\boldsymbol{R}^N$ is *smooth* if it is closed and convex; it has a non-empty interior; and there is a unique unit normal for each point on the boundary of W.

Proposition 6. *Suppose that the full support assumption, (IFR), and (PFR) are satisfied. Suppose also that public randomization is available. Then, for any smooth subset W of the interior of V^*, there is $\overline{\delta} \in (0,1)$ such that for any $\delta \in (\overline{\delta}, 1)$, the set W is stochastically ex-post self-generating. Hence for each $v \in W$, there is a public ex-post equilibrium which yields the payoff v regardless of the initial state ω.*

This proposition assumes public randomization, but it is dispensable. When public randomization is not available, we cannot use random blocks, so instead, we regard the infinite horizon as a series of T-period blocks. When T is large enough, we can show that there is a T-period strategy which approximates an extreme point of the feasible payoff set regardless of the belief. Likewise, there is a T-period strategy which minimaxes the opponent regardless of the belief. Then the rest of the proof is similar to the one with public randomization. See Appendix C for more details.

Also, the full support assumption is stronger than necessary. In Appendix D, we show that the folk theorem remains valid even if the full support assumption is replaced with a weaker condition. In particular, our result encompasses the folk theorem for the standard stochastic game provided by Fudenberg and Yamamoto (2011b) and Hörner, Sugaya, Takahashi, and Vieille (2011).

Appendix A: Cournot Example

Here we provide a precise description of the distribution of the prices $y = (y_1, y_2)$ in the Cournot example studied in Example 3.

Given a state ω and an action profile a, the price distribution is given by the following compound probability distribution. Specifically, the price vector y follows the distribution $F_1(\cdot|\omega, a)$ with probability $\frac{1}{50}$, the distribution $F_2(\cdot|\omega, a)$ with probability $\frac{19}{50}$, and the distribution $F_3(\cdot|\omega, a)$ with probability $\frac{3}{5}$. So we have $\pi_Y^\omega(y|a) = \frac{1}{50}F_1(y|\omega, a) + \frac{19}{50}F_2(\cdot|\omega, a) + \frac{3}{5}F_3(\cdot|\omega, a)$ for each ω, a, and y.

We choose the distribution F_1 as the uniform distribution over the set $Y = \{0, 10, \cdots, 50\}^2$, regardless of ω and a. Since there are 36 possible signals, this implies that each signal realizes with probability at least $\frac{1}{50} \cdot \frac{1}{36} = \frac{1}{1800}$, and thus the full support assumption holds in this example.

The distribution F_2 depends on the action profile a but not on the hidden state ω, and the possible realizations are only $y = (0, 0)$, $y = (0, 50)$, $y = (50, 0)$, or $y = (50, 50)$. The following table shows how the distribution F_2 changes for different a:

	L	M	H
L	$(0, \frac{3}{38}, \frac{3}{38}, \frac{32}{38})$	$(\frac{7}{38}, \frac{10}{38}, 0, \frac{21}{38})$	$(\frac{7}{38}, \frac{22}{38}, 0, \frac{9}{38})$
M	$(\frac{7}{38}, 0, \frac{10}{38}, \frac{21}{38})$	$(\frac{29}{38}, 0, 0, \frac{9}{38})$	$(\frac{21}{38}, \frac{10}{38}, 0, \frac{7}{38})$
H	$(\frac{7}{38}, 0, \frac{22}{38}, \frac{9}{38})$	$(\frac{21}{38}, 0, \frac{10}{38}, \frac{7}{38})$	$(\frac{37}{38}, 0, 0, \frac{1}{38})$

For each cell, the first number is the probability of $y = (0, 0)$, the second is the one of $y = (0, 50)$, the third is of $y = (50, 0)$, and the last is of $y = (50, 50)$. For example, when the current action is $a = (L, L)$, the distribution F_2 yields $y = (0, 0)$ with probability $\frac{15}{38}$, and $y = (50, 50)$ with probability $\frac{23}{38}$.

The distribution F_3 depends both on the action profile a and the hidden state ω. For each (ω, a), it chooses some particular signal y (which depends on ω and a) with probability one. The following table describes the possible signal realization for each ω and a:

	L	M	H
L	$(40, 40)$	$(50, 40)$	$(30, 10)$
M	$(40, 50)$	$(30, 30)$	$(20, 10)$
H	$(10, 30)$	$(10, 20)$	$(20, 20)$

	L	M	H
L	$(30, 30)$	$(40, 30)$	$(20, 0)$
M	$(30, 40)$	$(20, 20)$	$(10, 0)$
H	$(0, 20)$	$(0, 10)$	$(10, 10)$

As usual, the left table is for the good state ω_G and the right table is for the bad state ω_B. For example, when $(\omega,a)=(\omega_G,LL)$, the signal $y=(40,40)$ is chosen with probability one. Note that when the price distribution is F_3, for each action profile a, the expected price for the good state ω_G is greater than that for the bad state ω_B by 10. This in turn implies that when the price distribution is π_Y, the expected price for the good state ω_G is greater than that for the bad state ω_B by 6, which is consistent with π_Y described in Section 3.

The above signal structure satisfies the full support assumption, and simple algebra shows that the expected price is indeed consistent with the tables given in Section 3. In what follows, we will explain that the distributions $\{(\pi_Y^\omega(y|a))_{y\in Y}\}_{(\omega,a)}$ are linearly independent so that all the three properties given in the bullet points in Section 3 are actually satisfied.

Take a real number $c(\omega,a)$ for each (ω,a) so that

$$\sum_{(\omega,a)\in\Omega\times A} c(\omega,a)\pi_Y^\omega(y|a)=0 \tag{16}$$

for each y. It is sufficient to show that $c(\omega,a)=0$ for all (ω,a). Note that $y=(0,40)$ can realize only when the distribution F_1 is used, so $\pi_Y^\omega((0,40)|a)=\frac{1}{1800}$ for each (ω,a). Plugging this into (16) for $y=(0,40)$, we have

$$\sum_{(\omega,a)\in\Omega\times A} c(\omega,a)=0. \tag{17}$$

Now consider $y=(40,40)$. This signal realizes with probability $\frac{1}{1800}+\frac{3}{5}=\frac{1081}{1800}$ for $(\omega,a)=(\omega_G,HH)$ and with probability $\frac{1}{1800}$ for all other (ω,a). Plugging this into (16) for $y=(40,40)$, we have

$$\frac{1081}{1800}c(\omega_G,HH)+\frac{1}{1800}\sum_{(\omega,a)\neq(\omega_G,HH)} c(\omega,a)=0.$$

Since (17) implies $\sum_{(\omega,a)\neq(\omega_G,HH)} c(\omega,a)=-c(\omega_G,HH)$,

$$\frac{1081}{1800}c(\omega_G,HH)-\frac{1}{1800}c(\omega_G,HH)=0.$$

This implies that $c(\omega_G,HH)=0$. A similar argument shows that $c(\omega,a)=0$ except (ω_B,HH), (ω_G,MM), (ω_B,MM), and (ω_G,LL). So we will focus on these four cases below.

Consider $y=(0,50)$. This signal realizes with probability $\frac{1}{1800}+\frac{3}{38}$ for (ω_B, HH), and with $\frac{1}{1800}$ for the other three cases. So the argument similar to the one above shows that $c(\omega_B, HH)=0$. Similarly, consider $y=(30,30)$. This signal realizes with probability $\frac{1}{1800}+\frac{3}{5}=\frac{1081}{1800}$ for $(\omega, a)=(\omega_G, MM)$ and with probability $\frac{1}{1800}$ for (ω_B, MM) and (ω_G, LL). (Now we do not need to consider (ω_B, HH), as $c(\omega_B, HH)=0$.) Hence we have $c(\omega_G, MM)=0$. Then (17) reduces to

$$c(\omega_B, MM)+c(\omega_G, LL)=0.$$

Obviously, the only way to satisfy this equation and (16) is to set $c(\omega_B, MM)=c(\omega_G, LL)=0$. Hence the signal distributions are indeed linearly independent.

Appendix B: Proofs

B.1 Proof of Proposition 1

We use the following lemma, which is Corollary 2 of Theorem 4.9 of Seneta (1981). It shows weak ergodicity of inhomogeneous Markov matrices:

Lemma B1. *Consider a stochastic process $\{\omega^t\}_{t=1}^{\infty}$ such that for each $t \geq 2$, given ω^{t-1}, the random variable $\omega^t \in \Omega$ follows a Markov chain with the matrix M^t with rows $(m^t(\omega^t|\omega^{t-1}))_{\omega^t \in \Omega}$ for each ω^{t-1}. Let $\mu^t(\mu)$ denote the induced probability distribution of ω^t when ω^1 follows a distribution $\mu \in \triangle\Omega$, that is, $\mu^t(\mu)=\mu M^2 \cdots M^t$. Suppose that the matrices $\{M^t\}_{t=2}^{\infty}$ are "uniformly Markov" in the sense that there is a constant $\beta \in (0,1)$ such that $\max_{\omega^t} \min_{\omega^{t-1}} m^t(\omega^t|\omega^{t-1}) \geq 1-\beta$ for each t. Then weak ergodicity obtains at a rate which is at least geometric with parameter β. That is, $|\mu^t(\mu)-\mu^t(\tilde{\mu})| \leq \beta^{t-1}$ for each t, μ, and $\tilde{\mu}$.*

The rest of the proof is similar to that of Lemma 3 of Connault (2015). Pick s_i and s_{-i} as stated. Pick an arbitrary public history $h^t=(y^1,\cdots,y^t)$. Let $(\alpha^1,\cdots,\alpha^t)$ be the sequence of (possibly mixed) action profiles induced by the strategy profile s, conditional on the public history h^t.

For each $\tilde{t} \leq t$, let $\Pr(\omega^{\tilde{t}+1}|\omega^{\tilde{t}},\alpha^{\tilde{t}},\cdots,\alpha^t,y^{\tilde{t}},\cdots,y^t)$ denote the probability of $\tilde{\omega}^{\tilde{t}+1}$ given that the state in period $\tilde{t}$ is $\omega^{\tilde{t}}$, players play the action sequence $(\alpha^{\tilde{t}},\cdots,\alpha^t)$ and observe the signal sequence $(y^{\tilde{t}},\cdots,y^t)$. Let $\Pr(\omega^{\tilde{t}},\alpha^{\tilde{t}},\cdots,\alpha^t,y^{\tilde{t}},\cdots,y^t)=(\Pr(\omega^{\tilde{t}+1}|\omega^{\tilde{t}},\alpha^{\tilde{t}},\cdots,\alpha^t,y^{\tilde{t}},\cdots,y^t))_{\omega^{\tilde{t}+1}\in\Omega}$, that is, $\Pr(\omega^{\tilde{t}},\alpha^{\tilde{t}},\cdots,\alpha^t,y^{\tilde{t}},\cdots,y^t)$ is the conditional distribution of $\omega^{\tilde{t}+1}$. Then construct a matrix $M^{\tilde{t}+1}$ by stacking these distributions over all $\omega^{\tilde{t}}$. Intuitively, this matrix $M^{\tilde{t}+1}$ maps a conditional distribution of the state $\omega^{\tilde{t}}$ in period $\tilde{t}$ given the action sequence $(\alpha^{\tilde{t}},\cdots,\alpha^t)$ and the signal sequence $(y^{\tilde{t}},\cdots,y^t)$ to a conditional distribution of the state $\omega^{\tilde{t}+1}$ in the next period. Hence we have $\mu_i(h^t|\mu,s) = f(\mu)M^2\cdots M^{t+1}$, where $f(\mu)$ denotes the conditional distribution of the state ω^1 given the initial distribution μ, the action sequence $(\alpha^1,\cdots,\alpha^t)$, and the signal sequence $(y^1,\cdots,y^t)$.

Under the full support assumption, for each $\tilde{t}\leq t$, $\omega^{\tilde{t}}$, and $\omega^{\tilde{t}+1}$, we have

$$\begin{aligned}
&\Pr(\omega^{\tilde{t}+1}|\omega^{\tilde{t}},\alpha^{\tilde{t}},\cdots,\alpha^t,y^{\tilde{t}},\cdots,y^t)\\
&=\frac{\Pr(\omega^{\tilde{t}+1},y^{\tilde{t}},\cdots,y^t|\omega^{\tilde{t}},\alpha^{\tilde{t}},\cdots,\alpha^t)}{\sum_{\omega^{\tilde{t}+1}\in\Omega}\Pr(\omega^{\tilde{t}+1},y^{\tilde{t}},\cdots,y^t|\omega^{\tilde{t}},\alpha^{\tilde{t}},\cdots,\alpha^t)}\\
&=\frac{\pi^{\omega^{\tilde{t}}}(y^{\tilde{t}},\omega^{\tilde{t}+1}|\alpha^{\tilde{t}})\Pr(y^{\tilde{t}+1},\cdots,y^t|\omega^{\tilde{t}+1},\alpha^{\tilde{t}+1},\cdots,\alpha^t)}{\sum_{\omega^{\tilde{t}+1}\in\Omega}\pi^{\omega^{\tilde{t}}}(y^{\tilde{t}},\omega^{\tilde{t}+1}|\alpha^{\tilde{t}})\Pr(y^{\tilde{t}+1},\cdots,y^t|\omega^{\tilde{t}+1},\alpha^{\tilde{t}+1},\cdots,\alpha^t)}\\
&\geq\frac{\overline{\pi}\Pr(y^{\tilde{t}+1},\cdots,y^t|\omega^{\tilde{t}+1},\alpha^{\tilde{t}+1},\cdots,\alpha^t)}{\sum_{\omega^{\tilde{t}+1}\in\Omega}\Pr(y^{\tilde{t}+1},\cdots,y^t|\omega^{\tilde{t}+1},\alpha^{\tilde{t}+1},\cdots,\alpha^t)}.
\end{aligned}$$

Here, when $\tilde{t}=t$, we let $\Pr(y^{\tilde{t}+1},\cdots,y^t|\omega^{\tilde{t}+1},\alpha^{\tilde{t}+1},\cdots,\alpha^t)=1$.

Let $v^{\tilde{t}+1}(\omega^{\tilde{t}+1})=\frac{\Pr(y^{\tilde{t}+1},\cdots,y^t|\omega^{\tilde{t}+1},\alpha^{\tilde{t}+1},\cdots,\alpha^t)}{\sum_{\omega^{\tilde{t}+1}\in\Omega}\Pr(y^{\tilde{t}+1},\cdots,y^t|\omega^{\tilde{t}+1},\alpha^{\tilde{t}+1},\cdots,\alpha^t)}$. It is easy to check that $v^{\tilde{t}+1}$ is a probability distribution over Ω. Also the above inequality implies that for each $\omega^{\tilde{t}}$ and $\omega^{\tilde{t}+1}$,

$$\Pr(\omega^{\tilde{t}+1}|\omega^{\tilde{t}},\alpha^{\tilde{t}},\cdots,\alpha^t,y^{\tilde{t}},\cdots,y^t)\geq\overline{\pi}v^{\tilde{t}+1}(\omega^{\tilde{t}+1}).$$

Now for each $\tilde{t}\leq t$, let $\hat{\omega}^{t+1}\in\arg\max_{\omega^{\tilde{t}+1}}v^{\tilde{t}+1}(\omega^{\tilde{t}+1})$. Then $v^{\tilde{t}+1}(\hat{\omega}^{t+1})\geq\frac{1}{|\Omega|}$, and thus letting $1-\beta=\frac{\overline{\pi}}{|\Omega|}$, we have

$$\Pr(\hat{\omega}^{\tilde{t}+1}|\omega^{\tilde{t}},\alpha^{\tilde{t}},\cdots,\alpha^t,y^{\tilde{t}},\cdots,y^t)\geq 1-\beta$$

for each $\omega^{\tilde{t}}$. Note that this inequality holds for each $\tilde{t}$, and thus the matrices $\{M^{\tilde{t}}\}_{\tilde{t}=2}^{t+1}$ are uniformly Markov. Then it follows from Lemma B1 that

$$|\mu_i(h^t|\mu,s)-\mu_i(h^t|\tilde{\mu},s)|=|f(\mu)M^2\cdots M^{t+1}-f(\tilde{\mu})M^2\cdots M^{t+1}|\leq\beta^t$$

for each μ and $\tilde{\mu}$. Since the parameter $1-\beta=\frac{\overline{\pi}}{|\Omega|}$ does not depend on the choice of t or h^t, we obtain the result.

B.2 Proof of Proposition 3

In Section 4.1, we have explained that Proposition 3 can be proved using the belief convergence theorem. In what follows, we will provide an alternative proof. This proof is (indirect but) simpler than the one which uses the belief convergence theorem. Also, it gives a better bound on the rate of convergence; specifically, it allows us to show that ε in Proposition 3 can be replaced with $O(1-\delta)$. (To obtain such a result, simply replace ε in the proof below with $O(1-\delta)$.)

The proof idea is as follows. Consider the case in which there are only two states ω_1 and ω_2, and let μ denote the probability on the state ω_1. Pick λ so that $\lambda_i = 1$ and $\lambda_j = 0$ for all $j \neq i$, and pick δ close to one. From Proposition 2, we know that the score is almost constant across all μ; for simplicity, we assume that the score is actually constant over all μ, as described by the flat line in Figure 11. Let v^* denote this constant score. Let $s^{\tilde{\mu}}$ be the optimal policy for the dummy belief $\tilde{\mu} = \frac{1}{2}$. Then we have

$$v^* = v_i^{\tilde{\mu}}(\delta, s^{\tilde{\mu}}) = \frac{1}{2} v_i^{\omega_1}(\delta, s^{\tilde{\mu}}) + \frac{1}{2} v_i^{\omega_2}(\delta, s^{\tilde{\mu}}), \tag{18}$$

that is, the score for the initial prior $\tilde{\mu} = \frac{1}{2}$ is equal to the average of the payoff when the profile s is played given the initial state ω_1 and the payoff when the profile s is played given the initial state ω_2.

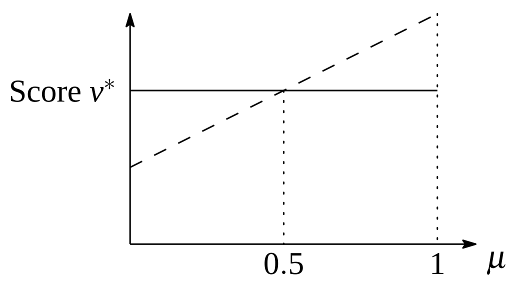

Figure 11: Payoff by the Strategy Profile s

Now, assume by contradiction that $s^{\tilde{\mu}}$ does not approximate the score for some belief μ; in particular, assume that $v_i^{\omega_2}(\delta, s^{\tilde{\mu}}) < v^*$ so that the payoff given the initial state (belief) ω_2 is lower than the score. Then from (18), we must have $v_i^{\omega_1}(\delta, s^{\tilde{\mu}}) > v^*$, that is, the strategy $s^{\tilde{\mu}}$ must yield a payoff higher than the score for the initial state ω_1. (See also Figure 11; the dashed line represents the payoff by playing $s^{\tilde{\mu}}$ for each belief μ.) However this is a contradiction, because any strategy profile cannot yield a payoff higher than the score. Hence we must have $v_i^{\omega_2}(\delta, s^{\tilde{\mu}}) = v^*$. Similarly we have $v_i^{\omega_1}(\delta, s^{\tilde{\mu}}) = v^*$, and so the profile $s^{\tilde{\mu}}$ yields the score regardless of the initial prior μ, as desired.

The formal proof consists of three steps. In the first step, we show that the idea above remains valid as long as the dummy belief $\tilde{\mu}$ is not too close to the boundary of the belief space. That is, we show that for any interior belief $\tilde{\mu}$ which assigns at least probability $\overline{\pi}$ on each state ω, the corresponding optimal policy $s^{\tilde{\mu}}$ approximates the score regardless of the true belief μ.

In the second step, we show that the same result holds for all dummy beliefs $\tilde{\mu}$. The proof relies on the full support assumption. Then in the third step, we show that the same result holds for any continuation strategy of $s^{\tilde{\mu}}$.

B.2.1 Step 1: Optimal Policy for Interior Beliefs

Pick $\varepsilon > 0$ arbitrarily. Proposition 2 ensures that there is $\overline{\delta} \in (0,1)$ such that for any λ, $\delta \in (\overline{\delta}, 1)$, $\tilde{\mu}$, and $\hat{\mu}$,

$$\left| \max_{\tilde{v} \in V^{\tilde{\mu}}(\delta)} \lambda \cdot \tilde{v} - \max_{\hat{v} \in V^{\hat{\mu}}(\delta)} \lambda \cdot \hat{v} \right| < \overline{\pi}\varepsilon - \frac{(1-\delta)\overline{\pi}\overline{g}}{\delta}. \tag{19}$$

Take such $\overline{\delta}$, and then pick λ and $\delta \in (\overline{\delta}, 1)$ arbitrarily. As shown by Yamamoto (2016), given λ and δ, the score $\max_{\tilde{v} \in V^{\tilde{\mu}}(\delta)} \lambda \cdot \tilde{v}$ is convex with respect to the initial prior $\tilde{\mu}$ and hence maximized when the prior $\tilde{\mu}$ assigns probability one to some state ω. Let ω denote such a state.

For each $\tilde{\mu}$, let $s^{\tilde{\mu}}$ be the optimal policy, that is, $s^{\tilde{\mu}}$ achieves the score toward the direction λ given the initial prior $\tilde{\mu}$. Pick $\tilde{\mu}$ such that $\tilde{\mu}(\tilde{\omega}) \geq \overline{\pi}$ for each $\tilde{\omega}$, that is, pick an interior belief $\tilde{\mu}$ which assigns at least $\overline{\pi}$ on each state. We show that the optimal policy $s^{\tilde{\mu}}$ approximates the score regardless of the true belief μ.

From (19), we have

$$\left| \max_{v \in V^{\omega}(\delta)} \lambda \cdot v - \lambda \cdot v^{\tilde{\mu}}(\delta, s^{\tilde{\mu}}) \right| < \overline{\pi}\varepsilon - \frac{(1-\delta)\overline{\pi}\overline{g}}{\delta}.$$

Since $\lambda \cdot v^{\tilde{\mu}}(\delta, s^{\tilde{\mu}}) = \sum_{\tilde{\omega} \in \Omega} \tilde{\mu}(\tilde{\omega}) \lambda \cdot v^{\tilde{\omega}}(\delta, s)$, we obtain

$$\left| \sum_{\tilde{\omega} \in \Omega} \tilde{\mu}(\tilde{\omega}) \left(\max_{v \in V^{\omega}(\delta)} \lambda \cdot v - \lambda \cdot v^{\tilde{\omega}}(\delta, s^{\tilde{\mu}}) \right) \right| < \overline{\pi}\varepsilon - \frac{(1-\delta)\overline{\pi}\overline{g}}{\delta}.$$

Using $\max_{v\in V^{\omega}(\delta)}\lambda\cdot v \geq \max_{v\in V^{\tilde{\omega}}(\delta)}\lambda\cdot v \geq \lambda\cdot v^{\tilde{\omega}}(\delta,s^{\tilde{\mu}})$, we obtain

$$\tilde{\mu}(\tilde{\omega})\left|\max_{v\in V^{\omega}(\delta)}\lambda\cdot v-\lambda\cdot v^{\tilde{\omega}}(\delta,s^{\tilde{\mu}})\right|<\overline{\pi}\varepsilon-\frac{(1-\delta)\overline{\pi}\overline{g}}{\delta}$$

for each $\tilde{\omega}$. Dividing both sides by $\tilde{\mu}(\tilde{\omega})$ and using $\frac{\overline{\pi}}{\tilde{\mu}(\tilde{\omega})}\leq 1$, we have

$$\left|\max_{v\in V^{\omega}(\delta)}\lambda\cdot v-\lambda\cdot v^{\tilde{\omega}}(\delta,s^{\tilde{\mu}})\right|<\varepsilon-\frac{(1-\delta)\overline{g}}{\delta}$$

for each $\tilde{\omega}$.

Now, pick an arbitrary μ. Multiplying both sides of the above inequality by $\mu(\tilde{\omega})$ and summing over all $\tilde{\omega}$,

$$\left|\max_{v\in V^{\omega}(\delta)}\lambda\cdot v-\lambda\cdot v^{\mu}(\delta,s^{\tilde{\mu}})\right|<\varepsilon-\frac{(1-\delta)\overline{g}}{\delta}. \tag{20}$$

Since $\max_{v\in V^{\omega}(\delta)}\lambda\cdot v \geq \max_{v\in V^{\mu}(\delta)}\lambda\cdot v \geq \lambda\cdot v^{\mu}(\delta,s^{\tilde{\mu}})$,

$$\left|\max_{v\in V^{\mu}(\delta)}\lambda\cdot v-\lambda\cdot v^{\mu}(\delta,s^{\tilde{\mu}})\right|<\varepsilon-\frac{(1-\delta)\overline{g}}{\delta}.$$

This shows that if we choose the dummy belief $\tilde{\mu}$ as above, then the strategy $s^{\tilde{\mu}}$ approximates the score regardless of the true belief μ.

B.2.2 Step 2: Optimal Policy for General Beliefs

Now consider an arbitrary belief $\tilde{\mu}\in\triangle\Omega$, and consider the corresponding optimal policy $s^{\tilde{\mu}}$. Pick an arbitrary true belief μ. Then

$$\lambda\cdot v^{\mu}(\delta,s^{\tilde{\mu}})=(1-\delta)\lambda\cdot g^{\mu}(s^{\tilde{\mu}}(h^0))+\delta\sum_{y\in Y}\pi_Y^{\mu}(y|s^{\tilde{\mu}}(h^0))\lambda\cdot v^{\mu(y)}(\delta,s^{\tilde{\mu}}|_y)$$

where $\mu(y)$ is the posterior belief given that the initial belief is μ and players play $s^{\tilde{\mu}}(h^0)$ and observe y in period one. Since the optimal policy $s^{\tilde{\mu}}$ is Markov, the continuation strategy profile $s^{\tilde{\mu}}|_y$ is the optimal policy $s^{\hat{\mu}}$ for some belief $\hat{\mu}$, and the full support assumption ensures that this belief $\hat{\mu}$ is an interior belief so that $\hat{\mu}(\tilde{\omega})\geq\overline{\pi}$ for each $\tilde{\omega}$. Then from the result in the previous step, the second term in the right-hand side must approximate the score. Specifically, from (20),

$$\left|\max_{v\in V^{\omega}(\delta)}\lambda\cdot v-\sum_{y\in Y}\pi_Y^{\mu}(y|s^{\tilde{\mu}}(h^0))\lambda\cdot v^{\mu(y)}(\delta,s^{\tilde{\mu}}|_y)\right|<\varepsilon-\frac{(1-\delta)\overline{g}}{\delta}.$$

Hence

$$\left|\max_{v\in V^{\omega}(\delta)}\lambda\cdot v-\lambda\cdot v^{\mu}(\delta,s^{\tilde{\mu}})\right| \leq(1-\delta)\left|\max_{v\in V^{\omega}(\delta)}\lambda\cdot v-\lambda\cdot g^{\mu}(s^{\tilde{\mu}}(h^0))\right|$$
$$+\delta\left|\max_{v\in V^{\omega}(\delta)}\lambda\cdot v-\sum_{y\in Y}\pi_Y^{\mu}(y|s^{\tilde{\mu}}(h^0))\lambda\cdot v^{\mu(y)}(\delta,s^{\tilde{\mu}}|_y)\right|$$
$$<\delta\varepsilon<\varepsilon.$$

So the strategy $s^{\tilde{\mu}}$ approximates the score regardless of the true belief μ.

B.2.3 Step 3: Continuation Strategies

Now consider the continuation strategy $s^{\tilde{\mu}}|_{h^t}$ of the optimal policy. Since $s^{\tilde{\mu}}$ is Markov, the continuation strategy $s^{\tilde{\mu}}|_{h^t}$ is the optimal policy $s^{\hat{\mu}}$ for some belief $\hat{\mu}$. Then the result in the previous step ensures that this continuation strategy $s^{\tilde{\mu}}|_{h^t}$ approximates the score regardless of the true belief μ, as desired.

B.3 Proof of Proposition 4

Fix δ. For a given strategy s_{-i} and a prior μ, let $v_i^{\mu}(s_{-i})$ denote player i's best possible payoff; that is, let $v_i^{\mu}(s_{-i})=\max_{s_i\in S_i}v_i^{\mu}(\delta,s_i,s_{-i})$. This payoff function $v_i^{\mu}(s_{-i})$ is convex with respect to μ, because $v_i^{\mu}(\delta,s_i,s_{-i})$ is linear with respect to μ, and $v_i^{\mu}(s_{-i})$ is the upper envelop of the linear functions $v_i^{\mu}(\delta,s_i,s_{-i})$ for all s_i.

Lemma B2. *For each s_{-i}, $v_i^{\mu}(s_{-i})$ is convex with respect to μ.*

Let Δ be the set of beliefs μ such that $\mu(\omega)\geq\overline{\pi}$ for all ω. Intuitively, Δ is the set of beliefs which are not too close to the boundary of the belief space $\triangle\Omega$. Under the full support assumption, player i's posterior belief must be in the set Δ after any history.

Pick an arbitrary belief μ, and suppose that there is a minimax strategy s_{-i}^{μ} for this belief, that is, $\underline{v}_i^{\mu}(\delta)=\max_{s_i\in S_i}v_i^{\mu}(s_i,s_{-i}^{\mu})$. Take any public history h^t and the corresponding continuation strategy $s_{-i}^{\mu}|_{h^t}$ induced by this minimax strategy. Since s_{-i}^{μ} is a minimax strategy, the continuation strategy $s_{-i}^{\mu}|_{h^t}$ must give a lower payoff to player i than other strategies s_{-i}, at least for some player i's belief $\tilde{\mu}$. Specifically, we claim that for any strategy s_{-i}, there must be some belief $\tilde{\mu}\in\Delta$ such that $v_i^{\tilde{\mu}}(s_{-i}^{\mu}|_{h^t})\leq v_i^{\tilde{\mu}}(s_{-i})$.

To see that this result indeed holds, suppose not, so that there is s_{-i} such that $v_i^{\tilde{\mu}}(s^{\mu}_{-i}|_{h^t}) > v_i^{\tilde{\mu}}(s_{-i})$ for all $\tilde{\mu} \in \Delta$. Intuitively, this strategy s_{-i} gives a lower payoff to player i than $s^{\mu}_{-i}|_{h^t}$, regardless of player i's belief $\tilde{\mu} \in \Delta$. This means that s^{μ}_{-i} is not a minimax strategy; we can can lower player i's payoff by replacing the continuation strategy $s^{\mu}_{-i}|_{h^t}$ with s_{-i}. This is a contradiction, and thus the result follows.

In general, a minimax strategy s^{μ}_{-i} may not exist, since we take the infimum with respect to s_{-i} in the definition of the minimax payoff. In such a case, we let s^{μ}_{-i} be a strategy which approximates the minimax payoff. The following lemma ensures that the above result holds if we choose this strategy s^{μ}_{-i} carefully.

Lemma B3. *For each μ, there is a public strategy s^{μ}_{-i} such that*

$$\left| \underline{v}_i^{\mu}(\delta) - v_i^{\mu}(s^{\mu}_{-i}) \right| < 1-\delta. \tag{21}$$

and such that for any $t \geq 1$, for any h^t, and for any public strategy s_{-i}, there is $\tilde{\mu} \in \Delta$ satisfying

$$v_i^{\tilde{\mu}}(s^{\mu}_{-i}|_{h^t}) < v_i^{\tilde{\mu}}(s_{-i}) + 1 - \delta. \tag{22}$$

Note that we consider the case in which δ is close to one, so $1-\delta$ is small. The first condition (21) ensures that s^{μ}_{-i} approximates the minimax payoff. The second condition (22) asserts that there is no strategy s_{-i} which yields a lower payoff to player i than $s^{\mu}_{-i}|_{h^t}$ for all beliefs $\tilde{\mu} \in \Delta$.

Proof. Fix μ, and take s^{μ}_{-i} such that (21) holds. This s^{μ}_{-i} may not satisfy the second condition (22) in the lemma. We will modify s^{μ}_{-i} so that (22) holds.

Suppose that the second condition (22) is not satisfied for $t=1$, i.e., suppose that there is h^1 and s_{-i} such that

$$v_i^{\tilde{\mu}}(s^{\mu}_{-i}|_{h^1}) \geq v_i^{\tilde{\mu}}(s_{-i}) + 1 - \delta. \tag{23}$$

for all $\tilde{\mu} \in \Delta$. Fix such h^1, and let $S_{-i}(h^1)$ be the set of all strategies s_{-i} which satisfy (23) for all $\tilde{\mu} \in \Delta$. Choose $s^*_{-i} \in S_{-i}(h^1)$ so that

$$\left| \inf_{s_{-i} \in S_{-i}(h^1)} v_i^{\mu^U}(s_{-i}) - v_i^{\mu^U}(s^*_{-i}) \right| < 1-\delta$$

where $\mu^U = (\frac{1}{|\Omega|}, \cdots, \frac{1}{|\Omega|})$.

Now, consider the following strategy $\tilde{s}^{\mu}_{-i}$, which is a modification of s^{μ}_{-i}:

- Follow s^{μ}_{-i}, unless the history reaches h^1
- If h^1 is reached, they play s^*_{-i} (instead of $s^{\mu}_{-i}|_{h^1}$) in the rest of the game.

That is, we replace the continuation strategy after the history h^1 with s^*_{-i}.

We claim that this modified strategy $\tilde{s}^{\mu}_{-i}$ satisfies (22) for h^1, while it still satisfies (21). We first show that (22) holds for h^1. Since $\tilde{s}^{\mu}_{-i}|_{h^1} = s^*_{-i}$, it is sufficient to show that for any s_{-i}, there is $\tilde{\mu} \in \Delta$ such that

$$v_i^{\tilde{\mu}}(s^*_{-i}) < v_i^{\tilde{\mu}}(s_{-i}) + 1 - \delta. \tag{24}$$

Consider the case in which $s_{-i} \in S_{-i}(h^1)$. Then by the definition of s^*_{-i}, (24) is satisfied for $\tilde{\mu} = \mu^U$. Consider the case in which $s_{-i} \notin S_{-i}(h^1)$. Then by the definition of $S_{-i}(h^1)$, there is $\tilde{\mu} \in \Delta$ such that $v_i^{\tilde{\mu}}(s^{\mu}_{-i}|_{h^1}) < v_i^{\tilde{\mu}}(s_{-i}) + 1 - \delta$. (24) holds for such $\tilde{\mu}$, because $s^*_{-i} \in S_{-i}(h^1)$ so that $v_i^{\tilde{\mu}}(s^*_{-i}) < v_i^{\tilde{\mu}}(s^{\mu}_{-i}|_{h^1})$.

To see that the modified strategy $\tilde{s}^{\mu}_{-i}$ satisfies (21), recall that s^*_{-i} satisfies (23) for all $\tilde{\mu} \in \Delta$. That is, s^*_{-i} yields a lower payoff to player i than $s^{\mu}_{-i}|_{h^1}$, as long as the current belief is in the set Δ. This implies that if the opponents play $\tilde{s}^{\mu}_{-i}$ instead of s^{μ}_{-i}, player i's continuation payoff is lowered once the history h^1 realizes; and this is true regardless of player i's posterior belief in period two. Since the full support assumption ensures that the probability of h^1 is positive, we have

$$v_i^{\tilde{\mu}}(a_i, s^{\mu}_{-i}) > v_i^{\tilde{\mu}}(a_i, \tilde{s}^{\mu}_{-i})$$

for any $\tilde{\mu}$ and a_i, where $v_i^{\tilde{\mu}}(a_i, s_{-i})$ represents player i's payoff when the initial prior is $\tilde{\mu}$, the opponents play s_{-i}, and player i chooses a_i in period one and then plays a best reply after that. This implies that

$$v_i^{\tilde{\mu}}(\tilde{s}^{\mu}_{-i}) = \max_{a_i \in A_i} v_i^{\tilde{\mu}}(a_i, \tilde{s}^{\mu}_{-i}) \leq \max_{\tilde{a}_i \in A_i} v_i^{\tilde{\mu}}(\tilde{a}_i, s^{\mu}_{-i}) = v_i^{\tilde{\mu}}(s^{\mu}_{-i})$$

for all $\tilde{\mu}$. Then it is obvious that the modified strategy $\tilde{s}^{\mu}_{-i}$ satisfies (21), as the original strategy s^{μ}_{-i} satisfies it.

We can modify the continuation strategy $s_{-i}|_{\tilde{h}^1}$ for each one-period history $\tilde{h}^1$ in the same way, so that (22) holds for all one-period histories h^1. Then induction; we apply a similar argument to each history h^2 to obtain a strategy which satisfies (22) for all h^0, h^1, and h^2, and so on. This proves the lemma. *Q.E.D.*

Pick μ and h^t, and consider the corresponding strategy $s^{\mu}_{-i}|_{h^t}$. Then the payoff $v_i^{\tilde{\mu}}(s^{\mu}_{-}|_{h^t})$ is convex with respect to the initial belief $\tilde{\mu}$. For each μ, $t \geq 0$, and $h^t \in H^t$, let

$$\overline{v}_i(s^{\mu}_{-i}|_{h^t}) = \max_{\tilde{\mu} \in \triangle\Omega} v_i^{\tilde{\mu}}(s^{\mu}_{-}|_{h^t}).$$

That is, $\overline{v}_i(s^{\mu}_{-i}|_h)$ is the highest payoff attained by the convex function $v_i^{\tilde{\mu}}(s^{\mu}_{-i}|_{h^t})$. Note that different (μ, h^t) induce different convex curves $v_i^{\tilde{\mu}}(s^{\mu}_{-i}|_{h^t})$, and hence different highest payoffs $\overline{v}_i(s^{\mu}_{-i}|_{h^t})$. Take the supremum of these highest payoffs, and choose (μ^*, h^*) to approximate the supremum, that is,

$$\left| \sup_{\mu \in \triangle\Omega} \sup_{h \in H} \overline{v}_i(s^{\mu}_{-i}|_h) - \overline{v}_i(s^{\mu^*}_{-i}|_{h^*}) \right| < 1 - \delta. \tag{25}$$

We call $\overline{v}_i(s^{\mu^*}_{-i}|_{h^*})$ the *maximal value*, because it approximates $\sup_{\mu \in \triangle\Omega} \sup_{h \in H} \overline{v}_i(s^{\mu}_{-i}|_h)$, which is greater than any payoffs attained by any convex curves.

Since $v_i^{\tilde{\mu}}(s^{\mu^*}_{-i}|_{h^*})$ is convex, it is maximized when $\tilde{\mu}$ is an extreme point; i.e., it is maximized when the initial prior puts probability one on some state ω. Let ω denote this state, that is, $v_i^{\omega}(s^{\mu^*}_{-i}|_{h^*}) \geq v_i^{\tilde{\mu}}(s^{\mu^*}_{-i}|_{h^*})$ for all $\tilde{\mu}$.

In the rest of the proof, we will show that for any (μ, h^t), the corresponding convex curve $v_i^{\hat{\mu}}(s^{\mu}_{-i}|_{h^t})$ is almost constant, and approximates the maximal score for all beliefs $\hat{\mu}$. That is, if the opponents play the minimax strategy s^{μ}_{-i}, after every history h^t, player i's continuation payoff is approximately equal to the maximal score regardless of her posterior belief $\hat{\mu}$. This implies that the minimax payoff given a common prior μ approximates the maximal score regardless of μ, and hence the result follows.

Our proof consists of three steps. In the first step, we provide a sufficient condition for the convex curve $v_i^{\hat{\mu}}(s^{\mu}_{-i}|_{h^t})$ to be almost flat. Specifically, we show that given the opponents' strategy $s^{\mu}_{-i}|_{h^t}$, if the corresponding convex curve $v_i^{\tilde{\mu}}(s^{\mu}_{-i}|_{h^t})$ approximates the maximal score for some interior belief $\tilde{\mu}$, then the curve is almost flat and the payoff $v_i^{\hat{\mu}}(s^{\mu}_{-i}|_{h^t})$ approximates the maximal score for all beliefs $\hat{\mu}$. The proof technique is very similar to the one used in Yamamoto (2016).

In the second step, we show that there is some (μ, h^t) such that the corresponding convex curve $v_i^{\tilde{\mu}}(s^{\mu}_{-i}|_{h^t})$ is almost flat and approximates the maximal score for all initial beliefs $\tilde{\mu}$. The proof uses the sufficient condition derived in the first step.

In the third step, we show that the same result holds for all (μ, h^t), that is, for any (μ, h^t), the corresponding convex curve $v_i^{\tilde{\mu}}(s^{\mu}_{-i}|_{h^t})$ approximates the maximal score for all initial beliefs $\tilde{\mu}$. By letting $\tilde{\mu} = \mu$ and $h^t = h^0$, this implies that player i's minimax payoff approximates the maximal value regardless of the initial prior; hence the minimax payoff is insensitive to the initial prior.

B.3.1 Step 1: Almost Flat Convex Curve

The following lemma gives a sufficient condition for a convex curve to be almost flat. The proof is very similar to Lemma C8 of Yamamoto (2016) and hence omitted.

Lemma B4. *Pick any μ, $\tilde{\mu}$, and h^t, and let Ω^* be the support of $\tilde{\mu}$. Let $p = \min_{\tilde{\omega} \in \Omega^*} \tilde{\mu}(\tilde{\omega})$, which measures the distance from $\tilde{\mu}$ to the boundary of $\triangle\Omega^*$. Then for each $\hat{\mu} \in \triangle\Omega^*$,*

$$\left| \overline{v}_i(s_{-i}^{\mu^*}|_{h^*}) + (1-\delta) - v_i^{\hat{\mu}}(s_{-i}^{\mu}|_{h^t}) \right| \leq \frac{\left| \overline{v}_i(s_{-i}^{\mu^*}|_{h^*}) + (1-\delta) - v_i^{\tilde{\mu}}(s_{-i}^{\mu}|_{h^t}) \right|}{p}.$$

This lemma ensures that if the convex curve $v_i^{\tilde{\mu}}(s_{-i}^{\mu}|_{h^t})$ approximates the maximal value for *some* interior belief $\tilde{\mu} \in \Delta$, then the curve is almost flat and approximates the maximal value for *all* beliefs $\tilde{\mu}$. To see this, pick some interior belief $\tilde{\mu} \in \Delta$, and suppose that

$$\left| \overline{v}_i(s_{-i}^{\mu^*}|_{h^*}) + (1-\delta) - v_i^{\tilde{\mu}}(s_{-i}^{\mu}|_{h^t}) \right| < \varepsilon$$

where ε is a positive number close to zero. That is, assume that the curve $v_i^{\tilde{\mu}}(s_{-i}^{\mu}|_{h^t})$ approximates the maximal score $\overline{v}_i(s_{-i}^{\mu^*}|_{h^*})$ for some $\hat{\mu}$. Then since $\tilde{\mu} \in \Delta$, we have $p \geq \overline{\pi}$, and thus the above lemma implies that

$$\left| \overline{v}_i(s_{-i}^{\mu^*}|_{h^*}) + (1-\delta) - v_i^{\hat{\mu}}(s_{-i}^{\mu}|_{h^t}) \right| \leq \frac{\varepsilon}{p} \leq \frac{\varepsilon}{\overline{\pi}}$$

for all $\hat{\mu} \in \triangle\Omega$. So the payoff $v_i^{\hat{\mu}}(s_{-i}^{\mu}|_{h^t})$ indeed approximates the maximal value for *all* beliefs $\hat{\mu}$.

In the above argument, it is important that $\tilde{\mu} \in \Delta$. That is, $\tilde{\mu}$ should not be too close to the boundary of the belief space. This ensures that the parameter p is at least $\overline{\pi}$, so that $\frac{\varepsilon}{p}$ is small.

B.3.2 Step 2: Some Convex Curve Approximates the Maximal Value

The following lemma shows that there is (μ, h^t) such that the corresponding convex curve $v_i^{\tilde{\mu}}(s_{-i}^{\mu}|_{h^t})$ is almost flat and approximates the maximal value uniformly in $\tilde{\mu} \in \triangle\Omega$. The lemma shows that such (μ, h^t) can be found by letting $\mu = \mu^*$ and $h^t = (h^*, y)$ for some y. Let $C = \frac{2\overline{g}}{\pi^2}$ and let $\tilde{C} = \frac{1}{\pi^2}$.

Lemma B5. *For each y and $\tilde{\mu}$,*

$$\left|\overline{v}_i(s_{-i}^{\mu^*}|_{h^*}) + (1-\delta) - v_i^{\tilde{\mu}}(s_{-i}^{\mu^*}|_{(h^*,y)})\right| \leq \frac{1-\delta}{\delta}C + (1-\delta)\tilde{C}.$$

To prove this lemma, it is sufficient to find an interior belief $\tilde{\mu} \in \Delta$ such that the payoff $v_i^{\tilde{\mu}}(s_{-i}^{\mu^*}|_{(h^*,y)})$ approximates the maximal value. Indeed, if there is such an interior belief $\tilde{\mu}$, then Lemma B4 ensures that the convex curve $v_i^{\tilde{\mu}}(s_{-i}^{\mu^*}|_{(h^*,y)}$ is almost flat and approximates the maximal value for all $\tilde{\mu}$.

To find such an interior belief $\tilde{\mu}$, suppose that the current state is ω and it is common knowledge. Suppose also that the opponents play $s_{-i}^{\mu^*}|_{h^*}$ from now on, and player i takes a best reply. By the definition of $s_{-i}^{\mu^*}|_{h^*}$ and ω, player i's payoff approximates the maximal value. Now, suppose that no one deviates today and the signal y is observed. Letting $\tilde{\mu}$ be player i's posterior belief in period two, her continuation payoff from period two is denoted by $v_i^{\tilde{\mu}}(s_{-i}^{\mu^*}|_{(h^*,y)})$. We can show that this continuation payoff approximates the maximal value; the proof technique is very similar to the ones presented in Section 5.2.2 in Yamamoto (2016). So this belief $\tilde{\mu}$ satisfies the desired property, and hence we obtain the lemma.

Proof. Pick an arbitrary y. Suppose that the initial state is ω and that the opponents play $s_{-i}^{\mu^*}|_{h^*}$. Suppose that player i chooses a pure best reply strategy s_i^*. Let $\alpha^* = (s_i^*(h_i^0), s_{-i}^{\mu^*}|_{h^*}(h_{-i}^0))$ denote the action profile in period one, and let $\tilde{\mu}(\tilde{y})$ be player i's posterior in period two given that players observe $\tilde{y}$ in period one. Then

$$\overline{v}_i(s_{-i}^{\mu^*}|_{h^*}) = (1-\delta)g_i^{\omega}(\alpha^*) + \delta \sum_{\tilde{y} \in Y} \pi_Y^{\omega}(\tilde{y}|\alpha^*) v_i^{\tilde{\mu}(\tilde{y})}(s_{-i}^{\mu^*}|_{(h^*,\tilde{y})}).$$

Since $g_i^{\omega}(\alpha^*) \leq \overline{g}$,

$$\overline{v}_i(s_{-i}^{\mu^*}|_{h^*}) \leq (1-\delta)\overline{g} + \delta \sum_{\tilde{y} \in Y} \pi_Y^{\omega}(\tilde{y}|\alpha^*) v_i^{\tilde{\mu}(\tilde{y})}(s_{-i}^{\mu^*}|_{(h^*,\tilde{y})}).$$

From (25), we have

$$v_i^{\tilde{\mu}(\tilde{y})}(s_{-i}^{\mu^*}|_{(h^*,\tilde{y})}) \leq \overline{v}_i(s_{-i}^{\mu^*}|_{h^*}) + (1-\delta) \tag{26}$$

for each $\tilde{y}$. Plugging this into the above inequality,

$$\begin{aligned}\overline{v}_i(s_{-i}^{\mu^*}|_{h^*}) \leq &(1-\delta)\overline{g} + \delta\pi_Y^\omega(y|\alpha^*)v_i^{\tilde{\mu}(y)}(s_{-i}^{\mu^*}|_{(h^*,y)}) \\ &+ \delta(1-\pi_Y^\omega(y|\alpha^*))\left\{\overline{v}_i(s_{-i}^{\mu^*}|_{h^*}) + (1-\delta)\right\}.\end{aligned}$$

Since (26) holds and $\pi_Y^\omega(\tilde{y}|\alpha^*) \geq \overline{\pi}$,

$$\overline{v}_i(s_{-i}^{\mu^*}|_{h^*}) \leq (1-\delta)\overline{g} + \delta\overline{\pi}v_i^{\tilde{\mu}(y)}(s_{-i}^{\mu^*}|_{(h^*,y)}) + \delta\left(1-\overline{\pi}\right)\left\{\overline{v}_i(s_{-i}^{\mu^*}|_{h^*}) + (1-\delta)\right\}.$$

Subtracting $(1-\delta\overline{\pi})\overline{v}_i(s_{-i}^{\mu^*}|_{h^*}) - \delta\overline{\pi}(1-\delta) + \delta\overline{\pi}v_i^{\tilde{\mu}(y)}(s_{-i}^{\mu^*}|_{(h^*,y)})$ from both sides,

$$\begin{aligned}&\delta\overline{\pi}\left\{\overline{v}_i(s_{-i}^{\mu^*}|_{h^*}) + (1-\delta) - v_i^{\tilde{\mu}(y)}(s_{-i}^{\mu^*}|_{(h^*,y)})\right\} \\ &\leq (1-\delta)(\overline{g} - \overline{v}_i(s_{-i}^{\mu^*}|_{h^*})) + \delta(1-\delta).\end{aligned}$$

Since the left-hand side is non-negative, taking the absolute value of both sides and dividing both sides by $\delta\overline{\pi}$,

$$\left|\overline{v}_i(s_{-i}^{\mu^*}|_{h^*}) + (1-\delta) - v_i^{\tilde{\mu}(y)}(s_{-i}^{\mu^*}|_{(h^*,y)})\right| \leq \frac{(1-\delta)|\overline{g} - \overline{v}_i(s_{-i}^{\mu^*}|_{h^*})|}{\delta\overline{\pi}} + \frac{1-\delta}{\overline{\pi}}.$$

Since $\overline{v}_i(s_{-i}^{\mu^*}|_{h^*}) \geq -\overline{g}$,

$$\left|\overline{v}_i(s_{-i}^{\mu^*}|_{h^*}) + (1-\delta) - v_i^{\tilde{\mu}(y)}(s_{-i}^{\mu^*}|_{(h^*,y)})\right| \leq \frac{(1-\delta)2\overline{g}}{\delta\overline{\pi}} + \frac{1-\delta}{\overline{\pi}}.$$

This inequality ensures that the convex curve induced by $s_{-i}^{\mu^*}|_{(h^*,y)}$ approximates the maximal value at the belief $\tilde{\mu}(y) \in \Delta$. Hence Lemma B4 implies that this convex curve is almost flat and approximates the maximal value regardless of the belief; in particular it shows that the desired inequality holds, because Ω^* and p in the statement of Lemma B4 satisfy $\Omega^* = \Omega$ and $p \geq \overline{\pi}$. *Q.E.D.*

B.3.3 Step 3: All Convex Curves Approximate the Maximal Value

In the previous step, we have seen that there is the opponents' strategy $s_{-i}^{\mu}|_{h^t}$ such that the corresponding convex curve $v_i^{\tilde{\mu}}(s_{-i}^{\mu}|_{h^t})$ is almost flat and approximates the maximal value uniformly in $\tilde{\mu} \in \triangle\Omega$. The next lemma shows that the same is true for all strategies $s^{\mu}_{-i}|_{h^t}$. That is, we show that for any (μ, h^t), the corresponding convex curve $v_i^{\tilde{\mu}}(s^{\mu}_{-i}|_{h^t})$ is almost flat and approximates the maximal value uniformly in $\tilde{\mu} \in \triangle\Omega$. Let $C' = \frac{C}{\pi}$ and $\tilde{C}' = \frac{\tilde{C}+1}{\pi}$.

Lemma B6. *For each μ, $\tilde{\mu}$, $t \geq 0$, and h^t,*

$$\left| \overline{v}_i(s^{\mu^*}_{-i}|_{h^*}) + (1-\delta) - v_i^{\tilde{\mu}}(s^{\mu}_{-i}|_{h^t}) \right| \leq \frac{1-\delta}{\delta} C' + (1-\delta)\tilde{C}'$$

To see the proof idea, pick some (μ, h^t). Lemma B3 implies that the strategy $s^{\mu^*}_{-i}|_{(h^*,y)}$ must give a lower payoff to player i than $s^{\mu}_{-i}|_{h^t}$, at least for some $\tilde{\mu} \in \Delta$. That is, there must be some belief $\tilde{\mu} \in \Delta$ such that player i's payoff against $s^{\mu^*}_{-i}|_{(h^*,y)}$ is lower than her payoff against $s^{\mu}_{-i}|_{h^t}$. Lemma B5 ensures that this former payoff approximates the maximal value; hence the latter payoff does so too, that is, player i's payoff against $s^{\mu^*}_{-i}|_{(h^*,y)}$ approximates the maximal score given the belief $\tilde{\mu}$. Then Lemma B4 ensures that the convex curve $v_i^{\tilde{\mu}}(s^{\mu}_{-i}|_{h^t})$ approximates the maximal value for all beliefs $\tilde{\mu}$.

Proof. Pick μ, t, h^t, and y arbitrarily. From Lemma B3, there is $\tilde{\mu} \in \Delta$ such that

$$v_i^{\tilde{\mu}}(s^{\mu^*}_{-i}|_{(h^*,y)}) < v_i^{\tilde{\mu}}(s^{\mu}_{-i}|_{h^t}) + 1 - \delta.$$

Since $v_i^{\tilde{\mu}}(s^{\mu}_{-i}|_{h^t}) < \overline{v}_i(s^{\mu^*}_{-i}|_{h^*}) + (1-\delta)$, this implies that

$$v_i^{\hat{\mu}}(s^{\mu^*}_{-i}|_{(h^*,y)}) - (1-\delta) < v_i^{\tilde{\mu}}(s^{\mu}_{-i}|_{h^t}) < \overline{v}_i(s^{\mu^*}_{-i}|_{h^*}) + (1-\delta).$$

Then we obtain

$$\begin{aligned}
&\left| \overline{v}_i(s^{\mu^*}_{-i}|_{h^*}) + (1-\delta) - v_i^{\tilde{\mu}}(s^{\mu}_{-i}|_{h^t}) \right| \\
&\leq \left| \overline{v}_i(s^{\mu^*}_{-i}|_{h^*}) + (1-\delta) - v_i^{\tilde{\mu}}(s^{\mu^*}_{-i}|_{(h^*,y)}) + (1-\delta) \right| \\
&\leq \frac{1-\delta}{\delta} C + (1-\delta)(\tilde{C}+1)
\end{aligned}$$

where the second inequality follows from Lemma B5. This implies that player i's best possible payoff $v_i^{\tilde{\mu}}(s^{\mu}_{-i}|_{h^t})$ given the initial prior $\tilde{\mu}$ and the opponents' strategy $s^{\mu}_{-i}|_{h^t}$ approximates the maximal value $\overline{v}_i(s^{\mu^*}_{-i}|_{h^*})$. Then since $\tilde{\mu} \in \Delta$, Lemma B4 leads to the desired result. *Q.E.D.*

Letting $\mu = \tilde{\mu}$ and $h = h^0$, the above lemma implies that

$$\left|\overline{v}_i(s_{-i}^{\mu^*}|_{h^*}) + (1-\delta) - v_i^{\mu}(s_{-i}^{\mu})\right| \leq \frac{1-\delta}{\delta}C' + (1-\delta)\tilde{C}'.$$

This, together with (21), implies that regardless of the initial prior μ, the minimax payoff $\underline{v}_i^{\mu}(\delta)$ approximates the maximal value $\overline{v}_i(s_{-i}^{\mu^*}|_{h^*})$. Hence the minimax payoffs are insensitive to the initial prior, as desired.

B.4 Proof of Proposition 5

Let s_{-i}^{μ} be as in the proof of Proposition 4 for each μ. Also, let Δ, $\overline{v}_i(s_{-i}^{\mu}|_{h^t})$, $v_i^{\mu}(s_{-i}|_{h^t})$, C', and $\tilde{C}'$ as in the proof of Proposition 4.

Pick an arbitrary dummy belief $\tilde{\mu}$. Consider the corresponding $s_{-i}^{\tilde{\mu}}$, and let $s_i^{\tilde{\mu}}$ be a pure-strategy best reply to $s_{-i}^{\tilde{\mu}}$ given the belief $\tilde{\mu}$. By the definition, this strategy profile $s^{\tilde{\mu}}$ approximates the minimax payoff if the true belief μ equals $\tilde{\mu}$. In what follows, we show that this profile $s^{\tilde{\mu}}$ satisfies the desired inequalities (14) and (15).

Note first that from Lemma B6, $v_i^{\mu}(s_{-i}^{\tilde{\mu}}|_{h^t})$ approximates the maximal value for each μ, t, and h^t. This proves (15), because the minimax payoff $\underline{v}_i^{\mu}(\delta)$ approximates the maximal value regardless of the initial prior μ, as shown in the proof of Proposition 4.

To prove (14), we use the following lemma.

Lemma B7. *Pick $\tilde{\mu}$ and $s^{\tilde{\mu}}$ as stated above. Pick any μ, $t \geq 0$, and h^t, and let Ω^* be the support of μ. Let $p = \min_{\tilde{\omega} \in \Omega^*} \mu(\tilde{\omega})$, which measures the distance from μ to the boundary of $\triangle\Omega^*$. Then for each $s_i \in \arg\max_{\tilde{s}_i \in S_i} v_i^{\mu}(\delta, \tilde{s}_i, s_{-i}^{\tilde{\mu}}|_{h^t})$ and $\hat{\mu} \in \triangle\Omega^*$,*

$$\left|\overline{v}_i(s_{-i}^{\mu^*}|_{h^*}) + (1-\delta) - v_i^{\hat{\mu}}(\delta, s_i, s_{-i}^{\tilde{\mu}}|_{h^t})\right| \leq \frac{\left|\overline{v}_i(s_{-i}^{\mu^*}|_{h^*}) + (1-\delta) - v_i^{\mu}(s_{-i}^{\tilde{\mu}}|_{h^t})\right|}{p}.$$

To interpret this lemma, suppose that a strategy s_i is a best reply to $s_{-i}^{\tilde{\mu}}|_{h^t}$ given *some* interior belief $\mu \in \Delta$, and its payoff approximates the maximal value. Then the lemma ensures that the payoff by this strategy s_i approximates the maximal value for *all* beliefs $\hat{\mu}$.

Proof. The proof is very similar to that of Lemma C8 of Yamamoto (2016). Replace $\overline{v}_i(s_{-i}^{\mu^*})$, $\tilde{\mu}$, and s_{-i}^{μ} in Yamamoto (2016) with $\overline{v}_i(s_{-i}^{\mu^*}|_{h^*})$, μ, $s_{-i}^{\tilde{\mu}}|_{h^t}$, respectively. Then we can prove the lemma just as Yamamoto (2016) derives (20). *Q.E.D.*

Now we show that (14) holds for any h^t with $t \geq 1$. Pick such h^t, and let $\mu = \mu_i(h^t|\tilde{\mu}, h^t)$. Under the full support assumption, the posterior belief $\mu = \mu_i(h^t|\tilde{\mu}, s^{\tilde{\mu}})$ puts at least probability $\overline{\pi}$ on each state, and hence $\Omega^* = \Omega$ and $p \geq \overline{\pi}$. Also, Lemma B6 ensures that $v_i^{\mu_i(h^t|\tilde{\mu}, s^{\tilde{\mu}})}(s_{-i}^{\tilde{\mu}}|_{h^t})$ approximates the maximal value, that is,

$$\left| \overline{v}_i(s_{-i}^{\mu^*}|_{h^*}) + (1-\delta) - v_i^{\mu_i(h^t|\tilde{\mu}, s^{\tilde{\mu}})}(s_{-i}^{\tilde{\mu}}|_{h^t}) \right| \leq \frac{1-\delta}{\delta} C' + (1-\delta)\tilde{C}'.$$

Plugging this and $p \geq \overline{\pi}$ into the inequality in the above lemma, we obtain

$$\left| \overline{v}_i(s_{-i}^{\mu^*}|_{h^*}) + (1-\delta) - v_i^{\hat{\mu}}(\delta, s^{\tilde{\mu}}|_{h^t}) \right| \leq \frac{1}{\overline{\pi}} \left(\frac{1-\delta}{\delta} C' + (1-\delta)\tilde{C}' \right).$$

for all $\hat{\mu} \in \triangle\Omega$. This implies that (14) holds for this h^t, as the minimax payoff approximates the maximal score for all μ.

Also, we can show that (14) holds for h^0. The proof is very similar to Step 2 in the proof of Proposition 3, and hence omitted.

B.5 Proof of Proposition 6

We begin with providing the self-generation theorem, which shows that a ball W is supported by public ex-post equilibria if it is self-generating. The proof is very similar to Abreu, Pearce, and Stacchetti (1990) and hence omitted.

Definition B1. A pair (s, v) of a public strategy profile and a payoff vector is *stochastically ex-post enforceable with respect to* (δ, p) if there is a function $w : H \to \boldsymbol{R}^N$ such that (2) holds for all ω and i, and such that (3) holds for all ω, i, and $\tilde{s}_i$.

Definition B2. A subset W of $\boldsymbol{R}^N$ is *stochastically ex-post self-generating with respect to* (δ, p) if for each $v \in W$, there is a public strategy profile s and $w : H \to W$ such that (s, v) is stochastically ex-post enforceable with respect to (δ, p) using w.

Proposition B1. *Assume that public randomization is available, and fix δ. If W is bounded and stochastically ex-post self-generating with respect to (δ, p) for some p, then for each payoff vector $v \in W$, there is a public ex-post equilibrium which yields the payoff v regardless of the initial state ω.*

Pick an arbitrary smooth subset W of the interior of V^*. Thanks to the proposition above, to prove the folk theorem, it is sufficient to show that this set W is stochastically ex-post self-generating for patient players.

Pick an arbitrary point $v \in W$, and some positive numbers $\varepsilon > 0$ and $K > 0$. For each direction $\lambda \in \boldsymbol{R}^N$ with $|\lambda| = 1$, define the set $G_{v,\lambda,\delta}$ as in Figure 12. Formally, let

$$G_{v,\lambda,\delta} = \{\tilde{v} \in \boldsymbol{R}^N | \lambda \cdot v \geq \lambda \cdot \tilde{v} + (1-\delta), \ |v - \tilde{v}| < (1-\delta)K\}.$$

As λ changes, the set $G_{v,\lambda,\delta}$ changes and orbits the point v. Also, as δ increases, the set $G_{v,\lambda,\delta}$ shrinks and approaches the point v. Indeed, by the definition, the set $G_{v,\lambda,\delta}$ is in the $(1-\delta)K$-neighborhood of v.

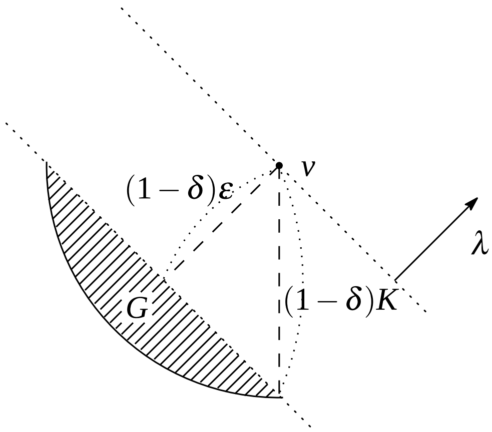

Figure 12: Set $G_{v,\lambda,\delta}$

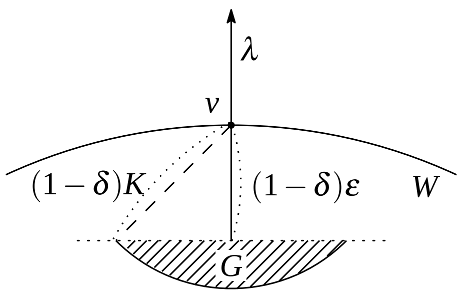

Figure 13: $G_{v,\lambda,\delta} \subseteq W$

As illustrated in Figure 13, for each payoff $v \in W$, there is at least one direction λ such that the set $G_{v,\lambda,\delta}$ is included in W. So if the payoff v is enforceable using continuation payoffs in this set $G_{v,\lambda,\delta}$, it is enforceable using continuation payoffs in the set W.

Generalizing this idea, Fudenberg and Yamamoto (2011b) show that *uniform decomposability* is sufficient for a set W to be self-generating with patient players.[13] It turns out that the same result holds in our setup; the following is the definition of uniform decomposability in our model. Here we write $G_{v,\lambda,\varepsilon,K,\delta}$ instead of $G_{v,\lambda,\delta}$, in order to emphasize that the set G depends on the parameters ε and K.

Definition B3. A subset W of $\boldsymbol{R}^N$ is *uniformly ex-post decomposable with respect to p* if there are $\varepsilon > 0$, $K > 0$, and $\overline{\delta} \in (0,1)$ such that for all $v \in W$, $\delta \in (\overline{\delta},1)$, and λ, there is a public strategy profile s and $w : H \to G_{v,\lambda,\varepsilon,K,\delta}$ such that (s,v) is stochastically ex-post enforceable with respect to (δ,p) using w.

In words, uniform decomposability requires that each payoff $v \in W$ can be achievable using continuation payoffs in the set $G_{v,\lambda,\varepsilon,K,\delta}$, regardless of the direction λ. The next lemma shows that uniform decomposability is indeed sufficient for the set W to be self-generating for patient players. The proof is similar to Fudenberg and Yamamoto (2011b), and hence omitted.

Lemma B8. *Suppose that a smooth and bounded subset W of $\boldsymbol{R}^N$ is uniformly ex-post decomposable with respect to p. Then there is $\overline{\delta} \in (0,1)$ such that for any payoff vector $v \in W$ and for any $\delta \in (\overline{\delta},1)$, there is a public ex-post equilibrium which yields the payoff v for any initial state ω.*

In what follows, we show that the set W is uniformly ex-post decomposable. That is, we show that each payoff $v \in W$ is enforceable using continuation payoffs in the set $G_{v,\lambda,\varepsilon,K,\delta}$, regardless of λ.

Throughout the proof, we use the following terminologies. A direction λ is *regular* if it has at least two non-zero components, and is *coordinate* if it has exactly one non-zero component. In other words, λ is a coordinate direction if $|\lambda_i| = 1$ for some i (because it automatically implies $\lambda_j = 0$ for all $j \neq i$).

For each strategy profile s and direction λ, let

$$\hat{v}(\delta,s,\lambda) = \max_{\omega\in\Omega} \lambda \cdot v^{\omega}(\delta,s) - \min_{\tilde{\omega}\in\Omega} \lambda \cdot v^{\tilde{\omega}}(\delta,s),$$

which measures how much the initial state can influence the welfare level $\lambda \cdot v^{\omega}(\delta,s)$ toward the direction λ. Proposition 3 ensures that $\hat{v}(\delta,s,\lambda)$ can be arbitrarily small when s is the optimal policy for some dummy belief $\tilde{\mu}$. Similarly, for

[13] This is a counterpart to the "local decomposability lemma" of FLM for infinitely repeated games. For more discussions, see Fudenberg and Yamamoto (2011b).

each δ and s, let

$$\hat{v}_i(\delta,s)=\max_{\omega\in\Omega}v_i^{\omega}(\delta,s)-\min_{\tilde{\omega}\in\Omega}v_i^{\tilde{\omega}}(\delta,s),$$

which measures how much the initial state can influence player i's payoff by the strategy profile s. Note that $\hat{v}_i(\delta,s)=\hat{v}_i(\delta,s,\lambda)$ for the coordinate direction λ with $\lambda_i=1$.

B.5.1 Step 1: Bound on Deviation Payoffs

In Section 3.6, we have claimed that the gain G^t by deviating from the prescribed strategy s^v in period t decreases with respect to t, at a geometric rate. Here we formally prove this result in the general model.

Consider the infinite-horizon game with the initial prior μ and the discount factor δ. Pick a public strategy profile s, where s_i is a pure strategy. Suppose that we are currently in period t, and that the past public history is h^{t-1}. Suppose that no one has deviated from s so far. If player i deviates in the continuation game, her gain is

$$G_i^t(\delta,\mu,s,h^{t-1})=\max_{\tilde{s}_i\in S_i}v_i^{\mu_i(h^{t-1}|\mu,s)}(\delta,\tilde{s}_i,s_{-i}|_{h^{t-1}})-v_i^{\mu_i(h^{t-1}|\mu,s)}(\delta,s|_{h^{t-1}}),$$

as in Section 3.6.

The following lemma shows that for a class of strategy profiles s, the gain G_i^t above decreases with respect to t, at a rate at least geometric with the parameter β. This implies that the sum of the gains,

$$G_i^*(\delta,s)=\sum_{t=1}^{\infty}\max_{\mu\in\triangle\Omega}\max_{h^{t-1}\in H^{t-1}}G_i^t(\delta,\mu,s,h^{t-1}),$$

is finite.

Lemma B9. *Suppose that the full support assumption holds. Then for each δ, μ, $\tilde{\mu}$, $t\geq 1$, public history h^{t-1}, public strategy s_{-i}, and pure public strategy $s_i\in\arg\max_{\tilde{s}_i\in S_i}v_i^{\tilde{\mu}}(\delta,\tilde{s}_i,s_{-i})$,*

$$G_i^t(\delta,\mu,s,h^{t-1})\leq\frac{\beta^{t-1}}{\overline{\pi}}\max_{\hat{\mu}\in\triangle\Omega}\left(\max_{\tilde{s}_i\in S_i}v_i^{\hat{\mu}}(\delta,\tilde{s}_i,s_{-i}|_{h^{t-1}})-v_i^{\hat{\mu}}(\delta,s|_{h^{t-1}})\right).$$

where β is chosen as in Proposition 1. This result implies that

$$G_i^*(\delta,s)\leq\frac{1}{(1-\beta)\overline{\pi}}\sup_{h^{t-1}\in H}\max_{\hat{\mu}\in\triangle\Omega}\left(\max_{\tilde{s}_i\in S_i}v_i^{\hat{\mu}}(\delta,\tilde{s}_i,s_{-i}|_{h^{t-1}})-v_i^{\hat{\mu}}(\delta,s|_{h^{t-1}})\right).$$

To interpret the above lemma, let s be the optimal policy which achieves player i's best payoff for some dummy belief $\tilde{\mu}$, just as in Section 3.6. Proposition 3 ensures that this strategy s (and its continuation strategy $s|_{h^t}$) is a pseudo-ergodic strategy which approximates player i's best payoff regardless of the true belief. That is, for any small $\varepsilon > 0$, when δ is large enough, we have $\max_{\tilde{s}_i \in S_i} v_i^{\hat{\mu}}(\delta, \tilde{s}_i, s_{-i}|_{h^t}) - v_i^{\hat{\mu}}(\delta, s|_{h^t}) < \overline{\pi}\varepsilon$ uniformly in t, h^t, and $\hat{\mu}$. Substituting this to the inequalities in the above lemma, we obtain (9) for each t, and hence $G_i^* \leq \frac{\varepsilon}{1-\beta}$. The same result holds when s is a minimax strategy for some dummy belief $\tilde{\mu}$, because Proposition 5 ensures that given any small $\varepsilon > 0$, when δ is large enough, we have $\max_{\tilde{s}_i \in S_i} v_i^{\hat{\mu}}(\delta, \tilde{s}_i, s_{-i}|_{h^t}) - v_i^{\hat{\mu}}(\delta, s|_{h^t}) < \overline{\pi}\varepsilon$ uniformly in t, h^t, and $\hat{\mu}$. This result is useful when we consider the problem associated with the "negative coordinate direction."

In what follows, we will prove the above lemma. We first provide a preliminary result. Pick the opponents' strategy s_{-i}. Suppose that player i's current belief is μ, but she is asked to play a strategy s_i which is a best reply to s_{-i} given a dummy belief $\tilde{\mu} \neq \mu$. Assume that this dummy belief is an interior belief so that $\tilde{\mu}(\omega) \geq \overline{\pi}$ for all ω. Since this strategy s_i is not necessarily a best reply given the true belief μ, player i can possibly increase her payoff by deviating from s_i; this gain is represented by

$$\max_{\tilde{s}_i \in S_i} v_i^{\mu}(\delta, \tilde{s}_i, s_{-i}) - v_i^{\mu}(\delta, s).$$

The following lemma provides a bound on this gain. When the true belief μ approaches $\tilde{\mu}$, the gain converges to zero at least linearly in $|\mu - \tilde{\mu}|$.

Lemma B10. *For each δ, for each μ, for each $\tilde{\mu}$ with $\tilde{\mu}(\omega) \geq \overline{\pi}$ for all ω, for each public strategy s_{-i}, and for each pure strategy $s_i \in \arg\max_{\tilde{s}_i \in S_i} v_i^{\tilde{\mu}}(\delta, \tilde{s}_i, s_{-i})$, we have*

$$\max_{\tilde{s}_i \in S_i} v_i^{\mu}(\delta, \tilde{s}_i, s_{-i}) - v_i^{\mu}(\delta, s) \leq \frac{|\mu - \tilde{\mu}|}{\overline{\pi}} \max_{\hat{\mu} \in \triangle\Omega} \left(\max_{\tilde{s}_i \in S_i} v_i^{\hat{\mu}}(\delta, \tilde{s}_i, s_{-i}) - v_i^{\hat{\mu}}(\delta, s) \right).$$

To illustrate the proof idea, consider the case in which there are only two states, ω_1 and ω_2. Then each belief μ is represented by a single number; let μ denote the probability on the state ω_1. Pick the opponents' strategy s_{-i} arbitrarily. We know from Lemma B2 that player i's best payoff $\max_{\tilde{s}_i \in S_i} v_i^{\mu}(\delta, \tilde{s}_i, s_{-i})$ against this strategy s_{-i} is convex with respect to the belief μ, as in Figure 14. Pick an arbitrary dummy belief $\tilde{\mu}$ as stated in the lemma, and let s_i be a best reply to s_{-i} given this dummy belief $\tilde{\mu}$. The dashed line in the figure shows player i's payoff $v_i^{\mu}(\delta, s)$ achieved by this strategy s_i for each true belief μ. This dashed line must intersect with the curve at $\mu = \tilde{\mu}$ because s_i is a best reply to s_{-i} given the initial prior $\mu = \tilde{\mu}$. Also this dashed line must be below the curve, because for each current belief μ, the curve gives the highest payoff for player i against s_{-i}. Taken together, the dashed line must be tangential to the curve at $\mu = \tilde{\mu}$ as in the figure.

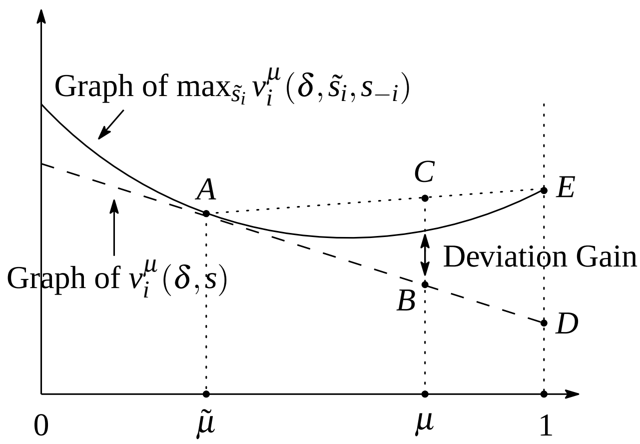

Figure 14: Bound on the Gain

Now, pick a true belief μ arbitrarily; without loss of generality, we assume $\mu > \tilde{\mu}$. Given this true belief μ, the gain by deviating from s_i is represented by the distance between the curve and the dashed line at μ. As the figure shows, since the curve is convex, this distance is less than the length $\overline{BC}$. (Here, $\overline{XY}$ represents the distance between the two points X and Y.) Note that this length $\overline{BC}$ is equal to $\frac{\mu-\tilde{\mu}}{1-\tilde{\mu}}\overline{DE}$, because that the two triangles ABC and ADE in the figure are similar with the ratio of corresponding sides $\frac{\mu-\tilde{\mu}}{1-\tilde{\mu}}$. Hence the gain by deviating from s_i is at most

$$\frac{\mu - \tilde{\mu}}{1 - \tilde{\mu}}\overline{DE}.$$

This immediately implies the lemma, because $1 - \tilde{\mu} \geq \overline{\pi}$ and

$$\overline{DE} \leq \max_{\hat{\mu} \in \triangle\Omega} \left(\max_{\tilde{s}_i \in S_i} v_i^{\hat{\mu}}(\delta, \tilde{s}_i, s_{-i}) - v_i^{\hat{\mu}}(\delta, s) \right).$$

The proof for the state space Ω is as follows:

Proof. Fix i and δ. For each s_{-i}, let $v_i^{\mu}(s_{-i})$ be as in the proof of Proposition 4, that is, $v_i^{\mu}(s_{-i}) = \max_{s_i \in S_i} v_i^{\mu}(\delta, s_i, s_{-i})$. As Lemma B2 shows, $v_i^{\mu}(s_{-i})$ is convex with respect to μ.

Take μ, $\tilde{\mu}$, s_{-i}, s_i as stated in the lemma. If $\mu = \tilde{\mu}$, the result obviously holds. So assume that $\mu = \tilde{\mu}$, and then pick a boundary point μ^* of $\triangle\Omega$ such that $\mu = \kappa\mu^* + (1-\kappa)\tilde{\mu}$ for some $\kappa \in (0,1]$.

As in the proof of Proposition 4, let $v_i^{\mu}(s_{-i})$ denote the best payoff for player i when the initial prior is μ and the opponents play s_{-i}. That is, let $v_i^{\mu}(s_{-i}) = \max_{\tilde{s}_i \in S_i} v_i^{\mu}(\delta, \tilde{s}_i, s_{-i})$. Then we have

$$\begin{aligned} v_i^{\mu}(s_{-i}) - v_i^{\mu}(\delta, s) \leq & \kappa v_i^{\mu^*}(s_{-i}) + (1-\kappa) v_i^{\tilde{\mu}}(s_{-i}) - v_i^{\mu}(\delta, s) \\ = & \kappa v_i^{\mu^*}(s_{-i}) + (1-\kappa) v_i^{\tilde{\mu}}(s_{-i}) - \kappa v_i^{\mu^*}(\delta, s) - (1-\kappa) v_i^{\tilde{\mu}}(\delta, s) \end{aligned}$$

Here the inequality follows from the fact that $v_i^{\mu}(s_{-i})$ is convex with respect to μ, and the second equality from the fact that $v_i^{\mu}(\delta, s)$ is linear with respect to μ. Since s_i is a best reply to s_{-i} given the belief $\tilde{\mu}$, we have $v_i^{\tilde{\mu}}(s_{-i}) = v_i^{\tilde{\mu}}(\delta, s)$. Plugging this into the above inequality, we have

$$v_i^{\mu}(s_{-i}) - v_i^{\mu}(\delta, s) \leq \kappa \left(v_i^{\mu^*}(s_{-i}) - v_i^{\mu^*}(\delta, s) \right) \leq \kappa \max_{\hat{\mu} \in \triangle\Omega} \left(v_i^{\hat{\mu}}(s_{-i}) - v_i^{\hat{\mu}}(\delta, s) \right).$$

Since $\mu = \kappa\mu^* + (1-\kappa)\tilde{\mu}$, $\kappa = \frac{|\mu - \tilde{\mu}|}{|\mu^* - \tilde{\mu}|}$. Now, we know that $|\mu^* - \tilde{\mu}| \geq \overline{\pi}$, because μ^* is a boundary point while $\tilde{\mu}(\omega) \geq \overline{\pi}$ for all ω by the assumption. Hence, we have $\kappa \leq \frac{|\mu - \tilde{\mu}|}{\overline{\pi}}$. Plugging this into the previous inequality, we obtain the result. *Q.E.D.*

Now we prove Lemma B9. For $t = 1$, it is obvious that

$$G_i^1(\delta, \mu, s, h^0) \leq \frac{1}{\overline{\pi}} \max_{\hat{\mu} \in \triangle\Omega} \left(\max_{\tilde{s}_i \in S_i} v_i^{\hat{\mu}}(\delta, \tilde{s}_i, s_{-i}) - v_i^{\hat{\mu}}(\delta, s) \right).$$

For $t \geq 2$, by replacing μ, $\tilde{\mu}$, s in Lemma B10 with $\mu(h^{t-1}|\mu, s)$, $\mu(h^{t-1}|\tilde{\mu}, s)$, and $s|_{h^{t-1}}$, we obtain

$$G_i^t(\delta, \mu, s, h^t) \leq \frac{|\mu(h^{t-1}|\mu, s) - \mu(h^{t-1}|\tilde{\mu}, s)|}{\overline{\pi}} \max_{\hat{\mu} \in \triangle\Omega} \left(\max_{\tilde{s}_i \in S_i} v_i^{\hat{\mu}}(\delta, \tilde{s}_i, s_{-i}|_{h^{t-1}}) - v_i^{\hat{\mu}}(\delta, s|_{h^{t-1}}) \right)$$

Here the full support assumption ensures that $\mu(h^{t-1}|\tilde{\mu}, s)$ is indeed an interior belief which puts at least $\overline{\pi}$ on each state ω. Proposition 1 ensures that $|\mu(h^{t-1}|\mu, s) - \mu(h^{t-1}|\tilde{\mu}, s)| \leq \beta^{t-1}$, so we obtain the desired inequality.

B.5.2 Step 2: Strategies with Cross-State Full-Rank Conditions

Let S^{PFR} be the set of all public strategy profiles s such that the induced action profile $s(h^t)$ has cross-state pairwise full rank for each history h^t and such that $|\{\alpha|\alpha = s(h^t)\ \ \exists t \exists h^t\}| \leq |A|$. The second constraint requires that s cannot induce more than $|A|$ different actions.

Also, for each $\eta > 0$, let $S^{\text{IFR}}(\eta)$ be the set of all public strategy profiles s such that given any history h^t, any action profile within the η-neighborhood of $s(h^t)$ has cross-state individual full rank; i.e., for any α which does not have cross-state pairwise full rank, we have $|s(h^t) - \alpha| \geq \eta$.

As in FLM, under (PFR), the set of action profiles α which have cross-state pairwise full rank is open and dense in the set of all action profiles, that is, any action profile can be approximated by an action profile α which has cross-state pairwise full rank. This in turn implies that any pure public strategy s^* can be approximated by some strategy $s \in S^{\text{PFR}}$. Note that the constraint $|\{\alpha|\alpha = s(h^t)\ \ \exists t \exists h^t\}| \leq |A|$ is indeed satisfied by letting $s(h^t) = s(\tilde{h}^{\tilde{t}})$ whenever $s^*(h^t) = s^*(\tilde{h}^{\tilde{t}})$. Similarly, any (possibly mixed) strategy can be approximated by some strategy in the set $S^{\text{IFR}}(\eta)$ for sufficiently small η.

Recall that our assumptions, (PFR) and (IFR), are weaker than requiring all mixed action profiles to have cross-state pairwise full-rank condition. Accordingly, the optimal policy s (which achieves the score for some direction λ) may involve some action profile which does not satisfy the full-rank condition. However, the above discussion suggests that we can "perturb" this optimal policy a bit so that the resulting strategy still approximates the score and satisfies the full-rank condition. Similarly, a minimax strategy profile can be perturbed so that the resulting strategy satisfies the cross-state individual full-rank condition. Formally, we obtain the following lemma.

Lemma B11. *Suppose that the full support assumption holds, and that (IFR) and (PFR) hold. Then for any smooth subset W of the interior of V^*, there is $\eta > 0$ such that for any $C > 0$, there are $\varepsilon > 0$, $p \in (0,1)$, and $\overline{\delta} \in (0,1)$ such that for each $\delta \in (\overline{\delta}, 1)$, the following properties hold:*

(i) For every regular direction λ, there is a strategy profile $s \in S^{PFR} \cap S^{IFR}(\eta)$

such that for each ω,

$$\lambda \cdot v^{\omega}(p\delta, s) > \max_{v \in W} \lambda \cdot v + C\hat{v}(p\delta, s, \lambda) + \varepsilon.$$

(ii) *For each i, there is a pure public strategy profile s such that for each ω,*

$$v_i^{\omega}(p\delta, s) > \max_{v \in W} v_i + C\hat{v}_i(p\delta, s) + CG_i^*(p\delta, s) + \varepsilon$$

and

$$G_i^*(p\delta, s) < \varepsilon.$$

(iii) *For each i, there is a public strategy profile $s \in S^{IFR}(\eta)$ such that s_i is a pure strategy and such that for each ω,*

$$v_i^{\omega}(p\delta, s) < \min_{v \in W} v_i - C\hat{v}_i(p\delta, s) - CG_i^*(p\delta, s) - \varepsilon$$

and

$$G_i^*(p\delta, s) < \varepsilon.$$

Clause (i) asserts that for each regular direction λ, there is a pseudo-ergodic strategy s which approximates the score regardless of the state ω and satisfies appropriate full-rank conditions. To see this, let $s^{\tilde{\mu}}$ be the optimal policy for some dummy belief $\tilde{\mu}$. We know that this strategy $s^{\tilde{\mu}}$ is a pseudo-ergodic strategy which approximates the score regardless of the state ω; this implies that the strategy $s^{\tilde{\mu}}$ satisfies the inequality in clause (i), because $\lambda \cdot v^{\omega}(p\delta, s^*) > \max_{v \in W} \lambda \cdot v$ for each ω and $\hat{v}(p\delta, s^*, \lambda) \approx 0$. However, some action profiles used by the strategy $s^{\tilde{\mu}}$ may not satisfy full-rank conditions, so we may have $s^{\tilde{\mu}} \notin S^{\mathrm{PFR}} \cap S^{\mathrm{IFR}}(\eta)$. Clause (i) ensures that in such a case, we can perturb this strategy $s^{\tilde{\mu}}$ so that the resulting strategy s still satisfies the inequality and $s \in S^{\mathrm{PFR}} \cap S^{\mathrm{IFR}}(\eta)$.

Clause (ii) considers the case with the coordinate direction λ with $\lambda_i = 1$. The strategy profile s in the clause (ii) is simply the optimal policy for some dummy belief μ. (We do not need to perturb it.) This strategy s indeed satisfies the inequalities in clause (ii), because Proposition 3 and Lemma B9 ensure that $\hat{v}_i(p\delta, s)$ and $G_i^*(p\delta, s)$ approximate zero. Note also that $s \in S^{\mathrm{IFR}}(\eta)$, because s is a pure strategy profile and (IFR) holds.

Clause (iii) implies that there is a strategy profile which approximately minimaxes player i regardless of the state ω. To see this, let $s^{\tilde{\mu}}$ be the minimax strategy profile given some dummy belief $\tilde{\mu}$. This strategy $s^{\tilde{\mu}}$ satisfies the inequalities in clause (iii); indeed, Proposition 5 and Lemma B9 ensure that $\hat{v}_i(p\delta,s)$ and $G_i^*(p\delta,s)$ approximate zero, and that $v_i^{\omega}(p\delta,s^{\tilde{\mu}}) < \min_{v\in W} v_i$ for each ω However, the minimax strategy $s^{\tilde{\mu}}$ may use actions which do not satisfy cross-state individual full rank, so we may have $s \notin S^{\mathrm{IFR}}(\eta)$. Clause (iii) ensures that in such a case, we can perturb this strategy $s^{\tilde{\mu}}$ so that the resulting strategy s still satisfies the inequalities and $s \in S^{\mathrm{IFR}}(\eta)$.

In what follows, we prove the above lemma. The proof is more involved than that in FLM, since a perturbation of an action in some period t influences both the stage-game payoff in period t and the continuation payoff from period $t+1$ on through the distribution of (ω^{t+1},y^t). Readers who are less interested in such technical issues may want to skip it and go to Step 3.

We begin with presenting some preliminary lemmas; roughly, they show that the scores and the minimax payoffs can be approximated by perturbed strategies which satisfy appropriate full-rank conditions.

The first lemma shows that when η is not too large, for every direction λ, the score can be approximated by a strategy $s \in S^{\mathrm{PFR}} \cap S^{\mathrm{IFR}}(\eta)$.

Lemma B12. *Suppose that the full support assumption, (IFR), and (PFR) hold. Then there is $\overline{\eta} > 0$ such that for any $\varepsilon > 0$ and $\eta \in (0,\overline{\eta})$, there is $\overline{\delta} \in (0,1)$ such that for any λ and $\delta \in (\overline{\delta},1)$, there is $s \in S^{PFR} \cap S^{IFR}(\eta)$ such that for each μ,*

$$\left| \lambda \cdot v^{\mu}(\delta,s) - \max_{v\in V^{\mu}(\delta)} \lambda \cdot v \right| < \varepsilon.$$

Proof. Take $\overline{\eta} > 0$ so that any (possibly mixed) action profile α which is $\overline{\eta}$-close to some pure action profile has cross-state individual full rank. The existence of such $\overline{\eta}$ is guaranteed, since (IFR) implies that any action profile which approximates some pure action profile has cross-state individual full rank.

As shown in FLM, under (PFR), the set of action profiles which have cross-state pairwise full rank is open and dense in the set of all action profiles. Hence we can approximate the score using strategies in the set S^{PFR}; in particular, Proposition 3 ensures that for any $\varepsilon > 0$ and $\eta \in (0,\overline{\eta})$, there is $\overline{\delta} \in (0,1)$ such that for each λ and $\delta \in (\overline{\delta},1)$, there is $s \in S^{\mathrm{PFR}}$ such that

$$\left| \lambda \cdot v^{\mu}(\delta,s) - \max_{v\in V^{\mu}(\delta)} \lambda \cdot v \right| < \varepsilon$$

for each μ and such that for each history h^t, the action profile $s(h^t)$ is η-close to some pure action profile. By the definition of $\overline{\eta}$, this strategy profile s is in the set $S^{\text{IFR}}(\eta)$, and hence the result follows. *Q.E.D.*

The next three lemmas are about the minimax payoffs by perturbed strategies. For each $\eta > 0$, let $S_{-i}^{\text{IFR}}(\eta)$ be the set of all public strategies s_{-i} such that for each pure public strategy s_i and history h^t, any action profile within the η-neighborhood of $s(h^t)$ has cross-state individual full rank. Then let

$$\underline{v}_i^{\mu}(\delta,\eta) = \inf_{s_{-i}\in S_{-i}^{\text{IFR}}(\eta)} \max_{s_i\in S_i} v_i^{\mu}(\delta,s).$$

In words, $\underline{v}_i^{\mu}(\delta,\eta)$ is the minimax payoff when the opponents' play is restricted to $S_{-i}^{\text{IFR}}(\eta)$. If $S_{-i}^{\text{IFR}}(\eta)$ is empty, then let $\underline{v}_i^{\mu}(\delta,\eta) = \infty$. The following lemmas show that Propositions 4 and 5 remain valid even when we consider the restricted strategy space $S_{-i}^{\text{IFR}}(\eta)$. Note also that the convergence rate is uniform in η.

Lemma B13. *Suppose that the full support assumption holds. Then for each $\varepsilon > 0$, there is $\overline{\delta} \in (0,1)$ such that $|\underline{v}_i^{\mu}(\delta,\eta) - \underline{v}_i^{\tilde{\mu}}(\delta,\eta)| < \varepsilon$ for any $\delta \in (\overline{\delta},1)$, μ, $\tilde{\mu}$, and $\eta > 0$ with $S_{-i}^{IFR}(\eta) \neq \emptyset$.*

Proof. The proof of Proposition 4 is valid without any change even when we consider the restricted strategy space $S_{-i}^{\text{IFR}}(\eta)$. In particular, the bounds in the lemmas in the proof do not depend on η. Hence the result follows. *Q.E.D.*

Lemma B14. *Suppose that the full support assumption holds. Then for each $\varepsilon > 0$, there is $\overline{\delta} \in (0,1)$ such that for each $\delta \in (\overline{\delta},1)$, for each $\tilde{\mu}$, and for each $\eta > 0$ with $S_{-i}^{IFR}(\eta) \neq \emptyset$, there is $s_{-i}^{\tilde{\mu}} \in S_{-i}^{IFR}(\eta)$ such that for each pure strategy $s_i^{\tilde{\mu}} \in \arg\max_{\tilde{s}_i\in S_i} v_i^{\tilde{\mu}}(\delta,\tilde{s}_i,s_{-i}^{\tilde{\mu}})$,*

$$\left| v_i^{\mu}(\delta,s^{\tilde{\mu}}|_{h^t}) - \underline{v}_i^{\mu}(\delta,\eta) \right| < \varepsilon \tag{27}$$

and

$$\max_{\tilde{s}_i\in S_i} v_i^{\mu}(\delta,\tilde{s}_i,s_{-i}^{\tilde{\mu}}|_{h^t}) - v_i^{\mu}(\delta,s^{\tilde{\mu}}|_{h^t}) < \varepsilon \tag{28}$$

for each μ, t, and h^{t-1}.

Proof. The proof of Proposition 5 is valid without any change. *Q.E.D.*

As in FLM, under (IFR), the set of actions α_{-i} such that (a_i,α_{-i}) has cross-state individual full rank for all a_i is open and dense in the set of all mixed actions. Hence, for a fixed discount factor δ, the minimax payoff with the restricted strategy space $S_{-i}^{\text{IFR}}(\eta)$ approximates the minimax payoff $\underline{v}_i^{\mu}(\delta)$ as $\eta \to 0$. The following lemma strengthens this result; it shows that the limit minimax payoff $\underline{v}_i(\eta) = \lim_{\delta\to 1} \underline{v}_i^{\mu}(\delta,\eta)$ with the restricted strategy space approximates the limit minimax payoff as $\eta \to 0$. (The existence of $\lim_{\delta\to 1} \underline{v}_i^{\mu}(\delta,\eta)$ can be proved as in Yamamoto (2016). Lemma B13 ensures that this limit is independent of μ.)

Lemma B15. *Suppose that the full support assumption and (IFR) hold. Then for any $\varepsilon > 0$, there is $\overline{\eta} > 0$ such that $|\underline{v}_i(\eta) - \underline{v}_i| < \varepsilon$ for any $\eta \in (0,\overline{\eta})$ and μ.*

Proof. By the definition, $\underline{v}_i(\eta)$ is non-increasing in η and at least $\underline{v}_i$. Hence, it is sufficient to show that there is μ such that for any $\varepsilon > 0$, there is $\eta > 0$ such that $|\underline{v}_i(\eta) - \underline{v}_i| < \varepsilon$. In other words, it is sufficient to show that there are μ, $\varepsilon > 0$, $\eta > 0$ and $\overline{\delta} \in (0,1)$ such that for any $\delta \in (\overline{\delta},1)$, $|\underline{v}_i^{\mu}(\delta,\eta) - \underline{v}_i^{\mu}(\delta)| < \varepsilon$.

Pick μ and $\varepsilon > 0$ arbitrarily. Let C' and $\tilde{C}'$ be as in the proof of Proposition 4. Then pick $\eta > 0$ sufficiently small so that $2\eta|A_{-i}|(\overline{g} + \frac{C'+\tilde{C}'}{\pi}) < \frac{\varepsilon}{3}$.

Pick an arbitrary $\delta \in (0,1)$, and choose s_{-i}^{μ} as in Lemma B3. Then take a perturbed strategy $s_{-i}^{\infty} \in S_{-i}^{\text{IFR}}(\eta)$ such that after every history h^t, the action $s_{-i}^{\infty}(h^t)$ is 2η-close to $s_{-i}^{\mu}(h^t)$. Such a strategy s_{-i}^{∞} indeed exists because η is sufficiently small. Then for each t, let s_{-i}^{t} be the strategy such that only the actions up to period t are perturbed; i.e., $s_{-i}^{t}(h^{\tilde{t}}) = s_{-i}^{\infty}(h^{\tilde{t}})$ for each $\tilde{t} \leq t-1$ and $s_{-i}^{t}(h^{\tilde{t}}) = s_{-i}^{\mu}(h^{\tilde{t}})$ for each $\tilde{t} \geq t$. Let $s_{-i}^{0} = s_{-i}^{\mu}$. For each t, let s_i^t be a best reply to s_{-i}^{t} given the initial prior μ. Also let s_i^{∞} be a best reply to s_{-i}^{∞} given μ. Then we have

$$\begin{aligned}
\left|\underline{v}_i^{\mu}(\delta,\eta) - \underline{v}_i^{\mu}(\delta)\right| &\leq \left|v_i^{\mu}(\delta,s^{\infty}) - \underline{v}_i^{\mu}(\delta)\right| \\
&< \left|v_i^{\mu}(\delta,s^{\infty}) - \left\{v_i^{\mu}(\delta,s^{0}) - (1-\delta)\right\}\right| \\
&\leq \left|v_i^{\mu}(\delta,s^{\infty}) - v_i^{\mu}(\delta,s^{0})\right| + (1-\delta) \\
&\leq \sum_{t=1}^{\infty}\left|v_i^{\mu}(\delta,s^{t}) - v_i^{\mu}(\delta,s^{t-1})\right| + 2(1-\delta).
\end{aligned}$$

Here, the first inequality follows from $v_i^{\mu}(\delta,s^{\infty}) \geq \underline{v}_i^{\mu}(\delta,\eta) \geq \underline{v}_i^{\mu}(\delta)$. The second inequality follows from Lemma B3, which ensures $|\underline{v}_i^{\mu}(\delta) - v_i^{\mu}(\delta,s^0)| < 1-\delta$. The last inequality follows from the fact that for a fixed δ, $v_i^{\mu}(\delta,s^t)$ converges to $v_i^{\mu}(\delta,s^{\infty})$ as $t \to \infty$ due to discounting.

To complete the proof, it is sufficient to show that

$$\left|v_i^{\mu}(\delta,s^t)-v_i^{\mu}(\delta,s^{t-1})\right|\leq\frac{\delta^{t-1}(1-\delta)\varepsilon}{3}\tag{29}$$

for each t. Indeed, if so, then plugging (29) to the previous inequality,

$$\left|\underline{v}_i^{\mu}(\delta,\eta)-\underline{v}_i^{\mu}(\delta)\right|<\sum_{t=1}^{\infty}\frac{\delta^{t-1}(1-\delta)\varepsilon}{3}+2(1-\delta)=\frac{\varepsilon}{3}+2(1-\delta).$$

Since this inequality holds for every δ, by taking sufficiently large δ, we obtain the desired inequality $|\underline{v}_i^{\mu}(\delta,\eta)-\underline{v}_i^{\mu}(\delta)|<\varepsilon$.

So what remains is to prove (29). Note that s_{-i}^t and s_{-i}^{t-1} differ only in the action in period t, which influences the stage-game payoff in period t and the continuation payoff from period $t+1$ through the distribution of (y^t,ω^{t+1}). Accordingly, given player i's strategy s_i^t, if we change the opponents' strategy from s_{-i}^t to s_{-i}^{t-1}, it changes player i's payoff by

$$\begin{aligned}
&v_i^{\mu}(\delta,s^t)-v_i^{\mu}(\delta,s_i^t,s_{-i}^{t-1})\\
\leq&\sum_{h^{t-1}\in H^{t-1}}\Pr(h^{t-1}|\tilde{\mu},s^t)\sum_{a_{-i}\in A_{-i}}\left|s_{-i}^t(h^{t-1})[a_{-i}]-s_{-i}^{t-1}(h^{t-1})[a_{-i}]\right|\\
&\times\left\{\begin{array}{l}(1-\delta)\delta^{t-1}\left(\max\limits_{\omega,a}g_i^{\omega}(a)-\min\limits_{\tilde{\omega},\tilde{a}}g_i^{\tilde{\omega}}(\tilde{a})\right)\\ +\delta^t\left(\max\limits_{\omega^{t+1},y^t}v_i^{\omega^{t+1}}(\delta,s^t|_{(h^{t-1},y^t)})-\min\limits_{\tilde{\omega}^{t+1},\tilde{y}^t}v_i^{\tilde{\omega}^{t+1}}(\delta,s^t|_{(h^{t-1},\tilde{y}^t)})\right)\end{array}\right\}\\
\leq&2\eta|A_{-i}|\left\{(1-\delta)\delta^{t-1}\overline{g}+\delta^t(1-\delta)\left(\frac{C'}{\delta\overline{\pi}}+\frac{\tilde{C}'}{\overline{\pi}}\right)\right\}\\
\leq&(1-\delta)2\eta|A_{-i}|\left(\overline{g}+\frac{C'+\tilde{C}'}{\overline{\pi}}\right).
\end{aligned}$$

Here the first inequality follows because of the reasoning we argued; s_{-i}^t and s_{-i}^{t-1} differ only in the action in period t, which influences the stage-game payoff in period t and the continuation payoff from period $t+1$ through the distribution of (y^t,ω^{t+1}). The first term in the curly bracket captures how much the stage-game payoff in period t can change, and the second term captures how much the continuation payoff from period $t+1$ can change. (Given (y^t,ω^{t+1}), the continuation payoff from period $t+1$ is $v_i^{\omega^{t+1}}(\delta,s^t|_{(h^{t-1},y^t)})$.) To obtain the second inequality, we use the fact that the distance between $s_{-i}^t(h^{t-1})$ and $s_{-i}^{t-1}(h^{t-1})$ is at most 2η

so that $|s^t_{-i}(h^{t-1})[a_{-i}] - s^{t-1}_{-i}(h^{t-1})[a_{-i}]| \leq 2\eta$ for each a_{-i}. We also use Lemmas B6 and B7 to show that the difference between the maximum of $v_i^{\omega^{t+1}}(\delta, s^t|_{h^t})$ and the minimum is at most $\frac{1}{\overline{\pi}}\left(\frac{1-\delta}{\delta}C' + (1-\delta)\tilde{C}'\right)$. Indeed, Lemma B6 implies that

$$\left|\overline{v}_i(s^{\mu^*}_{-i}|_{h^*}) + (1-\delta) - v_i^{\mu_i(h^t|\mu,s^t)}(\delta, s^t|_{h^t})\right| \leq \frac{1-\delta}{\delta}C' + (1-\delta)\tilde{C}'$$

because $s^t_{-i}|_{h^t} = s^{\mu}_{-i}|_{h^t}$ and $s^t_i|_{h^t}$ is a best reply to $s^t_{-i}|_{h^t}$ given $\mu_i(h^t|\mu, s^t)$. Then Lemma B7 and the fact that $\mu_i(h^t|\mu, s^t)$ puts at least probability $\overline{\pi}$ on each state ensure that

$$\left|\overline{v}_i(s^{\mu^*}_{-i}|_{h^*}) + (1-\delta) - v_i^{\omega^{t+1}}(\delta, s^t|_{h^t})\right| \leq \frac{1}{\overline{\pi}}\left(\frac{1-\delta}{\delta}C' + (1-\delta)\tilde{C}'\right)$$

for all ω^{t+1}, as desired.

Plugging $2\eta|A_{-i}|(\overline{g} + \frac{C'+\tilde{C}'}{\overline{\pi}}) < \frac{\varepsilon}{3}$ into the above result, we have

$$\left|v_i^{\mu}(\delta, s^t) - v_i^{\mu}(\delta, s^t_i, s^{t-1}_{-i})\right| \leq \frac{(1-\delta)\delta^{t-1}\varepsilon}{3}. \tag{30}$$

A similar argument shows that

$$\left|v_i^{\mu}(\delta, s^{t-1}_i, s^t_{-i}) - v_i^{\mu}(\delta, s^{t-1})\right| \leq \frac{(1-\delta)\delta^{t-1}\varepsilon}{3}. \tag{31}$$

Now we are ready to verify (29). Suppose that $v_i^{\mu}(\delta, s^t) \geq v_i^{\mu}(\delta, s^{t-1})$. Then we must have $v_i^{\mu}(\delta, s^t) \geq v_i^{\mu}(\delta, s^{t-1}) \geq v_i^{\mu}(\delta, s^t_i, s^{t-1}_{-i})$, and thus (30) implies (29). Similarly, when $v_i^{\mu}(\delta, s^t) < v_i^{\mu}(\delta, s^{t-1})$, we have $v_i^{\mu}(\delta, s^{t-1}) > v_i^{\mu}(\delta, s^t) \geq v_i^{\mu}(\delta, s^{t-1}_i, s^t_{-i})$ so that (31) implies (29). *Q.E.D.*

Now we prove Lemma B11. Take W as stated. Take $\overline{\eta} > 0$ as in Lemma B12. From Lemma B15, there is $\eta \in (0, \overline{\eta})$ such that $\min_{v \in W} v_i > \underline{v}_i(\eta)$ for all i. Pick such η, and pick $C > 0$ arbitrarily. Then from Lemmas B9 and B14, there are $\varepsilon > 0$ and $p_1 \in (0,1)$ such that for each $p \in (p_1, 1)$ and i, there is $s_{-i} \in S^{\mathrm{IFR}}_{-i}(\eta)$ and a pure public strategy s_i such that for each ω,

$$\min_{v \in W} v_i > v_i^{\omega}(p, s) + C\hat{v}_i(p, s) + CG_i^*(p, s) + \varepsilon$$

and

$$G_i^*(p, s) < \varepsilon.$$

This s satisfies the condition stated in clause (iii) for $\delta = 1$.

Pick $\varepsilon > 0$ and p_1 as stated above. Since ε can be arbitrarily small, without loss of generality, we assume

$$\left| \max_{v \in V} \lambda \cdot v - \max_{v \in W} \lambda \cdot v \right| > (5C + 3)\varepsilon \tag{32}$$

for all λ.

From Proposition 6 of Yamamoto (2016), there is $p_2 \in (p_1, 1)$ such that

$$\left| \max_{v \in V} \lambda \cdot v - \max_{v \in V^{\omega}(p)} \lambda \cdot v \right| < \varepsilon \tag{33}$$

for all λ, $p \in (p_2, 1)$, and ω. Also, it follows from Lemma B12 that there is $p_3 \in (p_2, 1)$ such that for each λ and $p \in (p_3, 1)$, there is a pure public strategy profile $s \in S^{\text{PFR}} \cap S^{\text{IFR}}(\eta)$ such that for each ω,

$$\left| \lambda \cdot v^{\omega}(p, s) - \max_{v \in V^{\omega}(p)} \lambda \cdot v \right| < \varepsilon. \tag{34}$$

Take a regular direction λ and $p \in (p_3, 1)$ arbitrarily. Take $s \in S^{\text{PFR}} \cap S^{\text{IFR}}(\eta)$ as above. From (33) and (34), we have

$$\hat{v}(p, s, \lambda) = \max_{\omega \in \Omega} \lambda \cdot v^{\omega}(p, s) - \min_{\omega \in \Omega} \lambda \cdot v^{\omega}(p, s) < 4\varepsilon. \tag{35}$$

Then for each ω,

$$\begin{aligned}
\lambda \cdot v^{\omega}(p, s) &> \max_{v \in V^{\omega}(p)} \lambda \cdot v - \varepsilon \\
&> \max_{v \in V} \lambda \cdot v - 2\varepsilon \\
&> \max_{v \in W} \lambda \cdot v + (5C + 1)\varepsilon \\
&> \max_{v \in W} \lambda \cdot v + C\hat{v}(p, s, \lambda) + \varepsilon.
\end{aligned}$$

Here, the first inequality follows from (34), the second from (33), the third from (32), and the last from (35). So the strategy profile s satisfies the condition stated in clause (i) for $\delta = 1$.

From Lemma B9, we know that there is $p_4 \in (p_3, 1)$ such that for each $p \in (p_4, 1)$ and i, there is a pure public strategy profile s such that (34) holds for all ω

and for the positive coordinate direction λ with $\lambda_i = 1$, and such that $G_i^*(p,s) < \varepsilon$. This s satisfies the condition stated in clause (ii) for $\delta = 1$. Indeed, for each ω,

$$\begin{aligned}
v_i^\omega(p,s) &> \max_{v \in V^\omega(p)} v_i - \varepsilon \\
&> \max_{v \in V} v_i - 2\varepsilon \\
&> \max_{v \in W} v_i + (5C+1)\varepsilon \\
&> \max_{v \in W} v_i + C\hat{v}_i(p,s) + CG_i^*(p,s) + \varepsilon
\end{aligned}$$

Here, the first inequality follows from (34), the second from (33), the third from (32), and the last from (33) and $G_i^*(p,s) < \varepsilon$.

So far we have shown that the result holds when $p \in (p_4, 1)$ and $\delta = 1$. By continuity, the same result holds even if δ is slightly less than one. This completes the proof.

B.5.3 Step 3: Enforceability for Regular Directions

Recall that our goal is to show uniform enforceability of W, that is, we want to show that each payoff vector $v \in W$ is enforceable using continuation payoffs in the set $G_{v,\lambda,\varepsilon,K,\delta}$, regardless of the parameter (direction) λ. The following lemma shows that $v \in W$ is indeed enforceable for any regular direction λ. This result is a generalization of the one presented in Section 3.5.

Lemma B16. *For each $\eta > 0$, there is $C > 0$ such that for each regular direction λ, for each $p \in (0,1)$ and for each $s \in S^{PFR} \cap S^{IFR}(\eta)$, there is $K > 0$ such that for each $\delta \in (0,1)$ and for each $v \in V$, there is w such that*

(i) *(s,v) is stochastically ex-post enforceable with respect to (δ, p) by w,*

(ii) *For all t and h^t,*

$$\lambda \cdot w(h^t) \le \lambda \cdot v - \frac{1-\delta}{(1-p)\delta}\left(\min_{\omega \in \Omega} \lambda \cdot v^\omega(p\delta, s) - \lambda \cdot v - C\hat{v}(p\delta, s, \lambda)\right),$$

(iii) *$|v - w(h^t)| < \frac{1-\delta}{(1-p)\delta}K$ for all t and h^t.*

To interpret this lemma, pick a regular direction λ and pick p and δ close to one. For the sake of the exposition, let v be a boundary point of W with the unit normal λ (see Figure 15). Choose s as in Lemma B11(i) so that it approximates the score regardless of the state ω. The above lemma asserts that this pair (s,v) is enforceable using continuation payoffs in the shaded area in Figure 15, where $l^* = \min_{\omega \in \Omega} \lambda \cdot v^\omega(p\delta, s) - \lambda \cdot v - C\hat{v}(p\delta, s, \lambda)$. Note that this length l^* is approximately equal to the length l in the figure, because s approximates the same payoff regardless of ω so that $\hat{v}(p\delta, s, \lambda) \approx 0$.

Now, choose $\varepsilon > 0$ as in Lemma B11, and let $\tilde{\varepsilon} = \frac{\varepsilon}{1-p}$ and $\tilde{K} = \frac{2K}{1-p}$. Then the shaded area in the figure is included in the set $G_{v,\lambda,\tilde{\varepsilon},\tilde{K},\delta}$. Indeed, we have $\tilde{\varepsilon} < \frac{\varepsilon}{(1-p)\delta} < \frac{l^*}{(1-p)\delta}$ from Lemma B11(i), and we have $\tilde{K} > \frac{K}{(1-p)\delta}$ since δ is close to one. So the above lemma ensures that (s,v) is enforceable using continuation payoffs in the set $G_{v,\lambda,\tilde{\varepsilon},\tilde{K},\delta}$.

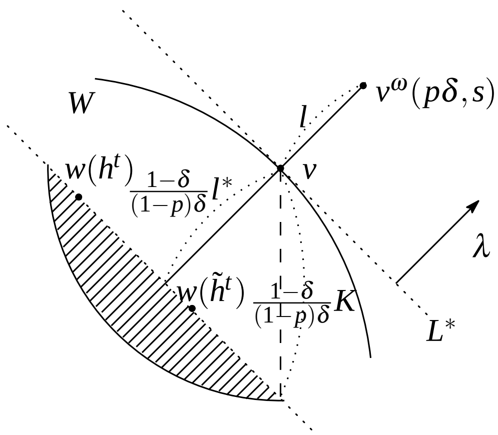

Figure 15: Continuation Payoffs for Regular Direction λ

In Section 3.5, we have shown that the promise-keeping condition (2) for the good state ω_G and the incentive compatibility condition (3) can be satisfied by moving continuation payoffs on the line L in Figure 10, which is a translate of the tangent line. Also, we have briefly explained that the promise-keeping condition (2) for the bad state ω_B can be satisfied by perturbing continuation payoffs a bit. Due to this perturbation, the resulting continuation payoffs are not on the line L, and this is the reason why we have a weak inequality rather than an equality in clause (ii) of the lemma above. The term $C\hat{v}(p\delta, s, \lambda)$ in clause (ii) measures the size of this perturbation: To satisfy the promise-keeping condition (2) for the bad state ω_B, we need to offset the difference in the block payoffs $v^{\omega_G}(p\delta,s)$ and $v^{\omega_B}(p\delta,s)$ at different states. Since this difference (toward the direction λ) is $\hat{v}(p\delta,s,\lambda)$, the size of the perturbation is bounded by $C\hat{v}(p\delta,s,\lambda)$ for some constant $C>0$. In the proof, we formally explain how to find such a perturbation.

Proof. Fix s, λ, p, δ, and v as stated. Pick ω^* such that $\omega^* \in \arg\max_{\omega\in\Omega} \lambda \cdot v^{\omega}(p\delta,s)$.

Consider a constant continuation payoff w such that the target payoff v is exactly achieved as the sum of the block payoff and the continuation payoff given the initial state is ω^*. That is, take w such that $w(h^t)=w^*$ for all t and h^t, where w^* solves

$$v=\frac{1-\delta}{1-p\delta}v^{\omega^*}(p\delta,s)+\left(1-\frac{1-\delta}{1-p\delta}\right)w^*. \tag{36}$$

Note that the specification of w^* here is exactly the same as that in Section 3.5. As in Section 3.5, this constant function satisfies the promise-keeping condition (2) for ω^*, but not for other states $\omega \neq \omega^*$. Also, it does not satisfy the incentive compatibility condition (3). In what follows, we will modify this constant function w to satisfy these requirements.

Specifically, we consider the continuation payoff such that

$$w(h^t)=w^*+z^t(h^t)+\tilde{z}(h^1)+\hat{z}(h^1)$$

for each t and h^{t+1}. Here we add three perturbation terms, z^{t+1}, $\tilde{z}$, and $\hat{z}$, to the constant continuation payoff w^*. As in Section 3.5, we will choose the term z^t so that any deviation in period t is deterred and the incentive compatibility condition (3) holds. Also, we will choose the terms $\tilde{z}^t$ and $\hat{z}$ so that the promise-keeping condition (2) holds for all ω. Note that $\tilde{z}^t$ and $\hat{z}$ depend only on the public signal in period one.

We begin with explaining how to choose the terms $\tilde{z}^t$ and $\hat{z}$. Recall that the constant continuation payoff $w(h^t)=w^*$ does not satisfy the promise-keeping condition (2) for $\omega \neq \omega^*$, because different initial states ω yield different block payoffs $v^{\omega}(p\delta,s)$. Note in particular that the welfare level $\lambda\cdot v^{\omega}(p\delta,s)$ of the block payoff (with respect to the direction λ) depends on the initial state ω. We will first choose the term $\hat{z}$ in order to offset this difference in the welfare level. That is, we will choose $\hat{z}$ so that if the continuation payoff is $w = w^* + \hat{z}$, then the payoff (the sum of the block payoff and the continuation payoff) is on the line L^* in Figure 15 regardless of the initial state ω. Of course, this needs not imply the promise-keeping condition (2), since different initial states may yield different payoffs on the line L^*. We will choose the term $\tilde{z}$ in order to offset this payoff difference.

Formally, we choose $\hat{z}$ in the following way. Pick an arbitrary player $i^* \in I$ such that $\lambda_{i^*} \neq 0$, and let $\hat{z} : Y \to \boldsymbol{R}^N$ be such that

$$\begin{aligned}\sum_{t=1}^{\infty} \delta^t p^{t-1}(1-p) \sum_{y \in Y} \pi^{\omega}(y|a_{i^*}, s_{-i^*}(h^0))\lambda_{i^*}\hat{z}_{i^*}(y) \\ = \frac{1-\delta}{1-p\delta}\left(\lambda \cdot v^{\omega^*}(p\delta, s) - \lambda \cdot v^{\omega}(p\delta, s)\right)\end{aligned} \tag{37}$$

for all ω and a_{i^*}, and

$$\hat{z}_i(y) = 0 \tag{38}$$

for all $i \neq i^*$ and y. From (38), the perturbation $\hat{z}$ influences the payoff of player i^* only. Hence its impact on the welfare level is $\lambda \cdot \hat{z}(y) = \lambda_{i^*}\hat{z}_{i^*}(y)$. (37) ensures that the expected value of this impact indeed offsets the difference $v^{\omega^*}(p\delta, s) - \lambda \cdot v^{\omega}(p\delta, s)$ in the welfare level. Note that the term $p^{t-1}(1-p)$ on the left-hand side is the probability that the random block ends after period t, and we take the expectation with respect to the random termination period t. The term $\frac{1-\delta}{1-p\delta}$ on the right-hand side is the coefficient on the block payoff.

Note that the above perturbation $\hat{z}$ does not influence players' incentives at all. Indeed, (38) implies that $\hat{z}$ does not influence the incentive of player $i \neq i^*$, and (37) ensures that the expected value of $\hat{z}_{i^*}$ does not depend on the action of player i^*.

To see the existence of such $\hat{z}_{i^*}$, note that (37) is equivalent to

$$\sum_{y \in Y} \pi^{\omega}(y|a_{i^*}, s_{-i}(h^0))\hat{z}_{i^*}(y) = \frac{1-\delta}{(1-p)\delta} \cdot \frac{\lambda \cdot v^{\omega^*}(p\delta, s) - \lambda \cdot v^{\omega}(p\delta, s)}{\lambda_{i^*}}$$

for all ω and a_{i^*}. So it is sufficient to show that this system of equations has a solution $\hat{z}_{i^*}$. Since $s \in S^{\text{IFR}}(\eta)$, the action profile $s(h^0)$ in period one has cross-state individual full rank for player i^*. This ensures that the coefficient matrix for the above system has full row rank, and hence it has a solution. Also, since the absolute value of the right-hand side is at most $\frac{(1-\delta)\hat{v}(p\delta,s,\lambda)}{(1-p)\delta|\lambda_{i^*}|}$, without loss of generality, we can assume that there is $C > 0$ such that

$$|\hat{z}(y)| \leq \frac{(1-\delta)C\hat{v}(p\delta,s,\lambda)}{(1-p)\delta|\lambda_{i^*}|} \tag{39}$$

for all δ, p, λ, and y. Multiplying both sides by $|\lambda_{i^*}|$, we have

$$|\lambda \cdot \hat{z}(y)| \leq \frac{1-\delta}{(1-p)\delta} C\hat{v}(p\delta,s,\lambda) \tag{40}$$

for all δ, p, λ, and y. Note that given $\eta > 0$, the constant $C > 0$ can be chosen independently of the choice of s.

The perturbation term $\hat{z}$ above ensures that the same welfare level is achieved for all initial states, but the actual payoff may still depend on the initial state. Let k_i^ω denote this payoff for player i given the initial state ω. That is, let k_i^ω be the sum of the block payoff and the continuation payoff when the continuation payoff function is $w = w^* + \hat{z}$:

$$k_i^\omega = \frac{1-\delta}{1-p\delta} v_i^\omega(p\delta,s) + \sum_{t=1}^{\infty} \delta^t p^{t-1}(1-p) \left(w_i^* + \sum_{y \in Y} \pi^\omega(y|s(h^0))\hat{z}_{i^*}(y) \right). \tag{41}$$

By the definition of w^* and $\hat{z}$, we have $k^{\omega^*} = v$, that is, the payoff k^{ω^*} given the initial state ω^* exactly achieves the target payoff v. Also, by the definition of $\hat{z}$, we have $\lambda \cdot k^\omega = \lambda \cdot k^{\omega^*}$ for all ω, so the payoff vector k^ω is on the line L^* in Figure 15 for any state ω.

Choose $\tilde{z} : Y \to \boldsymbol{R}^N$ such that

$$\sum_{t=1}^{\infty} \delta^t p^{t-1}(1-p) \sum_{y \in Y} \pi^\omega(y|a_i, s_{-i}(h^0))\tilde{z}_i(y) = k_i^{\omega^*} - k_i^\omega \tag{42}$$

for each ω, i, and a_i, and

$$\lambda \cdot \tilde{z}(y) = 0 \tag{43}$$

for each y. (42) ensures that the expected discounted value of $\tilde{z}$ offsets the difference between k_i^ω and $k_i^{\omega^*}$, and that the term $\tilde{z}$ does not influence each player's incentive. That is, if the continuation payoff is $w = w^* + \tilde{z} + \hat{z}$, the payoff v is exactly achieved regardless of the initial state ω (so the promise-keeping condition

(2) holds for all ω). (43) implies that the term $\tilde{z}(y)$ moves the continuation payoff $w(h^t)$ only toward directions orthogonal to λ. The existence of such $\tilde{z}$ follows from the fact that $\lambda \cdot k^{\omega} = \lambda \cdot k^{\omega^*}$ for all ω and that the action profile $s_{-i}(h^0)$ in period one has cross-state pairwise full rank. The proof is very similar to that of Lemmas 5.3 and 5.4 of FLM and hence omitted.

Since (42) is equivalent to

$$\sum_{y \in Y} \pi^{\omega}(y|a_i, s_{-i}(h^0))\tilde{z}_i(y) = \frac{1-p\delta}{(1-p)\delta}(k_i^{\omega^*} - k_i^{\omega}),$$

without loss of generality, we can assume that for a given λ, there is $\tilde{K} > 0$ such that

$$|\tilde{z}(y)| \leq \frac{1-p\delta}{(1-p)\delta}\tilde{K}\max_{\omega,i}|k_i^{\omega^*} - k_i^{\omega}|$$

for each δ, p, and y. Here the bound $\tilde{K}$ can be arbitrarily large when λ approaches a coordinate direction; see FLM for details. Note that

$$\begin{aligned}
&\left|k_i^{\omega^*} - k_i^{\omega}\right| \\
=&\frac{1-\delta}{1-p\delta}\left|v_i^{\omega^*}(p\delta,s) - v_i^{\omega}(p\delta,s)\right| \\
&+\sum_{t=1}^{\infty}\delta^t p^{t-1}(1-p)\left|\sum_{y^1\in Y}\pi^{\omega^*}(y^1|s(h^0))\hat{z}_{i^*}(y^1) - \sum_{y^1\in Y}\pi^{\omega}(y^1|s(h^0))\hat{z}_{i^*}(y^1)\right| \\
\leq&\frac{1-\delta}{1-p\delta}\left(\left|v_i^{\omega^*}(p\delta,s) - v_i^{\omega}(p\delta,s)\right| + \left|\frac{\lambda\cdot v^{\omega^*}(p\delta,s) - \lambda\cdot v^{\omega}(p\delta,s)}{\lambda_{i^*}}\right|\right) \\
\leq&\frac{1-\delta}{1-p\delta}\left(\overline{g} + \frac{\overline{g}}{|\lambda_{i^*}|}\right)
\end{aligned}$$

for each ω and i, where the first inequality uses (37) and (38). Plugging this into the previous inequality, we have

$$|\tilde{z}(y)| < \frac{1-\delta}{(1-p)\delta}\tilde{K}\left(\overline{g} + \frac{\overline{g}}{|\lambda_{i^*}|}\right). \tag{44}$$

So far, we have explained that the promise-keeping condition (2) can be satisfied for all states by choosing the perturbations $\hat{z}$ and $\tilde{z}$ appropriately. Next, we will show that the incentive compatibility condition (3) can be satisfied by choosing the perturbation z^t appropriately. This extends the analysis in Section 3.5 to the general setup.

Pick $t \geq 1$ and h^{t-1} arbitrarily. Let $s_i|_{(h^{t-1},a_i)}$ denote the strategy for the continuation game after h^{t-1}, which chooses action a_i in period t and then follows the prescribed strategy s. That is, $s_i|_{(h^{t-1},a_i)}(\tilde{h}^0) = a_i$ and $s_i|_{(h^{t-1},a_i)}(\tilde{h}^{\tilde{t}}) = s_i|_{h^{t-1}}(\tilde{h}^{\tilde{t}})$ for each $\tilde{t} \geq 1$ and $\tilde{h}^{\tilde{t}}$. Let $(z^t(h^{t-1},y))_{y\in Y}$ be such that

$$\begin{aligned}&\delta(1-p)\sum_{y\in Y}\pi^{\omega}(y|a_i,s_{-i}(h^{t-1}))z_i^t(h^{t-1},y)\\&=\frac{1-\delta}{1-p\delta}\left(v_i^{\omega}(p\delta,s|_{h^{t-1}})-v_i^{\omega}(p\delta,s_i|_{(h^{t-1},a_i)},s_{-i}|_{h^{t-1}})\right)\end{aligned} \tag{45}$$

for all ω, i, and a_i, and such that

$$\lambda\cdot z^t(h^{t-1},y)=0 \tag{46}$$

for all y. (45) implies that player i is indifferent over all actions a_i in period t, regardless of the current hidden state ω^t. Indeed, if player i deviates to a_i today, it changes the block payoff from $v_i^{\omega^t}(p\delta,s|_{h^{t-1}})$ to $v_i^{\omega^t}(p\delta,s_i|_{(h^{t-1},a_i)},s_{-i}|_{h^{t-1}})$, but (45) guarantees that this change is offset by the expected value of z_i^t. (Note also that z_i^t does not influence player i's incentive in earlier periods $\tilde{t}<t$, since (45) implies that the expected value of z^t is zero as long as player i does not deviate in period t.) (46) ensures that this z^t moves the continuation payoff w only toward directions orthogonal to λ, as in Section 3.5. The existence of such z^t follows from the fact that the action profile $s_{-i}(h^{t-1})$ in period t has cross-state pairwise full rank. Also, since (45) is equivalent to

$$\begin{aligned}&\sum_{y\in Y}\pi^{\omega}(y|a_i,s_{-i}(h^t))z_i^{t+1}(h^t,y)\\&=\frac{1-\delta}{\delta(1-p)(1-p\delta)}\left(v_i^{\omega}(p\delta,s|_{h^t})-v_i^{\omega}(p\delta,s_i|_{(h^t,a_i)},s_{-i}|_{h^t})\right),\end{aligned}$$

without loss of generality, we can assume that for a fixed λ, there is $\tilde{K}'>0$ such that

$$|z^t(h^t)|<\frac{1-\delta}{(1-p\delta)(1-p)\delta}\tilde{K}'\max_{\omega,i,h^{t-1},a_i}\left|v_i^{\omega}(p\delta,s|_{h^{t-1}})-v_i^{\omega}(p\delta,s_i|_{(h^{t-1},a_i)},s_{-i}|_{h^{t-1}})\right|$$

for all t, h^t, δ, and p. Here we can choose $\tilde{K}'$ uniformly in t and h^t, since $s\in S^{\text{PFR}}$ induce at most $|A|$ different actions. Since $v_i^{\omega}(p\delta,s|_{h^{t-1}})-v_i^{\omega}(p\delta,s_i|_{(h^{t-1},a_i)},s_{-i}|_{h^{t-1}})\leq \overline{g}$, we have

$$|z^t(h^t)|<\frac{1-\delta}{(1-p\delta)(1-p)\delta}\tilde{K}'\overline{g}. \tag{47}$$

Now we verify that the constructed w satisfies the clauses (i) through (iii) in Lemma B16. By the definition of z^t, ex-post incentive compatibility is satisfied each period. Also, by the definition of $\tilde{z}$ and $\hat{z}$, the payoff v is exactly achieved regardless of the initial state ω. Hence clause (i) follows.

To prove clause (ii), we arrange (36) and obtain

$$w^* = v - \frac{1-\delta}{(1-p)\delta}(v^{\omega^*}(p\delta,s) - v). \tag{48}$$

Then we have

$$\begin{aligned}\lambda \cdot w(h^t) &= \lambda \cdot w^* + \lambda \cdot z^t(h^t) + \lambda \cdot \tilde{z}(h^1) + \lambda \cdot \hat{z}(h^1) \\ &\leq \lambda \cdot v - \frac{1-\delta}{(1-p)\delta}(\lambda \cdot v^{\omega^*}(p\delta,s) - \lambda \cdot v - C\hat{v}_i(p\delta,s,\lambda)).\end{aligned}$$

Here the inequality comes from (40), (43), (46), and (48). This proves clause (ii).

To prove clause (iii), note that from (48),

$$w_i(h^t) = v_i - \frac{1-\delta}{(1-p)\delta}(v_i^{\omega^*}(p\delta,s) - v_i) + z_i^t(h^t) + \tilde{z}_i(h^1) + \hat{z}_i(h^1)$$

for all i. This implies that

$$|v - w(h^t)| \leq \frac{1-\delta}{(1-p)\delta}|v^{\omega^*}(p\delta,s) - v| + |z^t(h^t)| + |\tilde{z}(h^1)| + |\hat{z}(h^1)|.$$

Since $v \in V$, we have $|v^{\omega^*}(p\delta,s) - v| \leq \overline{g}$. This, together with (39), (44), and (47), implies that

$$|v - w(h^t)| \leq \frac{1-\delta}{(1-p)\delta}\left\{\overline{g} + \frac{\tilde{K}'\overline{g}}{1-p\delta} + \tilde{K}\left(\overline{g} + \frac{\overline{g}}{|\lambda_{i^*}|}\right) + \frac{C\hat{v}(p\delta,s,\lambda)}{|\lambda_{i^*}|}\right\}.$$

By the definition, $\hat{v}(p\delta,s,\lambda) \leq \overline{g}$. Hence, by letting $K > \overline{g}\{1 + \frac{\tilde{K}'}{1-p} + \tilde{K}(1 + \frac{1}{|\lambda_{i^*}|}) + \frac{C}{|\lambda_{i^*}|}\}$, we have clause (iii). *Q.E.D.*

B.5.4 Step 4: Enforceability for Positive Coordinate Direction

In this step, we consider enforceability for the positive coordinate direction (i.e., λ with $\lambda_i = 1$). The analysis here is an extension of that in Section 3.6.

Lemma B17. *Assume (IFR), and fix i. Then there is $C > 0$ such that $p \in (\overline{p}, 1)$, there is $K > 0$ such that for each pure public strategy profile s, for each $v \in V$, and for each δ, there is w such that* *(i)* (s,v) *is stochastically ex-post enforceable with respect to* (δ,p) *by* w*,*

(ii) *For all* t *and* h^t*,*

$$w_i(h^t) \leq v_i - \frac{1-\delta}{(1-p)\delta}\left(\min_{\omega\in\Omega} v_i^\omega(p\delta,s) - v_i - C\hat{v}_i(p\delta,s) - CG_i^*(p\delta,s)\right),$$

(iii) *For all* t *and* h^t*,*

$$|v - w(h^t)| < \frac{1-\delta}{(1-p)\delta}\left(K + CG_i^*(p\delta,s)\right).$$

Let λ be such that $\lambda_i = 1$. For the same of exposition, choose v as in Figure 16, that is, let v be a boundary point of W whose unit normal is λ. Choose s as in Lemma B11(ii), so that it approximates the score regardless of the initial state. The above lemma asserts that this (s,v) is enforceable using continuation payoffs in the shaded area in the figure 16, where $l^* = \min_{\omega\in\Omega} v_i^\omega(p\delta,s) - v_i - C\hat{v}_i(p\delta,s) - CG_i^*(p\delta,s)$ and $K^* = K + CG_i^*(p\delta,s)$. Note that this length l^* is approximately equal to the length l in the figure, because by the definition of s, both $\hat{v}_i(p\delta,s)$ and $G_i^*(p\delta,s)$ are approximately zero.

Choose $\varepsilon > 0$ as in Lemma B11, and let p be given. Let $\tilde{\varepsilon} = \frac{\varepsilon}{1-p}$ and $\tilde{K} = \frac{2(K+C\varepsilon)}{1-p}$. Then the shaded area in the figure is included in the set $G_{v,\lambda,\tilde{\varepsilon},\tilde{K},\delta}$, because $\tilde{\varepsilon} < \frac{\varepsilon}{(1-p)\delta} < \frac{l^*}{(1-p)\delta}$ and $\tilde{K} > \frac{2(K+C\varepsilon)}{(1-p)\delta} > \frac{K^*}{(1-p)\delta}$ for high δ. So the above lemma ensures that (s,v) is enforceable using continuation payoffs in the set $G_{v,\lambda,\tilde{\varepsilon},\tilde{K},\delta}$.

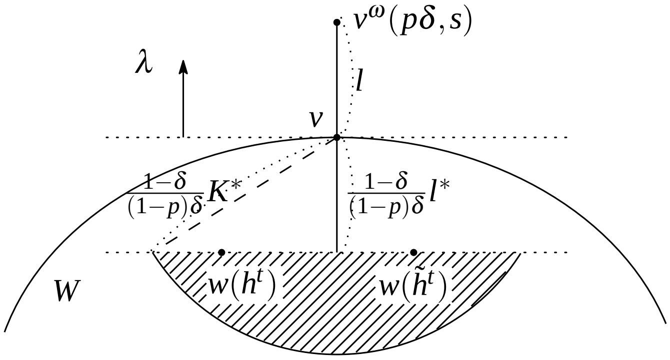

Figure 16: Continuation Payoffs for λ with $\lambda_i = 1$

Compared to Lemma B16, we have the new term $CG_i^*(p\delta, s)$ in the inequalities in (ii) and (iii). We need this term because we need to take care of player i's incentive compatibility: As explained in Section 3.6, player i has a profitable deviation unless her initial belief matches the dummy belief $\tilde{\mu}$. In order to deter such deviations, we need to move the continuation payoffs w vertically. $CG_i^*(p\delta, s)$ measures the size of this vertical move.

Proof. Fix i, s, p, δ, and v as stated. Pick ω^* such that $\omega^* \in \arg\max_{\omega \in \Omega} v_i^{\omega}(p\delta, s)$. Then pick the constant continuation function $w(h^t) = w^*$ as in the proof of Lemma B16. This constant continuation payoff does not satisfy the promise-keeping condition (2) for $\omega \neq \omega^*$ or the incentive compatibility condition (3). In what follows, we will modify this w to satisfy these requirements.

Specifically, for each $j \neq i$, we let

$$w_j(h^t) = w_j^* + z_j^t(h^t) + \tilde{z}_j(h^1) + \hat{z}_j(h^1),$$

as in the proof of Lemma B16. That is, we add three perturbation terms, z_j^t, $\tilde{z}_j$, and $\hat{z}_j$, to the constant value w_j^*. On the other hand, for player i, we let

$$w_i(h^t) = w_i^* + \sum_{\tilde{t}=1}^{t} z_i^{\tilde{t}}(h^{\tilde{t}}) + \tilde{z}_i(h^1) + \hat{z}_i(h^1).$$

Now the term z_j^t is replaced with the sum $\sum_{\tilde{t}=1}^{t} z_i^{\tilde{t}}$ of the perturbation terms, as in Section 3.6.

We first show that the promise-keeping condition (2) can be satisfied by choosing $\tilde{z}$ and $\hat{z}$ appropriately. This part is almost identical with the one in the proof of Lemma B16. We begin with $\hat{z}$. Let $i^* = i$, and let $\hat{z}$ be as in the proof of Lemma B16. That is, choose $\hat{z} : Y \to \boldsymbol{R}^N$ such that

$$\sum_{t=1}^{\infty} \delta^t p^{t-1}(1-p) \sum_{y \in Y} \pi^{\omega}(y|a_i, s_{-i}(h^0))\hat{z}_i(y) = \frac{1-\delta}{1-p\delta}\left(v_i^{\omega^*}(p\delta, s) - v_i^{\omega}(p\delta, s)\right)$$

for all ω and a_i, and

$$\hat{z}_j(y) = 0$$

for all $j \neq i$ and y. As in the proof of Lemma B16, this $\hat{z}$ ensures that the same welfare level is achieved regardless of the initial state ω. Recall that this $\hat{z}$ does not influence players' incentives, and that there is $\tilde{C} > 0$ such that

$$|\hat{z}(y)| \leq \frac{1-\delta}{(1-p)\delta}\tilde{C}\hat{v}_i(p\delta, s) \tag{49}$$

for all δ, p, λ, and y.

Similarly, we choose $\tilde{z}$ as in the proof of Lemma B16. That is, let $\tilde{z} : Y \to \boldsymbol{R}^N$ be such that

$$\sum_{t=1}^{\infty} \delta^t p^{t-1}(1-p) \sum_{y \in Y} \pi^{\omega}(y|a_j, s_{-j}(h^0))\tilde{z}_j(y) = \frac{1-\delta}{1-p\delta}\left(v_j^{\omega^*}(p\delta, s) - v_j^{\omega}(p\delta, s)\right)$$

for each ω, $j \neq i$, and a_j, and

$$\tilde{z}_i(y) = 0$$

for each y. This $\tilde{z}$ ensures that the payoff v is exactly achieved as the sum of the block payoff and the continuation payoff, regardless of the initial state ω. Hence the promise-keeping condition (2) holds for all ω. Note that this $\tilde{z}$ does not influence players' incentives, and that there is $\tilde{K} > 0$ such that

$$|\tilde{z}^1(y)| < \frac{1-\delta}{(1-p)\delta}\tilde{K}2\overline{g} \tag{50}$$

for all δ, p, and y, as in the proof of Lemma B16.

Next, we show that the incentive compatibility condition (3) can be satisfied by choosing z^t appropriately. We begin with considering the incentive problem of player $j \neq i$. Pick $t \geq 1$ and h^{t-1} arbitrarily, and let $s_j|_{(h^{t-1}, a_j)}$ be as in the proof of Lemma B16. Then we choose $(z_j^t(h^{t-1}, y))_{y \in Y}$ such that

$$\begin{aligned}
&\delta(1-p) \sum_{y \in Y} \pi^{\omega}(y|a_j, s_{-j}(h^{t-1}))z_j^t(h^{t-1}, y) \\
&= \frac{1-\delta}{1-p\delta}\left(v_j^{\omega}(p\delta, s|_{h^{t-1}}) - v_j^{\omega}(p\delta, s_j|_{(h^{t-1}, a_j)}, s_{-j}|_{h^{t-1}})\right)
\end{aligned}$$

for all ω and a_j. This z_j^t ensures that player $j \neq i$ is indifferent over all actions in period t, regardless of the current hidden state ω^t. As in the proof of Lemma B16, we can show that there is $\tilde{K}' > 0$ such that

$$|z_j^t(h^t)| < \frac{1-\delta}{(1-p\delta)(1-p)\delta}\tilde{K}'\overline{g} \tag{51}$$

for all t, h^t, δ, and p. Here we can choose $\tilde{K}'$ uniformly in t and h^t, since s is a pure strategy profile and hence induces at most $|A|$ different action profiles.

Now we consider the incentive problem of player i. This part extends the analysis in Section 3.6 to the general setup. Pick $t \geq 1$ and h^{t-1} arbitrarily. Then we choose $(z_i^t(h^{t-1},y))_{y\in Y}$ such that

$$\sum_{\tilde{t}=1}^{\infty} \delta^{\tilde{t}} p^{\tilde{t}-1}(1-p) \sum_{y\in Y} \pi_Y^{\omega}(y|a_i, s_{-i}(h^{t-1}))z_i^t(h^t)$$
$$= \begin{cases} 0 & \text{if } a_i = s_i(h^{t-1}) \\ -\dfrac{1-\delta}{1-p\delta} \displaystyle\sup_{\mu\in\triangle\Omega(s,h^{t-1})} G_i^t(p\delta,\mu,s,h^{t-1}) & \text{if } a_i \neq s_i(h^{t-1}) \end{cases} \tag{52}$$

for all ω and a_i. Here, $\Omega(s,h^{t-1})$ is the set of initial states ω such that the public history h^t realizes with positive probability given the strategy profile s. (Under the full support assumption, $\Omega(s,h^t) = \Omega$.) As discussed in Section 3.6, this z_i^t ensures that "play s_i until period $t-1$, then deviates in period t, and then play a best reply thereafter" is not profitable for player i regardless of the initial belief μ. Indeed, (52) implies that such a deviation decreases her payoff by at least

$$\frac{1-\delta}{1-p\delta} \sup_{\mu\in\triangle\Omega(s,h^{t-1})} G_i^t(p\delta,\mu,s,h^{t-1}),$$

which exceeds the gain $G_i^t(p\delta,\mu,s,h^{t-1})$. (Here we say "at least" because this is the effect of z_i^t only. If player i chooses an action different from the one induced by s in period $\tilde{t} \geq t+1$, it changes the expected value of $z_i^{\tilde{t}}$ and decreases the total payoff further.) The existence of such z_i^t is guaranteed since $s(h^{t-1})$ has cross-state individual full rank for player i. Without loss of generality, we can assume that there is $\tilde{C}' > 0$ such that

$$|z_i^t(h^{t-1},y)| < \frac{1-\delta}{(1-p)\delta}\tilde{C}' \sup_{\mu\in\triangle\Omega(s,h^{t-1})} G_i^t(p\delta,\mu,s,h^{t-1}) \tag{53}$$

for all t, h^{t-1}, y, δ, and p. Again, we can choose $\tilde{C}'$ uniformly in t and h^{t-1}, since s is a pure strategy profile and hence induces at most $|A|$ different action profiles.

This z^t ensures that the incentive compatibility (3) holds for all players. Also by the definition of $\tilde{z}$ and $\hat{z}$, the promise-keeping condition (2) is satisfied. Hence clause (i) holds.

To prove clause (ii), use (48) so that

$$w_i(h^t) = v_i - \frac{1-\delta}{(1-p)\delta}(v_i^\omega(p\delta,s) - v_i) + \sum_{\tilde{t}=1}^{t} z^{\tilde{t}}(h^{\tilde{t}}) + \tilde{z}_i(h^1) + \hat{z}_i(h^1).$$

Using (49), (53), and $\tilde{z}_i(h^1) = 0$,

$$w_i(h^t) \le v_i - \frac{1-\delta}{(1-p)\delta}\left(v_i^\omega(p\delta,s) - v_i - \tilde{C}'G_i^*(p\delta,s) - \tilde{C}\hat{v}_i(p\delta,s)\right).$$

So by setting $C = \max\{\tilde{C},\tilde{C}'\}$, clause (ii) follows. Also clause (iii) follows just as in the proof of Lemma B16. *Q.E.D.*

B.5.5 Step 5: Enforceability for Negative Coordinate Direction

Note that the lemma in the previous step considers only pure strategy profiles s. This is enough for the positive coordinate direction, since the score toward this direction is achieved by a pure strategy profile, and (IFR) guarantees that any pure action profile has cross-state individual full rank. On the other hand, when we consider the negative coordinate direction λ with $\lambda_i = -1$, we need to consider mixed strategies, as minimaxing player i may require mixture by the opponents. Since mixed actions may not have cross-state individual full rank even under (IFR), the statement of the following lemma, which concerns enforceability for the negative coordinate direction, is a bit more complicated than Lemma B17.

Lemma B18. *Fix i. For each $\eta > 0$, there is $C > 0$ such that for each $p \in (0,1)$, for each public strategy profile $s \in S^{IFR}(\eta)$ such that s_i is a pure strategy, there is $K > 0$ such that for each $v \in V$, for each δ, there is w such that*

(i) (s,v) is stochastically ex-post enforceable with respect to (δ,p) by w,

(ii) For all t and h^t,

$$w_i(h^t) \ge v_i + \frac{1-\delta}{(1-p)\delta}\left(v_i - \max_{\omega\in\Omega} v_i^\omega(p\delta,s) + C\hat{v}_i(p\delta,s) + CG_i^*(p\delta,s)\right),$$

(iii) For all t and h^t,

$$|v - w(h^t)| < \frac{1-\delta}{(1-p)\delta}\left(K + CG_i^*(p\delta,s)\right).$$

The interpretation of the above lemma is very similar to the one for the previous lemma. Choose v as in Figure 17. That is, let v be a boundary point of W whose unit normal is the negative coordinate direction λ with $\lambda_i = -1$. Choose s as in Lemma B11(iii), so player i is minimaxed regardless of the initial state. The above lemma asserts that this (s,v) is enforceable using continuation payoffs in the shaded area in Figure 17, where $l^* = v_i - \max_{\omega \in \Omega} v_i^{\omega}(p\delta, s) + C\hat{v}_i(p\delta, s) + CG_i^*(p\delta, s)$ and $K^* = K + CG_i^*(p\delta, s)$.

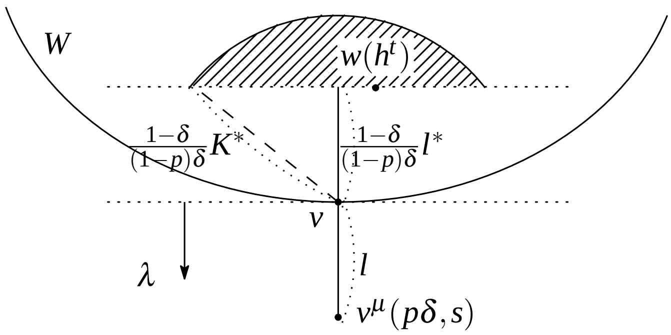

Figure 17: Continuation Payoffs for λ with $\lambda_i = -1$

The proof of the above lemma is very similar to that of Lemma B17, and hence omitted. In order to find bounds $\tilde{K}$ and $\tilde{C}$ which work uniformly in t and h^t, we use the fact that $s \in S^{\text{IFR}}(\eta)$ for some $\eta > 0$.

B.5.6 Step 6: Uniform Enforceability

Now we are ready to show that W is uniformly ex-post decomposable.

Lemma B19. *For any smooth subset W of the interior of V^*, there is $p \in (0,1)$ such that W is uniformly ex-post decomposable with respect to p.*

Proof. Pick W as stated. Pick η as stated in Lemma B11, and then pick $C > 0$ as in Lemmas B16 through B18. Fix $\tilde{\varepsilon}$, $p \in (0,1)$, and $\overline{\delta} \in (0,1)$ as stated in Lemma B11. (Here, $\tilde{\varepsilon}$ represents ε in Lemma B11.) Applying Lemmas B16 through B18 to the strategy profiles specified in Lemma B11, it follows that for each λ, there is $\tilde{K}_\lambda > 0$ such that for each $\delta \in (\overline{\delta}, 1)$ and $v \in W$, there is a strategy profile $s_{v,\lambda,\delta}$ and a function $w_{v,\lambda,\delta}$ such that

(i) $(s_{v,\lambda,\delta}, v)$ is enforced by $w_{v,\lambda,\delta}$ for (δ, p),

(ii) $\lambda \cdot w_{v,\lambda,\delta}(h^t) \leq \lambda \cdot v - \frac{(1-\delta)\tilde{\varepsilon}}{(1-p)\delta}$ for each t and h^t, and

(iii) $|v - w_{v,\lambda,\delta}(h^t)| < \frac{(1-\delta)}{(1-p)\delta}\tilde{K}_\lambda$ for each t and h^t.

Set $\varepsilon = \frac{\tilde{\varepsilon}}{2(1-p)}$, and for each λ, let $K_\lambda = \frac{\tilde{K}_\lambda}{(1-p)\overline{\delta}}$. Then it follows from (ii) and (iii) that $w_{v,\lambda,\delta}(h^t) \in G_{v,\lambda,2\varepsilon,K_\lambda,\delta}$ for all t and h^t. The rest of the proof is similar to that of Fudenberg and Yamamoto (2011b). *Q.E.D.*

Appendix C: Dispensability of Public Randomization

In this appendix, we show that the folk theorem remains valid even without public randomization. Since public randomization is not available, we cannot use random blocks anymore. Instead, here we consider equilibria in which the infinite horizon is divided into a series of T-period blocks. The following lemma shows that each extreme point of the limit feasible payoff set V can be approximated by the average payoff in the T-period block, when T is sufficiently large. Let $v_i^\mu(T,s)$ denote player i's average payoff in the T-period game with the initial prior μ and the strategy profile s, that is, let $v_i^\mu(T,s) = \frac{1}{T}\sum_{t=1}^T E[g_i^{\omega^t}(a^t)|\mu,s]$. (No discounting here.) Let $V^\mu(T)$ denote the feasible payoff set in the T-period game with the initial prior μ, that is, $V^\mu(T) = \text{co}\{v^\mu(T,s)|s \in S\}$.

Proposition C1. *Suppose that the full support assumption holds. Then there is* $K > 0$ *such that for any* T*,* μ*, and* λ*,*

$$\left| \max_{v \in V^\mu(T)} \lambda \cdot v - \max_{\tilde{v} \in V} \lambda \cdot \tilde{v} \right| \leq \frac{K}{T}.$$

This proposition implies that the feasible payoff set $V^\mu(T)$ for the T-period game converges to V as T goes to infinity. Hence any extreme point of V is approximated by a T-period game payoff with an appropriate strategy profile.

Proof. As shown in Yamamoto (2016), ε in Proposition 2 can be replaced with $(1-\delta)K$ for some $K > 0$, that is, there is $K > 0$ such that for each λ, μ, and δ,

$$\left| \max_{v \in V^\mu(\delta)} \lambda \cdot v - \max_{\tilde{v} \in V^{\tilde{\mu}}(\delta)} \lambda \cdot \tilde{v} \right| < (1-\delta)K. \tag{54}$$

So the difference in the scores induced by different initial priors is of order $1-\delta$. Pick such K.

Fix λ. Let $s(\delta,\mu)$ be a pure-strategy profile which achieves the score toward λ given the initial prior μ and the discount factor δ. Then for each s,

$$\lambda\cdot v^{\mu}(\delta,s(\delta,\mu))\geq\lambda\cdot v^{\mu}(\delta,s).$$

Since the right-hand side is decomposed into the payoffs in the first T periods and the continuation payoff from period $T+1$ on,

$$\begin{aligned}\lambda\cdot v^{\mu}(\delta,s(\delta,\mu))\geq&(1-\delta)\sum_{t=1}^{T}\delta^{t-1}E[\lambda\cdot g^{\omega^t}(a^t)|\mu,s]\\&+\delta^{T}E[\lambda\cdot v^{\mu^{T+1}}(\delta,s(\delta,\mu^{T+1}))|\mu,s]\end{aligned}$$

for each s. Using (54), we have

$$\begin{aligned}\lambda\cdot v^{\mu}(\delta,s(\delta,\mu))\geq&(1-\delta)\sum_{t=1}^{T}\delta^{t-1}E[\lambda\cdot g^{\omega^t}(a^t)|\mu,s]\\&+\delta^{T}\left\{\lambda\cdot v^{\mu}(\delta,s(\delta,\mu))-(1-\delta)K\right\}\end{aligned}$$

for each s. Subtracting $\delta^T\{\lambda\cdot v^{\mu}(\delta,s(\delta,\mu))-(1-\delta)K\}$ from both sides and dividing them by $1-\delta^T$,

$$\frac{1-\delta}{1-\delta^T}\sum_{t=1}^{T}\delta^{t-1}E[\lambda\cdot g^{\omega^t}(a^t)|\mu,s]\leq\lambda\cdot v^{\mu}(\delta,s(\delta,\mu))+\frac{1-\delta}{1-\delta^T}\delta^{T}K.$$

Since this inequality holds for all s, taking the maximum of the left-hand side with respect to s,

$$\max_{s\in S}\frac{1-\delta}{1-\delta^T}\sum_{t=1}^{T}\delta^{t-1}E[\lambda\cdot g^{\omega^t}(a^t)|\mu,s]\leq\lambda\cdot v^{\mu}(\delta,s(\delta,\mu))+\frac{1-\delta}{1-\delta^T}\delta^{T}K.$$

Taking $\delta\to1$, we obtain

$$\max_{v\in V^{\mu}(T)}\lambda\cdot v\leq\max_{v\in V}\lambda\cdot v+\frac{K}{T}.\tag{55}$$

On the other hand, we know that

$$\begin{aligned}\lambda\cdot v^{\mu}(\delta,s(\delta,\mu))=&(1-\delta)\sum_{t=1}^{T}\delta^{t-1}E[\lambda\cdot g^{\omega^t}(a^t)|\mu,s(\delta,\mu)]\\&+\delta^{T}E[\lambda\cdot v^{\mu^{T+1}}(s(\delta,\mu^{T+1}))|\mu,s(\delta,\mu)].\end{aligned}$$

From (54),

$$\begin{aligned}\lambda\cdot v^{\mu}(\delta,s(\delta,\mu)) \leq &(1-\delta)\sum_{t=1}^{T}\delta^{t-1}E[\lambda\cdot g^{\omega^t}(a^t)|\mu,s(\delta,\mu)] \\ &+\delta^T\{\lambda\cdot v^{\mu}(\delta,s(\delta,\mu))+(1-\delta)K\}.\end{aligned}$$

Subtracting $\delta^T\{\lambda\cdot v^{\mu}(\delta,s(\delta,\mu))+(1-\delta)K\}$ from both sides and dividing them by $1-\delta^T$,

$$\frac{1-\delta}{1-\delta^T}\sum_{t=1}^{T}\delta^{t-1}E[\lambda\cdot g^{\omega^t}(a^t)|\mu,s(\delta,\mu)] \geq \lambda\cdot v^{\mu}(\delta,s(\delta,\mu))-\frac{1-\delta}{1-\delta^T}\delta^T K.$$

Since $s(\delta,\mu)$ is not necessarily the maximizer of the left-hand side,

$$\max_{s\in S}\frac{1-\delta}{1-\delta^T}\sum_{t=1}^{T}\delta^{t-1}E[\lambda\cdot g^{\omega^t}(a^t)|\mu,s] \geq \lambda\cdot v^{\mu}(\delta,s(\delta,\mu))-\frac{1-\delta}{1-\delta^T}\delta^T K.$$

Taking $\delta\to 1$, we obtain

$$\max_{v\in V^{\mu}(T)}\lambda\cdot v \geq \max_{v\in V}\lambda\cdot v-\frac{K}{T}. \tag{56}$$

Combining (55) and (56),

$$\max_{v\in V}\lambda\cdot v-\frac{K}{T} \leq \max_{v\in V^{\mu}(T)}\lambda\cdot v \leq \max_{v\in V}\lambda\cdot v+\frac{K}{T}.$$

Hence the result follows. *Q.E.D.*

The next proposition extends Proposition 3 and shows that in the T-period block, the optimal policy for some dummy belief $\tilde{\mu}$ can approximate the score regardless of the true belief μ.

Proposition C2. *Suppose that the full support assumption holds. Then there is $K>0$ such that for each λ, for each T, for each μ, for each $\tilde{\mu}$, for each pure public strategy profile $s^{\tilde{\mu}}\in\arg\max_{s\in S}\lambda\cdot v^{\tilde{\mu}}(T,s)$, for each $t\in\{0,\cdots,T-1\}$, and for each h^t,*

$$\left|\max_{v\in V^{\mu}(T-t)}\lambda\cdot v-\lambda\cdot v^{\mu}(T-t,s^{\tilde{\mu}}|_{h^t})\right| \leq \frac{K}{T-t}.$$

Proof. Like Proposition 3, the proof consists of three steps. In the first step, we show the inequality for $t = 0$ and $\tilde{\mu}$ such that $\tilde{\mu}(\omega) \geq \overline{\pi}$ for all ω. In the second step, we show that the result holds for an arbitrary belief $\tilde{\mu}$. In the third step, we show that the result holds for any $t \geq 1$.

Proposition C1 ensures that there is $K > 0$ such that for any T, μ, and λ,

$$\left| \max_{v \in V^{\mu}(T)} \lambda \cdot v - \max_{\tilde{v} \in V} \lambda \cdot \tilde{v} \right| < \frac{\overline{\pi}(K - \overline{g})}{T}.$$

Choose such $K > 0$.

Pick λ arbitrarily, and for each μ and T, let $s^{\mu}(T)$ be the optimal policy for the T-period game with the initial prior μ. Pick $\tilde{\mu}$ such that $\tilde{\mu}(\omega) \geq \overline{\pi}$ for all ω. Then, we can show that

$$\left| \max_{\hat{\mu} \in \triangle\Omega} \max_{v \in V^{\hat{\mu}}(T)} \lambda \cdot v - \lambda \cdot v^{\mu}(T, s^{\tilde{\mu}}) \right| < \frac{K - \overline{g}}{T}. \tag{57}$$

for each μ and T. The proof is very similar to the derivation of (20) in Step 1 in the proof of Proposition 3; we only need to replace $\overline{\pi}\varepsilon - \frac{(1-\delta)\overline{\pi}\overline{g}}{\delta}$, $V^{\omega}(\delta)$, and $v^{\mu}(\delta, s)$ in the proof of Proposition 3 with $\frac{\overline{\pi}(K-\overline{g})}{T}$, $V^{\omega}(T)$, and $v^{\mu}(T, s)$, respectively. So the first step is done. Similarly, the third step is exactly the same as that of Proposition 3.

The second step is also very similar to that of Proposition 3. For any $\tilde{\mu}$,

$$\begin{aligned}
&\left| \max_{v \in V^{\mu}(T)} \lambda \cdot v - \lambda \cdot v^{\mu}(T, s^{\tilde{\mu}}) \right| \\
&= \left| \lambda \cdot v^{\mu}(T, s^{\mu}) - \lambda \cdot v^{\mu}(T, s^{\tilde{\mu}}) \right| \\
&\leq \frac{1}{T} \left| \lambda \cdot g^{\mu}(s^{\mu}(h^0)) - \lambda \cdot g^{\mu}(s^{\tilde{\mu}}(h^0) \right| + \frac{T-1}{T} \cdot \frac{K - \overline{g}}{T-1} \\
&\leq \frac{1}{T}\overline{g} + \frac{K - \overline{g}}{T} = \frac{K}{T}.
\end{aligned}$$

Here the first inequality follows from (57), which ensures that the difference in the continuation payoffs from period two is at most $\frac{K-\overline{g}}{T-1}$. *Q.E.D.*

Define player i's minimax payoff in the T-period game as

$$\underline{v}_i^{\mu}(T) = \min_{s_{-i} \in S_{-i}} \max_{s_i \in S_i} v_i^{\mu}(T, s).$$

The following proposition shows that this minimax payoff for the T-period game approximates the limit minimax payoff $\underline{v}_i$ for the infinite-horizon game as T goes to infinity.

Proposition C3. *Suppose that the full support assumption holds. Then there is $K > 0$ such that $|\underline{v}_i^{\mu}(T) - \underline{v}_i| \leq \frac{K}{T}$ for each i, T, and μ.*

Proof. For each μ and δ, choose $s_{-i}^{\delta,\mu}$ as in the proof of Proposition 4. (Here we write $s_{-i}^{\delta,\mu}$ instead of s_{-i}^{μ} to emphasize the dependence on δ.) Note that this strategy $s_{-i}^{\delta,\mu}$ approximates the minimax payoff for the initial prior μ and the discount factor δ. Then as shown in the proof of Proposition 4, there is $\tilde{K} > 0$ such that

$$\left| \underline{v}_i^{\hat{\mu}}(\delta) - \max_{s_i \in S_i} v_i^{\mu}(\delta, s_i, s_{-i}^{\delta,\tilde{\mu}}|_{h^t}) \right| < (1-\tilde{K})\delta$$

for each i, δ, μ, $\tilde{\mu}$, $\hat{\mu}$, t, and h^t. Choose such $\tilde{K}$.

Pick some μ. By the definition of $\tilde{K}$, we have

$$\underline{v}_i^{\mu}(\delta) \geq \max_{s_i \in S_i} v_i^{\mu}(\delta, s_i, s_{-i}^{\delta,\mu}) - (1-\delta)\tilde{K}.$$

Hence for each s_i,

$$\begin{aligned} \underline{v}_i^{\mu}(\delta) \geq & v_i^{\mu}(\delta, s_i, s_{-i}^{\delta,\mu}) - (1-\delta)\tilde{K} \\ \geq & (1-\delta)\sum_{t=1}^{T} \delta^{t-1} E[g_i^{\omega^t}(a^t)|\mu, s_i, s_{-i}^{\delta,\mu}] \\ & + \delta^T \{\underline{v}_i^{\mu}(\delta) - (1-\delta)\tilde{K}\} - (1-\delta)\tilde{K}. \end{aligned}$$

Here the second inequality follows from the definition of $\tilde{K}$, which ensures that the continuation payoff from period $T+1$ on is at least $\underline{v}_i^{\mu}(\delta) - (1-\delta)\tilde{K}$. Subtracting $\delta^T\{\underline{v}_i^{\mu}(\delta) - (1-\delta)\tilde{K}\} - (1-\delta)\tilde{K}$ from both sides and dividing them by $1-\delta^T$, we have

$$\frac{1-\delta}{1-\delta^T}\sum_{t=1}^{T} \delta^{t-1} E[g_i^{\omega^t}(a^t)|\mu, s_i, s_{-i}^{\delta,\mu}] \leq \underline{v}_i^{\mu}(\delta) + \frac{1-\delta}{1-\delta^T}(\delta^T \tilde{K} + \tilde{K}).$$

Since this inequality holds for all s_i, taking the maximum of the left-hand side over all s_i,

$$\max_{s_i \in S_i} \frac{1-\delta}{1-\delta^T}\sum_{t=1}^{T} \delta^{t-1} E[g_i^{\omega^t}(a^t)|\mu, s_i, s_{-i}^{\delta,\mu}] \leq \underline{v}_i^{\mu}(\delta) + \frac{1-\delta}{1-\delta^T}(\delta^T \tilde{K} + \tilde{K}).$$

Since $s_{-i}^{\delta,\mu}$ does not necessarily minimize the left-hand side,

$$\min_{s_{-i}\in S_{-i}^{\text{pub}}}\max_{s_i\in S_i}\frac{1-\delta}{1-\delta^T}\sum_{t=1}^{T}\delta^{t-1}E[g_i^{\omega^t}(a^t)|\mu,s]\leq \underline{v}_i^{\mu}(\delta)+\frac{1-\delta}{1-\delta^T}(\delta^T\tilde{K}+\tilde{K}).$$

Taking $\delta\to 1$, we obtain

$$\underline{v}_i^{\mu}(T)\leq \underline{v}_i+\frac{2\tilde{K}}{T}.$$

Now, pick μ and δ, and pick a public strategy s_{-i} which differs from $s_{-i}^{\delta,\mu}$ only in the play in the first T periods; i.e., choose s_{-i} such that $s_{-i}|_{h^T}=s_{-i}^{\mu}|_{h^T}$ for each h^T. (The play in the first T periods can be arbitrarily chosen.) Then,

$$\begin{aligned}\underline{v}_i^{\mu}(\delta)&\leq \max_{s_i\in S_i} v_i^{\mu}(\delta,s)\\ &\leq \max_{s_i\in S_i}\left[(1-\delta)\sum_{t=1}^{T}\delta^{t-1}E[g_i^{\omega^t}(a^t)|\mu,s]+\delta^T\{\underline{v}_i^{\mu}(\delta)+(1-\delta)\tilde{K}\}\right].\end{aligned}$$

Here the first inequality follows from the fact that s_{-i} is not necessarily the minimax strategy, and the second follows from the definition of $\tilde{K}$, which ensures that the continuation payoff from period $T+1$ on is at most $\underline{v}_i^{\mu}(\delta)+(1-\delta)\tilde{K}$. Subtracting $\delta^T\{\underline{v}_i^{\mu}(\delta)+(1-\delta)\tilde{K}\}$ from both sides and dividing them by $1-\delta^T$,

$$\max_{s_i\in S_i}\frac{1-\delta}{1-\delta^T}\sum_{t=1}^{T}\delta^{t-1}E[g_i^{\omega^t}(a^t)|\mu,s]\geq \underline{v}_i^{\mu}(\delta)-\frac{1-\delta}{1-\delta^T}\delta^T\tilde{K}.$$

Since this inequality holds for all public strategies s_{-i},

$$\min_{s_{-i}\in S_{-i}^{\text{pub}}}\max_{s_i\in S_i}\frac{1-\delta}{1-\delta^T}\sum_{t=1}^{T}\delta^{t-1}E[g_i^{\omega^t}(a^t)|\mu,s]\geq \underline{v}_i^{\mu}(\delta)-\frac{1-\delta}{1-\delta^T}\delta^T\tilde{K}.$$

Taking $\delta\to 1$, we obtain

$$\underline{v}_i^{\mu}(T)\geq \underline{v}_i-\frac{\tilde{K}}{T}.$$

Combining the above two inequalities,

$$\underline{v}_i-\frac{\tilde{K}}{T}\leq \underline{v}_i^{\mu}(T)\leq \underline{v}_i+\frac{2\tilde{K}}{T}.$$

So by setting $K=2\tilde{K}$, we obtain the result. *Q.E.D.*

The next proposition is a counterpart to Proposition 5. It shows that in the T-period block, the minimax strategy for some dummy belief $\tilde{\mu}$ can approximate the minimax payoff regardless of the true belief μ.

Proposition C4. *Suppose that the full support assumption holds. Then there is $K > 0$ such that for any i, for any T, and for any $\tilde{\mu}$, there is a public strategy s_{-i} such that for each pure strategy $s_i \in \arg\max_{\tilde{s}_i \in S_i} v_i^{\tilde{\mu}}(T, \tilde{s}_i, s_{-i}^{\tilde{\mu}})$,*

$$\left| v_i^{\mu}(T,s) - \underline{v}_i^{\mu}(T) \right| \leq \frac{K}{T}$$

and

$$\max_{\tilde{s}_i \in S_i} v_i^{\mu}(T-t, \tilde{s}_i, s_{-i}|_{h^t}) - v_i^{\mu}(T-t, s|_{h^t}) \leq \frac{K}{T-t}$$

for each μ, $t \in \{0, \cdots, T-1\}$, and h^t.

Proof. Choose $s_{-i}^{\delta,\mu}$ and $\tilde{K}$ as in the proof of Proposition C3 Pick μ, T, $t \in \{0, \cdots, T-1\}$, and h^t arbitrarily. Then by the definition of $\tilde{K}$, we have

$$\underline{v}_i^{\mu}(\delta) + (1-\delta)\tilde{K} \geq \max_{s_i \in S_i} v_i^{\omega}(\delta, s_i, s_{-i}^{\delta,\mu}|_{h^t})$$

for each ω. Then for each ω and s_i, we have

$$\begin{aligned} \underline{v}_i^{\mu}(\delta) + (1-\delta)\tilde{K} \geq & v_i^{\omega}(\delta, s_i, s_{-i}^{\delta,\mu}|_{h^t}) \\ \geq & (1-\delta) \sum_{\tilde{t}=1}^{T-t} \delta^{\tilde{t}-1} E[g_i^{\omega^{\tilde{t}}}(a^{\tilde{t}}) | \omega, s_i, s_{-i}^{\delta,\mu}|_{h^t}] \\ & + \delta^{T-t} \{ \underline{v}_i^{\mu}(\delta) - (1-\delta)\tilde{K} \}. \end{aligned}$$

Here the second inequality follows from the definition of $\tilde{K}$, which ensures that the continuation payoff from period $T-t+1$ on is at most $\overline{v}_i(\delta) - (1-\delta)\tilde{K}$. Subtracting $\delta^{T-t}\underline{v}_i^{\mu}(\delta) + (1-\delta)\tilde{K}$ from both sides and dividing them by $1-\delta^{T-t}$, we obtain

$$\underline{v}_i^{\mu}(\delta) \geq \frac{1-\delta}{1-\delta^{T-t}} \sum_{\tilde{t}=1}^{T-t} \delta^{\tilde{t}-1} E[g_i^{\omega^{\tilde{t}}}(a^{\tilde{t}}) | \omega, s_i, s_{-i}^{\delta,\mu}|_{h^t}] - \frac{1-\delta}{1-\delta^{T-t}} (\delta^{T-t}\tilde{K} + \tilde{K}).$$

Since this inequality holds for each ω, s_i, $t \in \{0, \cdots, T-1\}$, and h^t,

$$\underline{v}_i^{\mu}(\delta) \geq \max_{\substack{t \in \{0,\cdots,T-1\} \\ h^t \in H^t \\ \omega \in \Omega \\ s_i \in S_i}} \left\{ \begin{array}{l} \frac{1-\delta}{1-\delta^{T-t}} \sum_{\tilde{t}=1}^{T-t} \delta^{\tilde{t}-1} E[g_i^{\omega^{\tilde{t}}}(a^{\tilde{t}}) | \omega, s_i, s_{-i}^{\delta,\mu}|_{h^t}] \\ -\frac{1-\delta}{1-\delta^{T-t}} (\tilde{K}+1) \end{array} \right\}.$$

Since $s_{-i}^{\delta,\mu}$ is not necessarily the minimizer of the right-hand side,

$$\underline{v}_i^{\mu}(\delta) \geq \min_{s_{-i}\in S_{-i}} \max_{\substack{t\in\{0,\cdots,T-1\}\\ h^t\in H^t\\ \omega\in\Omega\\ s_i\in S_i}} \left\{ \begin{array}{l} \dfrac{1-\delta}{1-\delta^{T-t}} \displaystyle\sum_{\tilde{t}=1}^{T-t} \delta^{\tilde{t}-1} E[g_i^{\omega^{\tilde{t}}}(a^{\tilde{t}})|\omega, s_i, s_{-i}|_{h^t}] \\ -\dfrac{1-\delta}{1-\delta^{T-t}}(\tilde{K}+1) \end{array} \right\}.$$

Taking $\delta \to 1$, we obtain

$$\underline{v}_i \geq \min_{s_{-i}\in S_{-i}} \max_{\substack{t\in\{0,\cdots,T-1\}\\ h^t\in H^t\\ \omega\in\Omega\\ s_i\in S_i}} \left(v_i^{\omega}(T-t, s_i, s_{-i}|_{h^t}) - \frac{\tilde{K}+1}{T-t} \right).$$

Let s_{-i} be a solution to the problem on the right-hand side of the above inequality. Then

$$\underline{v}_i + \frac{\tilde{K}+1}{T-t} \geq \max_{s_i\in S_i} v_i^{\omega}(T-t, s_i, s_{-i}|_{h^t}).$$

for each ω, $t \in \{0,\cdots,T-1\}$, and h^t. Since $\max_{s_i\in S_i} v_i^{\mu}(T-t, s_i, s_{-i}|_{h^t})$ is convex with respect to μ, it is maximized by some extreme belief. This and the fact that the above inequality holds for all ω imply that

$$\underline{v}_i + \frac{\tilde{K}+1}{T-t} \geq \max_{s_i\in S_i} v_i^{\mu}(T-t, s_i, s_{-i}|_{h^t}) \geq \underline{v}_i^{\mu}(T-t).$$

for each μ, $t \in \{0,\cdots,T-1\}$, and h^t, where the second inequality follows from the fact that $s_{-i}|_{h^t}$ is not the minimax strategy for the $(T-t)$-period game with the initial prior μ.

Proposition C3 ensures that there is $\tilde{K}'$ such that $|\underline{v}_i^{\mu}(T) - \underline{v}| \leq \frac{\tilde{K}'}{T}$ for each T and μ. Choose such $\tilde{K}'$. Then from the above inequality, we have

$$\underline{v}_i^{\mu}(T-t) + \frac{\tilde{K}+\tilde{K}'+1}{T-t} \geq \max_{s_i\in S_i} v_i^{\mu}(T-t, s_i, s_{-i}|_{h^t}) \geq \underline{v}_i^{\mu}(T-t).$$

for each μ, $t \in \{0,\cdots,T-1\}$, and h^t. Note that $\tilde{K}$ and $\tilde{K}'$ do not depend on T. The rest of the proof is exactly the same as the proof of Proposition 5. *Q.E.D.*

Using the above results, we can prove the folk theorem using T-period block strategies. The following is the self-generation theorem for T-period block strategies:

Definition C1. A pair (s,v) of a public strategy profile and a payoff vector is *ex-post enforceable with respect to* (δ,T) if there is a function $w: H^T \to \boldsymbol{R}^N$ such that

$$v_i = (1-\delta)\sum_{t=1}^{T}\delta^{t-1}E[g_i^{\omega^t}(a^t)|\omega,s] + \delta^T E[w_i(h^T)|\omega,s]$$

for all ω and i, and

$$v_i \geq (1-\delta)\sum_{t=1}^{T}\delta^{t-1}E[g_i^{\omega^t}(a^t)|\omega,\tilde{s}_i,s_{-i}] + \delta^T E[w_i(h^T)|\omega,\tilde{s}_i,s_{-i}]$$

for all ω, i, and $\tilde{s}_i$.

Definition C2. A subset W of $\boldsymbol{R}^{N\times|\Omega|}$ is *ex-post self-generating with respect to* (δ,T) if for each $v \in W$, there is a public strategy profile s and $w: H \to W$ such that (s,v) is ex-post enforceable with respect to (δ,T) using w.

Proposition C5. *Fix δ. If W is bounded and ex-post self-generating with respect to (δ,T) for some T, then for each payoff vector $v \in W$, there is a public ex-post equilibrium which yields the payoff v regardless of the initial state ω.*

So to prove the folk theorem, it is sufficient to show that any smooth subset of the interior of V^* is ex-post self-generating. The rest of the proof is quite similar to that of Proposition 6 and hence omitted.

Appendix D: Relaxing the Full Support Assumption

In this appendix, we will show that the full support assumption is stronger than necessary for the folk theorem. More precisely, we will show that the folk theorem obtains as long as uniform and robust connectedness, the common support condition, and the relative interior condition hold.

D.1 Uniform Connectedness and Feasible Payoff Set

Yamamoto (2016) introduces the idea of *uniform connectedness*, which is a natural extension of the communicating state assumption for stochastic games with observable states (Dutta (1995)). Uniform connectedness is about a condition on how the *support* of the belief evolves over time, and requires that players can jointly drive the support from any set $\Omega^* \subseteq \Omega$ to any other set $\tilde{\Omega}^* \subset \Omega$ (except the case in which the set $\tilde{\Omega}^*$ is "transient" in the sense that the probability of the support being $\tilde{\Omega}^*$ is negligible in a distant future.) Yamamoto (2016) shows that if the game is uniformly connected, then the feasible payoff set is invariant to the initial prior in the limit as $\delta \to 1$.

To give the formal definition of uniform connectedness, the following notation is useful. When the initial prior μ and a pure action sequence $(a^1, \cdots, a^T)$ are given, the posterior belief μ^{T+1} in period $T+1$ can be regarded as a random variable, because μ^{T+1} is determined by a realized signal sequence $(y^1, \cdots, y^T)$ which is randomly drawn given μ and $(a^1, \cdots, a^T)$. (Here μ^{T+1} is common for all players, because they play pure actions each period.) Let $\Pr(\mu^{T+1} = \tilde{\mu}|\mu, a^1, \cdots, a^T)$ denote the probability that this posterior belief in period $T+1$ is $\tilde{\mu}$. Likewise, let $\Pr(\mu^{T+1} = \tilde{\mu}|\mu, s)$ denote the probability that the posterior belief in period $T+1$ is $\tilde{\mu}$ given that the initial prior is μ and players play a pure strategy profile s until period T.

Definition D1. A non-empty subset $\Omega^* \subseteq \Omega$ is *globally accessible* if there is $\pi^* > 0$ such that for any initial prior μ, there is a natural number $T \leq 4^{|\Omega|}$, an action sequence $(a^1, \cdots, a^T)$, and a belief $\tilde{\mu}$ whose support is included in Ω^* such that

$$\Pr(\mu^{T+1} = \tilde{\mu}|\mu, a^1, \cdots, a^T) \geq \pi^*.$$

In words, global accessibility of Ω^* requires that given any initial prior μ, players can move the support of the posterior belief to Ω^* or its subset with positive probability, and this probability is bounded away from zero uniformly in μ. As explained in Yamamoto (2016), restricting attention to $T \leq 4^{|\Omega|}$ is without loss of generality.

Definition D2. A subset $\Omega^* \subseteq \Omega$ is *uniformly transient* if it is not globally accessible and for any pure strategy profile s and for any μ whose support is Ω^*, there is a natural number $T \leq 2^{|\Omega|}$ and a belief $\tilde{\mu}$ whose support is globally accessible such that $\Pr(\mu^{T+1} = \tilde{\mu}|\mu, s) > 0$.

In words, uniform transience of Ω^* implies that if the support of the current belief is Ω^*, then regardless of future actions, the support of the posterior belief cannot stay there forever and must reach some globally accessible set with positive probability at some point. As discussed in Yamamoto (2016), uniformly transient sets are "not essential" in the sense that the probability of the support being in a uniformly transient set is almost negligible in a distant future.

Our assumption, uniform connectedness, requires that each subset Ω^* be either globally accessible or uniformly transient.

Definition D3. A stochastic game is *uniformly connected* if each subset Ω^* is globally accessible or uniformly transient.

Uniform connectedness is weaker than the full support assumption, and is satisfied in a wide range of economic examples, including the ones discussed in Section 2.1.

Our Proposition 2 is valid even if the full support assumption is replaced with uniform connectedness; see Yamamoto (2016). Likewise, we can show that Proposition 3 is valid without the full support assumption, as stated below. The only difference from Proposition 3 is that the dummy belief $\tilde{\mu}$ must be an interior belief here. The proof is very similar to Step 1 in the proof of Proposition 3 and hence omitted.

Proposition D1. *Suppose that the game is uniformly connected. Then for each* $\varepsilon > 0$*, there is* $\overline{\delta} \in (0,1)$ *such that for each* λ*, for each* $\delta \in (\overline{\delta},1)$*, for each* $\tilde{\mu}$ *with* $\tilde{\mu}(\omega) \in [\overline{\pi}, 1-\overline{\pi}]$ *for all* ω*, for each pure public strategy profile* $s^{\tilde{\mu}}$ *with* $\lambda \cdot v^{\tilde{\mu}}(\delta, s^{\tilde{\mu}}) = \max_{v \in V^{\tilde{\mu}}(\delta)} \lambda \cdot v$*, for each* $t \geq 0$*, for each* h^t*, and for each* $\mu \in \triangle\Omega(\tilde{\mu}, s, h^t)$

$$\left| \max_{v \in V^{\mu}(\delta)} \lambda \cdot v - \lambda \cdot v^{\mu}(\delta, s^{\tilde{\mu}}|_{h^t}) \right| < \varepsilon$$

where $\Omega(\tilde{\mu}, s, h^t)$ *is the set of states which can happen with positive probability in period* $t+1$ *given the initial prior* $\tilde{\mu}$*, the strategy profile* s*, and the history* h^t*.*

D.2 Robust Connectedness and Minimax Payoffs

Proposition 4 shows that the limit minimax payoff is invariant to the initial prior, under the full support assumption. Here we show that the full support assumption can be replaced with a set of weaker conditions, called robust connectedness, the common support condition, and the relative interior condition.

We begin with presenting the common support condition.

Definition D4. The *common support condition* holds if for each ω, $\tilde{\omega}$, a, $\tilde{a}$, and y such that $\pi^{\omega}(y,\tilde{\omega}|a) > 0$ and $\pi_Y^{\omega}(y|\tilde{a}) > 0$, we have $\pi^{\omega}(y,\tilde{\omega}|\tilde{a}) > 0$.

In words, the common support condition requires that if the state $\tilde{\omega}$ can happen tomorrow with positive probability given the current state ω, the current action profile a, and the current signal y, then the same is true for any different action profile $\tilde{a}$ which induce the signal y with positive probability. An important consequence of this condition is that the support of each player's posterior belief does not depend on the past actions; that is, once the initial prior μ and the signal sequence $(y^1,\cdots,y^t)$ are given, the support of each player i's posterior belief μ_i^{t+1} in period $t+1$ is uniquely determined (although the belief itself may depend on the past actions). Accordingly, as long as players have the same initial prior μ, the support of the posterior belief is common across all players and is common knowledge. For each μ and h^t, let $\Omega(\mu,h^t) \subseteq \Omega$ denote this support. (Let $\Omega(\mu,h^0) = \text{supp}\mu$. Also let $\Omega(\mu,h^t) = \emptyset$ if h^t never happens given the initial prior μ for any strategy profile s.)

Next, we present the relative interior condition:

Definition D5. The *relative interior condition* holds if for each ω, $\tilde{\omega}$, $\hat{\omega}$, a, and y such that $\pi^{\omega}(y,\tilde{\omega}|a) > 0$ and $\pi_Y^{\hat{\omega}}(y|a) > 0$, we have $\pi^{\hat{\omega}}(y,\tilde{\omega}|a) > 0$.

In words, the relative interior condition requires that (conditional on the current action profile a and the current signal y), different states ω and $\hat{\omega}$ induce the same set of the next states $\tilde{\omega}$. When this condition is satisfied, each player i's posterior belief in period $t \geq 2$ is always in the relative interior. To be precise, take a pure public strategy s_i, a public strategy s_{-i}, and an initial prior μ arbitrarily. Take an arbitrary history h^t which can happen with positive probability given μ and s, and let $\mu_i(h^t|\mu,s) \in \triangle\Omega$ be player i's posterior belief after h^t. Then the relative interior condition ensures that $\mu_i(h^t|\mu,s)[\omega] \geq \overline{\pi}$ or $\mu_i(h^t|\mu,s)[\omega] = 0$ for all ω; that is, the belief $\mu_i(h^t|\mu,s)$ is in the relative interior of the set $\triangle\Omega(\mu,h^t)$.

Lastly, we present the robust connectedness assumption, which is introduced by Yamamoto (2016). Roughly speaking, it requires that the opponents can drive the support of player i's belief from any set to any other set. The definition here is a bit different from the one in Yamamoto (2016), because we assume that actions are not observable. We first define robust accessibility and transience.

Definition D6. A non-empty subset $\Omega^* \subseteq \Omega$ is *robustly accessible despite player i* if there is $\pi^* > 0$ such that for any initial prior μ, for any pure strategy s_i, there is a natural number $T \leq 4^{|\Omega|}$ and a public history h^T such that $\Pr(h^T|\mu, s_i, \alpha_{-i}^1, \cdots, \alpha_{-i}^T) \geq \pi^*$ and $\Omega(\mu, h^T) = \Omega^*$, where $(\alpha_{-i}^1, \cdots, \alpha_{-i}^T)$ is the action sequence which mixes all actions equally each period.

Robust accessibility of Ω^* ensures that given any initial prior μ, the opponents can move the support to the set Ω^* regardless of player i's play, by mixing all actions equally each period. When actions are observable (i.e., when the public signal y reveals the action profile today), the above definition reduces to that of Yamamoto (2016).

Definition D7. A subset $\Omega^* \subseteq \Omega$ is *transient given player i* if it is not robustly accessible and there is $\pi^* > 0$ such that for any μ whose support is Ω^*, and for any public strategy $s_{-i} \in S_{-i}^*$, there is a natural number $T \leq 4^{|\Omega|}$ and a belief $\tilde{\mu}$ whose support is robustly accessible such that $\Pr(\mu_i^{T+1} = \tilde{\mu}|\mu, \alpha_i^1, \cdots, \alpha_i^T, s_{-i}) \geq \pi^*$, where $(\alpha_i^1, \cdots, \alpha_i^T)$ is the action sequence which mixes all actions equally each period.

Transience of Ω^* requires that if the current support is Ω^*, player i can force the support to move to a different set at some point, regardless of the opponents' play.

Now we are ready to state the definition of robust connectedness.

Definition D8. The game is *robustly connected* if for each i, each non-empty subset $\Omega^* \subseteq \Omega$ is either robustly accessible despite player i or transient given player i.

The following proposition extends Proposition 4 and shows that when the above conditions are satisfied and δ is sufficiently large, the minimax payoffs are similar across all priors μ. The proof is given in at the end of this appendix.

Proposition D2. *Suppose that the common support condition and the relative interior condition hold. Suppose also that the game is robustly connected. Then for each $\varepsilon > 0$, there is $\overline{\delta} \in (0,1)$ such that $|\underline{v}_i^{\mu}(\delta) - \underline{v}_i^{\tilde{\mu}}(\delta)| < \varepsilon$ for each i, $\delta \in (\overline{\delta}, 1)$, μ, and $\tilde{\mu}$.*

The next proposition extends Proposition 5 and shows that the minimax strategy profile $s^{\tilde{\mu}}$ for some dummy belief $\tilde{\mu}$ approximates the minimax payoff regardless of the true belief μ. The difference from Proposition 5 is that now the dummy belief $\tilde{\mu}$ must be an interior belief. The proof is very similar to that of Proposition 5 and hence omitted.

Proposition D3. *Suppose that the common support condition and the relative interior condition hold. Suppose also that the game is robustly connected. Then for each $\varepsilon > 0$, there is $\overline{\delta} \in (0,1)$ such that for any i, for any $\delta \in (\overline{\delta}, 1)$, and for any $\tilde{\mu}$ with $\tilde{\mu}(\omega) \in [\overline{\pi}, 1-\overline{\pi}]$ for all ω, there is a public strategy $s_{-i}^{\tilde{\mu}}$ such that for each player i's pure strategy $s_i^{\tilde{\mu}} \in \arg\max_{\tilde{s}_i \in S_i} v_i^{\tilde{\mu}}(\delta, \tilde{s}_i, s_{-i}^{\tilde{\mu}})$,*

$$\left| v_i^{\mu}(\delta, s^{\tilde{\mu}}|_{h^t}) - \underline{v}_i^{\mu}(\delta) \right| < \varepsilon$$

and

$$\max_{\tilde{s}_i \in S_i} v_i^{\mu}(\delta, \tilde{s}_i, s_{-i}^{\tilde{\mu}}|_{h^t}) - v_i^{\mu}(\delta, s^{\tilde{\mu}}|_{h^t}) < \varepsilon$$

for each $t \geq 0$, h^t, and $\mu \in \triangle\Omega(\tilde{\mu}, h^t)$.

D.3 Belief Convergence Theorem

In the last subsection, we have seen that the relative interior condition is useful to derive invariance of the limit minimax payoffs. As we show in the following proposition, it is also useful to obtain the belief convergence theorem.

Proposition D4. *Suppose that the relative interior condition holds, and let $\beta = 1 - \frac{\overline{\pi}}{|\Omega|} \in (0,1)$. Then for each i, pure public strategy s_i, public strategy s_{-i}, μ, $\tilde{\mu}$, $t \geq 0$, and h^t such that $\Pr(h^t|\mu, s) > 0$ and $\Pr(h^t|\tilde{\mu}, s) > 0$,*

$$\left| \mu_i(h^t|\mu, s) - \mu_i(h^t|\tilde{\mu}, s) \right| \leq \beta^t.$$

Once we obtain the belief convergence theorem, it is easy to extend Lemma B9. The proof is omitted because it is almost identical with that of Lemma B9.

Lemma D1. *Suppose that the relative interior condition holds. Then for each δ, for each μ, for each $\tilde{\mu}$, for each $t \geq 1$, for each public history h^{t-1} such that*

$\Pr(h^{t-1}|\mu,s) > 0$ *and* $\Pr(h^{t-1}|\tilde{\mu},s) > 0$*, for each public strategy* s_{-i}*, and for each pure public strategy* $s_i \in \arg\max_{\tilde{s}_i \in S_i} v_i^{\tilde{\mu}}(\delta, \tilde{s}_i, s_{-i})$*, we have*

$$G_i^t(\delta,\mu,s,h^{t-1}) \leq \frac{\beta^{t-1}}{\overline{\pi}} \max_{\hat{\mu} \in \triangle\Omega(\tilde{\mu},h^{t-1})} \left(\max_{\tilde{s}_i \in S_i} v_i^{\hat{\mu}}(\delta,\tilde{s}_i,s_{-i}|_{h^{t-1}}) - v_i^{\hat{\mu}}(\delta,s|_{h^{t-1}}) \right),$$

and hence

$$G_i^*(\delta,s) \leq \frac{1}{(1-\beta)\overline{\pi}} \sup_{h^t \in H^*} \max_{\hat{\mu} \in \triangle\Omega(\tilde{\mu},h^t)} \left(\max_{\tilde{s}_i \in S_i} v_i^{\hat{\mu}}(\delta,\tilde{s}_i,s_{-i}|_{h^t}) - v_i^{\hat{\mu}}(\delta,s|_{h^t}) \right).$$

Here, we let $\max_{x \in X} f(x) = -\infty$ *if* X *is an empty set.*

D.4 Folk Theorem

So far, we have seen that Propositions 2, 3, 4, and 5, and Lemma B9 are valid even if the full support assumption does not hold. Accordingly, the following folk theorem holds. The proof is very similar to that of Proposition 6 and hence omitted. (We can show that Lemma B11, which is used in the proof of Proposition 6, remains valid even if the full support support assumption is not satisfied. The proof idea is similar to Proposition D2 and hence omitted.)

Proposition D5. *Suppose that the common support condition, the relative interior condition, uniform and robust connectedness, (IFR), and (PFR) are satisfied. Suppose also that public randomization is available. Then, for any smooth subset* W *of the interior of* V^**, there is* $\overline{\delta} \in (0,1)$ *such that for any* $\delta \in (\overline{\delta},1)$*, the set* W *is stochastically ex-post self-generating. Hence for each* $v \in W$*, there is a public ex-post equilibrium which yields the payoff* v *regardless of the initial state* ω*.*

The assumptions made in the above folk theorem (uniform and robust connectedness, the common support condition, and the relative interior condition) are satisfied in various economic examples. The following lemma shows that if the state is observable (i.e., $Y = \Omega \times \Omega \times Y^*$) and the state evolution is irreducible in the sense of Fudenberg and Yamamoto (2011b), then these assumptions are satisfied.

Lemma D2. *Suppose that the state is observable and the state evolution is irreducible. Then uniform and robust connectedness, the common support condition, and the relative interior condition hold.*

Proof. Since the state is observable, given a signal y in period one, the state ω in period two is common knowledge across players. This implies that clause (ii) in the definition of robust connectedness, as well as the common support condition, and the relative interior condition. Also, since the state evolution is irreducible, it is each to check that clause (i) in the definition of robust connectedness is satisfied. *Q.E.D.*

The next lemma shows that if the state is observable with delay (i.e., $Y = \Omega \times Y^*$) and the state evolution has a full support in the sense that $\pi^{\omega}(y, \tilde{\omega}|a) > 0$ for all ω, $\tilde{\omega}$, a, and y such that $\pi_y^{\omega}(y|a) > 0$, then the assumptions are satisfied.

Lemma D3. *Suppose that the state is observable with delay and the state evolution has a full support. Then robust connectedness, the common support condition, and the relative interior condition hold.*

D.5 Ex-Post Equilibria to Sequential Equilibria

When the full support assumption does not hold, some deviations can be observable, and accordingly there can be a Nash equilibrium payoff which is not achievable by any sequential equilibria. Here, we show that the sequential-equilibrium folk theorem holds when some additional assumptions are satisfied.

For each (i, j) with $i \neq j$ and for each α, let $\tilde{\Pi}_{ij}(\alpha)$ be a matrix with rows $\pi_Y^{\omega}(a_i, a_j, \alpha_{-ij})$ for all $\omega \in \Omega$, $a_i \in A_i$, and $a_j \in A_j$. In words, the matrix $\Pi_i(\alpha)$ is a collection of the marginal distributions of the public signal y induced by joint deviations by players i and j.

Definition D9. An action profile α has *strong full rank for* (i, j) if the matrix $\tilde{\Pi}_{ij}(\alpha)$ has rank equal to $|\Omega| \times |A_i| \times |A_j|$. An action profile α has *strong full rank* if it has strong full rank for all (i, j) with $i \neq j$.

Strong full rank requires that any joint deviation by players i and j must be distinguished by a public signal y. Very roughly, this condition is used to ensure that a modification of player j's actions at off-path histories does not influence player i's incentive. We have the following sequential-equilibrium folk theorem. The proof can be found at the end of this appendix.

Proposition D6. *Suppose that the common support condition, the relative interior condition, and uniform and robust connectedness are satisfied. Suppose also that public randomization is available, and each pure action profile has strong full rank. Then, for any interior point v of V^*, there is $\overline{\delta} \in (0,1)$ such that for any initial prior μ and $\delta \in (\overline{\delta}, 1)$, there is a sequential equilibrium which yields the payoff of v.*

Unfortunately, strong full rank is demanding, and it rules out many potential applications. For example, suppose that a two-player games in which the action set is $A_i = \{C, D\}$ for each i. Suppose that the game is symmetric so that the signal distribution induced by the profile (C,D) is the same as that by the profile (D,C). Then the pure action profile (C,C) cannot have strong full rank, and hence the assumption in Proposition D6 does not hold. (On the other hand, (PFR) can be satisfied in this symmetric game, as a mixed action profile α may have pairwise full rank.)

The next proposition shows that the sequential-equilibrium folk theorem is still valid when players can communicate. Suppose that at the end of each period t, each player i can send a public message $m_i \in M_i$. Assume that $M_i = A_i$; that, after each period, each player i can reveal her own action. Since the communication considered here is a cheap talk, each player may misreport to increase her continuation payoff; however, as the following proposition shows, we can construct a sequential equilibrium in which everyone reports her information truthfully after every history.

Proposition D7. *Suppose that the common support condition, the relative interior condition, uniform and robust connectedness, (IFR), and (PFR) are satisfied. Suppose also that public randomization is available and that players can communicate each period. Then, for any interior point v of V^*, there is $\overline{\delta} \in (0,1)$ such that for any initial prior μ and $\delta \in (\overline{\delta}, 1)$, there is a sequential equilibrium with the payoff of v in which everyone reports truthfully after every history.*

D.6 Proof of Proposition D2

Fix δ. Let $v_i^{\mu}(s_{-i})$ be as in the proof of Proposition 4. Let $\Delta(\mu, h^t)$ be the set of all beliefs $\tilde{\mu}$ such that $\tilde{\mu}(\omega) \geq \overline{\pi}$ for each $\omega \in \Omega(\mu, h^t)$ and $\tilde{\mu}(\omega) = 0$ for other ω.

Intuitively, any belief $\mu in \Delta(\mu, h^t)$ is in the relative interior of the set $\triangle\Omega(\mu, h^t)$, and is not too close to the boundary. Under the common support condition and the relative interior condition, given the initial prior μ and the public history h^t, player i's posterior belief must be in the set $\Delta(\mu, h^t)$ regardless of her past private actions.

Assume for now that there is a public strategy s_{-i}^{μ} which exactly achieves the minimax payoff for μ. Then after every history h^t, the continuation strategy $s_{-i}^{\mu}|_{h^t}$ should punish player i sufficiently harshly compared to any other strategy s_{-i}; more precisely, for each s_{-i}, there must be some belief $\tilde{\mu} \in \Delta(\mu, h^t)$ such that $v_i^{\tilde{\mu}}(s_{-i}^{\mu}|_{h^t}) \leq v_i^{\tilde{\mu}}(s_{-i})$. Indeed, if not and there is s_{-i} such that $v_i^{\tilde{\mu}}(s_{-i}^{\mu}|_{h^t}) > v_i^{\tilde{\mu}}(s_{-i})$ for all $\tilde{\mu} \in \Delta(\mu, h^t)$, then the strategy s_{-i}^{μ} is not the minimax strategy because the opponents can lower player i's payoff by replacing the continuation strategy $s_{-i}^{\mu}|_{h^t}$ with s_{-i}.

The following lemma shows that the same result holds even if there is no strategy which exactly achieves the minimax payoff. The proof is very similar to Lemma B3 and hence omitted.

Lemma D4. *For each μ, there is a public strategy s_{-i}^{μ} such that*

$$\left| \underline{v}_i^{\mu}(\delta) - v_i^{\mu}(s_{-i}^{\mu}) \right| < 1 - \delta. \tag{58}$$

and such that for any $t \geq 1$, for any h^t, and for any public strategy s_{-i}, there is $\tilde{\mu} \in \Delta(\mu, h^t)$ satisfying

$$v_i^{\tilde{\mu}}(s_{-i}^{\mu}|_{h^t}) < v_i^{\tilde{\mu}}(s_{-i}) + 1 - \delta. \tag{59}$$

For each μ, choose s_{-i}^{μ} as in the above lemma. Pick μ and h^t, and consider the corresponding strategy $s_{-i}^{\mu}|_{h^t}$. Then the payoff $v_i^{\tilde{\mu}}(s_{-}^{\mu}|_{h^t})$ is convex with respect to the initial belief $\tilde{\mu}$. In what follows, when we say the *convex curve* $v_i^{\tilde{\mu}}(s_{-}^{\mu}|_{h^t})$ or the *convex curve* induced by $s_{-}^{\mu}|_{h^t}$, it refers to the convex function $v_i^{\tilde{\mu}}(s_{-}^{\mu}|_{h^t})$ whose domain is restricted to $\tilde{\mu} \in \triangle\Omega(\mu, h^t)$. So when $\Omega(\mu, h^t) = \Omega$, the convex curve represents player i's payoff $v_i^{\tilde{\mu}}(s_{-}^{\mu}|_{h^t})$ for each initial belief $\tilde{\mu} \in \triangle\Omega$. On the other hand, when $\Omega(\mu, h^t) = \{\omega\}$, the convex curve is simply a scalar $v_i^{\omega}(s_{-}^{\mu}|_{h^t})$.

For each μ, $t \geq 0$, and $h^t \in H^t$, let

$$\overline{v}_i(s_{-i}^{\mu}|_{h^t}) = \max_{\tilde{\mu} \in \triangle\Omega(\mu, h^t)} v_i^{\tilde{\mu}}(s_{-}^{\mu}|_{h^t}).$$

That is, $\overline{v}_i(s^{\mu}_{-i}|_h)$ is the highest payoff attained by the convex function $v^{\tilde{\mu}}_i(s^{\mu}_{-i}|_{h^t})$. Note that different (μ, h^t) induce different strategies $s^{\mu}_{-i}|_{h^t}$, and hence different convex curves, and hence different highest payoffs $\overline{v}_i(s^{\mu}_{-i}|_{h^t})$. Take the supremum of these highest payoffs, and choose (μ^*, h^*) to approximate the supremum, that is,

$$\left| \sup_{\mu \in \triangle\Omega} \sup_{h \in H} \overline{v}_i(s^{\mu}_{-i}|_h) - \overline{v}_i(s^{\mu^*}_{-i}|_{h^*}) \right| < 1 - \delta. \tag{60}$$

We call $\overline{v}_i(s^{\mu^*}_{-i}|_{h^*})$ the *maximal value*, because it approximates $\sup_{\mu \in \triangle\Omega} \sup_{h \in H} \overline{v}_i(s^{\mu}_{-i}|_h)$, which is greater than any payoffs attained by any convex curves.

Since $v^{\mu}_i(s^{\mu^*}_{-i}|_{h^*})$ is convex, it is maximized when μ is an extreme point; i.e., it is maximized when the initial prior puts probability one on some state $\omega \in \Omega(\mu^*, h^*)$. Let ω denote this state.

The rest of the proof consists of three steps. In the first step, we show that there is a belief μ^{**} such that the minimax payoff for the initial prior μ^{**} approximates the maximal value. The proof is very similar to Steps 1 through 3 of the proof of Proposition 4.

In the second step, we show that for any (μ, h^t) such that $\Omega(\mu, h^t)$ is robustly accessible, the corresponding convex curve $v^{\tilde{\mu}}_i(s^{\mu}_{-}|_{h^t})$ is almost flat and approximates the maximal value. The result in the first step plays an important role here.

Then in the third step, we show that for any (μ, h^t) such that $\Omega(\mu, h^t)$ is transient, the corresponding convex curve $v^{\tilde{\mu}}_i(s^{\mu}_{-}|_{h^t})$ is almost flat and approximates the maximal value. This and the result in the second step ensure that all the convex curves are almost flat and approximate the maximal value, which implies that all the minimax payoffs approximate the maximal value.

D.6.1 Step 1: Minimax Payoff for μ^{**}

The following lemma extends Lemma B5 to the case in which the full support assumption does not hold; it shows that there is the opponents' strategy $s^{\mu}_{-i}|_{h^t}$ such that the corresponding convex curve $v^{\tilde{\mu}}_i(s^{\mu}_{-i}|_{h^t})$ is almost flat and approximates the maximal value uniformly in $\tilde{\mu} \in \triangle\Omega(\mu, h^t)$. The lemma also shows that such a strategy can be obtained by letting $\mu = \mu^*$ and $h^t = (h^*, y)$ for some y. Let $C = \frac{2\overline{g}}{\overline{\pi}^2}$ and let $\tilde{C} = \frac{1}{\overline{\pi}^2}$. Let s^*_i be player i's best reply when the initial state is ω and the opponents play $s^{\mu^*}_{-i}|_{h^*}$.

Lemma D5. *Pick y such that $\pi_Y^{\omega}(y|s_i^*(h^0), s_{-i}^{\mu^*}(h^*)) > 0$. Then for any $\tilde{\mu} \in \triangle\Omega(\mu^*, (h^*, y))$,*

$$\left|\overline{v}_i(s_{-i}^{\mu^*}|_{h^*}) + (1-\delta) - v_i^{\tilde{\mu}}(s_{-i}^{\mu^*}|_{(h^*,y)})\right| \leq \frac{1-\delta}{\delta}C + (1-\delta)\tilde{C}.$$

To prove this lemma, it is sufficient to find a relative interior belief $\tilde{\mu} \in \triangle\Omega(\mu^*, (h^*, y))$ such that the payoff $v_i^{\tilde{\mu}}(s_{-i}^{\mu^*}|_{(h^*,y)})$ approximates the maximal value; indeed, if there is such a relative interior belief, then Lemma B4 ensures that the convex curve $v_i^{\tilde{\mu}}(s_{-i}^{\mu^*}|_{(h^*,y)})$ is almost flat and approximates the maximal value for all $\tilde{\mu} \in \triangle\Omega(\mu^*, (h^*, y))$.

To find such a relative interior belief $\tilde{\mu}$, suppose that the current state is ω and that the opponents play $s_{-i}^{\mu^*}|_{h^*}$ from now on. Suppose that player i takes the best reply strategy s_i^*. By the definition, player i's payoff achieves the maximal value. Now, suppose that no one deviates today and the signal y is observed. Letting $\tilde{\mu}$ be player i's posterior belief in period two, her continuation payoff from period two is denoted by $v_i^{\tilde{\mu}}(s_{-i}^{\mu^*}|_{(h^*,y)})$. Then we can show that this continuation payoff approximates the maximal value; as in the proof of Lemma B4. Also, under the common support condition and the relative interior condition, the support of the posterior belief today is solely determined by the public signal y in the last period; hence we have $\Omega(\omega, y) = \Omega(\mu^*, (h^*, y))$, and the belief $\tilde{\mu}$ is indeed in the relative interior of $\triangle\Omega(\mu^*, (h^*, y))$. (Indeed, we can show that $\tilde{\mu} \in \Delta(\mu^*, (h^*, y))$.) Hence this belief $\tilde{\mu}$ satisfies all the desired conditions. The formal proof of the lemma is omitted, as it is very similar to that of Lemma B5.

The above lemma shows that the convex curve induced by $s_{-i}^{\mu^*}|_{(h^*,y)}$ is almost flat and approximates the maximal value. The next lemma extends this result; it shows that for any (μ, h^t) such that $\Omega(\mu, h^t) = \Omega(\mu^*, (h^*, y))$, the corresponding convex curve is almost flat and approximates the maximal value. The proof is very similar to Lemma B6 and hence omitted. Let $C' = \frac{C}{\underline{\underline{\pi}}}$ and $\tilde{C}' = \frac{\tilde{C}+1}{\underline{\underline{\pi}}}$.

Lemma D6. *Pick y such that $\pi_Y^{\omega}(y|s_i^*(h^0), s_{-i}^{\mu^*}(h^*)) > 0$, and then pick (μ, h^t) such that $\Omega(\mu, h^t) = \Omega(\mu^*, (h^*, y))$. Then for each $\tilde{\mu} \in \triangle\Omega(\mu, h^t)$,*

$$\left|\overline{v}_i(s_{-i}^{\mu^*}|_{h^*}) + (1-\delta) - v_i^{\tilde{\mu}}(s_{-i}^{\mu}|_{h^t})\right| \leq \frac{1-\delta^T}{\delta^T}C' + (1-\delta)\tilde{C}'.$$

Pick y as in the lemma, and pick an arbitrary belief $\mu^{**} \in \triangle\Omega(\mu^*, (h^*, y))$. Letting $\mu = \tilde{\mu} = \mu^{**}$ and $h^t = h^0$, the lemma ensures that the minimax payoff given the initial prior μ^{**} approximates the maximal value. That is,

$$\left|\overline{v}_i(s_{-i}^{\mu^*}|_{h^*}) + (1-\delta) - v_i^{\mu^{**}}(s_{-i}^{\mu^{**}})\right| \leq \frac{1-\delta^T}{\delta^T}C' + (1-\delta)\tilde{C}'. \tag{61}$$

D.6.2 Step 2: Convex Curves when $\Omega(\mu, h^t)$ is Robustly Accessible

Choose $\pi^* > 0$ so that the condition stated in the definition of robust accessibility and transience is satisfied.

Pick a pair (μ, h^t) such that the set $\Omega(\mu, h^t)$ is robustly accessible. Intuitively, if the initial prior is μ and the past public history is h^t, player i's posterior belief must be in the set $\triangle\Omega(\mu, h^t)$ regardless of her private history. In particular, under the common support condition and the relative interior condition, her posterior belief must be a relative interior belief so that it must be in the set $\Delta(\mu, h^t)$.

Suppose that the initial prior is μ^{**} and that the opponents play the following strategy $\tilde{s}_{-i}(\mu, h^t)$:

- Randomize all actions equally likely until player i's posterior belief reaches the set $\Delta(\mu, h^t)$.
- Once it happens, then play $s_{-i}^{\mu}|_{h^t}$ in the rest of the game.

Intuitively, $\tilde{s}_{-i}(\mu, h^t)$ asks the opponents to randomize all actions equally and wait until player i's posterior belief reaches the set $\Delta(\mu, h^t)$; and once it happens, they switch the play to $s_{-i}^{\mu}|_{h^t}$ in the rest of the game. This strategy is well-defined, because the common support condition ensures that the support of player i's posterior belief is common knowledge after every public history, and the relative interior condition ensures that her posterior belief is always a relative interior belief.

Suppose that player i takes a best reply. Since $\Omega(\mu, h^t)$ is robustly accessible and the relative interior condition holds, the switch to $s_{-i}^{\mu}|_{h^t}$ must happen in finite time with probability one. Hence for δ close to one, player i's expected payoff is approximated by the expected continuation payoff after the switch. Since the belief at the time of the switch is in the set $\Delta(\mu, h^t)$, this continuation payoff is at most

$$K_i(\mu, h^t) = \max_{\tilde{\mu} \in \Delta(\mu, h^t)} v_i^{\tilde{\mu}}(s_{-i}^{\mu}|_{h^t}).$$

So player i's payoff against the above strategy $\tilde{s}_{-i}(\mu,h^t)$ is approximately at most $K_i(\mu,h^t)$. Formally, we have the following lemma. The proof is very similar to that of Yamamoto (2016) and hence omitted.

Lemma D7. *For each (μ,h^t) such that $\Omega(\mu,h^t)$ is robustly accessible,*

$$v_i^{\mu^{**}}(\tilde{s}_{-i}(\mu,h^t)) \leq K_i(\mu,h^t) + \frac{(1-\delta^{4^{|\Omega|}})2\overline{g}}{\pi^*}.$$

For now, ignore the term $\frac{(1-\delta^{4^{|\Omega|}})2\overline{g}}{\pi^*}$ because it is approximately zero when δ is close to one. Then the above lemma ensures that the payoff $K_i(\mu,h^t)$ is at least $v_i^{\mu^{**}}(\tilde{s}_{-i}(\mu,h^t))$, which must be at least the minimax payoff $v_i^{\mu^{**}}(s_{-i}^{\mu^{**}})$ due to the fact that $\tilde{s}_{-i}(\mu,h^t)$ is not necessarily the minimax strategy. On the other hand, the payoff $K_i(\mu,h^t)$ cannot exceed the maximal value. Hence the payoff $K_i(\mu,h^t)$ is between the minimax payoff $v_i^{\mu^{**}}(s_{-i}^{\mu^{**}})$ and the maximal value. Now, from Step 1, we know that these two bounds are close each other; hence the payoff $K_i(\mu,h^t) = \max_{\tilde{\mu}\in\Delta(\mu,h^t)} v_i^{\tilde{\mu}}(s_{-i}^{\mu}|_{h^t})$ approximates the maximal value. That is, the convex curve $v_i^{\tilde{\mu}}(s_{-i}^{\mu}|_{h^t})$ approximates the maximal value for some belief $\tilde{\mu} \in \Delta(\mu,h^t)$. Then Lemma B4 ensures that the convex curve is almost flat over the space $\tilde{\mu} \in \triangle\Omega(\mu.h^t)$ and that the payoff $v_i^{\tilde{\mu}}(s_{-i}^{\mu}|_{h^t})$ approximates the maximal value for all beliefs $\tilde{\mu} \in \triangle\Omega(\mu.h^t)$. Formally, we obtain the following lemma. Let $C'' = \frac{C'}{\overline{\pi}}$ and $\tilde{C}'' = \frac{\tilde{C}'+1}{\overline{\pi}}$.

Lemma D8. *For each (μ,h^t) such that $\Omega(\mu,h^t)$ is robustly accessible and for each $\tilde{\mu} \in \triangle\Omega(\mu,h^t)$,*

$$\left|\overline{v}_i(s_{-i}^{\mu^*}|_{h^*}) + (1-\delta) - v_i^{\tilde{\mu}}(s_{-i}^{\mu}|_{h^t})\right| < \frac{(1-\delta^{4^{|\Omega|}})2\overline{g}}{\pi^*\overline{\pi}} + \frac{1-\delta^T}{\delta^T}C'' + (1-\delta)\tilde{C}''.$$

Proof. From (58), we know that

$$v_i^{\mu^{**}}(s_{-i}^{\mu^{**}}) - (1-\delta) \leq v_i^{\mu^{**}}(\tilde{s}_{-i}(\mu,h^t)).$$

Combining it with Lemma D7, we have

$$v_i^{\mu^{**}}(s_{-i}^{\mu^{**}}) - (1-\delta) \leq K_i(\mu,h^t) + \frac{(1-\delta^{4^{|\Omega|}})2\overline{g}}{\pi^*}.$$

Using (61), we obtain

$$\overline{v}_i(s_{-i}^{\mu^*}|_{h^*}) - \frac{(1-\delta^{4|\Omega|})2\overline{g}}{\pi^*} - \frac{1-\delta^T}{\delta^T}C' - (1-\delta)\tilde{C}' \leq K_i(\mu,h^t).$$

This and $K_i(\mu,h^t) \leq \overline{v}_i(s_{-i}^{\mu^*}|_{h^*}) + (1-\delta)$ imply that

$$\left|\overline{v}_i(s_{-i}^{\mu^*}|_{h^*}) + (1-\delta) - K_i(\mu,h^t)\right| \leq \frac{(1-\delta^{4|\Omega|})2\overline{g}}{\pi^*} + \frac{1-\delta^T}{\delta^T}C' + (1-\delta)(\tilde{C}'+1).$$

So the value $K_i(\mu,h^t)$ is close to the maximal value, that is, the convex curve $v_i^{\tilde{\mu}}(s_{-i}^{\mu}|_{h^t})$ approximates the maximal value for some relative interior belief $\tilde{\mu} \in \Delta(\mu,h^t)$. Then Lemma B4 ensures the result. *Q.E.D.*

D.6.3 Step 3: Convex Curves when $\Omega(\mu,h^t)$ is Transient

Pick a pair (μ,h^t) such that $\Omega(\mu,h^t)$ is transient. Suppose that the initial prior is $\tilde{\mu} \in \triangle\Omega(\mu,h^t)$ and that the opponents play $s_{-i}^{\mu}|_{h^t}$. Suppose that player i plays the following strategy $\tilde{s}_i$:

- Randomize all actions equally likely until the support of player i's posterior belief reaches a globally accessible set.
- Once it happens, then play a best reply in the rest of the game.

That is, player i switches her play to a best reply once the support of her belief reaches a globally accessible set.

Since the game is robustly connected, the switch must happen in finite time with probability one. Hence for δ close to one, player i's expected payoff is approximated by the expected continuation payoff after the switch. By the definition, the opponents' strategy at the time of the switch is $s_{-i}^{\hat{\mu}}|_{\hat{h}}$ for some $(\hat{\mu},\hat{h})$ such that $\Omega(\hat{\mu},\hat{h})$ is robustly accessible; then from Lemma D8, player i's continuation payoff after the switch approximates the maximal value, which in turn implies that her overall payoff approximates the maximal value. Formally, we have the following lemma. The proof is very similar to that of Yamamoto (2016) and hence omitted.

Lemma D9. *For each (μ,h^t) such that $\Omega(\mu,h^t)$ is transient and for each $\tilde{\mu} \in \triangle\Omega(\mu,h^t)$,*

$$\left|\overline{v}_i(s_{-i}^{\mu^*}|_{h^*}) + (1-\delta) - v_i^{\tilde{\mu}}(\delta,\tilde{s}_i,s_{-i}^{\mu}|_{h^t})\right| < \frac{(1-\delta^{4|\Omega|})4\overline{g}}{\pi^*\overline{\pi}} + \frac{1-\delta^T}{\delta^T}C'' + (1-\delta)\tilde{C}''.$$

Note that the strategy $\tilde{s}_i$ is not a best reply given the belief $\tilde{\mu} \in \triangle\Omega(\mu, h^t)$ and the opponents' strategy $s_{-i}^{\mu}|_{h^t}$. When player i chooses a best reply, her payoff (weakly) increases and hence becomes closer to the maximal value. Hence we have

$$\left|\overline{v}_i(s_{-i}^{\mu^*}|_{h^*}) + (1-\delta) - v_i^{\tilde{\mu}}(s_{-i}^{\mu}|_{h^t})\right| < \frac{(1-\delta^{4^{|\Omega|}})4\overline{g}}{\pi^*\overline{\pi}} + \frac{1-\delta^T}{\delta^T}C'' + (1-\delta)\tilde{C}''.$$

for each (μ, h^t) such that $\Omega(\mu, h^t)$ is transient and for each $\tilde{\mu} \in \triangle\Omega(\mu, h^t)$. This result and the result in the previous step show that all the convex curves $v_i^{\tilde{\mu}}(s_{-i}^{\mu}|_{h^t})$ are almost flat and approximate the maximal value, and hence all the minimax payoffs approximate the maximal value.

D.7 Proof of Proposition D6

In the proof of Lemma B17, the function z_j^{t+1} is chosen in such a way that player j is indifferent over all actions in period $t+1$ regardless of the current hidden state ω, given that the opponents play the prescribed action $s_{-j}(h^t)$. With strong full rank, we can modify this z_j^{t+1} so that the following condition holds:

$$\begin{aligned}
&\delta(1-p)\sum_{y\in Y}\pi^{\omega}(y|a_i, a_j, s_{-ij}(h^t))z_j^{t+1}(h^t, y) \\
&= \frac{1-\delta}{1-p\delta}\left(v_j^{\omega}(p\delta, s|_{h^t}) - v_j^{\omega}(p\delta, s_i|_{(h^t,a_i)}, s_j|_{(h^t,a_j)}, s_{-ij}|_{h^t})\right)
\end{aligned}$$

for all ω, a_i, and a_j. With this modification, player j is now indifferent over all actions in period $t+1$ regardless of the current hidden state ω, even if player i deviates from the prescribed action $s_i(h^t)$.

Pick an arbitrary target payoff v from an interior point of V^*, Then from Proposition 6, for sufficiently large δ, there is a public ex-post equilibrium s which achieves v regardless of the initial prior. Pick an arbitrary initial prior μ. In what follows, we will modify this strategy s and construct a sequential equilibrium for this initial prior μ.

With an abuse of notation, let H_i^t denote the set of all player i's private histories with length t which can be reached given μ and some $\tilde{s} \in S$. Let $H_i^t = \cup_{t=0}^{\infty} H_i^t$. In general, H_i^t may not coincide with $(A_i \times Y \times [0,1])^t$, since some sequence

$(a_i^\tau, y^\tau, z^\tau)_{\tau=1}^t$ may not be reachable given the initial prior μ. But these are redundant histories which should not show up in the game tree, so without loss of generality, we ignore these histories.

Recall that our public ex-post s has the random block structure. For each $\varepsilon > 0$, consider a perturbed strategy profile s^ε such that in each period t, each player i uses the original equilibrium strategy $s_i(h^{t-1})$ with probability $1-\varepsilon^{N^{t-1}}$, and mixes all actions equally likely with the remaining probability. So each action is chosen with at least probability $\frac{\varepsilon^{N^{t-1}}}{|A_i|}$ in period t. Intuitively, here we choose the perturbation probability in such a way that the probability that someone makes a single "mistake" in the current period is significantly smaller than the probability that all the opponents make mistakes in all the past periods.[14] This in turn implies that the probability that someone makes a single "mistake" during the current random block is significantly smaller than the probability that all the opponents make mistakes in all periods before the current block.

For each player i's history h_i^t, the perturbed strategy profile s^ε uniquely determines her belief $\zeta_i^\varepsilon(h_i^t)$ about the current state ω^{t+1} and the opponents' history h_{-i}^t. Let $\zeta_i(h_i^t)$ be the limit of $\zeta_i^\varepsilon(h_i^t)$ as $\varepsilon \to 0$.

Given s, let $\tilde{H}_i^t$ be the set of all histories $h_i^t = (a_i^\tau, y^\tau, z^\tau)_{\tau=1}^t$ that can happen when no one deviates from s during the current random block (but we allow any deviations in previous blocks). In other words, $h_i^t \notin \tilde{H}_i^t$ if it happens only when there have been deviations during the current random block. Then the belief system ζ satisfies the following properties:

- For each $h_i^t \in \tilde{H}_i^t$, the corresponding belief $\zeta_i(h_i^t)$ assigns probability one on the event that nobody has deviated from s during the current random block.

- For each $h_i^t \notin \tilde{H}_i^t$, the corresponding belief $\zeta_i(h_i^t)$ assigns probability one on the event that someone has deviated from s during the current random block.

As one can see from the proof of Lemma B19, each random block is associated with some target payoff, which in turn determines the corresponding direction λ. For each public history h^t, let $\lambda(h^t)$ denote the direction for the current random block.

[14] To see this, suppose that we are in period t now. The probability that all players except i make mistakes in all periods in the past is $\varepsilon^{(N-1)\sum_{\tilde{t}=1}^{t-1} N^{\tilde{t}-1}} = \varepsilon^{N^{t-1}-1}$, which is significantly smaller than the probability that someone makes a mistake in the current period, $\varepsilon^{N^{t-1}}$.

Now, we modify the strategy profile s and construct a new strategy profile s^* and a belief system ζ^* inductively. In each step t, we specify actions $s^*(h_i^{t-1})$ in period t and beliefs $\zeta^*(h_i^{t-1})$ at the beginning of the next period.

- Step 1: We do not modify the play in period one, and let $s_i^*(h_i^0) = s(h_i^0)$, and let $\zeta_i^*(h_i^1) = \zeta_i(h_i^1)$ for each h^1.

- Step t: Note that the actions up to period $t-1$ are already given, and the beliefs $\zeta_i^*(h_i^{t-1})$ at the beginning of period t are also given for each h^{t-1}. Now we choose the actions $s_i^*(h_i^{t-1})$ in period t as follows:

  – If $\lambda(h_i^{t-1})$ is regular, then let $s_i^*(h_i^{t-1}) = s_i(h_i^{t-1})$.
  – If $\lambda(h_i^{t-1})$ is a coordinate direction with $|\lambda_j| = 1$ for some $j \neq i$, then let $s_i^*(h_i^{t-1}) = s_i(h_i^{t-1})$.
  – If $\lambda(h_i^{t-1})$ is a coordinate direction with $|\lambda_i| = 1$ and $h_i^{t-1} \in \tilde{H}_i^{t-1}$, then let $s_i^*(h_i^{t-1}) = s_i(h_i^{t-1})$.
  – If $\lambda(h_i^{t-1})$ is a coordinate direction with $|\lambda_i| = 1$ and $h_i^{t-1} \notin \tilde{H}_i^{t-1}$, then let $s_i^*(h_i^{t-1}) = \tilde{s}_i(h^0)$ for some pure strategy $\tilde{s}_i \in \arg\max v_i^{\mu_i}(\delta, \tilde{s}_i, s_{-i}|_{h^{t-1}})$ where $\mu_i = \mathrm{marg}_{\triangle\Omega}\zeta_i^*(h_i^{t-1})$.

  Then we choose the beliefs $\zeta_i^*(h_i^t)$ at the beginning of period $t+1$ as follows: Let $s^{*,t}$ be the strategy profile for the t-period game we have defined so far, and let $s^{*,t,\varepsilon}$ be the perturbation of this profile $s^{*,t}$ where the perturbation probability is chosen as above. Then for each history h_i^t with length t, let $\zeta_i^{*,\varepsilon}(h_i^t)$ be the posterior belief induced by the perturbed strategy $s^{*,t,\varepsilon}$, and then let $\zeta_i^*(h_i^t) = \lim_{\varepsilon\to 0}\zeta_i^{*,\varepsilon}(h_i^t)$.

Note that $\zeta_i^*(h_i^t)$ can be different from the original belief $\zeta_i(h_i^t)$ only if $\lambda(h_i^{t-1})$ is a coordinate direction with $|\lambda_i| = 1$ and $h_i^{t-1} \notin \tilde{H}_i^{t-1}$.

We claim that the pair (s^*, ζ^*) constitutes a sequential equilibrium for the initial prior μ. Since ζ^* is consistent with s^*, it is sufficient to show that s^* is sequentially rational.

Note first that s_i^* differs from the original equilibrium s only for the actions $s_i^*(h^{t-1})$ after histories $h_i^{t-1} \notin \tilde{H}_i^{t-1}$. This implies that, regardless of the past history, once the current block terminates and players go to the next block, the continuation strategy induced by s^* and the one by s yield the same outcome distribution and thus the same continuation payoffs.

Now we check player i's incentive. Suppose that the direction λ corresponding to the current block is regular. By the construction, the strategy profile s^* and s induce the same actions during the current block. Then from the proof of Lemma B16, player i is indifferent over all actions regardless of the hidden state ω, after every public history within the block. (Here, the continuation payoff $w(h^t)$ is indeed achieved, as s^* and s yield the same continuation payoff from the next block.) So player i's play in this block is sequentially rational.

Suppose next that λ corresponding to the current block is a coordinate direction with $|\lambda_j| = 1$ for some $j \neq i$. In this case, player j's play induced by s^* can be different from that of s. However, as explained at the beginning of this proof, the continuation payoffs w_i are chosen in such a way that after every public history within the block, player i is indifferent over all actions regardless of the hidden state ω and regardless of player j's action in the current period. Hence, again, player i's play in this block is sequentially rational.

Finally, suppose that λ corresponding to the current block is a coordinate direction with $|\lambda_i| = 1$. By the construction, the strategy s^*_{-i} and s_{-i} induce the same actions during the current block, so player i's incentive problem is exactly the same as the one in Lemmas B17 and B18. Suppose that player i's past history is $h_i^{t-1} \in \tilde{H}_i^{t-1}$. In this case, the corresponding belief $\zeta_i^*(h_i^{t-1})$ assigns probability one on the event that no one has deviated during the current block game. Then player i's incentive problem is exactly the same as the one studied in the proofs of Lemma Lemmas B17 and B18, and hence player i is willing to continue to play $s_i^*(h^{t-1}) = s_i(h^{t-1})$ today, regardless of her belief at the beginning of the current block game. Now suppose that $h_i^{t-1} \notin \tilde{H}_i^{t-1}$. Since $s_{-i}|_{h^{t-1}} = s^*_{-i}|_{h^{t-1}}$, by the construction, choosing $s_i^*(h_i^{t-1})$ is optimal.

D.8 Proof of Proposition D7

The proof is the same as that of Proposition D6, except the modification of the payment z_j^{t+1}. In the proof of Proposition D6, z_j^{t+1} is modified in such a way that player j is indifferent over all actions in period $t+1$ regardless of player i's current action, and this modification relies on the fact that each pure action profile has strong full rank. Here we show that a similar modification is possible without strong full rank, as long as players can communicate.

Suppose that for each $j \neq i$, the payment z_j^{t+1} depends on the public signals $h^{t+1} = (y^1, \cdots, y^{t+1})$ and on player i's message $m_i \in A_i$ in period $t+1$. We consider z_j^{t+1} which satisfies the following condition:

$$\begin{aligned}
&\delta(1-p)\sum_{y\in Y}\pi^{\omega}(y|a_i, a_j, s_{-ij}(h^t))z_j^{t+1}(h^t, y, m_i = a_i) \\
&= \frac{1-\delta}{1-p\delta}\left(v_j^{\omega}(p\delta, s|_{h^t}) - v_j^{\omega}(p\delta, s_i|_{(h^t,a_i)}, s_j|_{(h^t,a_j)}, s_{-ij}|_{h^t})\right)
\end{aligned}$$

for all ω, a_i, and a_j, This system of equations indeed has a solution, since each pure action profile has cross-state individual full rank. Note that strong full rank is not needed here, since z_j^{t+1} now depends on m_i.

Intuitively, the above condition says that player j is indifferent over all actions in period $t+1$ regardless of the current hidden state ω, and of player i's current action, as long as player i reports her action truthfully. And player i indeed reports her action truthfully after every history, since her report does not influence her own payoff. Hence player $j \neq i$ is indifferent over all actions after every history, even if we modify player i's action at off-path histories.

References

Abreu, D., D. Pearce, and E. Stacchetti (1990): “Toward a Theory of Discounted Repeated Games with Imperfect Monitoring,” *Econometrica* 58, 1041-1063.

Athey, S., and K. Bagwell (2001): “Optimal Collusion with Private Information,” *RAND Journal of Economics* 32, 428-465.

Athey, S., and K. Bagwell (2008): “Collusion with Persistent Cost Shocks,” *Econometrica* 76, 493-540.

Connault, B. (2015): “Hidden Rust Models,” *mimeo*.

Dutta, P. (1995): “A Folk Theorem for Stochastic Games,” *Journal of Economic Theory* 66, 1-32.

Escobar, J., and J. Toikka (2013) “Efficiency in Games with Markovian Private Information,” *Econometrica* 81, 1887-1934.

Fudenberg, D., and D.K. Levine (1994): “Efficiency and Observability in Games with Long-Run and Short-Run Players,” *Journal of Economic Theory* 62, 103-135.

Fudenberg, D., D.K. Levine, and E. Maskin (1994): “The Folk Theorem with Imperfect Public Information,” *Econometrica* 62, 997-1040.

Fudenberg, D., and Y. Yamamoto (2010): “Repeated Games where the Payoffs and Monitoring Structure are Unknown,” *Econometrica* 78, 1673-1710.

Fudenberg, D., and Y. Yamamoto (2011a): “Learning from Private Information in Noisy Repeated Games,” *Journal of Economic Theory* 146, 1733-1769.

Fudenberg, D., and Y. Yamamoto (2011b): “The Folk Theorem for Irreducible Stochastic Games with Imperfect Public Monitoring,” *Journal of Economic Theory* 146, 1664-1683.

Green, E.J., and R.H. Porter (1984): “Noncooperative Collusion under Imperfect Price Information,” *Econometrica* 52, 87-100.

Hörner, J., and S. Lovo (2009): “Belief-Free Equilibria in Games with Incomplete Information,” *Econometrica* 77, 453-487.

Hörner, J., S. Lovo, and T. Tomala (2011): “Belief-Free Equilibria in Games with Incomplete Information: Characterization and Existence,” *Journal of Economic Theory* 146, 1770-1795.

Hörner, J., T. Sugaya, S. Takahashi, and N. Vieille (2011): "Recursive Methods in Discounted Stochastic Games: an Algorithm for $\delta \to 1$ and a Folk Theorem," *Econometrica* 79, 1277-1318.

Hörner, J., S. Takahashi, and N. Vieille (2015): "Truthful Equilibria in Dynamic Bayesian Games," *Econometrica* 83, 1795-1848.

Kandori, M., and H. Matsushima (1998): "Private Observation, Communication and Collusion," *Econometrica* 66, 627-652.

Koren, G. (1992): "Two-Person Repeated Games where Players Know Their Own Payoffs," *mimeo*.

Levin, J. (2003): "Relational Incentive Contracts," *American Economic Review* 93, 835-857.

Radner, R., R. Myerson, and E. Maskin (1986): "An Example of a Repeated Partnership Game with Discounting and with Uniformly Inefficient Equilibria," *Review of Economic Studies* 53, 59-70.

Sekiguchi, T. (1997): "Efficiency in Repeated Prisoner's Dilemma with Private Monitoring," *Journal of Economic Theory* 76, 345-361.

Seneta, E. (1981): *Non-negative Matrices and Markov Chains*, Springer.

Shalev, J. (1994): "Nonzero-Sum Two-Person Repeated Games with Incomplete Information and Known-Own Payoffs," *Games and Economic Behavior* 7, 246-259.

Shapley, L. (1953): "Stochastic Games," *Proceedings of the National Academy of Sciences of the United States of America* 39, 1095-1100.

Wiseman, T. (2005): "A Partial Folk Theorem for Games with Unknown Payoff Distributions," *Econometrica* 73, 629-645.

Wiseman, T. (2012) "A Partial Folk Theorem for Games with Private Learning," *Theoretical Economics* 7, 217-239.

Yamamoto, Y. (2014): "Individual Learning and Cooperation in Noisy Repeated Games," *Review of Economic Studies* 81, 473-500.

Yamamoto, Y. (2016): "Stochastic Games with Hidden States," *mimeo*.