# PIER Working Paper 15-019

"Stochastic Games with Hidden States"
Second Version

by

Yuichi Yamamoto

http://ssrn.com/abstract=2614965

# Stochastic Games with Hidden States[*]

Yuichi Yamamoto[†]

First Draft: March 29, 2014
This Version: June 1, 2015

**Abstract**

This paper studies infinite-horizon stochastic games in which players observe payoffs and noisy public information about a hidden state each period. Public randomization is available. We find that, very generally, the feasible and individually rational payoff set is invariant to the initial prior about the state in the limit as the discount factor goes to one. We also provide a recursive characterization of the equilibrium payoff set and establish the folk theorem.

*Journal of Economic Literature* Classification Numbers: C72, C73.

Keywords: stochastic game, hidden state, connectedness, stochastic self-generation, folk theorem.

---

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

1

# 1  Introduction

When agents have a long-run relationship, underlying economic conditions may change over time. A leading example is a repeated Bertrand competition with stochastic demand shocks. Rotemberg and Saloner (1986) explore optimal collusive pricing when random demand shocks are i.i.d. each period. Haltiwanger and Harrington (1991), Kandori (1991), and Bagwell and Staiger (1997) further extend the analysis to the case in which demand fluctuations are cyclic or persistent. A common assumption of these papers is that demand shocks are publicly observable *before* firms make their decisions in each period. This means that in their model, firms can perfectly adjust their price contingent on the true demand today. However, in the real world, firms often face uncertainty about the market demand when they make decisions. Firms may be able to learn the current demand shock through their sales *after* they make decisions; but then in the next period, a new demand shock arrives, and hence they still face uncertainty about the true demand. In such a situation, firms need to estimate the true demand in order to figure out the optimal pricing each period, and they may want to "experiment" in order to obtain better information about the future economic condition. This paper develops a general framework which incorporates such uncertainty, and investigates how uncertainty influences long-run incentives.

Specifically, we consider a new class of stochastic games in which the state of the world is hidden information. At the beginning of each period $t$, a hidden state $\omega^t$ (booms or slumps in the Bertrand model) is given, and players have some posterior belief $\mu^t$ about the state. Players simultaneously choose actions, and then a public signal $y$ and the next hidden state $\omega^{t+1}$ are randomly drawn. After observing the signal $y$, players updates their posterior belief using Bayes' rule, and then go to the next period. The signal $y$ can be informative about both the current and next states, which ensures that our formulation accommodates a wide range of economic applications.

Throughout the paper, we assume that actions are perfectly observable. In this case, there is no private information and thus players have the same posterior belief $\mu^t$ about the current state $\omega^t$ after every history. Then this posterior belief $\mu^t$ can be regarded as a common state variable, and thus our model reduces to a stochastic game with *observable* states $\mu^t$. This is a great simplification, but

still the model is not as tractable as one may expect; a problem is that there are infinitely many possible posterior beliefs, so we need to consider a stochastic game with *infinite* states. This is in a sharp contrast with past work which assumes a *finite* state space (Dutta (1995), Fudenberg and Yamamoto (2011b), and Hörner, Sugaya, Takahashi, and Vieille (2011)).[1]

A main problem of having infinite states is that a Markov chain over infinite states can be "badly" behaved in the long run. When we consider a finite-state Markov chain, some states must be *positive recurrent* in the sense that the state will return to the current state in finite time with probability one. Intuitively, positive recurrence ensures that the Markov chain is "well-behaved" in the long run; in particular, under a mild condition, the Markov chain is *ergodic* so that the state in a distant future is not influenced by the initial state. Using this property, Dutta (1995) shows that the feasible payoff set for patient players, who care only a distant future, is invariant to the initial state. All the existing techniques on stochastic games rely on this invariance result. In contrast, when a Markov chain has infinite states, states may not be positive recurrent, and accordingly, it is well-known that an infinite-state Markov chain is not ergodic in many cases.[2] Hence, *a priori*, there is no reason to expect the belief evolution to be ergodic.

Nonetheless, we find that the invariance result extends to our setup under a mild condition. Specifically, we show that if the game is *connected*, then the feasible payoff set is invariant to the initial belief in the limit as the discount factor goes to one. We also show that the limit minimax payoff is invariant to the initial belief under a stronger assumption, *strong connectedness*. Our proof is substantially different from that of the literature, since the techniques which refer to ergodic theorems are not applicable due to the infinite state space.

Our assumption, connectedness, is a condition about how the *support* of the belief evolves over time; it requires that players can jointly drive the support of the belief from any set $\Omega^*$ to any other set $\Omega^{**}$, unless the set $\Omega^{**}$ is "not essential" in

---

[1]Hörner, Takahashi, and Vieille (2011) consider stochastic games with infinite states, but they assume that the limit equilibrium payoff set is identical for all initial states, that is, they assume a sort of ergodicity. There is also an extensive literature on the existence of Markov strategy equilibria for the infinite-state case. See recent work by Duggan (2012) and Levy (2013), and an excellent survey by Dutta and Sundaram (1998). In contrast to the literature, this paper considers a general class of equilibria which are not necessarily Markovian.

[2]There are some well-known sufficient conditions for ergodicity of infinite-state Markov chains, but these conditions are not satisfied in our model. See Appendix G for discussions.

the sense that the probability of the support being $\Omega^{**}$ in a distant future is almost negligible. (Here, $\Omega^*$ and $\Omega^{**}$ denote subsets of the whole state space $\Omega$.) Intuitively, this property ensures that the evolution of the support of the belief is well-behaved in the long run. As discussed in Section 4.3, connectedness is regarded as a natural generalization of irreducibility commonly assumed in the literature to the hidden-state case. We also show that connectedness can be replaced with an even weaker condition, called *asymptotic connectedness*. Asymptotic connectedness is satisfied for generic games, as long as the underlying state evolution is irreducible.

As noted, connectedness is a condition on the evolution of the support of the belief, and thus it is much weaker than assuming the belief evolution itself to be well-behaved. Nonetheless, connectedness is enough to establish the result we want. To illustrate the idea, think about a one-player game. Since public randomization is available, the feasible payoff set is an interval and hence determined by the maximal and minimal payoffs. Let $f(\mu)$ be the maximal payoff with the initial prior $\mu$ in the limit as the discount factor goes to one, and let $\mu^*$ be a belief which maximizes $f(\mu)$. Our key finding is roughly as follows:

(∗) If there is a belief $\mu$ such that $f(\mu)$ is equal to $f(\mu^*)$, then $f(\tilde{\mu})$ is also equal to $f(\mu^*)$ for *every* belief $\tilde{\mu}$ with the same support as $\mu$.

That is, once we can find a belief $\mu$ with support $\Omega^*$ which satisfies the above property, it gives a uniform bound on $f(\tilde{\mu})$ for *all* beliefs $\tilde{\mu}$ with support $\Omega^*$. The result (∗) greatly simplifies our problem, because it implies that in order to prove that the maximal feasible payoff $f(\mu)$ is invariant to *all* beliefs $\mu$, it suffices to find *one* belief $\mu$ with the above property for each subset $\Omega^*$. And we can indeed find such $\mu$ for each support $\Omega^*$, using the fact that connectedness ensures that players can jointly drive the support from any set to other sets, and the fact that $f$ is a solution to a dynamic programming equation.

The second main result of the paper is the folk theorem, that is, we show that any feasible and individually rational payoffs are achieved by sequential equilibria as long as players are patient enough and the game is strongly connected. As an intermediate result, we provide a recursive characterization of the equilibrium payoff set, which generalizes self-generation of Abreu, Pearce, and Stacchetti (1990). Taking into account the fact that the state evolution is not necessarily ergodic, we

decompose payoffs in a way different than Abreu, Pearce, and Stacchetti (1990), and by doing so, an equilibrium payoff can be regarded as the sum of a payoff approximating the Pareto-efficient frontier and of an expected continuation payoff. This structure is reminiscent of that of the standard repeated games, in which an equilibrium payoff is the sum of a stage-game payoff in period one, which is often on the Pareto-efficient frontier, and of a continuation payoff. Hence we can generalize the proof idea of Fudenberg, Levine, and Maskin (1994, hereafter FLM) to our setup, and can establish the folk theorem.

Stochastic games are proposed by Shapley (1953). Dutta (1995) characterizes the feasible and individually rational payoffs for patient players, and proves the folk theorem for the case of observable actions. Fudenberg and Yamamoto (2011b) and Hörner, Sugaya, Takahashi, and Vieille (2011) extend the folk theorem to games with public monitoring. All these papers assume that the state of the world is publicly observable at the beginning of each period.[3]

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 current state each period. An important assumption in their model is that the state of the world is a collection of players' private information. They look at equilibria in which players report their private information truthfully, so the state is perfectly revealed before they choose actions.[4] In contrast, in this paper, players have only limited information about the true state and 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 all assume that the state of the world is fixed at the beginning of the game and does not change over time. Since the state influences the distribution of a public signal each period, players can (almost) perfectly learn the true state by aggregating all the past public signals. In contrast, in our model, the state changes as time goes and thus players never learn the true state perfectly.

---

[3]Independently of this paper, Renault and Ziliotto (2014) also study stochastic games with hidden states, but they focus only on an example in which multiple states are absorbing.

[4]An exception is Sections 4 and 5 of Hörner, Takahashi, and Vieille (2015); they consider equilibria in which some players do not reveal information and the public belief is used as a state variable. But their analysis relies on the independent private value assumption.

# 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 $\omega^1$ from a finite set $\Omega$. 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 $\omega^t$ is not observable to players, and let $\mu \in \triangle \Omega$ be the common prior about $\omega^1$.

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 perfectly observable, and in addition players observe a public signal $y$ from a finite set $Y$. Then players go to the next period $t + 1$, with a (hidden) state $\omega^{t+1}$. The distribution of $y$ and $\omega^{t+1}$ depends on the current state $\omega^t$ 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 $\omega$ *and* the next state $\tilde{\omega}$, because the distribution of $y$ may depend on $\omega$ and $y$ may be correlated with $\tilde{\omega}$. Let $\pi_Y^\omega(y|a)$ denote the marginal probability of $y$. Assume that public randomization $z$, which follows the uniform distribution on $[0, 1]$, is available at the end of each period.

Player $i$'s payoff in period $t$ is a function of the current action profile $a$ and the current public signal $y$, and is denoted by $u_i(a, y)$. Then her expected stage-game payoff conditional on the current state $\omega$ and the current action profile $a$ is $g_i^\omega(a) = \sum_{y \in Y} \pi_Y^\omega(y|a) u_i(a, y)$. Here the hidden state $\omega$ 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 $\overline{g}_i = \max_{\omega, a} |2 g_i^\omega(a)|$, 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 $Y = \Omega \times \Omega$ and suppose that $\pi^\omega(y, \tilde{\omega}|a) = 0$ for $y = (y_1, y_2)$ 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. Suppose also that $u_i(a, y)$ is does not depend on the second component $y_2$, so that stage-game payoffs are influenced by the current state only. Since the signal in the previous period perfectly reveals

6

the current state, in this model players know the state $\omega^t$ *before* they choose an action profile $a^t$. This is exactly the standard stochastic games studied in the literature.

- *Delayed observation.* Let $Y = \Omega$ and assume that $\pi_Y^\omega(y|a) = 1$ for $y = \omega$. That is, assume that the current signal $y^t$ reveals the current state $\omega^t$. This is the case in which players observe the state *after* they choose their actions $a^t$. In what follows, this class of games is referred to as stochastic games with delayed observations.

- *Observable and unobservable states.* Assume that $\omega$ consists of two components, $\omega_O$ and $\omega_U$, and that the signal $y^t$ perfectly reveals the first component of the next state, $\omega_O^{t+1}$. Then we can interpret $\omega_O$ as an observable state and $\omega_U$ as an unobservable state. One of the examples which fits this formulation is a duopoly market in which firms face uncertainty about the demand, and their cost function depends on their knowledge, know-how, or experience. The firms' experience can be described as an observable state variable as in Besanko, Doraszelski, Kryukov, and Satterthwaite (2010), and the uncertainty about the market demand as an unobservable state.

In the infinite-horizon stochastic game, players have a common discount factor $\delta \in (0,1)$. Let $(\omega^\tau, a^\tau, y^\tau, z^\tau)$ be the state, the action profile, the public signal, and the public randomization in period $\tau$. Then the history up to period $t \geq 1$ is denoted by $h^t = (a^\tau, y^\tau, z^\tau)_{\tau=1}^t$. Let $H^t$ denote the set of all $h^t$ for $t \geq 1$, and let $H^0 = \{\emptyset\}$. Let $H = \bigcup_{t=0}^\infty H^t$ be the set of all possible histories. A strategy for player $i$ is $s_i = (s_i^t)_{t=1}^\infty$ such that $s_i^t : H^{t-1} \to \triangle A_i$ is a measurable mapping for each $t$. To simplify the notation, given any strategy $s_i$ and history $h^t$, let $s_i(h^t)$ denote the action after period $h^t$, i.e., $s_i(h^t) = s_i^{t+1}(h^t)$. Let $S_i$ be the set of all strategies for player $i$, and let $S = \times_{i \in I} S_i$. Also let $S_i^*$ be the set of all player $i$'s strategies which do not use public randomization, and let $S^* = \times_{i \in I} S_i^*$. Given a strategy $s_i$ and history $h^t$, let $s_i|_{h^t}$ be the continuation strategy induced by $s_i$ after history $h^t$.

Let $v_i^\omega(\delta, s)$ denote player $i$'s average payoff in the stochastic game when the initial prior puts probability one on $\omega$, the discount factor is $\delta$, and players play strategy profile $s$. That is, let $v_i^\omega(\delta, s) = E[(1-\delta) \sum_{t=1}^\infty \delta^{t-1} g_i^{\omega^t}(a^t) | \omega, s]$. Similarly, let $v_i^\mu(\delta, s)$ denote player $i$'s average payoff when the initial prior is $\mu$. Note

that for each initial prior $\mu$, discount factor $\delta$, and $s_{-i} \in S^*_{-i}$, player $i$'s best re-ply $s_i \in S^*_i$ exists; see Appendix F for the proof. Let $v^\omega(\delta, s) = (v^\omega_i(\delta, s))_{i \in I}$ and $v^\mu(\delta, s) = (v^\mu_i(\delta, s))_{i \in I}$.

## 2.2 Alternative Interpretation: Belief as a State Variable

In each period $t$, each player forms a belief $\mu^t$ about the current hidden state $\omega^t$. Since players have the same initial prior $\mu$ and the same information $h^{t-1}$, the posterior belief $\mu^t$ is also the same across all players. Then we can regard this belief $\mu^t$ as a common state variable; that is, our model can be interpreted as a stochastic game with *observable states* $\mu^t$.

With this interpretation, the model can be re-written as follows. In period one, the belief is simply the initial prior; $\mu^1 = \mu$. In period $t \geq 2$, players use Bayes' rule to update the belief; given $\mu^{t-1}$, $a^{t-1}$, and $y^{t-1}$, let

$$\mu^t(\tilde{\omega}) = \frac{\sum_{\omega \in \Omega} \mu^{t-1}(\omega) \pi^\omega(y^{t-1}, \tilde{\omega}|a^{t-1})}{\sum_{\omega \in \Omega} \mu^{t-1}(\omega) \pi^\omega_Y(y^{t-1}|a^{t-1})}$$

for each $\tilde{\omega}$. Given this (common) belief $\mu^t$, players chooses actions $a^t$, and then observe a signal $y$ according to the distribution $\pi^{\mu^t}_Y(y|a) = \sum_{\omega \in \Omega} \mu^t(\omega) \pi^\omega_Y(y|a)$. Public randomization $z \sim U[0, 1]$ is also observed. Player $i$'s expected stage-game payoff given $\mu^t$ and $a^t$ is $g^{\mu^t}_i(a^t) = \sum_{\omega \in \Omega} \mu^t(\omega) g^\omega_i(a^t)$.

Now we give the definition of sequential equilibria. Let $\zeta : H \to \triangle \Omega$ be a belief system; i.e., $\zeta(h^t)$ is the posterior about $\omega^{t+1}$ after history $h^t$. A belief system $\zeta$ is *consistent with the initial prior* $\mu$ if there is a completely mixed strategy profile $s$ such that $\zeta(h^t)$ is derived by Bayes' rule in all on-path histories of $s$. Since actions are observable, given the initial prior $\mu$, a consistent belief is unique at each information set which is reachable by some strategy. (So essentially there is a unique belief system $\zeta$ consistent with $\mu$.) A strategy profile $s$ is a *sequential equilibrium* in the stochastic game with the initial prior $\mu$ if $s$ is sequentially rational given the belief system $\zeta$ consistent with $\mu$.

## 3 Example: Stochastic Bertrand Competition

Consider two firms which produce a homogeneous (undifferentiated) product. In each period, each firm $i$ chooses one of the three prices: A high price ($a^H_i = 2$),

a low price ($a_i^L = 1$), or a Nash equilibrium price ($a_i^* = 0$). Here $a_i^* = 0$ is called "Nash equilibrium price," since we assume that the production cost is zero; this ensures that there is a unique Nash equilibrium in the static game and each firm charges $a_i^* = 0$ in the equilibrium. To simplify the notation, let $a^H = (a_1^H, a_2^H)$, $a^L = (a_1^L, a_2^L)$, and $a^* = (a_1^*, a_2^*)$.

There is a persistent demand shock and an i.i.d. demand shock. The persistent demand shock is captured by a hidden state $\omega$, which follows a Markov process. Specifically, in each period, the state is either a boom ($\omega = \omega^H$) or a slump ($\omega = \omega^L$), and after each period, the state stays at the current state with probability 0.9. We assume that the current action (price) does not influence the state evolution. Let $\mu \in (0,1)$ be the probability of $\omega^H$ in period one.

Due to the i.i.d. demand shock, the aggregate demand of the product is stochastic, and its distribution depends on the current economic condition $\omega$ and on the effective price $\min\{a_1, a_2\}$. For simplicity, assume that the aggregate demand $y$ takes one of the two values, $y^H = 10$ and $y^L = 1$. Assume that its distribution is

$$
(\pi_Y^\omega(y^H|a), \pi_Y^\omega(y^L|a)) = \begin{cases} (0.9, 0.1) & \text{if } \omega = \omega^H \text{ and } \min\{a_1, a_2\} = 1 \\ (0.8, 0.2) & \text{if } \omega = \omega^L \text{ and } \min\{a_1, a_2\} = 1 \\ (0.8, 0.2) & \text{if } \omega = \omega^H \text{ and } \min\{a_1, a_2\} = 2 \\ (0.1, 0.9) & \text{if } \omega = \omega^L \text{ and } \min\{a_1, a_2\} = 2 \\ (1, 0) & \text{if } \min\{a_1, a_2\} = 0 \end{cases}.
$$

Intuitively, the high price $a^H$ is a "risky" option in the sense that the expected demand is high (the probability of $y^H$ is 0.8) if the current economy is in a boom but is extremely low (the probability of $y^H$ is only 0.1) if the current economy is in a slump. On the other hand, the low price $a^L$ is a "safe" option in the sense that the expected demand is not very sensitive to the underlying economic condition. If the effective price is zero, the probability of $y^H$ is one regardless of the current state $\omega$. We assume that the realized demand $y$ is public information. Assume also that $y$ and the next state $\tilde{\omega}$ are independently drawn.

This is the Bertrand model, and a firm with a lower price takes the whole market share. Accordingly, firm $i$'s current profit is $u_i(a, y) = a_i y$ if $a_i < a_{-i}$. and $u_i(a, y) = 0$ if $a_i > a_{-i}$. If $a_i = a_{-i}$, the firms share the market equally and $u_i(a, y) = \frac{a_i y}{2}$. Given $\omega$ and $a$, let $g_i^\omega(a) = \sum_{y \in Y} \pi_Y^\omega(y|a) u_i(a, y)$ be the expected profit of firm $i$, and let $g^\omega(a) = g_1^\omega(a) + g_2^\omega(a)$ be the total profit. An

9

easy calculation shows that $g^{\omega^H}(a^H) = 16.4$, $g^{\omega^H}(a^L) = 9.1$, $g^{\omega^L}(a^H) = 3.8$, and $g^{\omega^L}(a^L) = 8.2$. So the high price $a^H$ yields higher total profits than the low price $a^L$ if it is in a boom, while the low price $a^L$ is better if it is in a slump. Also, letting $g^\mu(a) = \mu g^{\omega^H}(a) + (1-\mu)g^{\omega^L}(a)$ be the total profit given $\mu$ and $a$, it is easy to see that $g^\mu(a)$ is maximized by the high price $a^H$ if $\mu \geq \frac{44}{117} \approx 0.376$, and by the low price $a^L$ if $\mu \leq \frac{44}{117}$. Let $\mu^* = \frac{44}{117}$ represent this threshold.

Now, consider the infinite-horizon model where the discount factor is $\delta \in (0,1)$. What is the optimal collusive pricing in this model, i.e., what strategy profile $s$ maximizes the expectation of the discounted sum of the total profit, $\sum_{t=1}^{\infty} \delta^{t-1} g^{\omega^t}(a^t)$? To answer this question, let $f(\mu)$ be the maximized value given the initial prior $\mu$, that is, $f(\mu) = \max_{s \in S} E[\sum_{\delta=1}^{\infty} \delta^{t-1} g^{\omega^t}(a^t)|\mu, s]$. From the principle of optimality, the function $f$ must solve

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

where $\tilde{\mu}(\mu, a, y)$ is the belief in period two given that the initial prior is $\mu$ and players play $a$ and observe $y$ in period one. Intuitively, (1) says that the total profit $f(\mu)$ consists of today's profit $g^\mu(a)$ and the expectation of the future profits $f(\tilde{\mu}(\mu, a, y))$, and that the current action should maximize it.

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 1 shows the value function $f$ for $\delta = 0.7$. As one can see, the value function $f$ is upward sloping, which means that the total profit becomes larger when the initial prior becomes more optimistic.

Figure 2 shows the optimal policy. (In the vertical axis, 0 means the low price $a^L$, while 1 means the high price $a^H$). It shows that the optimal policy is a simple cut-off rule; the optimal action is the low price $a^L$ when the current belief $\mu$ is less than $\mu^{**}$, and is the high price $a^H$ otherwise, with the threshold value $\mu^{**} \approx 0.305$. This threshold value $\mu^{**}$ is lower than that for the static game, $\mu^* \approx 0.376$. That is, when the current belief is $\mu \in (\mu^{**}, \mu^*)$, the firms choose the high price which does not maximize the current profit. Note that this is so *even though actions do not influence the state evolution*. Why is this the case?

A key is that choosing the high price provides better information about the
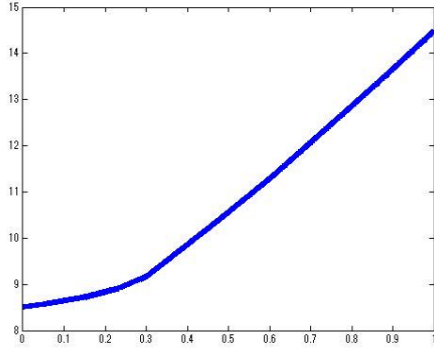
Figure 1: Value Function
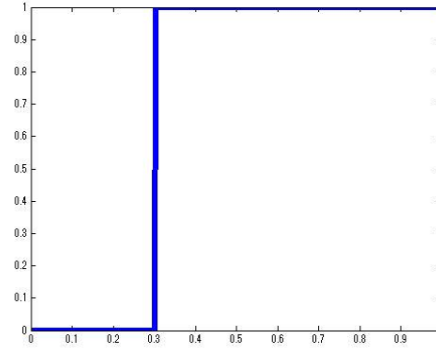*x*-axis: belief $\mu$. *y*-axis: payoffs.

Figure 2: Optimal Policy
*x*-axis: belief $\mu$. *y*-axis: actions.

hidden state $\omega$ than the low price, in Blackwell's sense.[5] To see this, for each $a$, let $\Pi(a)$ denote the two-by-two matrix with rows $(\pi_Y^\omega(y^H|a), \pi_Y^\omega(y^L|a))$ for each $\omega$. Then we have

$$\Pi(a^L) = \Pi(a^H) \begin{pmatrix} \frac{13}{14} & \frac{1}{14} \\ \frac{11}{14} & \frac{3}{14} \end{pmatrix},$$

that is, $\Pi(a^L)$ is the product of $\Pi(a^H)$ and a *stochastic matrix* in which each row is a probability distribution. This shows that $\Pi(a^L)$ is a *garbling* of $\Pi(a^H)$ (see Kandori (1992)), and in this sense, the public signal $y$ given the low price $a^L$ is less informative than that given the high price.

When the current belief is $\mu \in (\mu^{**}, \mu^*)$, the current profit is maximized by choosing the low price $a^L$. However, by choosing the high price $a^H$ today, the firms can obtain better information and can make a better estimation about the hidden state tomorrow. This yields higher expected profits in the continuation game, and when $\mu \in (\mu^{**}, \mu^*)$, this effect dominates the decrease in the current profit. Hence the high price is chosen in the optimal policy.

In this example, the efficient payoff $f(\mu)$ can be achieved by a trigger strategy. Consider the strategy profile in which the firms follow the optimal policy above, but switch to "forever $a^*$" once there is a deviation from the optimal policy. Let us check firm $i$'s incentive. In the punishment phase, firm $i$ has no reason to deviate from $a^*$, since "playing $a^*$ forever" is a Markov strategy equilibrium in this model.

---

[5]See Hao, Iwasaki, Yokoo, Joe, Kandori, and Obara (2012) for the case in which lower prices yield better information about the hidden state.

(Indeed, when the opponent chooses $a^*$ forever, even if firm $i$ deviates, its payoff is zero.) In the collusive phase, if the optimal policy specifies the low price today, firm $i$ has no reason to deviate because any deviation yields the payoff of zero. So consider the case in which the optimal policy specifies the high price today. If firm $i$ deviates, its current payoff is at most $g_i^{\omega^H}(a_i^L, a_{-i}^H) = 9.1$, and its continuation payoff is zero. So the overall payoff is at most $(1-\delta)9.1 + \delta \cdot 0 = 2.73$. On the other hand, if firm $i$ does not deviate, its payoff is at least $\min_{\mu \in [\mu^*, 1]} \frac{f(\mu)}{2} \geq 4$. Hence the above strategy profile is an equilibrium.

A couple of remarks are in order. First, the firms do "experiments" in this efficient equilibrium. As argued, when the current belief is $\mu \in (\mu^{**}, \mu^*)$, the firms choose the high price $a^H$ in order to obtain better information, although it does not maximize the current expected payoff.

Second, the equilibrium construction here is misleadingly simple, since it relies on the existence of a Markov strategy equilibrium in which $a^*$ is charged forever. In general, when the state space is infinite, the existence of Markov strategy equilibria is not guaranteed (see Duggan (2012) and Levy (2013)), and accordingly, it is not obvious how to punish a deviator in an equilibrium. Indeed, the proof of our folk theorem is non-constructive.

Third, the solution to (1) depends on the discount factor $\delta$. Figure 3 illustrates how the value function changes when the firms become more patient; it gives the value functions for $\delta = 0.9$, $\delta = 0.99$, and $\delta = 0.999$. The optimal policies are still cut-off rules, and the cut-off value is $\mu = 0.285$ for $\delta = 0.9$, $\mu = 0.276$ for $\delta = 0.99$, and $\mu = 0.275$ for $\delta = 0.999$. Note that the cut-off value becomes lower (so the high price is chosen more frequently) when the discount factor increases. The reason is that when the firms become patient, they care future profits more seriously, and thus information about the hidden state tomorrow is more valuable.

As one can see from the figure, when the firms become patient, the value function becomes almost flat, that is, the firms' initial prior has almost no impact on the total profit. This property is not specific to this example; we will show in Lemma 5 that if the game is *connected*, then the feasible payoff set does not depend on the initial prior in the limit as the discount factor goes to one. This result plays an important role when we prove the folk theorem.
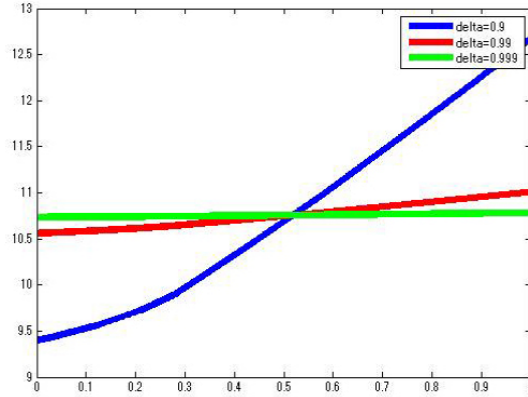
Figure 3: Value Functions for High $\delta$
$x$-axis: belief $\mu$. $y$-axis: payoffs.

# 4 Connected Stochastic Games

In general, stochastic games can be very different from infinitely repeated games. Indeed, the irreversibility created by absorbing states can support various sorts of backward induction arguments with no real analog in the infinitely repeated setting. To avoid such a problem, most of the existing papers assume irreducibility of the state evolution, which rules out absorbing states (Dutta (1995), Fudenberg and Yamamoto (2011b), and Hörner, Sugaya, Takahashi, and Vieille (2011)).

Since we consider a new environment in which the state $\omega$ is hidden, we need to identify an appropriate condition which parallels the notion of irreducibility in the standard model. We find that one of such conditions is *connectedness*, which imposes a restriction on how the *support* of the posterior belief evolves over time.

## 4.1 Full Support Assumption

Connectedness is satisfied in a wide range of examples, including the ones presented in Section 2.1. But its definition is a bit complex, and hence it would be desirable to have a simple sufficient condition for connectedness. One of such conditions is the full support assumption.

**Definition 1.** The state transition function has a *full support* if $\pi^\omega(y, \tilde{\omega}|a) > 0$ for

all $\omega$, $\tilde{\omega}$, $a$, and $y$ such that $\pi_Y^\omega(y|a) > 0$.

In words, the full support assumption holds if any state $\tilde{\omega}$ can happen to-morrow given any current state $\omega$, action profile $a$, and signal $y$. An important consequence of this assumption is that players' posterior belief is always in the interior of $\triangle\Omega$; that is, after every history, the posterior belief $\mu^t$ assigns positive probability to each state $\omega$. Note that we do not require a full support with respect to $y$, so some signal $y$ may not occur for some state $\omega$ and some action profile $a$. As a result, the full support assumption can be satisfied for games with delayed observations, in which the signal $y$ does not have a full support.

In general, the full support assumption is much stronger than connectedness, and it rules out many economic applications. For example, the full support assumption is never satisfied if observable and unobservable state coexist.

## 4.2 Connectedness

In this subsection, we describe the idea of connectedness. In particular, we illustrate how it is related to irreducibility, which is commonly assumed in the literature on stochastic games with observable states.

The idea of irreducibility is introduced by Dutta (1995), and it is named by Fudenberg and Yamamoto (2011b). Irreducibility requires that each state $\tilde{\omega}$ be reachable from any state $\omega$ in finite time. Formally, $\tilde{\omega}$ is *accessible from* $\omega$ if there is a natural number $T$ and an action sequence $(a^1, \cdots, a^T)$ such that

$$\Pr(\omega^{T+1} = \tilde{\omega}|\omega, a^1, \cdots, a^T) > 0, \tag{2}$$

where $\Pr(\omega^{T+1} = \tilde{\omega}|\omega, a^1, \cdots, a^T)$ denotes the probability that the state in period $T+1$ is $\tilde{\omega}$ given that the initial state is $\omega$ and players play the action sequence $(a^1, \cdots, a^T)$ for the first $T$ periods. $\tilde{\omega}$ is *globally accessible* if it is accessible from any state $\omega \in \Omega$. Irreducibility requires the following property:[6]

**Definition 2.** The state evolution is *irreducible* if each state $\tilde{\omega}$ is globally accessible.

---

[6] Irreducibility of Fudenberg and Yamamoto (2011b) is stronger than the one presented here, but for our purpose (i.e., the invariance of the feasible payoff set), this weaker requirement is enough. In this paper, their condition is stated as *robust irreducibility*; see Section 5.3.

That is, irreducibility says that there is a path from $\omega$ to $\tilde{\omega}$ for any pair of states. Very roughly speaking, this property ensures that the evolution of the state is well-behaved in the long-run.

A natural extension of irreducibility to our model is to require global accessibility of all posterior beliefs $\mu^t$, because $\mu^t$ is an observable state in our model. A belief $\tilde{\mu} \in \triangle\Omega$ is *globally accessible* if for any initial prior $\mu$, there is $T$ and $(a^1, \cdots, a^T)$ such that

$$\Pr(\mu^{T+1} = \tilde{\mu}|\mu, a^1, \cdots, a^T) > 0.$$

Here, $\Pr(\mu^{T+1} = \tilde{\mu}|\mu, a^1, \cdots, a^T)$ denotes the probability that the posterior belief in period $T+1$ is $\mu^{T+1} = \tilde{\mu}$ given that the initial prior is $\mu$ and players play the action sequence $(a^1, \cdots, a^T)$. A naive generalization of irreducibility is to require each belief $\tilde{\mu} \in \triangle\Omega$ to be globally accessible.

Unfortunately, such a condition is too demanding and never satisfied. The problem is that there are infinitely many possible beliefs $\mu$ and thus there is no reason to expect recurrence; i.e., the posterior belief may not return to the current belief in finite time.[7] So we need to find a condition which is weaker than global accessibility of $\mu$ but still parallels irreducibility of the standard model.

A key is to look at the evolution of the *support* of $\mu^t$, rather than the evolution of $\mu^t$ itself. As will be explained in Section 5.2, all we need for our result is that the evolution of the *support* of the posterior belief is well-behaved in the long run. This suggests us to consider global accessibility of the support of the belief:

**Definition 3.** A non-empty subset $\Omega^* \subseteq \Omega$ is *globally accessible* if there is $\pi^* > 0$ such that for any initial prior $\mu$, there is a natural number $T \le 4^{|\Omega|}$, an action sequence $(a^1, \cdots, a^T)$, and a belief $\tilde{\mu}$ whose support is included in $\Omega^*$ such that

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

In words, $\Omega^* \subseteq \Omega$ is globally accessible if given any initial prior $\mu$, the support of the posterior belief $\mu^{T+1}$ can be a subset of $\Omega^*$ with probability at least $\pi^* > 0$ when players play some appropriate action sequence $(a^1, \cdots, a^T)$. A couple of remarks are in order. First, the above condition differs from (2) in that the former

---

[7]Formally, there always exists a belief $\mu$ which is not globally accessible, because given an initial belief, only countably many beliefs are reachable.

requires that there be a lower bound $\pi^* > 0$ on the probability of the posterior belief reaching $\tilde{\mu}$, while the latter does not. The reason why $\pi^*$ does not show up in (2) is that when states are observable, possible initial states are finite and thus the existence of a lower bound $\pi^* > 0$ is obvious. On the other hand, here we explicitly assume the existence of the bound $\pi^*$, since there are infinitely many initial priors $\mu$.[8]

Second, the restriction $T \leq 4^{|\Omega|}$ in the definition above is without loss of generality. That is, if there is $\tilde{T} > 4^{|\Omega|}$ which satisfies the condition stated above, then there is $T \leq 4^{|\Omega|}$ which satisfies the same condition. See Appendix A for details.

To state the definition of connectedness, we need to introduce one more idea, *transience*. We first give its definition and then discuss why we need it. Let $\Pr(\mu^{T+1} = \tilde{\mu} | \mu, s)$ denote the probability that the posterior belief in period $T + 1$ is $\mu^{T+1} = \tilde{\mu}$ given that the initial prior is $\mu$ and players play the strategy profile $s$. We would like to emphasize that the restriction $T \leq 2^{|\Omega|}$ in the definition below is without loss of generality; see Appendix A for details.

**Definition 4.** A subset $\Omega^* \subseteq \Omega$ is *transient* if it is not globally accessible and for any pure strategy profile $s \in S^*$ and for any $\mu$ whose support is $\Omega^*$, 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, transience of $\Omega^*$ implies that if the support of the current belief is $\Omega^*$, 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.[9] As shown in Lemma 11 in Appendix A, this property implies that if the support of the current belief is transient, then the support cannot return to the current one forever with positive probability. Hence, if $\Omega^*$ is transient, the time during which the support of the posterior belief stays at $\Omega^*$ is almost negligible in the long run. In other words, the existence of transient sets $\Omega^*$ does not influence the long-run behavior of the support of the posterior belief. Hence, if each

---

[8]Replacing the action sequence $(a^1, \cdots, a^T)$ in the definition with a strategy profile $s$ does not weaken the condition; that is, as long as there is a strategy profile which satisfies the condition stated in the definition, we can find an action sequence which satisfies the same condition.

[9]The strategy profile $s$ in Definition 4 cannot be replaced with an action sequence $(a^1, \cdots, a^T)$. This is in sharp contrast with global accessibility, in which both $(a^1, \cdots, a^T)$ and $s$ give the same condition.

subset $\Omega^* \subseteq \Omega$ is either globally accessible or transient, we can expect that the evolution of the support should be well-behaved in the long-run, on the analogy of irreducibility for games with observable states. This condition is precisely our assumption, connectedness:

**Definition 5.** A stochastic game is *connected* if each subset $\Omega^* \subseteq \Omega$ is globally accessible or transient.

The above definition is stated using the posterior belief $\mu^t$. In Appendix A, we will give an equivalent definition of connectedness based on primitives. It is stated as a condition on the distribution of the next state *and* the distribution of the public signal; this is a natural consequence from the fact that the evolution of the posterior belief is determined by the interaction of the evolution of the underlying state $\omega$ and of players' public signal $y$. Using this definition, one can check if a given game is connected or not in finitely many steps.

Connectedness is weaker than requiring that all subsets $\Omega^*$ be globally accessible, since some sets can be transient. This difference is important, because requiring global accessibility of all subsets $\Omega^*$ is too demanding in most applications. To see this, take a singleton set $\Omega^* = \{\omega\}$. For this set to be globally accessible, given any initial prior, the posterior belief $\mu^t$ must puts probability one on this state $\omega$ at some period $t$. However this can happen only if the signal $y$ reveals the next state, and such an assumption is violated in most applications.

## 4.3   When is the Game Connected?

Now we will explain that connectedness is satisfied in a wide range of examples. First of all, as argued, connectedness is satisfied whenever the full support assumption holds. To see this, note that under the full support assumption, the support of the posterior belief is the whole space $\Omega$ after every history. This implies that $\Omega$ is globally accessible, and other subsets are transient. Hence the game is connected. We record this result as a lemma:

**Lemma 1.** *If the state transition function has a full support, then the game is connected.*

While the full support assumption is satisfied in many applications, it is still stronger than connectedness. One of the examples in which the game is connected

but the full support assumption does not hold is stochastic games with observable states. So extending the full support assumption to connectedness is necessary if we want to establish a general theory which subsumes existing models as a special example. The following lemma shows that in stochastic games with observable states, connectedness reduces to a condition which is weaker than irreducibility. $\omega$ is *transient* if for any pure strategy profile $s$, the state must reaches from $\omega$ to some globally accessible state $\tilde{\omega}$ within $|\Omega|$ periods with positive probability.

**Lemma 2.** *In stochastic games with observable states, the game is connected if each state $\omega$ is globally accessible or transient.*

The proof of the lemma is straightforward; it is obvious that a singleton set $\{\omega\}$ with globally accessible $\omega$ is globally accessible, and other sets $\Omega^*$ are transient.

Next, consider the case in which the state is observable with delay. In this model, the full support assumption is satisfied if any state can happen tomorrow with positive probability. On the other hand, if the state evolution is deterministic, the full support assumption is violated. The following lemma shows that connectedness is satisfied even with a deterministic state evolution, as long as it is irreducible. The proof is given in Appendix E.

**Lemma 3.** *In stochastic games with delayed observations, the game is connected if each state $\omega$ is globally accessible or transient.*

In some applications, observable and unobservable states coexist. The full support assumption is never satisfied in such an environment, due to the observable component of the state. The next lemma shows that connectedness can be satisfied even in such a case. Recall that $\omega_O$ denotes an observable state and $\omega_U$ denotes an unobservable state. Let $\pi_\Omega^\omega(\tilde{\omega}|a)$ be the marginal distribution of the next state $\tilde{\omega}$ given the current state $\omega$ and the action profile $a$. Let $\pi_{\Omega_U}^\omega(\tilde{\omega}_U|a,y)$ be the conditional probability of the unobservable state $\tilde{\omega}_U$ given that the current state is $\omega$, the current action is $a$, and the signal $y$ is observed. The state evolution is *fully stochastic* if $\pi_\Omega^\omega(\tilde{\omega}|a) > 0$ and $\pi_{\Omega_U}^\omega(\tilde{\omega}_U|a,y) > 0$ for all $\omega$, $\tilde{\omega}$, $a$, and $y$. Intuitively, this condition says that any observable state can happen tomorrow with positive probability, and that players cannot rule out the possibility of any unobservable state conditional on any signal $y$. Note that we *do not* assume that

the evolutions of $\omega_O$ and $\omega_U$ are independent, so the distribution of the next observable state may depend on the current unobservable state. Hence the evolution of the observable state $\omega_O$ here can be quite different from the one for the standard stochastic game.

**Lemma 4.** *Suppose that observable and unobservable states coexist. The game is connected if the state evolution is fully stochastic.*

From Lemmas 2 and 3, we know that irreducibility of the underlying state is sufficient for connectedness, if states are observable (possibly with delay). Unfortunately, this result does not hold if states are not observable; irreducibility may not imply connectedness when states are hidden. See Example 2 in Appendix A, in which the state follows a deterministic cycle (and hence irreducible) but the game is not connected.

**Remark 1.** Although the game is not connected, we can still show that the folk theorem holds in Example 2 in Appendix A. A key is that connectedness is stronger than necessary, and it can be replaced with a weaker condition, called *asymptotic connectedness*. (See Appendix C for the definition.) The example satisfies asymptotic connectedness. More generally, as Lemma 14 in Appendix C shows, the game is asymptotically connected for generic signal structures as long as the state evolution is irreducible. This means that irreducibility of the underlying state "almost" implies connectedness.

# 5   Feasible and Individually Rational Payoffs

## 5.1   Invariance of Scores

Let $V^\omega(\delta)$ be the set of feasible payoffs when the initial state is $\omega$ and the discount factor is $\delta$, i.e., let $V^\omega(\delta) = \text{co}\{v^\omega(\delta,s)|s \in S^*\}$. Likewise, let $V^\mu(\delta)$ be the set of feasible payoffs when the initial prior is $\mu$. Note that the feasible payoff set depends on $\delta$, as the state $\omega$ changes over time.

Let $\Lambda$ be the set of directions $\lambda \in \mathbf{R}^N$ with $|\lambda| = 1$. For each direction $\lambda$, we compute the "score" using the following formula:

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

Note that this maximization problem indeed has a solution; see Appendix F for the proof. Roughly speaking, the score characterizes the boundary of the feasible payoff set $V^\mu(\delta)$ toward direction $\lambda$. For example, when $\lambda$ is the coordinate vector with $\lambda_i = 1$ and $\lambda_j = 0$ for all $j \neq i$, we have $\max_{v \in V^\mu(\delta)} \lambda \cdot v = \max_{v \in V^\mu(\delta)} v_i$, so the score represents the highest possible payoff for player $i$ in the feasible payoff set. Given a direction $\lambda$, let $f(\mu)$ be the score given the initial prior $\mu$. The function $f$ can be derived by solving the following Bellman equation:

$$f(\mu) = \max_{a \in A} \left[ (1-\delta)\lambda \cdot g^\mu(a) + \delta \sum_{y \in Y} \pi_Y^\mu(y|a) f(\tilde{\mu}(\mu,a,y)) \right] \quad (3)$$

where $\tilde{\mu}(\mu,a,y)$ is the belief in period two given that the initial prior is $\mu$ and players play $a$ and observe $y$ in period one. Note that (3) is a generalization of (1), which characterizes the best possible profit in the stochastic Bertrand model; indeed, when $\lambda = (\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}})$, (3) reduces to (1).

In Section 3, we have found that the total profit in the Bertrand model is insensitive to the initial prior when the discount factor is close to one. The following lemma generalizes this observation; it shows that if the game is connected and if $\delta$ is sufficiently large, the scores do not depend on the initial prior. This result implies that the feasible payoff sets $V^\mu(\delta)$ are similar across all initial priors $\mu$ when $\delta$ is close to one. The proof is given in Appendix E. [10]

**Lemma 5.** *Suppose that the game is connected. Then for each $\varepsilon > 0$, there is*

---

[10]We thank Johannes Hörner for pointing out that Lemma 5 strengthens the results in the literature of partially observable Markov decision process (POMDP). Whether the value function is constant or not in the limit as $\delta \to 1$ is an important question in the POMDP literature, since the constant value function ensures the existence of a solution to the dynamic programming equation with time-average payoffs. It turns out that connectedness is weaker than sufficient conditions found in the literature, including renewability of Ross (1968), reachability-detectability of Platzman (1980), and Assumption 4 of Hsu, Chuang, and Arapostathis (2006). (There is a minor error in Hsu, Chuang, and Arapostathis (2006); see Appendix H for more details.) So for anyone interested in a POMDP problem with time-average payoffs, connectedness is a condition which subsumes these existing conditions and is applicable to a broader class of games. Indeed, Examples 1 and 4 in this paper do not satisfy any assumptions above, but they are connected. (Also, Examples 2 and 3 do not satisfy the above assumptions, but they are asymptotically connected and hence Lemma 15 applies.) The only conditions which do not imply connectedness are Assumptions 2 and 5 of Hsu, Chuang, and Arapostathis (2006), but they are stated using the optimal policy and hence not tractable. For example, to check their assumptions in our setup, and we need to compute the optimal policy for each direction $\lambda$.

$\overline{\delta} \in (0,1)$ *such that for any* $\lambda \in \Lambda$, $\delta \in (\overline{\delta},1)$, $\mu$, *and* $\tilde{\mu}$,

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

Although it is not stated in the lemma, in the proof we show that the score converges at the rate of $1 - \delta$. That is, we can replace $\varepsilon$ in Lemma 5 with $O(1-\delta)$.

This lemma extends the invariance result of Dutta (1995) to the hidden-state case. The proof technique is different, because his proof essentially relies on ergodic theorems, which are not applicable to our model due to infinite states. In Appendix G, we explain why his technique does not apply to our model in more details. In the next subsection, we provide a sketch of the proof of Lemma 5 under the full support assumption, and discuss how to generalize it to connected stochastic games.

Now we define the "limit feasible payoff set." Lemma 9 in Appendix A shows that the score $\max_{v \in V^{\mu}(\delta)} \lambda \cdot v$ has a limit as $\delta \to 1$, so let $V^{\mu}$ be the set of all $v \in \boldsymbol{R}^{N}$ such that $\lambda \cdot v \leq \lim_{\delta \to 1} \max_{v \in V^{\mu}(\delta)} \lambda \cdot v$ for all $\lambda$. Lemma 5 above guarantees that this set $V^{\mu}$ is independent of $\mu$, so we denote it by $V$. This set $V$ is the limit feasible payoff set, in the sense that $V^{\mu}(\delta)$ approximates $V$ for all $\mu$ as long as $\delta$ is close to one; see Lemma 10 in Appendix A for details.

## 5.2 Proof Sketch

To illustrate the idea of our proof, consider the coordinate direction $\lambda$ with $\lambda_i = 1$ so that the score is simply player $i$'s highest possible payoff within the feasible payoff set. For simplicity, assume that $\pi$ has a full support, that is, $\pi^{\omega}(y, \tilde{\omega}|a) > 0$. Assume further that there are only two states; so the initial prior $\mu$ is represented by a real number between $[0,1]$. Let $s^{\mu}$ be the strategy profile which attains the score when the initial prior is $\mu$, i.e., let $s^{\mu}$ be such that $v_i^{\mu}(\delta, s^{\mu}) = \max_{v \in V^{\mu}(\delta)} v_i$.

As shown in Lemma 21 in Appendix E, the score $v_i^{\mu}(\delta, s^{\mu})$ is convex with respect to $\mu$. (The proof relies on the fact that player $i$'s payoff $v_i^{\mu}(\delta, s)$ is linear in a belief $\mu$ for a given $s$.) This implies that the score must be maximized by $\mu = 0$ or $\mu = 1$. Without loss of generality, assume that $\mu = 0$ is the maximizer, and let $\omega$ be the corresponding state. The curve in Figure 4 represents the score $v_i^{\mu}(\delta, s^{\mu})$ for each $\mu$; note that this is indeed a convex function and the value is
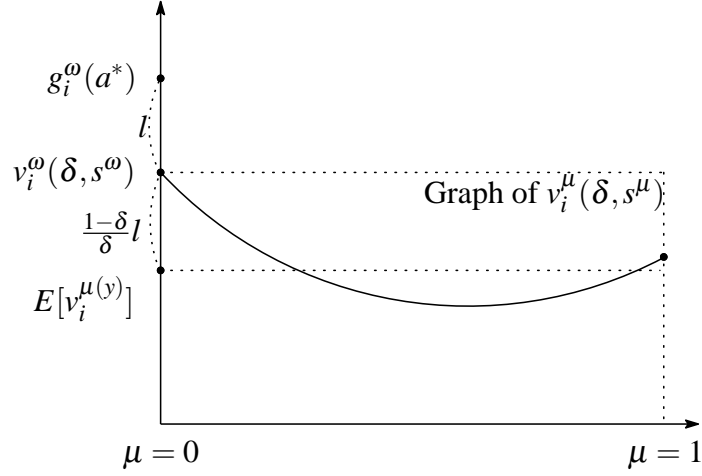
21

Figure 4: Score $v_i^\mu(\delta, s^\mu)$

maximized at $\mu = 0$. In what follows, the maximized value $v_i^\omega(\delta, s^\omega)$ is called maximal score. (Here the superscript $\omega$ means $\mu = 0$.)

The rest of the proof consists of two steps. In the first step, we show that there is a belief $\mu \in [\overline{\pi}, 1 - \overline{\pi}]$ such that the score $v_i^\mu(\delta, s^\mu)$ with the belief $\mu$ is close to the maximal score. This result follows from the fact that the score function is a solution to a dynamic programming equation.

The second step of the proof is more essential; it shows that if such a belief $\mu$ exists, then the score $v_i^{\tilde{\mu}}(\delta, s^{\tilde{\mu}})$ for *every* belief $\tilde{\mu}$ is close to the maximal score. This means that we do not need to compute the score for each belief separately; although the set of beliefs is continuous, if we can find *one* belief $\mu \in [\overline{\pi}, 1 - \overline{\pi}]$ which approximates the maximal score, we can bound the scores for *all* beliefs uniformly. The proof crucially relies on the convexity of the score function $v_i^\mu(\delta, s^\mu)$, which comes from the fact that the state variable $\mu$ is a belief.

### 5.2.1 Step 1: Existence of $\mu$ Approximating the Maximal Score

Recall that the score function is maximized at the belief $\mu = 0$, and $s^\omega$ is the strategy profile which achieves this maximal score. Let $a^*$ be the action profile in period one played by $s^\omega$. Let $\mu(y)$ be the posterior belief at the beginning of period two when the initial prior is $\mu = 0$ and the outcome in period one is $(a^*, y)$. Since the score $v_i^\omega(\delta, s^\omega)$ is the sum of the payoff $g_i^\omega(a^*)$ in period one and the

22

expectation of the continuation payoff $v_i^{\mu(y)}(\delta, s^{\mu(y)})$, we have

$$v_i^{\omega}(\delta, s^{\omega}) = (1 - \delta)g_i^{\omega}(a^*) + \delta E[v_i^{\mu(y)}(\delta, s^{\mu(y)})]$$

where $E$ is the expectation with respect to $y$ given that the initial state is $\omega$ and $a^*$ is chosen in period one. Equivalently,

$$v_i^{\omega}(\delta, s^{\omega}) - E[v_i^{\mu(y)}(\delta, s^{\mu(y)})] = \frac{1 - \delta}{\delta}(g_i^{\omega}(a^*) - v_i^{\omega}(\delta, s^{\omega})).$$

For simplicity, assume that $\omega$ and $s^{\omega}$ does not depend on $\delta$. (Lemma 22 in Appendix E shows that the result easily extends to the case in which they depend on $\delta$.) Then the above equality implies that

$$v_i^{\omega}(\delta, s^{\omega}) - E[v_i^{\mu(y)}(\delta, s^{\mu(y)})] = O\left(\frac{1 - \delta}{\delta}\right). \tag{4}$$

That is, the expected continuation payoff $E[v_i^{\mu(y)}(\delta, s^{\mu(y)})]$ is "close" to the maximal score $v_i^{\omega}(\delta, s^{\omega})$.

Now, we claim that the same result holds even if we take out the expectation operator; i.e., for each realization of $y$, the continuation payoff $v_i^{\mu(y)}(\delta, s^{\mu(y)})$ is close to the maximal score so that

$$v_i^{\omega}(\delta, s^{\omega}) - v_i^{\mu(y)}(\delta, s^{\mu(y)}) = O\left(\frac{1 - \delta}{\delta}\right). \tag{5}$$

To see this, note that

$$v_i^{\omega}(\delta, s^{\omega}) - E[v_i^{\mu(y)}(\delta, s^{\mu(y)})] = \sum_{y \in Y} \pi_Y^{\omega}(y|a^*)\{v_i^{\omega}(\delta, s^{\omega}) - v_i^{\mu(y)}(\delta, s^{\mu(y)})\}.$$

Since $v_i^{\omega}(\delta, s^{\omega})$ is the maximum score, the term in the curly brackets is non-negative for all $y$. Thus, if there is $y$ such that the term in the curly brackets is not of order $\frac{1-\delta}{\delta}$, then the right-hand side is not of order $\frac{1-\delta}{\delta}$. However this contradicts (4), and hence (5) holds for all $y$.

Pick an arbitrary $y$. Then the resulting posterior belief $\mu(y)$ satisfies the desired properties. Indeed, (5) ensures that the score for the belief $\mu(y)$ approximates the maximal score. Also, the full support of $\pi$ implies $\mu(y) \in [\overline{\pi}, 1 - \overline{\pi}]$.

23

### 5.2.2   Step 2: Uniform Bound

Take $\mu(y)$ as in the first step so that $\mu(y) \in [\overline{\pi}, 1 - \overline{\pi}]$ and the score approximates the maximal score if $\mu(y)$ is the initial prior. Consider the strategy profile $s^{\mu(y)}$, which achieves the score with the initial prior $\mu(y)$. The dashed line in Figure 5 represents player $i$'s payoff with this strategy profile $s^{\mu(y)}$ for each belief $\mu$. Note that it must be a line, because given a strategy profile $s$, the payoff for any interior belief $\mu \in (0, 1)$ is a convex combination of those for the boundary beliefs $\mu = 0$ and $\mu = 1$. The dashed line must be below the curve, since the curve gives the best possible payoff for each $\mu$. Also, the dashed line must intersect with the curve at $\mu = \mu(y)$, since the strategy profile $s^{\mu(y)}$ achieves the score at $\mu = \mu(y)$. Taken together, the dashed line must be tangential to the curve at $\mu = \mu(y)$, as described in the figure.
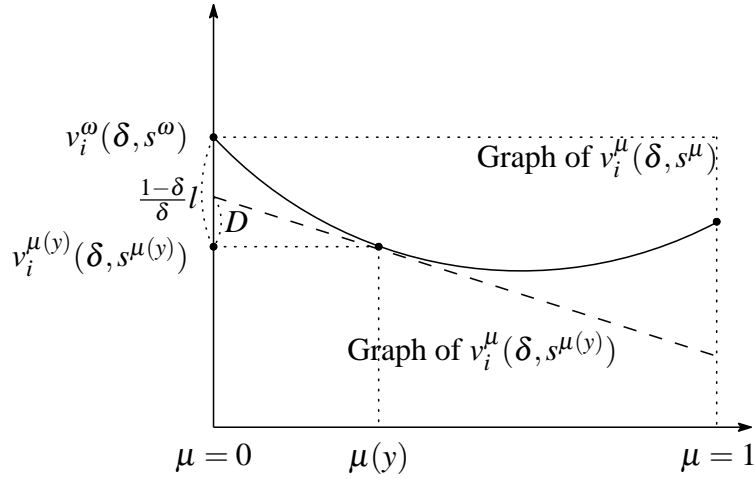


Figure 5: Payoff by $s^{\mu(y)}$

Suppose that the dashed line is downward-sloping, and let $D$ be as in the figure. In words, $D$ is the difference between the $y$-intercept of the dashed line and $v_i^{\mu(y)}(\delta, s^{\mu(y)})$. Since we have (5), the value $D$ is also of order $\frac{1-\delta}{\delta}$, as one can see from the figure. Then the slope of the dashed line, which is equal to $\frac{D}{\mu(y)}$, is also of order $\frac{1-\delta}{\delta}$. This implies that the dashed line is "almost flat" and thus the strategy profile $s^{\mu(y)}$ yields similar payoffs regardless of the initial prior $\mu$. In particular, since $v_i^{\mu(y)}(\delta, s^{\mu(y)})$ is close to $v_i^\omega(\delta, s^\omega)$, the dashed line is close to the horizontal line corresponding to the maximal score $v_i^\omega(\delta, s^\omega)$ regardless of the initial prior

24

$\mu$. Then the curve, which is between these two lines, is also close to the horizontal line corresponding to the maximal score $v_i^\omega(\delta, s^\omega)$ for all $\mu$. This implies that the score $v_i^\mu(\delta, s^\mu)$ approximates the maximal score for all $\mu$, which proves Lemma 5.

We can also show that the same result holds even if the dashed line is upward-sloping. To prove it, we use the payoffs at $\mu = 1$ rather than at $\mu = 0$ to bound the slope of the dashed line.

### 5.2.3  Discussion

In the above proof, in order to bound the slope of the dashed line, we use the full support assumption, which guarantees that $\mu(y) \in [\overline{\pi}, 1 - \overline{\pi}]$; i.e., the support of $\mu(y)$ must be the whole state space $\Omega$. Indeed, if $\mu(y)$ does not have a full support and is the boundary point $\mu(y) = 0$, the slope of the dashed line $\frac{D}{\mu(y)}$ is not necessarily of order $\frac{1-\delta}{\delta}$, even if $D$ is of order $\frac{1-\delta}{\delta}$.

In general, when we consider connected stochastic games, the full support assumption may not be satisfied, and hence the support of the posterior may not reach the whole state space $\Omega$. However, we can extend the idea presented in the second step to obtain the following result: If the score for some initial prior $\mu$ is close to the the maximal score,[11] then for *every* belief which has the same support as $\mu$, the corresponding score is also close to the maximal score. That is, if we can bound the score for some belief $\mu$ with support $\Omega^*$, then it gives a uniform bound on the scores for *all* beliefs with support $\Omega^*$.

This suggests that we may classify the set of all beliefs into groups with the same supports, and use the following "infection" argument:

- As a first step, we try to find a belief $\mu^*$ such that the score for $\mu^*$ is close to the maximal score. Let $\Omega^*$ be the support of $\mu^*$. Then the above result bounds the score for *every* belief with the support $\Omega^*$; that is, the scores for all beliefs with the support $\Omega^*$ are close to the maximal score.

- As a second step, we try to find a belief $\mu^{**}$ such that the score for $\mu^{**}$ is close to the score for some belief $\tilde{\mu}^*$ with support $\Omega^*$. Let $\Omega^{**}$ be the

---

[11]More precisely, we need that this belief $\mu$ is not too close to the boundary of $\triangle\Omega^*$, where $\Omega^*$ is the support of $\mu$. This parallels the fact that $\mu(y)$ in the above proof satisfies $\mu(y) \in [\overline{\pi}, 1 - \overline{\pi}]$ and hence does not approximate the boundary point $\mu = 0$ or $\mu = 1$.

support of $\mu^{**}$. From the first step, we already know that the latter score is close to the maximal score, and so is the former. This implies that the scores for all beliefs with the support $\Omega^{**}$ are also close to the maximal score.

In this way, we may try to bound the scores in order, group by group. Since there are only finitely many subsets $\Omega^*$, this process ends in finite steps, and it bounds the scores for all beliefs. The proof shows that this idea indeed works in connected stochastic games, because connectedness ensures that players can drive the support from any sets to others, which helps to find $\mu^*$ or $\mu^{**}$ stated above.

## 5.3 Minimax Payoffs

The *minimax payoff* to player $i$ in the stochastic game given the initial prior $\mu$ and discount factor $\delta$ is defined to be

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

Note that the minimizer $s_{-i}$ indeed exists (see Appendix F for the proof), and it is possibly a mixed strategy.

When the state is observable, Dutta (1995) shows that the minimax payoff has a limit as $\delta \to 1$ and its limit is invariant to the initial state $\omega$, by assuming a condition which we call *robust irreducibility*.[12] Robust irreducibility strengthens irreducibility in that it assures that any state $\tilde{\omega}$ can be reachable from any state $\omega$ *regardless of player $i$'s play*. More formally, $\tilde{\omega}$ is *robustly accessible from* $\omega$ if for each $i$, there is a (possibly mixed) action sequence $(\alpha_{-i}^1, \cdots, \alpha_{-i}^{|\Omega|})$ such that for any player $i$'s strategy $s_i$, there is a natural number $T \leq |\Omega|$ such that $\Pr(\omega^{T+1} = \tilde{\omega} | \omega, s_i, \alpha_{-i}^1, \cdots, \alpha_{-i}^T) > 0$. $\tilde{\omega}$ is *robustly globally accessible* if it is robustly globally accessible from $\omega$ for all $\omega$.

**Definition 6.** The state evolution is *robustly irreducible* if each state $\tilde{\omega}$ is robustly globally accessible.

When the state is observable, robust irreducibility ensures that the limit minimax payoff is invariant to the initial state. This paper extends this result to the

---

[12]As noted in footnote 6, this condition is called irreducibility in Fudenberg and Yamamoto (2011b).

hidden-state model. The assumption we make is *strong connectedness*, which strengthens connectedness and parallels the idea of robust irreducibility. Roughly, strong connectedness requires that players can drive the support of the belief from any sets to other sets regardless of player $i$'s play. The formal definition is stated in Appendix B, but we would like to emphasize that it is satisfied in a wide range of applications. For example, the game is strongly connected if the state evolution has a full support. Also Lemmas 2 through 4 hold even for strong connectedness, if the assumption is replaced with a stronger condition which ensures robustness to player $i$'s deviation. For example, in stochastic games with observable states, the game is strongly connected if the state evolution is robustly irreducible.

As shown in Appendix B, if the game is strongly connected, the limit minimax payoff exists and is invariant to the initial prior $\mu$. The proof of the invariance of the minimax payoff is quite different from that of the feasible payoff set. A new complication here is that the minimax payoff $\underline{v}_i^\mu(\delta)$ is not necessarily convex (or concave) with respect to $\mu$, since player $i$ maximizes the value while the opponents minimize it. In the proof, we take advantage of the fact that for each fixed strategy $s_{-i}$ of the opponents, player $i$'s best possible payoff $\max_{s_i \in S_i} v_i^\mu(s_i, s_{-i})$ is convex with respect to the initial prior $\mu$. This implies that if the opponents play a minimax strategy $s_{-i}^{\tilde{\mu}}$ for some *fixed* $\tilde{\mu}$, then player $i$'s best possible payoff is convex with respect to the initial prior $\mu$. Since the set of beliefs $\tilde{\mu}$ is continuous, there is a continuum of minimax strategies $\{s_{-i}^{\tilde{\mu}}\}_{\tilde{\mu} \in \triangle \Omega}$, and thus there is a continuum of convex functions, each of which is induced by some minimax strategy $s_{-i}^{\tilde{\mu}}$. In the proof, we bound this series of convex curves uniformly.

Let $\underline{v}_i^\mu = \lim_{\delta \to 1} \underline{v}_i^\mu(\delta)$ denote the limit minimax payoff with the initial prior $\mu$. Since this limit is invariant to $\mu$ for strongly connected stochastic games, we denote it by $\underline{v}_i$. Let $V^*$ denote the limit set of feasible and individually rational payoffs; that is, $V^*$ is the set of all feasible payoffs $v \in V$ such that $v_i \geq \underline{v}_i$ for all $i$.

# 6   Stochastic Self-Generation

The invariance results in the previous sections implies that after every history $h^t$, the feasible and individually rational payoff set in the continuation game is similar to the one for the original game. This suggests that dynamic programming can be helpful to characterize the equilibrium payoff set, as in the standard repeated

games. In this section, we show that this approach indeed works. Specifically, we introduce the notion of *stochastic self-generation*, which generalizes the idea of self-generation of Abreu, Pearce, and Stacchetti (1990).

In standard repeated games where the payoff functions are common knowledge, Abreu, Pearce, and Stacchetti (1990) show that for each discount factor $\delta$, the equilibrium payoff set is equal to the maximal self-generating payoff set. Their key idea is to decompose an equilibrium payoff into the stage-game payoff in period one and the continuation payoff from period two on. To be more concrete, let $s$ be a pure-strategy subgame-perfect equilibrium of some repeated game, and $a^*$ be the action profile chosen by $s$ in period one. Assume that actions are observable. Then the equilibrium payoff $v$ of $s$ must satisfy

$$v = (1 - \delta)g(a^*) + \delta w(a^*). \tag{6}$$

Here, $g(a^*)$ is the stage-game payoff vector given the action profile $a^*$, and $w(a^*)$ is the continuation payoff vector from period two given that $a^*$ is chosen in period one. Since $s$ is a subgame-perfect equilibrium, the continuation play is also subgame-perfect, which implies that $w(a^*)$ is chosen from the equilibrium payoff set $E(\delta)$. Also, when $s$ is an equilibrium which approximates the Pareto-efficient frontier, typically the action $a^*$ in period one yields a Pareto-efficient payoff. So (6) implies that the equilibrium payoff $v$ is *decomposable into a Pareto-efficient payoff and some continuation payoff chosen from the equilibrium payoff set*, as shown in Figure 6. FLM use this structure to establish a folk theorem.

Obviously, a similar payoff decomposition is possible in our model. Let $\mu$ be the initial prior and $s \in S^*$ be a sequential equilibrium which does not use public randomization. Then the equilibrium payoff $v$ must satisfy

$$v = (1 - \delta)g^\mu(a^*) + \delta \sum_{y \in Y} \pi_Y^\mu(y|a^*)w(a^*, y) \tag{7}$$

where $w(a^*, y)$ is the continuation payoff vector from period two on when the outcome in period one is $(a^*, y)$. However, decomposing the payoff like (7) is not very useful. A problem is that the stage-game payoff $g^\mu(a^*)$ is not necessarily on the Pareto-efficient frontier of the feasible payoff set $V$; this comes from the fact that the feasible payoff set in stochastic games is defined to be the set of long-run payoffs, rather than stage-game payoffs.
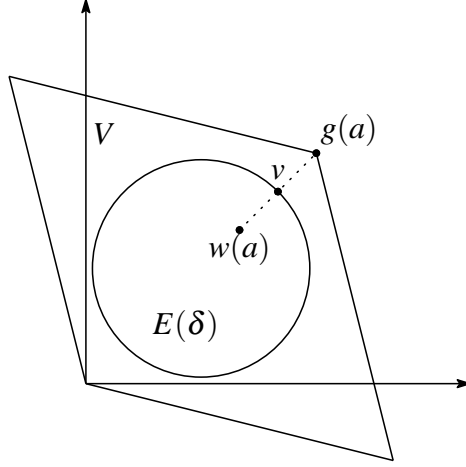
28

Figure 6: Payoff Decomposition

To fix this problem, we consider the following "random block" structure. The infinite horizon is divided into *blocks*, whose lengths are determined by public randomization $z \in [0,1]$. Specifically, at the end of each period $t$, players determine whether to continue the current block or not in the following way: Given some parameter $p \in (0,1)$, if $z^t \leq p$, the current block continues so that period $t+1$ is the next period of the current block. If $z^t > p$, then the current block terminates and the next period $t+1$ is regarded as the first period of the next block. In sum, the current block terminates with probability $1-p$ each period.

With this random block structure, equilibrium payoffs can be decomposed in the following way. Fix some initial prior $\mu$ and sequential equilibrium $s \in S$ arbitrarily, and let $v$ be the equilibrium payoff. For simplicity, assume that $s$ does not use public randomization; so in the following discussions, public randomization is used only for determining the length of random blocks. Since the equilibrium payoff is the sum of the payoffs during the first random block and the continuation payoff from the second random block, we have

$$v = (1-\delta)\sum_{t=1}^{\infty}(p\delta)^{t-1}E[g^{\omega^t}(a^t)|\mu,s] + (1-p)\sum_{t=1}^{\infty}p^{t-1}\delta^t E[w(h^t)|\mu,s] \quad (8)$$

where $w(h^t)$ be the continuation payoff vector after history $h^t$. Note that the first term on the right-hand side is the expected payoff during the first random block; the stage game payoff $g^{\omega^t}(a^t)$ in period $t$ is discounted by $(p\delta)^{t-1}$, since the

probability of period $t$ being reached before the termination of the first block is $p^{t-1}$. The second term on the right-hand side is the expected continuation payoff from the second random block; the term $(1-p)p^{t-1}$ represents the probability that the first block is terminated at the end of period $t$. Arranging the first term of the right-hand side, we obtain

$$v = \frac{1-\delta}{1-p\delta}v^\mu(p\delta,s) + (1-p)\sum_{t=1}^{\infty}p^{t-1}\delta^t E[w(h^t)|\mu,s].$$

This shows that the equilibrium payoff vector $v$ is decomposable into the following two terms; the first term on the right-hand side is *the payoff vector in the stochastic game with discount factor $p\delta$* (not $\delta$), and the second is the continuation payoff from the second random block. Intuitively, this comes from the fact that the "effective discount factor" in the random block is $p\delta$, due to the termination probability $1-p$.

Now, pick $p$ sufficiently large and then take $\delta$ close to one. Then $p\delta$ is close to one, and thus the payoff vector $v^\mu(p\delta,s)$ can approximate the efficient frontier of the limit feasible payoff set $V$ (with an appropriate choice of $s$). This implies that $v$ is a weighted average of some payoff approximating the efficient frontier and expected continuation payoffs, just as in Figure 6. Also, for a fixed $p$, the coefficient $\frac{1-\delta}{1-p\delta}$ on the term $v^\mu(p\delta,s)$ converges to zero when $\delta$ goes to one; hence a small variation in continuation payoffs is enough to provide appropriate incentives during the first random block. These properties are reminiscent of the payoff decomposition (6) for the standard repeated game, and we use them to establish the folk theorem.

Now we present a version of the self-generation theorem which decomposes payoffs in the above way. Consider an auxiliary dynamic game such that the game terminates at some period $t$, which is randomly determined by public randomization; that is, after every period $t$, the game terminates with probability $1-p$ (and proceeds to the next period with probability $p$), where $p \in (0,1)$ is a fixed parameter. Assume also that if the game terminates at the end of period $t$, each player $i$ receives some "bonus payment" $\frac{w_i(h^t)}{1-\delta}$ depending on the past history $h^t$. Intuitively, this bonus payment $w_i(h^t)$ corresponds to player $i$'s continuation payoff from the next block in the original stochastic game. Here the payment $w_i$ depends on $h^t$, which reflects the fact that players' continuation payoffs from the next block depend on the history during the current block. Given $\mu, \delta, p \in (0,1)$,

30

and $w : H \to \boldsymbol{R}^N$, let $\Gamma(\mu, \delta, p, w)$ denote this *stochastic termination game*. For each strategy profile $s$, player $i$'s expected average payoff in this stochastic termination game is precisely the right-hand side of (8). When we consider the stochastic termination game, we assume that the function $w$ does not depend on the past public randomization, in order to avoid a measurability problem.

**Definition 7.** A pair $(s, v)$ of a strategy profile and a payoff vector is *stochastically enforceable with respect to* $(\delta, \mu, p)$ if there is a function $w : H \to \boldsymbol{R}^N$ such that the following properties are satisfied:

(i) When players play the strategy profile $s$ in the stochastic termination game $\Gamma(\mu, \delta, p, w)$, the resulting payoff vector is exactly $v$.

(ii) $s$ is a sequential equilibrium in the stochastic termination game $\Gamma(\mu, \delta, p, w)$.

Intuitively, $s$ in the above definition should be interpreted as a strategy profile for the first block, and $v$ as players' equilibrium payoff in the original stochastic game. Stochastic enforceability guarantees that there are some continuation payoffs $w(h^t)$ from the second block so that players' incentive constraints for the first block are satisfied and the equilibrium payoff $v$ is indeed achieved.

Now we introduce the concept of stochastic self-generation, which is a counterpart to self-generation of Abreu, Pearce, and Stacchetti (1990).

**Definition 8.** A subset $W$ of $\boldsymbol{R}^N$ is *stochastically self-generating with respect to* $(\delta, p)$ if for each $v \in W$ and $\mu$, there are $s \in S$ and $w : H \to W$ such that $(s, v)$ is stochastically enforceable with respect to $(\delta, \mu, p)$ using $w$.

In words, for $W$ to be stochastically self-generating, each payoff $v \in W$ must be stochastically enforceable given any initial prior $\mu$, using some strategy profile $s$ and function $w$ which chooses continuation payoffs from the set $W$. Here we may use different strategy profile $s$ and continuation payoff $w$ for different priors $\mu$, since the posterior belief is a common state variable in our model.

The following result is an extension of the self-generation theorem of Abreu, Pearce, and Stacchetti (1990). The proof is similar to theirs and hence omitted.

**Proposition 1.** *Fix $\delta$. If $W$ is bounded and stochastically self-generating with respect to $(\delta, p)$ for some $p$, then for each payoff vector $v \in W$ and initial prior $\mu$, there is a sequential equilibrium with the payoff $v$.*

For stochastic games with observable (and finite) states, Hörner, Sugaya, Takahashi, and Vieille (2011) use the idea of "$T$-period generation," which decomposes a player's overall payoff into her average payoff in the $T$-period stochastic game and her continuation payoff, where $T$ is a fixed number. It is unclear if their idea works in our setup, because the belief evolution in our model may not be ergodic and thus an average payoff in the $T$-period stochastic game may not approximate the boundary of the set $V^*$. Fudenberg and Yamamoto (2011b) propose the concept of "return generation," which considers a stochastic game such that the game terminates when the state returns to the initial state. Unfortunately we cannot follow their approach, as the belief may not return to the initial belief forever.

Independently of this paper, Hörner, Takahashi, and Vieille (2015) also propose the same self-generation concept, which they call "random switching." However, their model and motivation are quite different from ours. They study repeated adverse-selection games in which players report their private information every period. In their model, a player's incentive to disclose her information depends on the impact of her report on her flow payoffs until the effect of the initial state vanishes. Measuring this impact is difficult in general, but it becomes tractable when the equilibrium strategy has the random switching property. That being said, they use stochastic self-generation in order to measure payoffs by misreporting. In contrast, in this paper, stochastic self-generation is used to decompose equilibrium payoffs in an appropriate way. Another difference between the two papers is the order of limits. They take the limits of $p$ and $\delta$ simultaneously, while we fix $p$ first and then take $\delta$ large enough.

**Remark 2.** Note that Proposition 1 does not rely on the fact that the state is a belief. Hence it applies to general stochastic games with infinite states, in which the state is not necessarily a belief.

## 7 Folk Theorem

Now we will establish the folk theorem for strongly connected stochastic games:

**Proposition 2.** *Suppose that the game is strongly connected. Then for any interior point v of $V^*$, there is $\overline{\delta} \in (0,1)$ such that for any $\delta \in (\overline{\delta},1)$ and for any initial prior $\mu$, there is a sequential equilibrium which yields the payoff of v.*

This proposition shows that the folk theorem holds under the full dimensional condition, $\dim V^* = N$. (If $\dim V^* < N$, there is no interior point of $V^*$.)

The proof builds on the techniques developed by FLM and Fudenberg and Yamamoto (2011b). From Proposition 1, it is sufficient to show that any "smooth" subset $W$ of the interior of $V^*$ is stochastically self-generating for $\delta$ close to one, that is, it is enough to show that each target payoff $v \in W$ is stochastically enforceable using continuation payoffs $w$ chosen from the set $W$. Construction of the continuation payoff function $w$ is more complex than those in FLM and Fudenberg and Yamamoto (2011b), since we need to consider the random block structure; but it is still doable. The formal proof can be found in Appendix D.

# 8   Conclusion

This paper considers a new class of stochastic games in which the state is hidden information. We find that if the game is strongly connected, then the feasible and individually rational payoff set is invariant to the initial belief in the limit as the discount factor goes to one. Then we develop the idea of stochastic self-generation, which generalizes self-generation of Abreu, Pearce, and Stacchetti (1990), and prove the folk theorem.

Throughout this paper, we assume that actions are perfectly observable. In an ongoing project, we try to extend the analysis to the case in which actions are not observable. When actions are not observable, each player has private information about her actions, and thus different players may have different beliefs. This implies that a player's belief is not public information and cannot be regarded as a common state variable; hence the model does not reduce to stochastic games with observable states. Accordingly, the analysis of the imperfect-monitoring case is quite different from that for the perfect-monitoring case.

# Appendix A: More on Connectedness

## A.1 Connectedness in Terms of Primitives

Here we provide the definition of global accessibility, transience, and connectedness in terms of primitives. We begin with global accessibility.

**Definition 9.** A subset $\Omega^* \subseteq \Omega$ is *globally accessible* if for each state $\omega \in \Omega$, there is a natural number $T \leq 4^{|\Omega|}$, an action sequence $(a^1, \cdots, a^T)$, and a signal sequence $(y^1, \cdots, y^T)$ such that the following properties are satisfied:[13]

(i) If the initial state is $\omega$ and players play $(a^1, \cdots, a^T)$, then the sequence $(y^1, \cdots, y^T)$ realizes with positive probability. That is, there is a state sequence $(\omega^1, \cdots, \omega^{T+1})$ such that $\omega^1 = \omega$ and $\pi^{\omega^t}(y^t, \omega^{t+1}|a^t) > 0$ for all $t \leq T$.

(ii) If players play $(a^1, \cdots, a^T)$ and observe $(y^1, \cdots, y^T)$, then the state in period $T+1$ must be in the set $\Omega^*$, regardless of the initial state $\hat{\omega}$ (possibly $\hat{\omega} \neq \omega$). That is, for each $\hat{\omega} \in \Omega$ and $\tilde{\omega} \notin \Omega^*$, there is no sequence $(\omega^1, \cdots, \omega^{T+1})$ such that $\omega^1 = \hat{\omega}$, $\omega^{T+1} = \tilde{\omega}$, and $\pi^{\omega^t}(y^t, \omega^{t+1}|a^t) > 0$ for all $t \leq T$.

As the following lemma shows, the definition of globally accessibility here is equivalent to the one presented in Section 4.2. The proof can be found in Appendix E.

**Lemma 6.** *Definitions 3 and 9 are equivalent.*

Let $\mathcal{O}$ be the set of all globally accessible $\Omega^* \subseteq \Omega$. The set $\mathcal{O}$ is non-empty, as the whole state space $\Omega^* = \Omega$ is always globally accessible. Indeed, when

---

[13] As claimed in Section 4.2, restricting attention to $T \leq 4^{|\Omega|}$ is without loss of generality. To see this, pick a subset $\Omega^* \subseteq \Omega$ and $\omega$ arbitrarily. Assume that there is a natural number $T > 4^{|\Omega|}$ so that we can choose $(a^1, \cdots, a^T)$ and $(y^1, \cdots, y^T)$ which satisfy (i) and (ii) in Definition 9. For each $t \leq T$ and $\tilde{\omega} \in \Omega$, let $\Omega^t(\tilde{\omega})$ be the support of the posterior belief given the initial state $\tilde{\omega}$, the action sequence $(a^1, \cdots, a^t)$, and the signal sequence $(y^1, \cdots, y^t)$. Since $T > 4^{|\Omega|}$, there are $t$ and $\tilde{t} > t$ such that $\Omega^t(\tilde{\omega}) = \Omega^{\tilde{t}}(\tilde{\omega})$ for all $\tilde{\omega}$. Now, consider the action sequence with length $T - (\tilde{t} - t)$, which is constructed by deleting $(a^{t+1}, \cdots, a^{\tilde{t}})$ from the original sequence $(a^1, \cdots, a^T)$. Similarly, construct the signal sequence with length $T - (\tilde{t} - t)$. Then these new sequences satisfy (i) and (ii) in Definition 9 in Appendix A. We can repeat this procedure to show the existence of sequences with length $T \leq 4^{|\Omega|}$ which satisfy (i) and (ii).

$\Omega^* = \Omega$, clause (ii) is vacuous and clause (i) is trivially satisfied by an appropriate choice of $(y^1, \cdots, y^T)$.

When the full support assumption holds, only the whole state space $\Omega$ is globally accessible, i.e., $\mathscr{O} = \{\Omega\}$. On the other hand, when the full support assumption is not satisfied, $\mathscr{O}$ may contain a proper subset $\Omega^* \subset \Omega$. Note that if some proper subset $\Omega^*$ is in the set $\mathscr{O}$, then by the definition of global accessibility, any superset $\Omega^{**} \supseteq \Omega^*$ is in the set $\mathscr{O}$. That is, any superset of $\Omega^*$ is globally accessible.

The following is the definition of transience in terms of primitives. With an abuse of notation, for each pure strategy profile $s \in S^*$ which does not use public randomization, let $s(y^1, \cdots, y^{t-1})$ denote the pure action profile induced by $s$ in period $t$ when the past signal sequence is $(y^1, \cdots, y^{t-1})$.

**Definition 10.** A singleton set $\{\omega\} \notin \mathscr{O}$ is *transient* if for any pure strategy profile $s \in S^*$, there is a globally accessible set $\Omega^* \in \mathscr{O}$, a natural number $T \leq 2^{|\Omega|}$, and a signal sequence $(y^1, \cdots, y^T)$ such that for each $\tilde{\omega} \in \Omega^*$, there is a state sequence $(\omega^1, \cdots, \omega^{T+1})$ such that $\omega^1 = \omega$, $\omega^{T+1} = \tilde{\omega}$, and $\pi^{\omega^t}(y^t, \omega^{t+1} | s(y^1, \cdots, y^{t-1})) > 0$ for all $t \leq T$.[14]

In words, $\{\omega\}$ is transient if the support of the belief cannot stay there forever given any strategy profile; that is, the support of the belief must reach some globally accessible set $\Omega^*$ at some point in the future.[15] It is obvious that the definition of transience above is equivalent to Definition 4 in Section 4.2, except that here we consider only singleton sets $\{\omega\}$.

Now we are ready to give the definition of connectedness:

**Definition 11.** A stochastic game is *connected* if each singleton set $\{\omega\}$ is globally accessible or transient.

---

[14]Restricting attention to $T \leq 2^{|\Omega|}$ is without loss of generality. To see this, suppose that there is a strategy profile $s$ and an initial prior $\mu$ whose support is $\Omega^*$ such that the probability that the support of the posterior belief reaches some globally accessible set within period $2^{|\Omega|}$ is zero. Then we can construct a strategy profile $\tilde{s}$ such that if the initial prior is $\mu$ and players play $\tilde{s}$, the support of the posterior belief never reaches a globally accessible set. The proof is standard and hence omitted.

[15]While we consider an arbitrary strategy profile $s \in S^*$ in the definition of transience, in order to check whether a set $\{\omega\}$ is transient or not, what matters is the belief evolution in the first $2^{|\Omega|}$ periods only, and thus we can restrict attention to $2^{|\Omega|}$-period pure strategy profiles, Hence the verification of transience of each set $\{\omega\}$ can be done in finite steps.

In this definition, we consider only singleton sets $\{\omega\}$. However, as the following lemma shows, if each singleton set $\{\omega\}$ is globally accessible or transient, then any subset $\Omega^* \subseteq \Omega$ is globally accessible or transient. Hence the above definition is equivalent to the one in Section 4.2.

**Lemma 7.** *If each singleton set $\{\omega\}$ is globally accessible or transient, then any subset $\Omega^* \subseteq \Omega$ is globally accessible or transient.*

## A.2 Limit Feasible Payoff Set

As shown in Lemma 5, connectedness implies the invariance of the feasible payoff set when $\delta$ is close to one. The following lemma shows that for each direction $\lambda$ and sufficiently large $\delta$, there is a strategy profile $s$ which yields a payoff vector approximating the score regardless of the initial prior $\mu$. The proof is found in Appendix E.

**Lemma 8.** *Suppose that the game is connected. Then for each $\varepsilon > 0$, there is $\overline{\delta} \in (0,1)$ such that for any $\lambda \in \Lambda$ and $\delta \in (\overline{\delta}, 1)$, there is a pure strategy profile $s \in S^*$ such that for all $\mu$,*

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

This lemma is used in the proof of the next lemma, which shows that for each initial prior $\mu$ and direction $\lambda$, the score converges to a limiting value as $\delta$ goes to one. This ensures that the set $V^{\mu}$ in Section 4.2 is well-defined. The proof can be found in Appendix E.[16]

**Lemma 9.** *If the game is connected, then $\lim_{\delta \to 1} \max_{v \in V^{\mu}(\delta)} \lambda \cdot v$ exists for each $\lambda$ and $\mu$.*

The next lemma shows that the convergence of the score function $\lambda \cdot v^{\mu}(\delta, s^{\mu})$ is uniform in $\lambda$. This implies that the set $V^{\mu}$ in Section 4.2 is indeed the limit feasible payoff set in the sense that the feasible payoff set $V^{\mu}(\delta)$ approximates $V^{\mu}$ when $\delta$ is close to one. The proof can be found in Appendix E.

---

[16]This lemma is a corollary of Theorem 2 of Rosenberg, Solan, and Vieille (2002), but for completeness, we provide a (simple and new) proof. We thank Johannes Hörner for pointing this out.

**Lemma 10.** *For each $\varepsilon$, there is $\overline{\delta} \in (0,1)$ such that for each $\lambda \in \Lambda$, $\delta \in (\overline{\delta}, 1)$, and $\mu$,*

$$\left| \max_{v \in V^\mu(\delta)} \lambda \cdot v - \lim_{\delta \to 1} \max_{v \in V^\mu(\delta)} \lambda \cdot v \right| < \varepsilon.$$

## A.3 Examples

As Lemmas 2 and 3 suggest, irreducibility of the underlying state is not necessary for connectedness. Here we present one of such examples; that is, in the example below, the game is connected although the underlying state is not irreducible.

**Example 1.** Suppose that there are three states, $\omega_1$, $\omega_2$, and $\omega_3$. If the current state is $\omega_1$, the state stays at $\omega_1$ with probability $\frac{1}{2}$ and moves to $\omega_2$ with probability $\frac{1}{2}$. If the current state is $\omega_2$, the state moves to $\omega_3$ for sure. If the current state is $\omega_3$, the state moves to $\omega_2$ for sure. Note that this state transition is not irreducible, as there is no path from $\omega_2$ to $\omega_1$. Assume that the signal space is $Y = \{y_0, y_1, y_2, y_3\}$, and that the signal $y$ is correlated with the next hidden state. Specifically, if the next state is $\omega_k$, players observe $y_0$ or $y_k$ with probability $\frac{1}{2}$-$\frac{1}{2}$, for each $k = 1, 2, 3$. Intuitively, $y_k$ reveals the next state $\omega_k$ for each $k = 1, 2, 3$, while $y_0$ does not reveal the state. In this example, it is easy to check that $\{\omega_2\}$ and $\{\omega_3\}$ are globally accessible, while $\{\omega_1\}$ is transient. Hence the game is connected.

From Lemmas 2 and 3, we know that irreducibility of the underlying state is sufficient for connectedness, if states are observable (possibly with delay). Unfortunately, this result does not hold if states are not observable; that is, irreducibility may not imply connectedness when states are hidden. We can show this through the following example, in which the state follows a deterministic cycle.

**Example 2.** Suppose that there are only two states, $\Omega = \{\omega_1, \omega_2\}$, and that the state evolution is a deterministic cycle; i.e., the state goes to $\omega_2$ for sure if the current state is $\omega_1$, and vice versa. Assume that $|A_i| = 1$ for each $i$, and that the public signal $y$ does not reveal the state $\omega$, that is, $\pi_Y^\omega(y|a) > 0$ for all $\omega$, $a$, and $y$. In this game, if the initial prior is fully mixed so that $\mu(\omega_1) > 0$ and $\mu(\omega_2) > 0$, then the posterior belief is also mixed. Hence only the whole state space $\Omega^* = \Omega$ is globally accessible. On the other hand, if the initial prior puts probability one on some state $\omega$, then the posterior belief puts probability one on $\omega$ in all odd periods

37

and on $\tilde{\omega} \neq \omega$ in all even periods. Hence the support of the posterior belief cannot reach the globally accessible set $\Omega^* = \Omega$, and thus each $\{\omega\}$ is not transient.

Now we present another example in which the game is not connected although the state evolution is irreducible. What is different from Example 2 is that the state evolution is not deterministic. (As will be explained in Appendix C, asymptotically connectedness is satisfied in both these examples.)

**Example 3.** Consider a machine with two states, $\omega_1$ and $\omega_2$. $\omega_1$ is a "normal" state and $\omega_2$ is a "bad" state. Suppose that there is only one player and that she has two actions, "operate" and "replace." If the machine is operated and the current state is normal, the next state will be normal with probability $p_1$ and will be bad with probability $1 - p_1$, where $p_1 \in (0, 1)$. If the machine is operated and the current state is bad, the next state will be bad for sure. If the machine is replaced, regardless of the current state, the next state will be normal with probability $p_2$ and will be bad with probability $1 - p_2$, where $p_2 \in (0, 1]$. There are three signals, $y_1$, $y_2$, and $y_3$. When the machine is operated, both the "success" $y_1$ and the "failure" $y_2$ can happen with positive probability; we assume that its distribution depends on the current hidden state and is not correlated with the distribution of the next state. When the machine is replaced, the "null signal" $y_3$ is observed regardless of the hidden state. Connectedness is not satisfied in this example, since $\{\omega_2\}$ is neither globally accessible nor transient. Indeed, when the support of the current belief is $\Omega$, it is impossible to reach the belief $\mu$ with $\mu(\omega_2) = 1$, which shows that $\{\omega_2\}$ is not globally accessible. Also $\{\omega_2\}$ is not transient, because if the current belief puts probability one on $\omega_2$ and "operate" is chosen forever, the support of the posterior belief is always $\{\omega_2\}$.

In some applications, there are action profiles which reveal the next state. If there is such an action profile, then the game is connected, as illustrated in the next example.

**Example 4.** Consider the machine replacement problem discussed above, but now assume that there are three actions; "operate," "replace," and "inspect." If the machine is inspected, the state does not change and a signal reveals the current state (hence the next state). Then it is easy to verify that each $\{\omega\}$ is globally accessible and thus the game is connected.

## A.4 Consequence of Transience

Lastly, we would like to present a lemma, which can be helpful to understand the interpretation of transience. As claimed in Section 4.2, if the support of the current belief is transient, then the support cannot return to the current one forever with positive probability. This in turn implies that the probability of the support of the belief being transient in period $T$ is almost negligible when $T$ is large enough. The following lemma verifies this claim formally. The proof is given in Appendix E. Let $X(\Omega^*|\mu,s)$ be the random variable $X$ which represents the first time in which the support of the posterior belief is $\Omega^*$ given that the initial prior is $\mu$ and players play $s$. That is, let

$$X(\Omega^*|\mu,s) = \inf\{T \geq 2 \text{ with } \mathrm{supp}\mu^T = \Omega^*|\mu,s\}.$$

Let $\Pr(X(\Omega^*|\mu,s) < \infty)$ denote the probability that the random variable is finite; i.e., it represents the probability that the support reaches $\Omega^*$ in finite time.

**Lemma 11.** *For any transient set $\Omega^*$, any initial prior $\mu$ whose support is $\Omega^*$, and any strategy profile $s$, the probability that the support returns to $\Omega^*$ in finite time is strictly less than one. That is,*

$$\Pr(X(\Omega^*|\mu,s) < \infty) < 1.$$

# Appendix B: Strong Connectedness

Here we present the definition of strong connectedness, which ensures the invariance of the minimax payoffs. To do so, we first need to introduce the idea of robust global accessibility and strong transience. We begin with robust global accessibility:

**Definition 12.** A non-empty subset $\Omega^* \subseteq \Omega$ is *robustly globally accessible* if there is $\pi^* > 0$ such that for any initial prior $\mu$ and for any $i$, there is an action sequence $(a^1_{-i}, \cdots, a^{4^{|\Omega|}}_{-i})$ such that for any player $i$'s strategy $s_i$, there is a natural number $T \leq 4^{|\Omega|}$ and a belief $\tilde{\mu}$ whose support is $\Omega^*$ such that

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

Robust global accessibility of $\Omega^*$ differs from global accessibility in two respects. First, robust global accessibility requires that the support of the belief should reach $\Omega^*$ regardless of player $i$'s play. Second, the support of the resulting belief $\tilde{\mu}$ must be precisely equal to $\Omega^*$; global accessibility requires only that the support of the posterior belief $\tilde{\mu}$ be a subset of $\Omega^*$.

Next, we define strong transience.

**Definition 13.** A subset $\Omega^* \subseteq \Omega$ is *strongly transient* if it is not robustly globally accessible and there is $\pi^* > 0$ such that for any pure strategy profile $s \in S^*$ and for any $\mu$ whose support is $\Omega^*$, there is a natural number $T \leq 2^{|\Omega|}$ and a belief $\tilde{\mu}$ whose support is robustly globally accessible such that

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

Again, strong transience differs from transience in two respects. First, the support of the posterior belief $\tilde{\mu}$ must be robustly globally accessible. Second, we now require that the probability that the support of the belief reaches a robust globally accessible set is at least $\pi^*$ regardless of the initial prior with support $\Omega^*$.

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

**Definition 14.** The game is *strongly connected* if the following conditions hold:

(i) Each non-empty subset $\Omega^* \subseteq \Omega$ is robustly globally accessible or strongly transient.

(ii) For each $\omega$, for each $\mu$ whose support is $\Omega$, and for each pure strategy profile $s \in S^*$, there is a natural number $T \leq 4^{|\Omega|}$ and a history $h^T$ such that the probability of $h^T$ given $(\omega, s)$ is positive, and the support of the posterior belief induced by $\omega$ and $h^T$ is identical with that induced by $\mu$ and $h^T$.

Clause (i) is a natural extension of connectedness. Clause (ii) is new, which says that the supports of the two posterior beliefs induced by different posteriors $\omega$ and $\mu$ must merge at some point, regardless of the play. Note that clause (ii) is trivially satisfied in many examples; for example, if the state evolution has a full support, then the support of the posterior belief is $\Omega$ regardless of the initial belief, and hence clause (ii) holds.

The following lemma, which parallels Lemmas 5 and 8, is a consequence of strong connectedness. The proof is given in Appendix E.

**Lemma 12.** *Suppose that the game is strongly connected. Then for each $\varepsilon > 0$, there is $\overline{\delta} \in (0,1)$ such that the following properties hold:*

*(i) For any $\delta \in (\overline{\delta}, 1)$, $\mu$, and $\tilde{\mu}$*

$$\left| \underline{v}_i^\mu(\delta) - \underline{v}_i^{\tilde{\mu}}(\delta) \right| < \varepsilon.$$

*(ii) For any $\delta \in (\overline{\delta}, 1)$, there is a strategy profile $s \in S^*$ such that for all $\mu$,*

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

*and*

$$\left| \max_{\tilde{s}_i \in S_i} v_i^\mu(\delta, \tilde{s}_i, s_{-i}) - \underline{v}_i^\mu(\delta) \right| < \varepsilon.$$

Clause (i) asserts that when $\delta$ is sufficiently large, the minimax payoffs are similar across all priors $\mu$. Clause (ii) ensures that when $\delta$ is sufficiently large, there is a strategy profile $s$ such that regardless of the initial prior $\mu$, the generated payoff approximates the minimax payoff and player $i$'s play $s_i$ is approximately optimal. The existence of such a strategy profile is useful in the proof of Lemma 13.

The next lemma shows that the limit minimax payoff exists. See Appendix E for the proof.

**Lemma 13.** *Suppose that the game is strongly connected. Then $\lim_{\delta \to 1} \underline{v}_i^\mu(\delta)$ exists.*

# Appendix C: Relaxing Connectedness

Here we show that connectedness is stronger than necessary for the invariance of the feasible payoff set; we show that *asymptotic connectedness*, which is weaker than connectedness, is sufficient.

To illustrate the idea of asymptotic connectedness, consider Example 2 in Section 4.2, where the state $\omega$ is unobservable and follows a deterministic cycle. Suppose that the signal distribution is different at different states and does not depend on the action profile, that is, $\pi_Y^{\omega_1}(\cdot|a) = \pi_1$ and $\pi_Y^{\omega_2}(\cdot|a) = \pi_2$ for all $a$, where $\pi_1 \neq \pi_2$. Suppose that the initial state is $\omega_1$. Then the true state must be $\omega_1$ in all

odd periods, and be $\omega_2$ in all even periods. Hence if we consider the empirical distribution of the public signals in odd periods, it should approximate $\pi_1$ with probability close to one, by the law of large numbers. Similarly, if the initial state is $\omega_2$, the empirical distribution of the public signals in odd periods should approximate $\pi_2$. This implies that players can eventually learn the current state by aggregating the past public signals, regardless of the initial prior $\mu$. Hence for $\delta$ close to one, the feasible payoff set should be similar across all $\mu$, i.e., Lemma 5 should remain valid in this example, even though the game is not connected.

The point in this example is that, while the singleton set $\{\omega_1\}$ is not globally accessible, it is *asymptotically accessible* in the sense that at some point in the future, the posterior belief puts a probability arbitrarily close to one on $\omega_1$, regardless of the initial prior. As will be explained, this property is enough to establish the invariance of the feasible payoff set. Formally, asymptotic accessibility is defined as follows:

**Definition 15.** A non-empty subset $\Omega^* \subseteq \Omega$ is *asymptotically accessible* if for any $\varepsilon > 0$, there is a natural number $T$ and $\pi^* > 0$ such that for any initial prior $\mu$, there is a natural number $T^* \leq T$ and an action sequence $(a^1, \cdots, a^{T^*})$ such that $\Pr(\mu^{T^*} = \tilde{\mu}|\mu, a^1, \cdots, a^{T^*}) \geq \pi^*$ for some $\tilde{\mu}$ with $\sum_{\omega \in \Omega^*} \tilde{\mu}(\omega) \geq 1 - \varepsilon$.

Asymptotic accessibility of $\Omega^*$ requires that given any initial prior $\mu$, there is an action sequence $(a^1, \cdots, a^{T^*})$ so that the posterior belief can approximate a belief whose support is $\Omega^*$. Here the length $T^*$ of the action sequence may depend on the initial prior, but it must be uniformly bounded by some natural number $T$.

As argued above, each singleton set $\{\omega\}$ is asymptotically accessible in Example 2. In this example, the state changes over time, and thus if the initial prior puts probability close to zero on $\omega$, then the posterior belief in the second period will put probability close to one on $\omega$. This ensures that there is a uniform bound $T$ on the length $T^*$ of the action sequence.

Similarly, the set $\{\omega_2\}$ in Example 3 is asymptotically accessible, although it is not globally accessible. To see this, suppose that the machine is operated every period. Then $\omega_2$ is the unique absorbing state, and hence there is some $T$ such that the posterior belief after period $T$ attaches a very high probability on $\omega_2$ regardless of the initial prior (at least after some signal realizations). This is precisely asymptotic accessibility of $\{\omega_2\}$.

The definition of asymptotic accessibility is stated in terms of the posterior belief, not primitives. However, as explained above, in some applications, checking whether each set $\Omega^*$ is asymptotically accessible is a relatively simple task. Also, as will be explained in Lemma 14 below, there is a simple sufficient condition for asymptotic connectedness.

Note that $\Omega^*$ is asymptotically accessible whenever it is globally accessible, and hence the whole state space $\Omega^* = \Omega$ is always asymptotically accessible. Let $\tilde{\mathscr{O}}$ be the set of all asymptotically accessible sets. Next, we give the definition of asymptotic transience.

**Definition 16.** A singleton set $\{\omega\} \notin \tilde{\mathscr{O}}$ is *asymptotically transient* if there is $\tilde{\pi}^*$ such that for any $\varepsilon > 0$, there is a natural number $T$ such that for each pure strategy profile $s \in S^*$, there is an asymptotically accessible set $\Omega^* \in \tilde{\mathscr{O}}$, a natural number $T^* \leq T$, and a belief $\tilde{\mu}$ such that $\Pr(\mu^{T^*} = \tilde{\mu}|\omega, s) > 0$, $\sum_{\tilde{\omega} \in \Omega^*} \tilde{\mu}(\tilde{\omega}) \geq 1 - \varepsilon$, and $\tilde{\mu}(\tilde{\omega}) \geq \tilde{\pi}^*$ for all $\tilde{\omega} \in \Omega^*$.

In words, asymptotic transience of $\{\omega\}$ requires that if the support of the current belief is $\{\omega\}$, then regardless of the future play, with positive probability, the posterior belief $\mu^{T^*} = \tilde{\mu}$ approximates a belief whose support $\Omega^*$ is globally accessible. Asymptotic transience is weaker than transience in two respects. First, a global accessible set $\Omega^*$ in the definition of transience is replaced with an asymptotic transient set $\Omega^*$. Second, the support of the posterior $\tilde{\mu}$ is not necessarily identical with $\Omega^*$; it is enough if $\tilde{\mu}$ assigns probability at least $1 - \varepsilon$ on $\Omega^*$.[17]

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

**Definition 17.** A stochastic game is *asymptotically connected* if each singleton set $\{\omega\}$ is asymptotically accessible or asymptotically transient.

The definition of asymptotic connectedness is very similar to connectedness; the only difference is that global accessibility in the definition of connectedness is replaced with asymptotic accessibility. Asymptotic connectedness is weaker than

---

[17]Asymptotic transience is different from asymptotic transience in that the last requirement, $\tilde{\mu}(\tilde{\omega}) \geq \tilde{\pi}^*$, ensures that the posterior belief $\tilde{\mu}$ is not close to the boundary of $\triangle \Omega^*$. So the posterior belief $\tilde{\mu}$ must be really an interior point of $\triangle \Omega^*$. Note that this requirement is automatically satisfied in $\tilde{\mu}$ in the definition of transience, by setting $\tilde{\pi}^* = \overline{\pi}^{2^{|\Omega|}}$.

connectedness, and indeed, the game is asymptotically connected but not connected in Examples 2 and 3. Also, asymptotic connectedness is satisfied in an important class of games, as the following lemma shows. Let $\pi_Y^\omega(a) = (\pi_Y^\omega(y|a))_{y \in Y}$ be the marginal distribution of $y$ given $\omega$ and $a$.

**Lemma 14.** *The game is asymptotically connected if the following properties hold:*

- *The state evolution is irreducible.*

- *Signals do not reveal the current state, that is, for each $a$ and $y$, if $\pi_Y^\omega(y|a) > 0$ for some $\omega$, then $\pi_Y^{\tilde{\omega}}(y|a) > 0$ for all $\tilde{\omega}$.*

- *Signals do not reveal the next state, that is, for each $\omega$, $\tilde{\omega}$, $\hat{\omega}$, $a$, $\tilde{y}$, and $\hat{y}$, if $\pi^\omega(\tilde{y}, \tilde{\omega}|a) > 0$ and $\pi^\omega(\hat{y}, \hat{\omega}|a) > 0$, then $\pi^\omega(\tilde{y}, \hat{\omega}|a) > 0$.*

- *For each $a$, the signal distributions $\{\pi_Y^\omega(a), \omega \in \Omega\}$ are linearly independent.*

The lemma shows that if the state evolution is irreducible and $|Y| \geq |\Omega|$, then for "generic" signal structures, asymptotic connectedness is satisfied. The proof of the lemma can be found in Appendix E.

As shown in Lemma 5, connectedness ensures invariance of the feasible payoff set, which plays an important role in our equilibrium analysis. The following lemma shows that the invariance result remains valid as long as the game is asymptotically connected.[18] The proof can be found in Appendix E.

**Lemma 15.** *Suppose that the game is asymptotically connected. Then for each $\varepsilon > 0$, there is $\overline{\delta} \in (0, 1)$ such that the following properties hold:*

*(i) For any $\lambda \in \Lambda$, $\delta \in (\overline{\delta}, 1)$, $\mu$, and $\tilde{\mu}$,*

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

---

[18]However, unlike Lemma 5, we do not know the rate of convergence, and in particular, we do not know if we can replace $\varepsilon$ in the lemma with $O(1 - \delta)$.

*(ii)* *For any $\lambda \in \Lambda$ and $\delta \in (\overline{\delta}, 1)$, there is a pure strategy profile $s \in S^*$ such that for all $\mu$,*

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

With this lemma, it is easy to see that Proposition 2 is valid for asymptotically connected stochastic games. That is, the folk theorem holds as long as the game is asymptotically connected.

In the same spirit, we can show that strong connectedness is stronger than necessary for the invariance of the limit minimax payoff. Indeed, the following condition, *asymptotic strong connectedness*, is weaker than strong connectedness but sufficient for the invariance result. The proof is omitted, as it is a combination of those of Lemmas 12 and 15.

**Definition 18.** A non-empty subset $\Omega^* \subseteq \Omega$ is *asymptotically robustly accessible* if there is $\tilde{\pi}^* > 0$ such that for any $\varepsilon > 0$, there is a natural number $T$ and $\pi^* > 0$ such that for any initial prior $\mu$ and for any $i$, there is an action sequence $(a_{-i}^1, \cdots, a_{-i}^T)$ such that for any player $i$'s strategy $s_i$, there is a natural number $T^* \leq T$ and a belief $\tilde{\mu}$ such that $\Pr(\mu^{T+1} = \tilde{\mu} | \mu, a^1, \cdots, a^T) \geq \pi^*$, $\sum_{\omega \in \Omega^*} \tilde{\mu}(\omega) \geq 1 - \varepsilon$, and $\tilde{\mu}(\omega) \geq \tilde{\pi}^*$ for all $\omega \in \Omega^*$.

**Definition 19.** A subset $\Omega^* \subseteq \Omega$ is *asymptotically strongly transient* if it is not asymptotically robustly accessible and for any $\varepsilon > 0$, there is a natural number $T$ and $\pi^* > 0$ such that for any pure strategy profile $s \in S^*$ and for any $\mu$ whose support is $\Omega^*$, there is an asymptotically accessible set $\Omega^* \in \tilde{\mathcal{O}}$, a natural number $T^* \leq T$, and a belief $\tilde{\mu}$ such that $\Pr(\mu^{T+1} = \tilde{\mu} | \mu, s) \geq \pi^*$ and $\sum_{\tilde{\omega} \in \Omega^*} \tilde{\mu}(\tilde{\omega}) \geq 1 - \varepsilon$,

**Definition 20.** The game is *asymptotically strongly connected* if the following conditions hold:

(i) Each non-empty subset $\Omega^* \subseteq \Omega$ is asymptotically robustly accessible or asymptotically strongly transient.

(ii) There is $\tilde{\pi}^* > 0$ such that for any $\varepsilon > 0$, there is a natural number $T$ such that for each $\omega$, for each $\mu$ whose support is $\Omega$, and for each pure strategy profile $s \in S^*$, there is a natural number $T^* \leq T$ and a history $h^{T^*}$ such that the probability of $h^T$ given $(\omega, s)$ is positive, and such that the posterior

45

belief $\tilde{\mu}$ induced by $\omega$ and $h^T$ satisfies $\sum_{\tilde{\omega} \in \Omega^*} \tilde{\mu}(\tilde{\omega}) \geq 1 - \varepsilon$ and $\tilde{\mu}(\tilde{\omega}) \geq \tilde{\pi}^*$ for each $\tilde{\omega} \in \Omega^*$, where $\Omega^*$ is the support of the posterior induced by $\mu$ and $h^T$.

# Appendix D: Proof of Proposition 2

Here we prove the folk theorem for strongly connected stochastic games.

**Definition 21.** A subset $W$ of $\mathbf{R}^N$ 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 $W$.[19]

Pick an arbitrary smooth subset $W$ of the interior of $V^*$. From Proposition 1, it is sufficient to show that each target payoff $v \in W$ is stochastically enforceable using continuation payoffs $w$ chosen from the set $W$. As argued by FLM, a key step is to show enforceability of boundary points $v$ of $W$. Indeed, if we can show that all boundary points $v$ of $W$ are enforceable, then it is relatively easy to check that all interior points of $W$ are also enforceable.

So pick an arbitrary boundary point $v \in W$. We want to prove that $v$ is enforceable using continuation payoffs $w(h^t)$ in $W$ for all sufficiently large $\delta$. A sufficient condition for this is that there are some real numbers $\varepsilon > 0$ and $K > 0$ such that $v$ is enforceable using continuation payoffs in the set $G$ in Figure 7 for all $\delta$. Formally, letting $\lambda$ be a unit normal to $W$ at $v$, the set $G$ in the figure refers to the set of all payoff vectors $\tilde{v}$ such that $\lambda \cdot v \geq \lambda \cdot \tilde{v} + (1 - \delta)\varepsilon$ and such that $\tilde{v}$ is within $(1 - \delta)K$ of $v$. Note that the set $G$ depends on the discount factor $\delta$, but if $W$ is smooth, this set $G$ is always in the interior of the set $W$ for any $\delta$ close to one. (See Fudenberg and Levine (1994) for a formal proof.) Thus if continuation payoffs $w(h^t)$ are chosen from the set $G$ for each $\delta$, they are in the set $W$, as desired.

The concept of *uniform decomposability* of Fudenberg and Yamamoto (2011b) formalizes the above idea, and here we extend it to our setup. Given any $v$, $\lambda$, $\varepsilon > 0$, $K > 0$, and $\delta \in (0,1)$, let $G_{v,\lambda,\varepsilon,K,\delta}$ be the set of all $\tilde{v}$ such that $\lambda \cdot v \geq \lambda \cdot \tilde{v} + (1 - \delta)\varepsilon$ and such that $\tilde{v}$ is within $(1 - \delta)K$ of $v$. When $v$ is a boundary

---

[19]A sufficient condition for each point on the boundary of $W$ to have a unique unit normal is that the boundary is a $C^2$-submanifold of $\mathbf{R}^N$.
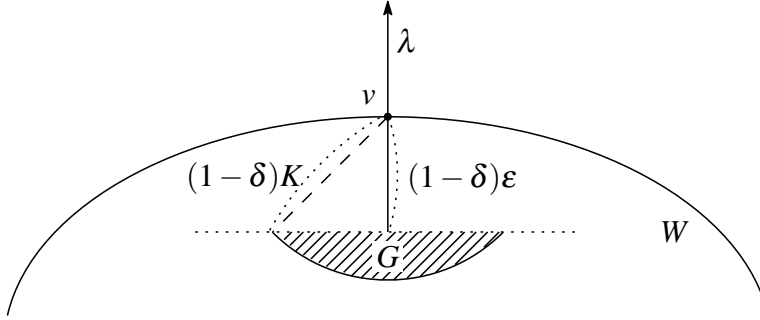
Figure 7: Continuation payoffs are chosen from $G$.

point of $W$ and $\lambda$ is the corresponding unit normal, this set $G_{v,\lambda,\varepsilon,K,\delta}$ is exactly the set $G$ in Figure 7.

**Definition 22.** A subset $W$ of $\boldsymbol{R}^N$ is *uniformly decomposable with respect to p* if there are $\varepsilon > 0$, $K > 0$, and $\overline{\delta} \in (0,1)$ such that for each $v \in W$, $\delta \in (\overline{\delta},1)$, $\lambda \in \Lambda$, and $\mu$, there are $s \in S$ and $w : H \to G_{v,\lambda,\varepsilon,K,\delta}$ such that $(s,v)$ is stochastically enforceable with respect to $(\delta,\mu,p)$ using $w$.

In words, uniform decomposability requires that any target payoff $v \in W$ is stochastically enforceable using continuation payoffs in the set $G_{v,\lambda,\varepsilon,K,\delta}$. The following lemma shows uniform decomposability is sufficient for the set $W$ to be self-generating for sufficiently large $\delta$.[20] The proof is similar to Fudenberg and Yamamoto (2011b) and hence omitted.

**Lemma 16.** *Suppose that a smooth and bounded subset $W$ of $\boldsymbol{R}^N$ is uniformly decomposable with respect to p. Then there is $\overline{\delta} \in (0,1)$ such that for any payoff vector $v \in W$, for any $\delta \in (\overline{\delta},1)$, and for any $\mu$, there is a sequential equilibrium which yields the payoff v.*

In what follows, we will show that any smooth $W$ in the interior of $V^*$ is uniformly decomposable. A direction $\lambda$ is *regular* if it has at least two non-zero components, and is *coordinate* if it has exactly one non-zero component. The next lemma is an extension of Theorem 5.1 of FLM. Very roughly speaking, it shows that a boundary point $v$ of $W$ with a regular unit normal vector $\lambda$ is enforceable using continuation payoffs in the set $G$. The proof can be found in Appendix E.

---

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

**Lemma 17.** *For each regular direction $\lambda$ and for each $p \in (0,1)$, there is $K > 0$ such that for each $\mu$, for each $s \in S^*$, for each $\delta \in (0,1)$, and for each $v \in V$, there is $w$ such that*

   *(i) $(s,v)$ is stochastically enforceable with respect to $(\delta, \mu, p)$ by $w$,*

   *(ii) $\lambda \cdot w(h^t) = \lambda \cdot v - \frac{1-\delta}{(1-p)\delta}(\lambda \cdot v^\mu(p\delta, s) - \lambda \cdot v)$ for all $t$ and $h^t$, and*

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

This lemma applies to all target payoffs $v \in V$, but for the sake of the exposition, let $v$ be a boundary point of $W$ with a regular unit normal $\lambda$. Assume also that $p$ and $\delta$ are close to one. Let $s$ be a strategy profile approximating the boundary of $V$ toward direction $\lambda$ when the discount factor is $p\delta$. Since $v$ is in the interior of $V^*$, we have $\lambda \cdot v^\mu(p\delta, s) - \lambda \cdot v = l > 0$. Clause (i) says that such a pair $(s,v)$ is enforceable using some continuation payoffs $w$. Also, clauses (ii) and (iii) ensure that these continuation payoffs are in the set $G$ in Figure 7, by letting $\varepsilon = l$. Indeed, clause (ii) reduces to $\lambda \cdot w(h^t) = \lambda \cdot v - \frac{1-\delta}{(1-p)\delta}l$, and thus all the continuation payoffs must be in the shaded area in Figure 8. (In particular, clause (ii) says that continuation payoffs can be chosen from a hyperplane orthogonal to $\lambda$. This is reminiscent of the idea of "utility transfers across players" of FLM.)
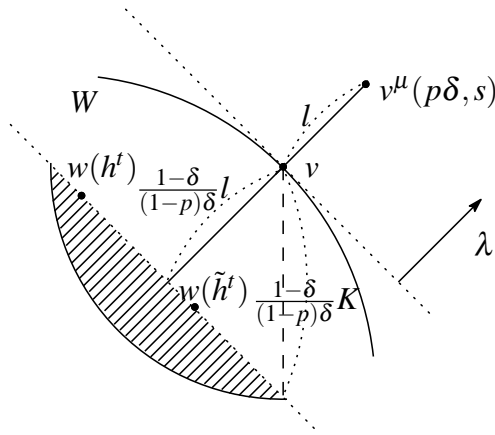


Figure 8: Continuation Payoffs for Regular Direction $\lambda$

Note that in the above lemma, the rate of convergence (the constant $K$ in the lemma) depends on direction $\lambda$; indeed this constant $K$ can become arbitrarily

large as $\lambda$ approaches a coordinate vector. However, in Lemma 20 we will extract a finite set of direction vectors so the dependence of the constant on $\lambda$ will not cause problems.

The next lemma extends Lemma 5.2 of FLM, which considers enforceability for coordinate directions. The proof can be found in Appendix E.

**Lemma 18.** *For each $p \in (0,1)$, there is $K > 0$ such that for each $\mu$, $s_{-i} \in S_{-i}^*$, $v \in V$, $\delta$, and $s_i \in \arg\max_{\tilde{s}_i \in S_i^*} v_i^\mu(p\delta, \tilde{s}_i, s_{-i})$, there is $w$ such that*

*(i) $(s,v)$ is stochastically enforceable with respect to $(\delta, \mu, p)$ by $w$,*

*(ii) $w_i(h^t) = v_i - \frac{1-\delta}{(1-p)\delta}(v_i^\mu(p\delta, s) - v_i)$ for all $t$ and $h^t$, and*

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

To see the implication of the above lemma, let $v$ be a boundary point of $W$ such that the corresponding unit normal is a coordinate vector with $\lambda_i = 1$. That is, let $v$ be a payoff vector which gives the best possible payoff to player $i$ within the set $W$. Let $s$ be a strategy profile which approximates the best payoff for player $i$ within the feasible payoff set $V$, so that $v_i^\mu(p\delta, s) - v_i = l > 0$. Clause (i) says that such a pair $(s,v)$ is enforceable using some continuation payoffs $w$. Clauses (ii) and (iii) ensure that continuation payoffs are in the shaded area in Figure 9.



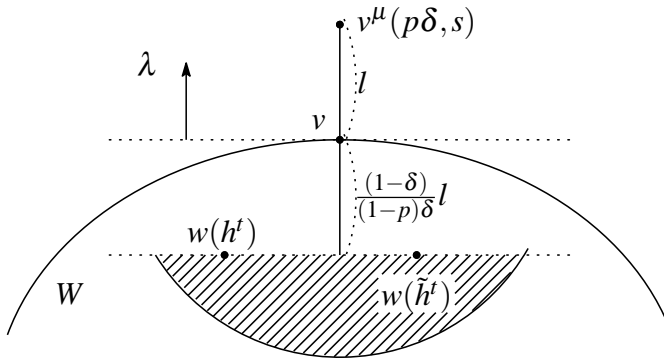Figure 9: Continuation Payoffs for $\lambda$ with $\lambda_i = 1$

Likewise, let $v$ be a boundary point of $W$ such that the corresponding unit normal is a coordinate vector with $\lambda_i = -1$. Consider $s$ which approximates the limit minimax payoff to player $i$ so that $v_i - v_i^\mu(p\delta, s) = l > 0$. Then clauses (ii) and (iii) ensure that the continuation payoffs are in the shaded area in Figure 10.
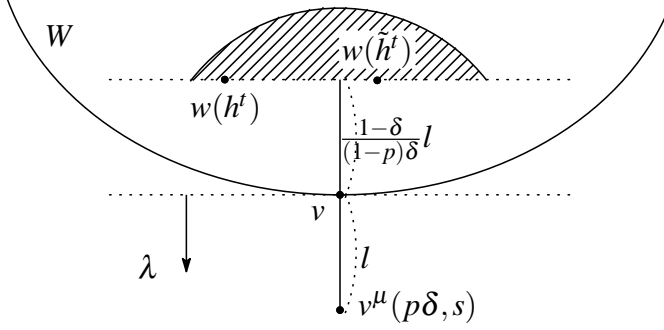
Figure 10: Continuation Payoffs for $\lambda$ with $\lambda_i = -1$

To prove the above lemma, it is important that $s_i$ is a best reply to $s_{-i}$ given $\mu$ and $p\delta$; this property ensures that player $i$'s incentive constraints are satisfied by a constant continuation function $w_i$, so that clause (ii) is satisfied.

The next lemma guarantees that for $p$ close to one, there are strategy profiles which approximate the boundary of $V^*$ in the stochastic game with the discount factor $p$. Intuitively, these strategies are the ones we have considered in Figures 8 through 10. The proof is found in Appendix E.

**Lemma 19.** *Suppose that the game is connected. Then for any smooth subset $W$ of the interior of $V^*$, there are $\varepsilon > 0$ and $p \in (0,1)$ such that the following properties hold:*

*(i) For every regular $\lambda$ and $\mu$, there is a strategy profile $s \in S^*$ such that*

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

*(ii) For each $i$ and for each $\mu$, there is a strategy $s_{-i} \in S^*_{-i}$ such that*

$$\max_{s_i \in S_i} v_i^\mu(p,s) > \max_{v \in W} v_i + \varepsilon.$$

*(iii) For each $i$ and for each $\mu$, there is a strategy $s_{-i} \in S^*_{-i}$ such that*

$$\max_{s_i \in S_i} v_i^\mu(p,s) < \min_{v \in W} v_i - \varepsilon.$$

Now we are in a position to prove uniform decomposability of a smooth subset $W$ of the interior of $V^*$. The proof can be found in Appendix E.

50

**Lemma 20.** *For any smooth subset $W$ of the interior of $V^*$, there is $p \in (0,1)$ such that $W$ is uniformly decomposable with respect to $p$.*

This lemma, together with Lemma 16, establishes Proposition 2.

**Remark 3.** When the game is not (asymptotically) connected, the limit feasible payoff set $V^\mu$ may depend on $\mu$. However, even in this case, the following result holds. Let $\underline{v}_i^\mu$ be the limit superior of player $i$'s minimax payoff with the initial prior $\mu$, as $\delta \to 1$. Let $V^{*\mu}$ be the set of $v \in V^\mu$ such that $v_i \geq \underline{v}_i^\mu$ for all $i$, and let $V^*$ be the intersection of $V^{*\mu}$ over all $\mu$. Then for any interior point $v$ of $V^*$ and any initial prior $\mu$, there is a sequential equilibrium with the payoff $v$ when $\delta$ is large enough. That is, if a payoff vector $v$ is feasible and individually rational regardless of the initial prior $\mu$, then it is achieved by some equilibrium.

# Appendix E: Proofs of Lemmas

## E.1   Proof of Lemma 3

It is obvious that if $\omega$ is transient, then $\{\omega\}$ is transient. Fix an arbitrary $\{\omega\}$ such that $\omega$ is globally accessible yet $\{\omega\}$ is not globally accessible. It is sufficient to show that $\{\omega\}$ is transient. To do so, fix arbitrary $a^*$ and $y^*$ such that $\pi_Y^\omega(y^*|a^*) > 0$, and let $\Omega^*$ be the set of all $\tilde{\omega}$ such that $\pi^\omega(y^*, \tilde{\omega}|a^*) > 0$. It is sufficient to show that $\Omega^*$ is globally accessible, as it implies that $\{\omega\}$ is transient.

Fix an arbitrary initial prior $\mu$, and take an arbitrary $\omega^*$ such that $\mu(\omega^*) \geq \frac{1}{|\Omega|}$. Since the state evolution is irreducible, there is an action sequence $(a^1, \cdots, a^T)$ with $T \leq |\Omega|$ such that the probability of the state in period $T+1$ being $\omega$ is positive conditional on the initial state $\omega^*$ and the action sequence $(a^1, \cdots, a^T)$. Suppose that, given the initial prior $\mu$, players play this action sequence until period $T$ and then $a^*$ in period $T+1$. Then in period $T+1$, the true state can be $\omega$, so that with positive probability, the signal $y^*$ realizes. Then the support of the posterior belief is $\Omega^*$, since the signal $y$ reveals that the current state is $\omega$. Note that the probability of this event is at least

$$\mu(\omega^*)\Pr(\omega^{T+1} = \omega|\omega^*, a^1, \cdots, a^T)\pi_Y^\omega(y^*|a^*) \geq \frac{\overline{\pi}^{T+1}}{|\Omega|}.$$

51

Since the lower bound $\frac{\overline{\pi}^{T+1}}{|\Omega|} > 0$ is the same across all initial priors, the set $\Omega$ is globally accessible.

## E.2  Proof of Lemma 5

Fix $\delta$ and $\lambda$. For each $\mu$, let $s^{\mu}$ be a pure-strategy profile which solves $\max_{s \in S} \lambda \cdot v(\delta, s)$. Without loss of generality we assume $s^{\mu} \in S^*$.

Note that given each initial prior $\mu$, the score is equal to $\lambda \cdot v^{\mu}(\delta, s^{\mu})$. The following lemma shows that the score is convex with respect to $\mu$.

**Lemma 21.** $\lambda \cdot v^{\mu}(\delta, s^{\mu})$ *is convex with respect to* $\mu$.

*Proof.* Take $\mu$ and $\tilde{\mu}$, and take an arbitrary $\kappa \in (0,1)$. Let $\hat{\mu} = \kappa\mu + (1 - \kappa)\tilde{\mu}$. Then we have

$$\lambda \cdot v^{\hat{\mu}}(\delta, s^{\hat{\mu}}) = \kappa\lambda \cdot v^{\mu}(\delta, s^{\hat{\mu}}) + (1 - \kappa)\lambda \cdot v^{\tilde{\mu}}(\delta, s^{\hat{\mu}})$$
$$\leq \kappa\lambda \cdot v^{\mu}(\delta, s^{\mu}) + (1 - \kappa)\lambda \cdot v^{\tilde{\mu}}(\delta, s^{\tilde{\mu}}),$$

which implies the convexity. *Q.E.D.*

Since $\lambda \cdot v^{\mu}(\delta, s^{\mu})$ is convex, there is $\omega$ such that

$$\lambda \cdot v^{\omega}(\delta, s^{\omega}) \geq \lambda \cdot v^{\mu}(\delta, s^{\mu}) \tag{9}$$

for all $\mu$. (Here the superscript $\omega$ refers to the initial prior $\mu$ which assigns probability one to the state $\omega$.) Pick such a $\omega$.

For each $T$ and signal sequence $(y^1, \cdots, y^T)$, let $\pi(y^1, \cdots, y^T)$ denote the probability that the sequence $(y^1, \cdots, y^T)$ appears when the initial state is $\omega$ and players play $s^{\omega}$. Since $s^{\omega}$ is pure and the initial belief puts probability one on $\omega$, we have $\pi(y^1, \cdots, y^T) \geq \overline{\pi}^T$ for all $(y^1, \cdots, y^T)$ with $\pi(y^1, \cdots, y^T) > 0$. For each such $(y^1, \cdots, y^T)$, let $\mu(y^1, \cdots, y^T)$ be the posterior belief $\mu^{T+1}$ given the initial state $\omega$, the strategy profile $s^{\omega}$, and the signal sequence $(y^1, \cdots, y^T)$. The following lemma bounds the difference between the maximal score $\lambda \cdot v^{\omega}(\delta, s^{\omega})$ and the score when the initial prior is $\mu = \mu(y^1, \cdots, y^T)$. Let $C(T) = \frac{2\overline{g}}{\overline{\pi}^T}$

**Lemma 22.** *For each $T$ and $(y^1, \cdots, y^T)$ with $\pi(y^1, \cdots, y^T) > 0$,*

$$\left| \lambda \cdot v^{\omega}(\delta, s^{\omega}) - \lambda \cdot v^{\mu(y^1, \cdots, y^T)}(\delta, s^{\mu(y^1, \cdots, y^T)}) \right| \leq \frac{1 - \delta^T}{\delta^T} C(T).$$

*Proof.* From the principle of optimality, we have

$$\lambda \cdot v^\omega(\delta, s^\omega) = (1-\delta) \sum_{t=1}^{T} \delta^{t-1} E[\lambda \cdot g^{\omega^t}(a^t) | \omega^1 = \omega, s^\omega]$$
$$+ \delta^T \sum_{(y^1, \cdots, y^T) \in Y^T} \pi(y^1, \cdots, y^T) \lambda \cdot v^{\mu(y^1, \cdots, y^T)}(\delta, s^{\mu(y^1, \cdots, y^T)}).$$

Since $(1-\delta) \sum_{t=1}^{T} \delta^{t-1} E[\lambda \cdot g^{\omega^t}(a^t) | \omega^1 = \omega, s^\omega] \leq (1-\delta^T) \bar{g}$, we have

$$\lambda \cdot v^\omega(\delta, s^\omega) \leq (1-\delta^T) \bar{g}$$
$$+ \delta^T \sum_{(y^1, \cdots, y^T) \in Y^T} \pi(y^1, \cdots, y^T) \lambda \cdot v^{\mu(y^1, \cdots, y^T)}(\delta, s^{\mu(y^1, \cdots, y^T)}).$$

This inequality, together with (9), implies that

$$\lambda \cdot v^\omega(\delta, s^\omega) \leq (1-\delta^T) \bar{g} + \delta^T \pi(y^1, \cdots, y^T) \lambda \cdot v^{\mu(y^1, \cdots, y^T)}(\delta, s^{\mu(y^1, \cdots, y^T)})$$
$$+ \delta^T (1 - \pi(y^1, \cdots, y^T)) \lambda \cdot v^\omega(\delta, s^\omega)$$

for each $(y^1, \cdots, y^T)$. Arranging, we have

$$\lambda \cdot v^\omega(\delta, s^\omega) - \lambda \cdot v^{\mu(y^1, \cdots, y^T)}(\delta, s^{\mu(y^1, \cdots, y^T)}) \leq \frac{(1-\delta^T)(\bar{g} - \lambda \cdot v^\omega(\delta, s^\omega))}{\delta^T \pi(y^1, \cdots, y^T)}.$$

for each $(y^1, \cdots, y^T)$ such that $\pi(y^1, \cdots, y^T) > 0$. Since $\pi(y^1, \cdots, y^T) \geq \bar{\pi}^T$,

$$\lambda \cdot v^\omega(\delta, s^\omega) - \lambda \cdot v^{\mu(y^1, \cdots, y^T)}(\delta, s^{\mu(y^1, \cdots, y^T)}) \leq \frac{(1-\delta^T)(\bar{g} - \lambda \cdot v^\omega(\delta, s^\omega))}{\delta^T \bar{\pi}^T}.$$

Using (9) and $\lambda \cdot v^\omega(\delta, s^\omega) \geq -\bar{g}$, the result follows. *Q.E.D.*

Since the game is connected, $\{\omega\}$ is either globally accessible or transient. When it is transient, there is a natural number $T \leq 2^{|\Omega|}$ and a signal sequence $(y^1, \cdots, y^T)$ such that if the initial state is $\omega$ and players play $s^\omega$, then the signal sequence $(y^1, \cdots, y^T)$ appears with positive probability and the support of the resulting posterior belief $\mu(y^1, \cdots, y^T)$ is a globally accessible set $\Omega^* \in \mathscr{O}$. Take such $T$, $(y^1, \cdots, y^T)$, and $\Omega^*$. To simplify our notation, let $\mu^* = \mu(y^1, \cdots, y^T)$. Since $\mu^*$ is a belief induced by the initial state $\omega$ in period $T+1$, we have $\mu^*(\tilde{\omega}) \geq \bar{\pi}^T$ for each $\tilde{\omega} \in \Omega^*$. Let $C = \frac{C(T)}{\bar{\pi}^T}$.

When $\{\omega\}$ is globally accessible, let $\Omega^* = \{\omega\}$, $T = 0$, $\mu^* \in \triangle \Omega$ with $\mu^*(\omega) = 1$ and $C = 2\bar{g}$.

Let $s^* = s^{\mu^*}$, that is, $s^*$ is the strategy profile which achieves the score when the initial prior is $\mu^*$. Lemma 22 shows that this strategy $s^*$ approximates the maximal score $\lambda \cdot v^\omega(\delta, s^\omega)$ when the initial prior is $\mu^*$. The following lemma shows that this strategy $s^*$ can still approximate the maximal score even if the initial prior $\mu^*$ is replaced with some state $\tilde{\omega}$ in the set $\Omega^*$.

**Lemma 23.** *For each $\tilde{\omega} \in \Omega^*$, we have*

$$\left| \lambda \cdot v^\omega(\delta, s^\omega) - \lambda \cdot v^{\tilde{\omega}}(\delta, s^*) \right| \leq \frac{1 - \delta^T}{\delta^T} C.$$

*Proof.* When $\{\omega\}$ is globally accessible, $\tilde{\omega} \in \Omega^*$ implies $\tilde{\omega} = \omega$ so that the lemma obviously follows. Hence we focus on the case in which $\{\omega\}$ is not globally accessible. Take $\tilde{\omega} \in \Omega^*$ arbitrarily. Then we have

$$\lambda \cdot v^{\mu^*}(\delta, s^*) = \sum_{\hat{\omega} \in \Omega^*} \mu^*[\hat{\omega}] \lambda \cdot v^{\hat{\omega}}(\delta, s^*)$$

$$\leq \mu^*[\tilde{\omega}] \lambda \cdot v^{\tilde{\omega}}(\delta, s^*) + \sum_{\hat{\omega} \neq \tilde{\omega}} \mu^*[\hat{\omega}] \lambda \cdot v^{\hat{\omega}}(\delta, s^{\hat{\omega}})$$

Using (9),

$$\lambda \cdot v^{\mu^*}(\delta, s^*) \leq \mu^*[\tilde{\omega}] \lambda \cdot v^{\tilde{\omega}}(\delta, s^*) + (1 - \mu^*[\tilde{\omega}]) \lambda \cdot v^\omega(\delta, s^\omega).$$

Arranging,

$$\mu^*[\tilde{\omega}](\lambda \cdot v^\omega(\delta, s^\omega) - \lambda \cdot v^{\tilde{\omega}}(\delta, s^*)) \leq \lambda \cdot v^\omega(\delta, s^\omega) - \lambda \cdot v^{\mu^*}(\delta, s^*).$$

Note that the term in the parenthesis in the left-hand side is non-negative. Similarly the right-hand side is non-negative. Hence

$$\left| \lambda \cdot v^\omega(\delta, s^\omega) - \lambda \cdot v^{\tilde{\omega}}(\delta, s^*) \right| \leq \frac{|\lambda \cdot v^\omega(\delta, s^\omega) - \lambda \cdot v^{\mu^*}(\delta, s^*)|}{\mu^*[\tilde{\omega}]}.$$

Since we have $\mu^*[\tilde{\omega}] \geq \overline{\pi}^T$,

$$\left| \lambda \cdot v^\omega(\delta, s^\omega) - \lambda \cdot v^{\tilde{\omega}}(\delta, s^*) \right| \leq \frac{|\lambda \cdot v^\omega(\delta, s^\omega) - \lambda \cdot v^{\mu^*}(\delta, s^*)|}{\overline{\pi}^T}.$$

Using Lemma 22, the result follows.                                                  *Q.E.D.*

The following lemma shows that the same result holds even if the initial state $\tilde{\omega} \in \Omega^*$ is replaced with some prior $\mu$ whose support is $\Omega^*$.

**Lemma 24.** *For each $\mu$ such that $\mu(\tilde{\omega}) = 0$ for all $\tilde{\omega} \notin \Omega^*$, we have*

$$|\lambda \cdot v^\omega(\delta, s^\omega) - \lambda \cdot v^\mu(\delta, s^*)| \leq \frac{1 - \delta^T}{\delta^T} C.$$

*Proof.* We have

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

Subtracting both sides from $\lambda \cdot v^\omega(\delta, s^\omega)$,

$$\lambda \cdot v^\omega(\delta, s^\omega) - \lambda \cdot v^{\tilde{\mu}}(\delta, s^*) = \sum_{\tilde{\omega} \in \Omega^*} \tilde{\mu}(\tilde{\omega})(\lambda \cdot v^\omega(\delta, s^\omega) - \lambda \cdot v^{\tilde{\omega}}(\delta, s^*)).$$

Then from Lemma 23, the result follows. *Q.E.D.*

Take $\pi^* > 0$ such that for each globally accessible set, $\pi^*$ satisfies the condition stated in Definition 3. (Such $\pi^*$ exists, since there are only finitely many subsets of $\Omega$.) Since $\Omega^*$ is globally accessible, given any initial prior $\mu$, there is a natural number $T(\mu) \leq 4^{|\Omega|}$ and an action sequence $\boldsymbol{a}(\mu) = (a^1(\mu), \cdots, a^{T(\mu)}(\mu))$ such that the support of the posterior belief at the beginning of period $T(\mu) + 1$ is a subset of $\Omega^*$ with probability at least $\pi^*$.

Now let $\mu^{**}$ be such that $\mu^{**}(\omega) = \frac{1}{|\Omega|}$ for each $\omega$. Given the initial prior $\mu^{**}$, consider the following strategy profile $s^{**}$:

- Let $\mu^{(1)} = \mu^{**}$.

- Players play the action sequence $\boldsymbol{a}(\mu^{(1)})$ for the first $T(\mu^{(1)})$ periods. Let $\mu^{(2)}$ be the posterior belief in period $T(\mu^{(1)}) + 1$.

- If the support of the posterior belief $\mu^{(2)}$ is a subset of $\Omega^*$, then players play $s^*$ in the rest of the game.

- If not, then players play $\boldsymbol{a}(\mu^{(2)})$ for the next $T(\mu^{(2)})$ periods. Let $\mu^{(3)}$ be the posterior belief after that.

- If the support of the posterior belief $\mu^{(3)}$ is a subset of $\Omega^*$, then players play $s^*$ in the continuation game.

- And so on.

Intuitively, for the first $T(\mu^{(1)})$ periods, players play an action sequence which can potentially induce a posterior belief whose support is a subset of $\Omega^*$. Let $\mu^{(2)}$ be the posterior belief, and if its support is indeed a subset of $\Omega^*$, they play the strategy profile $s^*$ in the continuation game. Lemma 24 guarantees that the score in the continuation payoff after the switch to $s^*$ approximates the maximal score $\lambda \cdot v^\omega(\delta, s^\omega)$. If the support of $\mu^{(2)}$ is not a subset of $\Omega^*$, once again players play an action sequence which can potentially induce a posterior belief whose support is a subset of $\Omega^*$. And they do the same thing over and over.

The following lemma shows that the score given the initial prior $\mu^{**}$ and the strategy profile $s^{**}$ approximates the maximal score $\lambda \cdot v^\omega(\delta, s^\omega)$ when $\delta$ is close to one. Let $\tilde{C} = \frac{4\overline{g}}{\pi^*}$.

**Lemma 25.** *We have*

$$\left| \lambda \cdot v^\omega(\delta, s^\omega) - \lambda \cdot v^{\mu^{**}}(\delta, s^{**}) \right| \le \frac{1 - \delta^T}{\delta^T} C + (1 - \delta^{4^{|\Omega|}}) \tilde{C}.$$

*Proof.* If $\frac{1-\delta^T}{\delta^T} C \ge \overline{g}$, then the result is obvious because we have $|\lambda \cdot v^\omega(\delta, s^\omega) - \lambda \cdot v^{\mu^{**}}(\delta, s^{**})| \le \overline{g}$. So in what follows, we assume that $\frac{1-\delta^T}{\delta^T} C < \overline{g}$,

Suppose that the initial prior is $\mu^{**}$ and players play the strategy profile $s^{**}$. By the definition of $s^{**}$, once the support of the posterior belief $\mu^{(k)}$ reaches a subset of $\Omega^*$ for some $k$, players switch their continuation play to $s^*$, and Lemma 24 ensures that the score in the continuation game is at least $\lambda \cdot v^\omega(\delta, s^\omega) - \frac{1-\delta^T}{\delta^T} C$. This implies that after the switch to $s^*$, the "per-period score" in the continuation game is at least $\lambda \cdot v^\omega(\delta, s^\omega) - \frac{1-\delta^T}{\delta^T} C$. To simplify the notation, let $v^*$ denote this payoff lower bound, that is, $v^* = \lambda \cdot v^\omega(\delta, s^\omega) - \frac{1-\delta^T}{\delta^T} C$. On the other hand, the per-period score before the switch to $s^*$ is at least $-2\overline{g}$, since $\lambda \cdot g^\omega(a) \le -2\overline{g}$ for all $\omega$ and $a$. For each $t$, let $\rho^t$ denote the probability that the switch to $s^*$ does not happen until the end of period $t$. Let $\rho^0 = 1$. Then from the above argument, we have

$$\lambda \cdot v^{\mu^{**}}(\delta, s^{**}) \ge (1 - \delta) \sum_{t=1}^{\infty} \delta^{t-1} \left\{ \rho^{t-1}(-2\overline{g}) + (1 - \rho^{t-1}) v^* \right\}. \qquad (10)$$

Recall that the length of the action sequence $\boldsymbol{a}(\mu)$ is at most $4^{|\Omega|}$ for each $\mu$. Recall also that the probability that the action sequence $\boldsymbol{a}(\mu)$ does not induce the switch to $s^*$ is at most $1 - \pi^*$ for each $\mu$. Hence we have

$$\rho^{n4^{|\Omega|}+k} \le (1 - \pi^*)^n$$

for each $n = 0, 1, \cdots$ and $k \in \{0, \cdots, 4^{|\Omega|} - 1\}$. This inequality, together with $-2\overline{g} \le v^*$, implies that

$$\rho^{n4^{|\Omega|}+k}(-2\overline{g}) + (1 - \rho^{n4^{|\Omega|}+k})v^* \ge (1 - \pi^*)^n(-2\overline{g}) + \{1 - (1 - \pi^*)^n\}v^*$$

for each $n = 0, 1, \cdots$ and $k \in \{0, \cdots, 4^{|\Omega|} - 1\}$. Plugging this inequality into (10), we obtain

$$\lambda \cdot v^{\mu^{**}}(\delta, s^{**}) \ge (1 - \delta) \sum_{n=1}^{\infty} \sum_{k=1}^{4^{|\Omega|}} \delta^{(n-1)4^{|\Omega|}+k-1} \left[ \begin{array}{c} -(1 - \pi^*)^{n-1}2\overline{g} \\ +\{1 - (1 - \pi^*)^{n-1}\}v^* \end{array} \right].$$

Since $\sum_{k=1}^{4^{|\Omega|}} \delta^{(n-1)4^{|\Omega|}+k-1} = \frac{\delta^{(n-1)4^{|\Omega|}}(1-\delta^{4^{|\Omega|}})}{1-\delta}$,

$$\begin{aligned} \lambda \cdot v^{\mu^{**}}(\delta, s^{**}) \ge & (1 - \delta^{4^{|\Omega|}}) \sum_{n=1}^{\infty} \delta^{(n-1)4^{|\Omega|}} \left[ \begin{array}{c} -(1 - \pi^*)^{n-1}2\overline{g} \\ +\{1 - (1 - \pi^*)^{n-1}\}v^* \end{array} \right] \\ = & -(1 - \delta^{4^{|\Omega|}}) \sum_{n=1}^{\infty} \{(1 - \pi^*)\delta^{4^{|\Omega|}}\}^{n-1}2\overline{g} \\ & + (1 - \delta^{4^{|\Omega|}}) \sum_{n=1}^{\infty} [(\delta^{4^{|\Omega|}})^{n-1} - \{(1 - \pi^*)\delta^{4^{|\Omega|}}\}^{n-1}]v^*. \end{aligned}$$

Since $\sum_{n=1}^{\infty} \{(1 - \pi^*)\delta^{4^{|\Omega|}}\}^{n-1} = \frac{1}{1-(1-\pi^*)\delta^{4^{|\Omega|}}}$ and $\sum_{n=1}^{\infty} (\delta^{4^{|\Omega|}})^{n-1} = \frac{1}{1-\delta^{4^{|\Omega|}}}$,

$$\lambda \cdot v^{\mu^{**}}(\delta, s^{**}) \ge -\frac{(1 - \delta^{4^{|\Omega|}})2\overline{g}}{1 - (1 - \pi^*)\delta^{4^{|\Omega|}}} + \frac{\delta^{4^{|\Omega|}}\pi^*}{1 - (1 - \pi^*)\delta^{4^{|\Omega|}}}v^*.$$

Subtracting both sides from $\lambda \cdot v^{\omega}(\delta, s^{\omega})$, we have

$$\begin{aligned} & \lambda \cdot v^{\omega}(\delta, s^{\omega}) - \lambda \cdot v^{\mu^{**}}(\delta, s^{**}) \\ & \le \frac{(1 - \delta^{4^{|\Omega|}})2\overline{g}}{1 - (1 - \pi^*)\delta^{4^{|\Omega|}}} + \frac{\delta^{4^{|\Omega|}}\pi^*(1 - \delta^T)C}{\{1 - (1 - \pi^*)\delta^{4^{|\Omega|}}\}\delta^T} - \frac{(1 - \delta^{4^{|\Omega|}})\lambda \cdot v^{\omega}(\delta, s^{\omega})}{1 - (1 - \pi^*)\delta^{4^{|\Omega|}}} \end{aligned}$$

Since $\lambda \cdot v^\omega(\delta, s^\omega) \geq -2\overline{g}$,

$$\lambda \cdot v^\omega(\delta, s^\omega) - \lambda \cdot v^{\mu^{**}}(\delta, s^{**})$$

$$\leq \frac{(1 - \delta^{4^{|\Omega|}})2\overline{g}}{1 - (1 - \pi^*)\delta^{4^{|\Omega|}}} + \frac{\delta^{4^{|\Omega|}}\pi^*(1 - \delta^T)C}{\{1 - (1 - \pi^*)\delta^{4^{|\Omega|}}\}\delta^T} + \frac{(1 - \delta^{4^{|\Omega|}})2\overline{g}}{1 - (1 - \pi^*)\delta^{4^{|\Omega|}}}$$

$$\leq \frac{(1 - \delta^{4^{|\Omega|}})4\overline{g}}{1 - (1 - \pi^*)} + \frac{\pi^*(1 - \delta^T)C}{\{1 - (1 - \pi^*)\}\delta^T}$$

$$= \frac{(1 - \delta^{4^{|\Omega|}})4\overline{g}}{\pi^*} + \frac{(1 - \delta^T)C}{\delta^T}$$

$$= \frac{1 - \delta^T}{\delta^T}C + (1 - \delta^{4^{|\Omega|}})\tilde{C}$$

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

The next lemma asserts that the strategy profile $s^{**}$ can approximate the maximal score regardless of the initial state $\tilde{\omega}$.

**Lemma 26.** *For each $\tilde{\omega} \in \Omega$, we have*

$$\left| \lambda \cdot v^\omega(\delta, s^\omega) - \lambda \cdot v^{\tilde{\omega}}(\delta, s^{**}) \right| \leq \frac{1 - \delta^T}{\delta^T}C|\Omega| + (1 - \delta^{4^{|\Omega|}})\tilde{C}|\Omega|.$$

*Proof.* The proof is very similar to that of Lemma 23. Replace $\Omega^*$, $\mu^*$, $s^*$, and $\overline{\pi}^T$ in the proof of Lemma 23 with $\Omega$, $\mu^{**}$, $s^{**}$, and $\frac{1}{|\Omega|}$, respectively. Then we have

$$\left| \lambda \cdot v^\omega(\delta, s^\omega) - \lambda \cdot v^{\tilde{\omega}}(\delta, s^{**}) \right| \leq |\Omega| \cdot |\lambda \cdot v^\omega(\delta, s^\omega) - \lambda \cdot v^{\mu^{**}}(\delta, s^{**})|$$

for each $\delta$ and $\tilde{\omega} \in \Omega$. Then from Lemma 25, the result follows. *Q.E.D.*

The next lemma implies that the strategy profile $s^{**}$ can approximate the maximal score regardless of the initial prior $\mu$.

**Lemma 27.** *For all $\mu \in \triangle\Omega$,*

$$|\lambda \cdot v^\omega(\delta, s^\omega) - \lambda \cdot v^\mu(\delta, s^{**})| \leq \frac{1 - \delta^T}{\delta^T}C|\Omega| + (1 - \delta^{4^{|\Omega|}})\tilde{C}|\Omega|.$$

*Proof.* The proof is very similar to that of Lemma 24. Replace $\Omega^*$ and $s^*$ in the proof of Lemma 24 with $\Omega$ and $s^{**}$, respectively. In the last step of the proof, use Lemma 26 instead of Lemma 23. *Q.E.D.*

From Lemma 27 and

$$\lambda \cdot v^{\omega}(\delta, s^{\omega}) \geq \lambda \cdot v^{\mu}(\delta, s^{\mu}) \geq \lambda \cdot v^{\mu}(\delta, s^{**}), \tag{11}$$

we have

$$|\lambda \cdot v^{\omega}(\delta, s^{\omega}) - \lambda \cdot v^{\mu}(\delta, s^{\mu})| \leq \frac{1 - \delta^T}{\delta^T} C|\Omega| + (1 - \delta^{4|\Omega|})\tilde{C}|\Omega|.$$

This implies the lemma, since $T \leq 2^{|\Omega|}$, $C = \frac{2\overline{g}}{\underline{\pi}^{T+1}}$ and $\tilde{C} = \frac{4\overline{g}}{\underline{\pi}^*}$ for any $\delta$ and $\lambda$.

## E.3   Proof of Lemma 6

We first show that global accessibility in Definition 9 implies the one in Definition 3. Take a set $\Omega^*$ which is globally accessible in the sense of Definition 9, and fix an arbitrarily initial prior $\mu$. Note that there is at least one $\omega$ such that $\mu(\omega) \geq \frac{1}{|\Omega|}$, so pick such $\omega$, and then pick $(a^1, \cdots, a^T)$ and $(y^1, \cdots, y^T)$ as stated in Definition 9. Suppose that the initial prior is $\mu$ and players play $(a^1, \cdots, a^T)$. Then clause (i) of Definition 9 guarantees that the signal sequence $(y^1, \cdots, y^T)$ appears with positive probability. Also, clause (ii) ensures that the support of the posterior belief $\mu^{T+1}$ after observing this signal sequence is a subset of $\Omega^*$, i.e., $\mu^{T+1}(\tilde{\omega}) = 0$ for all $\tilde{\omega} \notin \Omega^*$.[21] Note that the probability of this signal sequence $(y^1, \cdots, y^T)$ is at least

$$\mu(\omega) \Pr(y^1, \cdots, y^T | \omega, a^1, \cdots, a^T) \geq \frac{1}{|\Omega|} \overline{\pi}^T \geq \frac{1}{|\Omega|} \overline{\pi}^{4|\Omega|} > 0,$$

where $\Pr(y^1, \cdots, y^T | \omega, a^1, \cdots, a^T)$ denotes the probability of the signal sequence $(y^1, \cdots, y^T)$ given the initial state $\omega$ and the action sequence $(a^1, \cdots, a^T)$. This implies that global accessibility in Definition 9 implies the one in Definition 3, by letting $\pi^* \in (0, \frac{1}{|\Omega|} \overline{\pi}^{4|\Omega|})$.

Next, we show that the converse is true. Let $\Omega^*$ be a globally accessible set in the sense of Definition 3. Pick $\pi^* > 0$ as stated in Definition 3, and pick $\omega$ arbitrarily. Let $\mu$ be such that $\mu(\omega) = 1 - \frac{\pi^*}{2}$ and $\mu(\tilde{\omega}) = \frac{\pi^*}{2(|\Omega|-1)}$ for each $\tilde{\omega} \neq \omega$. Since $\Omega^*$ is globally accessible, we can choose an action sequence $(a^1, \cdots, a^T)$ and a belief $\tilde{\mu}$ whose support is included in $\Omega^*$ such that

$$\Pr(\mu^{T+1} = \tilde{\mu} | \mu, a^1, \cdots, a^T) \geq \pi^*. \tag{12}$$

---

[21]The reason is as follows. From Bayes' rule, $\mu^{T+1}(\tilde{\omega}) > 0$ only if $\Pr(y^1, \cdots, y^T, \omega^{T+1} = \tilde{\omega} | \hat{\omega}, a^1, \cdots, a^T) > 0$ for some $\hat{\omega}$ with $\mu(\hat{\omega}) > 0$. But clause (ii) asserts that the inequality does not hold for all $\hat{\omega} \in \Omega$ and $\tilde{\omega} \notin \Omega^*$.

Let $(y^1, \cdots, y^T)$ be the signal sequence which induces the posterior belief $\tilde{\mu}$ given the initial prior $\mu$ and the action sequence $(a^1, \cdots, a^T)$. Such a signal sequence may not be unique, so let $\hat{Y}^t$ be the set of these signal sequences. Then (12) implies that

$$\sum_{(y^1, \cdots, y^T) \in \hat{Y}^T} \Pr(y^1, \cdots, y^T | \mu, a^1, \cdots, a^T) \geq \pi^*.$$

Arranging,

$$\sum_{(y^1, \cdots, y^T) \in \hat{Y}^T} \sum_{\tilde{\omega} \in \Omega} \mu(\tilde{\omega}) \Pr(y^1, \cdots, y^T | \tilde{\omega}, a^1, \cdots, a^T) \geq \pi^*.$$

Plugging $\mu(\tilde{\omega}) = \frac{\pi^*}{2(|\Omega|-1)}$ and $\sum_{(y^1, \cdots, y^T) \in \hat{Y}^T} \Pr(y^1, \cdots, y^T | \tilde{\omega}, a^1, \cdots, a^T) \leq 1$ into this inequality,

$$\sum_{(y^1, \cdots, y^T) \in \hat{Y}^T} \mu(\omega) \Pr(y^1, \cdots, y^T | \omega, a^1, \cdots, a^T) + \frac{\pi^*}{2} \geq \pi^*$$

so that

$$\sum_{(y^1, \cdots, y^T) \in \hat{Y}^T} \mu(\omega) \Pr(y^1, \cdots, y^T | \omega, a^1, \cdots, a^T) \geq \frac{\pi^*}{2}.$$

Hence there is some $(y^1, \cdots, y^T) \in \hat{Y}^T$ which can happen with positive probability given the initial state $\omega$ and the action sequence $(a^1, \cdots, a^T)$. Obviously this sequence $(y^1, \cdots, y^T)$ satisfies clause (i) in Definition 9. Also it satisfies clause (ii) in Definition 9, since $(y^1, \cdots, y^T)$ induces the posterior belief $\tilde{\mu}$ whose support is $\Omega^*$, given the initial prior $\mu$ whose support is the whole space $\Omega$. Since $\omega$ can be arbitrarily chosen, the proof is completed.

## E.4   Proof of Lemma 7

By the definition of global accessibility, if $\{\omega\}$ is globally accessible, any superset $\Omega^* \supseteq \{\omega\}$ is globally accessible. So it is sufficient to show that if $\{\omega\}$ is transient, then any superset $\Omega^* \supseteq \{\omega\}$ is globally accessible or transient.

To prove this, take a transient set $\{\omega\}$ and a superset $\Omega^{**} \supseteq \{\omega\}$. Suppose that $\Omega^{**}$ is not globally accessible. In what follows, we will show that it is transient. Take a strategy $s \in S^*$ arbitrarily. Take $\Omega^* \in \mathcal{O}$, $T$, and $(y^1, \cdots, y^T)$, as stated in

Definition 10. Suppose that the initial belief is $\mu$ with the support $\Omega^{**}$, and that players play $s$. Then, since $\mu$ puts a positive probability on $\omega$, the signal sequence $(y^1, \cdots, y^T)$ realizes with positive probability and the support of the posterior belief $\mu^{T+1}$ must be a superset $\Omega^{***}$ of $\Omega^*$. Since $\Omega^*$ is globally accessible, so is the superset $\Omega^{***}$. This shows that $\Omega^{**}$ is transient, as $s$ can be arbitrary.

## E.5  Proof of Lemma 8

Set $s = s^{**}$ as in the proof of Lemma 5. Then from Lemma 27 and (11), we obtain

$$|\lambda \cdot v^\mu(\delta, s^{**}) - \lambda \cdot v^\mu(\delta, s^\mu)| \le \frac{1 - \delta^T}{\delta^T} C|\Omega| + (1 - \delta^{4^{|\Omega|}}) \tilde{C}|\Omega|.$$

Hence the result holds.

## E.6  Proof of Lemma 9

Take $\lambda$, $\mu$, and $\varepsilon > 0$ arbitrarily. Let $\overline{\delta} \in (0, 1)$ be such that

$$\left| \max_{v \in V^\mu(\delta)} \lambda \cdot v - \limsup_{\delta \to 1} \max_{v \in V^\mu(\delta)} \lambda \cdot v \right| < \frac{\varepsilon}{2}, \tag{13}$$

and such that there is $s^*$ with

$$\left| \max_{v \in V^\mu(\overline{\delta})} \lambda \cdot v - \lambda \cdot v^{\tilde{\mu}}(\overline{\delta}, s^*) \right| < \frac{\varepsilon}{2} \tag{14}$$

for all $\tilde{\mu}$. (The existence of such $\overline{\delta}$ is guaranteed by Lemma 8.) It suffices to show that

$$\limsup_{\delta \to 1} \max_{v \in V^\mu(\delta)} \lambda \cdot v - \max_{v \in V^\mu(\delta)} \lambda \cdot v < \varepsilon \tag{15}$$

for all $\delta \in (\overline{\delta}, 1)$.

Take $\delta \in (\overline{\delta}, 1)$ arbitrarily, and let $p = \frac{\overline{\delta}}{\delta}$. Consider the following strategy profile $s^{**}$: Play $s^*$ until the "random termination period" $t$ such that $z^t > p$ and $z^\tau \le p$ for all $\tau < t$. After the random termination period $t$, play $s^*$ once again from period $t + 1$, until the next random termination period, and so on. Intuitively, players revise their play with probability $1 - p$ in each period.

Let $\mu^* \in \arg\min_{\tilde{\mu}} \lambda \cdot v^{\tilde{\mu}}(\delta, s^{**})$. Suppose that $s^{**}$ is played in the stochastic game with initial prior $\mu^*$ and discount factor $\delta$. Then the corresponding score is

$$\lambda \cdot v^{\mu^*}(\delta, s^{**}) = (1-\delta)E\left[\sum_{t=1}^{\infty}(p\delta)^{t-1}\lambda \cdot g^{\omega^t}(a^t)\bigg|\mu^*, s^*\right]$$
$$+ (1-p)\sum_{t=1}^{\infty}p^{t-1}\delta^t E^t[\lambda \cdot v^{\mu(h^t)}(\delta, s^{**})|\mu^*, s^*]$$

where $\mu(h^t)$ denotes the posterior belief at the beginning of period $t+1$ given the initial prior $\mu^*$ and the past history $h^t$. Note that

$$E\left[\sum_{t=1}^{\infty}(p\delta)^{t-1}\lambda \cdot g^{\omega^t}(a^t)\bigg|\mu^*, s^*\right] = \frac{\lambda \cdot v^{\mu^*}(\overline{\delta}, s^*)}{1-p\delta}.$$

Hence we have

$$\lambda \cdot v^{\mu^*}(\delta, s^{**}) = \frac{(1-\delta)}{1-p\delta}\lambda \cdot v^{\mu^*}(\overline{\delta}, s^*)$$
$$+ (1-p)\sum_{t=1}^{\infty}p^{t-1}\delta^t E^t[\lambda \cdot v^{\mu(h^t)}(\delta, s^{**})|\mu^*, s^*].$$

Since $\mu^*$ minimizes $\lambda \cdot v^{\tilde{\mu}}(\delta, s^{**})$, we obtain

$$\lambda \cdot v^{\mu^*}(\delta, s^{**}) \geq \frac{(1-\delta)}{1-p\delta}\lambda \cdot v^{\mu^*}(\overline{\delta}, s^*) + (1-p)\sum_{t=1}^{\infty}p^{t-1}\delta^t\lambda \cdot v^{\mu^*}(\delta, s^{**})$$
$$= \frac{(1-\delta)}{1-p\delta}\lambda \cdot v^{\mu^*}(\overline{\delta}, s^*) + \frac{(1-p)\delta}{1-p\delta}\lambda \cdot v^{\mu^*}(\delta, s^{**}).$$

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

Then from (13) and (14), we have

$$\lambda \cdot v^{\mu^*}(\delta, s^{**}) > \limsup_{\delta \to 1}\max_{v \in V^{\omega}(\delta)}\lambda \cdot v - \varepsilon.$$

This implies (15), since we have

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

## E.7 Proof of Lemma 10

Note that $\lim_{\delta \to 1} \max_{v \in V^\mu(\delta)} \lambda \cdot v$ is continuous with respect to $\lambda$. Note also that $\{\max_{v \in V^\mu(\delta)} \lambda \cdot v\}_{(\delta, \mu)}$ is equi-Lipschitz continuous with respect to $\lambda$, since $V^\mu(\delta)$ is included in the bounded set $\times_{i \in I}[-\bar{g}_i, \bar{g}_i]$ for all $\delta$ and $\mu$. Hence, for each $\lambda$, there is an open set $U_\lambda \subset \Lambda$ containing $\lambda$ such that

$$\left| \lim_{\delta \to 1} \max_{v \in V^\mu(\delta)} \tilde{\lambda} \cdot v - \lim_{\delta \to 1} \max_{v \in V^\mu(\delta)} \lambda \cdot v \right| < \frac{\varepsilon}{3} \tag{16}$$

for all $\tilde{\lambda} \in U_\lambda$ and $\mu$, and such that

$$\left| \max_{v \in V^\mu(\delta)} \tilde{\lambda} \cdot v - \max_{v \in V^\mu(\delta)} \lambda \cdot v \right| < \frac{\varepsilon}{3} \tag{17}$$

for all $\tilde{\lambda} \in U_\lambda$, $\delta \in (0,1)$, and $\mu$. The family of open sets $\{U_\lambda\}_{\lambda \in \Lambda}$ covers the compact set $\Lambda$, so there is a finite subcover $\{U_\lambda\}_{\lambda \in \Lambda^*}$. Since the set $\Lambda^* \subset \Lambda$ is a finite set of directions $\lambda$, there is $\bar{\delta} \in (0,1)$ such that

$$\left| \lim_{\delta \to 1} \max_{v \in V^\mu(\delta)} \lambda \cdot v - \max_{v \in V^\mu(\delta)} \lambda \cdot v \right| < \frac{\varepsilon}{3}$$

for all $\lambda \in \Lambda^*$, $\delta \in (\bar{\delta}, 1)$, and $\mu$. Plugging (16) and (17) into this, we obtain

$$\left| \lim_{\delta \to 1} \max_{v \in V^\mu(\delta)} \lambda \cdot v - \max_{v \in V^\mu(\delta)} \lambda \cdot v \right| < \varepsilon.$$

for all $\lambda \in \Lambda$, $\delta \in (\bar{\delta}, 1)$, and $\mu$, as desired.

## E.8 Proof of Lemma 11

Pick a transient set $\Omega^*$ arbitrarily. To prove the lemma, suppose not so that there is a pure strategy profile $s \in S^*$ and a belief $\mu$ whose support is $\Omega^*$ such that

$$\Pr(X(\Omega^*|\mu, s) < \infty) = 1. \tag{18}$$

Pick such $s \in S^*$ and $\mu$.

Suppose that the initial prior is $\mu$ and players play $s$. Since $\Omega^*$ is transient, there is a natural number $T$ and a signal sequence $(y^1, \cdots, . y^T)$ with the following properties:

(i) The probability of the signal sequence $(y^1, \cdots, .y^T)$ given $\mu$ and $s$ is positive.

(ii) The support of the posterior belief has not yet returned to $\Omega^*$; that is, for each $t \leq T$, the support of the posterior belief $\tilde{\mu}$ given the initial prior $\mu$, the strategy profile $s$, and the signal realization $(y^1, \cdots, .y^t)$ is not $\Omega^*$.

(iii) The support of the posterior belief given $\mu$, $s$, and $(y^1, \cdots, .y^T)$ is globally accessible.

The existence of a signal sequence which satisfies (i) and (iii) is obvious, since $\Omega^*$ is transient. To see that there is a signal sequence which satisfies all three properties simultaneously, suppose not so that all signal sequences with (i) and (iii) do not satisfy (ii). That is, given that the strategy profile $s$ is played, the support of the posterior belief must return to $\Omega^*$ before it reaches a globally accessible set. Assume that the initial prior is $\mu$, and consider the following strategy profile $\tilde{s}$:

- Play the strategy profile $s$, until the support of the posterior belief returns to $\Omega^*$.

- Once the support returns to $\Omega^*$, then play the profile $s$ again, until the support of the posterior belief returns to $\Omega^*$ next time.

- And so on.

By the construction, if the initial prior is $\mu$ and players play $\tilde{s}$, the support of the posterior belief cannot reach a globally accessible set. This contradicts the fact that the set $\Omega^*$ is transient. Hence there must be a signal sequence which satisfies (i), (ii), and (iii) simultaneously. Take such a signal sequence $(y^1, \cdots, y^T)$, and let $\Omega^{**}$ be the support of the posterior belief after this signal sequence. Let $s|_{(y^1, \cdots, y^T)}$ be the continuation strategy profile after $(y^1, \cdots, y^T)$ induced by $s$. Note that this is well-defined, since $s$ is a pure strategy profile. Since the sequence $(y^1, \cdots, y^T)$ satisfies (ii) and since (18) holds, if the support of the current belief is $\Omega^{**}$ and players play $s|_{(y^1, \cdots, y^T)}$ in the continuation game, then the support will reach $\Omega^*$ with probability one. That is, for all $\hat{\mu}$ whose support is $\Omega^{**}$, we have

$$\Pr(X(\Omega^* | \hat{\mu}, s|_{(y^1, \cdots, y^T)}) < \infty) = 1.$$

This in turn implies that for each $\hat{\omega} \in \Omega^{**}$, there is a natural number $\hat{T}$, an action sequence $(\hat{a}^1, \cdots, \hat{a}^{\hat{T}})$, and a signal sequence $(\hat{y}^1, \cdots, \hat{y}^{\hat{T}})$ such that

$$\Pr(\hat{y}^1, \cdots, \hat{y}^{\hat{T}} | \hat{\omega}, \hat{a}^1, \cdots, \hat{a}^{\hat{T}}) > 0 \tag{19}$$

and such that

$$\Pr(\hat{y}^1, \cdots, \hat{y}^{\hat{T}}, \omega^{\hat{T}+1} | \omega^1, \hat{a}^1, \cdots, \hat{a}^{\hat{T}}) = 0 \tag{20}$$

for all $\omega^1 \in \Omega^{**}$ and $\omega^{\hat{T}+1} \notin \Omega^*$.

Pick $\omega$ arbitrarily. From (iii), the set $\Omega^{**}$ is globally accessible, and hence we can choose a natural number $\tilde{T} \le 4^{|\Omega|}$, an action sequence $(\tilde{a}^1, \cdots, \tilde{a}^{\tilde{T}})$, a signal sequence $(\tilde{y}^1, \cdots, \tilde{y}^{\tilde{T}})$, and a state $\hat{\omega}$ such that

$$\Pr(\tilde{y}^1, \cdots, \tilde{y}^{\tilde{T}}, \omega^{\tilde{T}+1} = \hat{\omega} | \omega, \tilde{a}^1, \cdots, \tilde{a}^{\tilde{T}}) > 0 \tag{21}$$

and such that

$$\Pr(\tilde{y}^1, \cdots, \tilde{y}^{\tilde{T}}, \omega^{\tilde{T}+1} | \omega^1, \tilde{a}^1, \cdots, \tilde{a}^{\tilde{T}}) = 0 \tag{22}$$

for all $\omega^1 \in \Omega$ and $\omega^{\tilde{T}+1} \notin \Omega^{**}$. Given this $\hat{\omega}$, choose $\hat{T}$, $(\hat{a}^1, \cdots, \hat{a}^{\hat{T}})$, and $(\hat{y}^1, \cdots, \hat{y}^{\hat{T}})$ so that (19) and (20) hold. Now, consider the action sequence

$$\boldsymbol{a} = (\tilde{a}^1, \cdots, \tilde{a}^{\tilde{T}}, \hat{a}^1, \cdots, \hat{a}^{\hat{T}})$$

and the signal sequence

$$\boldsymbol{y} = (\tilde{y}^1, \cdots, \tilde{y}^{\tilde{T}}, \hat{y}^1, \cdots, \hat{y}^{\hat{T}}).$$

Then from (19) and (21), we have

$$\Pr(\boldsymbol{y} | \omega, \boldsymbol{a}) > 0.$$

Also, from (20) and (22), we have

$$\Pr(\boldsymbol{y}, \omega^{\tilde{T}+\hat{T}+1} | \omega^1, \boldsymbol{a}) = 0$$

for all $\omega^1 \in \Omega$ and $\omega^{\tilde{T}+\hat{T}+1} \notin \Omega^*$. This shows that the sequences $\boldsymbol{a}$ and $\boldsymbol{y}$ satisfy (i) and (ii) in Definition 9 for $\omega$. Since such $\boldsymbol{a}$ and $\boldsymbol{y}$ exist for each $\omega \in \Omega$, the set $\Omega^*$ is globally accessible. (The length of the action sequence $\boldsymbol{a}$ above may be greater than $4^{|\Omega|}$, but as discussed in footnote 13, we can always find a sequence with length $T \le 4^{|\Omega|}$.) However this is a contradiction, as the set $\Omega^*$ is transient.

## E.9 Proof of Lemma 12

Fix $\delta$. For a given strategy $s_{-i}$ and a prior $\mu$, let $v_i^\mu(s_{-i})$ denote player $i$'s best possible payoff; that is, let $v_i^\mu(s_{-i}) = \max_{s_i \in S_i} v_i^\mu(\delta, s_i, s_{-i})$. As the following lemma shows, this payoff function $v_i^\mu(s_{-i})$ is convex with respect to $\mu$. The proof is very similar to Lemma 21 and hence omitted.

**Lemma 28.** *For each $s_{-i}$, $v_i^\mu(s_{-i})$ is convex with respect to $\mu$.*

Let $s^\mu$ denote the minimax strategy profile given the initial prior $\mu$. For each $\mu$, let $\overline{v}_i(s_{-i}^\mu) = \max_{\tilde{\mu} \in \triangle \Omega} v_i^{\tilde{\mu}}(s_{-i}^\mu)$ be the maximal value of the convex function $v_i^{\tilde{\mu}}(s_{-i}^\mu)$. Let $\mu^* \in \arg\max_{\mu \in \triangle \Omega} \overline{v}_i(s_{-i}^\mu)$. That is, $\overline{v}_i(s_{-i}^{\mu^*})$ is the maximal value when we consider a series of the convex functions $\{v_i^{\tilde{\mu}}(s_{-i}^\mu)\}_{\mu \in \triangle \Omega}$. Since $v_i^\mu(s_{-i}^{\mu^*})$ is convex, it is maximized when $\mu$ is an extreme point. Let $\omega$ denote this extreme point that is, $v_i^\omega(s_{-i}^{\mu^*}) \geq v_i^\mu(s_{-i}^{\mu^*})$ for all $\mu$. (This $\omega$ denotes the prior in which players believe that the initial state is $\omega$ for sure.)

Take a small $\eta > 0$, and let $\mu_\eta^*$ be the perturbation of the prior $\mu^*$ such that $\mu_\eta^* = (1-\eta)\mu^* + \eta(\frac{1}{|\Omega|}, \cdots, \frac{1}{|\Omega|})$. Note that $\mu_\eta^*$ assigns at least probability $\eta$ on each state.

Let $Z^\mu$ be the set of all (possibly mixed) action profiles which is chosen in period one by some minimax strategies associated with the initial prior $\mu$. ($Z^\mu$ is a set because there can be multiple minimax strategy associated with $\mu$.) Then a standard argument shows that $Z^\mu$ is upper hemi-continuous. This, together with the fact that $s_{-i}^\mu$ is Markov, implies that without loss of generality we can assume that $s_{-i}^{\mu^*}$ and $s_{-i}^{\mu_\eta^*}$ are "close" when $\eta$ is small. (This is a loose statement because we do not explain how to measure the distance between two strategies.) Accordingly we have the following lemma; the proof is standard and hence omitted.

**Lemma 29.** *For any $\delta \in (0,1)$ and $\varepsilon > 0$, there is $\overline{\eta} > 0$ such that for any $\eta \in (0, \overline{\eta})$,*

$$\left| \overline{v}_i(s_{-i}^{\mu^*}) - v_i^\omega(s_{-i}^{\mu_\eta^*}) \right| < \varepsilon.$$

*By letting $\varepsilon = 1 - \delta$, this implies that for any $\delta \in (0,1)$, there is $\overline{\eta} > 0$ such that for any $\eta \in (0, \overline{\eta})$,*

$$\left| \overline{v}_i(s_{-i}^{\mu^*}) - v_i^\omega(s_{-i}^{\mu_\eta^*}) \right| < 1 - \delta. \tag{23}$$

In what follows, we fix $\eta$ so that (23) holds. Let $s_i^*$ be player $i$'s best reply against $s_{-i}^{\mu_\eta^*}$ given the initial prior $\omega$. Since the game is strongly connected, there is a natural number $T$ and a history $h^T$ such that

- The history $h^T$ happens with probability at least $(\frac{\overline{\pi}}{|A|})^T$, given the initial state $\omega$ and the profile $(s_i^*, s_{-i}^{\mu_\eta^*})$, and

- The support of the posterior induced by the initial state $\omega$ and the history $h^T$ is identical with that induced by the initial prior $\mu_\eta^*$ and the history $h^T$.

Pick such $T$ and $h^T$. (Here, the minimum probability $(\frac{\overline{\pi}}{|A|})^T$ comes from the fact that clause (ii) in the definition of strong connectedness considers "all" strategy profiles $s$.) Then let $\mu(h^T|\mu_\eta^*)$ be the posterior in period $T+1$ given that the initial prior is $\mu^*$ and the history is $h^T$. Similarly, let $\mu(h^T|\omega)$ be the posterior given the initial prior $\omega$. By the definition of $h^T$, the supports of these two beliefs are the same; let $\Omega^*$ denote this support. Let $C = 2\overline{g}(\frac{|A|}{\overline{\pi}})^T$ and let $\tilde{C} = (\frac{|A|}{\overline{\pi}})^T$. The following lemma shows that player $i$'s continuation payoff induced by the initial state $\omega$, the profile $(s_i^*, s_{-i}^{\mu^*})$ and the history $h^T$ is close to the value $v_i^\omega(s_{-i}^{\mu^*})$.

**Lemma 30.** *For any* $\delta$,

$$\left| v_i^\omega(s_{-i}^{\mu^*}) - v_i^{\mu(h^T|\omega)}(s_{-i}^{\mu_\eta^*}|_{h^T}) \right| \leq \frac{1-\delta^T}{\delta^T} C + (1-\delta)\tilde{C}.$$

*Since* $s_{-i}^{\mu_\eta^*}|_{h^T} = s_{-i}^{\mu(h^T|\mu_\eta^*)}$, *this implies*

$$\left| v_i^\omega(s_{-i}^{\mu^*}) - v_i^{\mu(h^T|\omega)}(s_{-i}^{\mu(h^T|\mu_\eta^*)}) \right| \leq \frac{1-\delta^T}{\delta^T} C + (1-\delta)\tilde{C}.$$

*Proof.* Note that

$$v_i^\omega(s_{-i}^{\mu_\eta^*}) = (1-\delta) \sum_{t=1}^{T} \delta^{t-1} E[g_i^{\omega^t}(a^t)|\omega, s_i^*, s_{-i}^{\mu_\eta^*}]$$
$$+ \delta^T \sum_{\tilde{h}^T \in H^T} \Pr(\tilde{h}^T|\omega, s_i^*, s_{-i}^{\mu_\eta^*}) v_i^{\mu(\tilde{h}^T|\omega)}(s_{-i}^{\mu_\eta^*}|_{\tilde{h}^T}).$$

Since $(1-\delta)\sum_{t=1}^{T} \delta^{t-1} E[g_i^{\omega^t}(a^t)|\omega, s] \leq (1-\delta^T)\overline{g}$,

$$v_i^\omega(s_{-i}^{\mu_\eta^*}) \leq (1-\delta^T)\overline{g} + \delta^T \sum_{\tilde{h}^1} \Pr(\tilde{h}^T|\omega, s_i^*, s_{-i}^{\mu_\eta^*}) v_i^{\mu(\tilde{h}^T|\omega)}(s_{-i}^{\mu_\eta^*}|_{\tilde{h}^T}).$$

From (23) and $\bar{v}_i(s_{-i}^{\mu^*}) \geq v_i^{\mu(\tilde{h}^T|\omega)}(s_{-i}^{\mu^*_\eta}|_{\tilde{h}^T})$, we know that

$$v_i^{\mu(\tilde{h}^T|\omega)}(\delta, s_{-i}^{\mu^*_\eta}|_{\tilde{h}^T}) \leq v_i^\omega(s_{-i}^{\mu^*_\eta}) + (1-\delta) \tag{24}$$

for all $\tilde{h}^T \in H^T$. Plugging this into the above inequality,

$$\begin{aligned}
v_i^\omega(s_{-i}^{\mu^*_\eta}) \leq &(1-\delta^T)\bar{g} + \delta^T \Pr(h^T|\omega, s_i^*, s_{-i}^{\mu^*_\eta}) v_i^{\mu(h^T|\omega)}(s_{-i}^{\mu^*_\eta}|_{h^T}) \\
&+ \delta^T(1 - \Pr(h^T|\omega, s_i^*, s_{-i}^{\mu^*_\eta}))\left\{ v_i^\omega(s_{-i}^{\mu^*_\eta}) + (1-\delta) \right\}.
\end{aligned}$$

Since (24) holds and $\Pr(h^T|\omega, s_i^*, s_{-i}^{\mu^*_\eta}) \geq (\frac{\bar{\pi}}{|A|})^T$,

$$\begin{aligned}
v_i^\omega(s_{-i}^{\mu^*_\eta}) \leq &(1-\delta^T)\bar{g} + \delta^T \left(\frac{\bar{\pi}}{|A|}\right)^T v_i^{\mu(h^T|\omega)}(s_{-i}^{\mu^*_\eta}|_{h^T}) \\
&+ \delta^T\left\{ 1 - \left(\frac{\bar{\pi}}{|A|}\right)^T \right\}\left\{ v_i^\omega(s_{-i}^{\mu^*_\eta}) + (1-\delta) \right\}.
\end{aligned}$$

Subtracting $\{1 - \delta^T(\frac{\bar{\pi}}{|A|})^T\}v_i^\omega(s_{-i}^{\mu^*_\eta}) - \delta^T(\frac{\bar{\pi}}{|A|})^T(1-\delta) - \delta^T(\frac{\bar{\pi}}{|A|})^T v_i^{\mu(h^T|\omega)}(s_{-i}^{\mu^*_\eta}|_{h^T})$ from both sides,

$$\begin{aligned}
\delta^T \left(\frac{\bar{\pi}}{|A|}\right)^T &\left\{ v_i^\omega(s_{-i}^{\mu^*_\eta}) + (1-\delta) - v_i^{\mu(h^T|\omega)}(s_{-i}^{\mu^*_\eta}|_{h^T}) \right\} \\
&\leq (1-\delta^T)(\bar{g} - v_i^\omega(s_{-i}^{\mu^*_\eta})) + \delta^T(1-\delta).
\end{aligned}$$

Dividing both sides by $\delta^T(\frac{\bar{\pi}}{|A|})^T$,

$$\begin{aligned}
&v_i^\omega(s_{-i}^{\mu^*_\eta}) + (1-\delta) - v_i^{\mu(h^T|\omega)}(s_{-i}^{\mu^*_\eta}|_{h^T}) \\
&\leq \frac{|A|^T(1-\delta^T)(\bar{g} - v_i^\omega(s_{-i}^{\mu^*_\eta}))}{\delta^T \bar{\pi}^T} + (1-\delta)\left(\frac{|A|}{\bar{\pi}}\right)^T.
\end{aligned}$$

From (23), we know that $v_i^\omega(s_{-i}^{\mu^*_\eta}) + 1 - \delta \geq v_i^\omega(s_{-i}^{\mu^*})$, and thus

$$v_i^\omega(s_{-i}^{\mu^*}) - v_i^{\mu(h^T|\omega)}(s_{-i}^{\mu^*_\eta}|_{h^T}) \leq \frac{|A|^T(1-\delta^T)(\bar{g} - v_i^\omega(s_{-i}^{\mu^*_\eta}))}{\delta^T \bar{\pi}^T} + (1-\delta)\left(\frac{|A|}{\bar{\pi}}\right)^T.$$

By the definition of $\mu^*$, the left-hand side is positive. Note also that $v_i^\omega(s_{-i}^{\mu^*_\eta}) \geq -\bar{g}$. Hence the result follows. $\qquad\qquad$ *Q.E.D.*

The second statement in the above lemma ensures that player $i$'s maximal payoff given the initial prior $\mu(h^T|\omega)$ and the opponents' strategy $s_{-i}^{\mu(h^T|\mu_\eta^*)}$ is close to the value $\bar{v}_i(s_{-i}^{\mu^*})$. The next lemma shows that the same is true even if the initial prior $\mu(h^T|\omega)$ is replaced with $\tilde{\omega} \in \Omega^*$. Let $C' = \frac{C}{\underline{\pi}^T}$ and $\tilde{C}' = \frac{\tilde{C}}{\underline{\pi}^T}$.

**Lemma 31.** *For each $\tilde{\omega} \in \Omega^*$,*

$$\left| \bar{v}_i(s_{-i}^{\mu^*}) - v_i^{\tilde{\omega}}(s_{-i}^{\mu(h^T|\mu_\eta^*)}) \right| \leq \frac{1-\delta^T}{\delta^T}C' + (1-\delta)\tilde{C}'$$

*Proof.* Pick an arbitrary $\tilde{\omega} \in \Omega^*$. Then we have

$$v_i^{\mu(h^T|\omega)}(s_{-i}^{\mu(h^T|\mu_\eta^*)})$$

$$= \sum_{\hat{\omega} \in \Omega} \mu(h^T|\omega)[\hat{\omega}] v_i^{\hat{\omega}}(\delta, s_i^*|_{h^T}, s_{-i}^{\mu(h^T|\mu_\eta^*)})$$

$$\leq \mu(h^T|\omega)[\tilde{\omega}] v_i^{\tilde{\omega}}(\delta, s_i^*|_{h^T}, s_{-i}^{\mu(h^T|\mu_\eta^*)}) + (1 - \mu(h^T|\omega)[\tilde{\omega}])\bar{v}_i(s_{-i}^{\mu^*}).$$

Here the equality follows from $v_i^{\mu(h^T|\omega)}(s_{-i}^{\mu(h^T|\mu_\eta^*)}) = v_i^{\mu(h^T|\omega)}(\delta, s_i^*|_{h^T}, s_{-i}^{\mu(h^T|\mu_\eta^*)})$, and the inequality from $v_i^{\hat{\omega}}(\delta, s_i^*|_{h^T}, s_{-i}^{\mu(h^T|\mu_\eta^*)}) \leq \bar{v}_i(s_{-i}^{\mu^*})$. Arranging,

$$\mu(h^T|\omega)[\tilde{\omega}]\left( \bar{v}_i(s_{-i}^{\mu^*}) - v_i^{\tilde{\omega}}(\delta, s_i^*|_{h^T}, s_{-i}^{\mu(h^T|\mu_\eta^*)}) \right) \leq \bar{v}_i(s_{-i}^{\mu^*}) - v_i^{\mu(h^T|\omega)}(s_{-i}^{\mu(h^T|\mu_\eta^*)}).$$

The term in the parenthesis in the left-hand side is non-negative. Similarly, the right-hand side is non-negative. Hence

$$\left| \bar{v}_i(s_{-i}^{\mu^*}) - v_i^{\tilde{\omega}}(\delta, s_i^*|_{h^T}, s_{-i}^{\mu(h^T|\mu_\eta^*)}) \right| \leq \frac{\left| \bar{v}_i(s_{-i}^{\mu^*}) - v_i^{\mu(h^T|\omega)}(s_{-i}^{\mu(h^T|\mu_\eta^*)}) \right|}{\mu(h^T|\omega)[\tilde{\omega}]}.$$

Using $\mu(h^T|\omega)[\tilde{\omega}] \geq \bar{\pi}^T$ and Lemma 30,

$$\left| \bar{v}_i(s_{-i}^{\mu^*}) - v_i^{\tilde{\omega}}(\delta, s_i^*|_{h^T}, s_{-i}^{\mu(h^T|\mu_\eta^*)}) \right| \leq \frac{1-\delta^T}{\delta^T}C' + (1-\delta)\tilde{C}'. \qquad (25)$$

Since $v_i^{\tilde{\omega}}(s_{-i}^{\mu(h^T|\mu_\eta^*)})$ is at least $v_i^{\tilde{\omega}}(\delta, s_i^*|_{h^T}, s_{-i}^{\mu(h^T|\mu_\eta^*)})$ but at most $\bar{v}_i(s_{-i}^{\mu^*})$, the result follows. $\qquad\qquad$ *Q.E.D.*

The next lemma shows that the same result holds for all initial prior $\mu$ whose support is $\Omega^*$. The proof is identical with that of Lemma 24 and hence omitted.

**Lemma 32.** *For each $\mu$ such that $\mu(\tilde{\omega}) = 0$ for all $\tilde{\omega} \notin \Omega^*$,*

$$\left| \bar{v}_i(s_{-i}^{\mu^*}) - v_i^{\mu}(s_{-i}^{\mu(h^T|\mu_\eta^*)}) \right| \leq \frac{1-\delta^T}{\delta^T} C' + (1-\delta)\tilde{C}'.$$

Letting $\mu = \mu(h^T|\mu_\eta^*)$, the above lemma implies that the minimax payoff given the initial prior $\mu(h^T|\mu_\eta^*)$ is close to the value $\bar{v}_i(s_{-i}^{\mu^*})$; that is,

$$\left| \bar{v}_i(s_{-i}^{\mu^*}) - v_i^{\mu(h^T|\mu_\eta^*)}(s_{-i}^{\mu(h^T|\mu_\eta^*)}) \right| \leq \frac{1-\delta^T}{\delta^T} C' + (1-\delta)\tilde{C}'.$$

Now let $\overline{\mu}$ be the prior which maximizes the minimax payoff. Note that such a maximizer exists, because player $i$'s payoff $v_i^{\mu}(\delta, s)$ is continuous with respect to $\mu$ and so is the minimax payoff $\underline{v}_i^{\mu}(\delta)$. By the definition of $\overline{\mu}$, the minimax payoff given $\overline{\mu}$ is at least the minimax payoff given the prior $\mu(h^T|\mu_\eta^*)$ and at most the value $\bar{v}_i(s_{-i}^{\mu^*})$. Then from the above inequality, we obtain the following lemma, which says that the minimax payoff given $\overline{\mu}$ is also close to the value $\bar{v}_i(s_{-i}^{\mu^*})$:

**Lemma 33.** *We have*

$$\left| \bar{v}_i(s_{-i}^{\mu^*}) - v_i^{\overline{\mu}}(s_{-i}^{\overline{\mu}}) \right| \leq \frac{1-\delta^T}{\delta^T} C' + (1-\delta)\tilde{C}'.$$

Pick $\pi^* > 0$ such that for each robustly globally accessible set, $\pi^*$ satisfies the condition stated in the definition of robust global accessibility, and such that for each strongly transient set, $\pi^*$ satisfies the condition stated in the definition of strong transience.

Let $\underline{\mu}$ be the prior which minimizes the minimax payoff, and let $\underline{\Omega}$ be the support of $\underline{\mu}$. The set $\underline{\Omega}$ is either robustly globally accessible or strongly transient. If $\underline{\Omega}$ is strongly transient, then there is a robustly globally accessible set $\Omega^{**}$, a natural number $\tilde{T}$, and a history $h^{\tilde{T}}$ such that the history $h^{\tilde{T}}$ happens with probability at least $\pi^*$ given the initial prior $\underline{\mu}$ and the profile $s^{\underline{\mu}}$, and such that the support of the posterior in period $\tilde{T}+1$ induced by the initial prior $\underline{\mu}$ and the history $h^{\tilde{T}}$ is $\Omega^{**}$. Pick such $\Omega^{**}$, $\tilde{T}$, and $h^{\tilde{T}}$, and let $\mu^{**}$ be the posterior induced by $\underline{\mu}$ and $h^{\tilde{T}}$.

If $\underline{\Omega}$ is robustly globally accessible, then set $\Omega^{**} = \underline{\Omega}$, $\tilde{T} = 0$, and $\mu^{**} = \underline{\mu}$.

The following lemma shows that the minimax payoffs for $\underline{\mu}$ and $\mu^{**}$ are close. Let $C'' = \frac{2\overline{g}}{\pi^*}$.

**Lemma 34.** *We have*

$$\left| v_i^{\mu^{**}}(s_{-i}^{\mu^{**}}) - v_i^{\mu}(s_{-i}^{\mu}) \right| \leq \frac{1-\delta^{\tilde{T}}}{\delta^{\tilde{T}}} C''.$$

*Proof.* When the set $\underline{\Omega}$ is robustly globally accessible, the proof is obvious. When $\underline{\Omega}$ is strongly transient, the proof is identical with that of Lemma 22. Replace $\lambda \cdot v^{\omega}(\delta, s^{\omega})$ in Lemma 22 with $v_i^{\mu}(s_{-i}^{\mu})$, $\lambda \cdot v^{\mu(y^1, \cdots, y^T)}(\delta, s^{\mu(y^1, \cdots, y^T)})$ with $\lambda \cdot v^{\omega}(\delta, s^{\omega})$, and $(y^1, \cdots, y^T)$ with $h^{\tilde{T}}$, respectively. *Q.E.D.*

Since $\Omega^{**}$ is robustly globally accessible, for any initial prior $\mu$, there is an action sequence $\boldsymbol{a}_{-i}(\mu) = (a_{-i}^1(\mu), \cdots, a_{-i}^{4^{|\Omega|}}(\mu))$ such that for any player $i$'s strategy $s_i$, there is a natural number $T \leq 4^{|\Omega|}$ such that after $T$ periods, the support of the posterior is equal to $\Omega^{**}$ with probability at least $\pi^*$. Note also that this posterior belief puts at least $\hat{\pi}^*$ on each state $\omega \in \Omega^{**}$, as the following lemma shows.

**Lemma 35.** *If a subset $\Omega^{**}$ is robustly globally accessible, then there is $\pi^* > 0$ such that for any initial prior $\mu$ and for any i, there is an action sequence $(a_{-i}^1, \cdots, a_{-i}^{4^{|\Omega|}})$ such that for any player i's strategy $s_i$, there is a natural number $T \leq 4^{|\Omega|}$ and a belief $\tilde{\mu}$ such that*

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

*and such that $\tilde{\mu}(\omega) \geq \frac{1}{|\Omega|} \overline{\pi}^{4^{|\Omega|}}$ for all $\omega \in \Omega^{**}$ and $\tilde{\mu}(\omega) = 0$ for other $\omega$.*

*Proof.* Like global accessibility, there is an alternative definition of robust global accessibility; a set $\Omega^{**}$ is robustly globally accessible if for each state $\omega \in \Omega$, there is an action sequence $(a_{-i}^1, \cdots, a_{-i}^{4^{|\Omega|}})$ such that for any $s_i$, there is a natural number $T \leq 4^{|\Omega|}$ and a signal sequence $(y^1, \cdots, y^T)$ such that the following properties are satisfied:

(i) If the initial state is $\omega$, player $i$ plays $s_i$, and the opponents play $(a_{-i}^1, \cdots, a_{-i}^T)$, then the sequence $(y^1, \cdots, y^T)$ realizes with positive probability.

(ii) If player $i$ plays $s_i$, the opponents play $(a_{-i}^1, \cdots, a_{-i}^T)$, and the signal sequence $(y^1, \cdots, y^T)$ realizes, then the state in period $T+1$ must be in the set $\Omega^{**}$, regardless of the initial state $\hat{\omega}$ (possibly $\hat{\omega} \neq \omega$).

71

(iii) If the initial state is $\omega$, player $i$ plays $s_i$, the opponents play $(a_{-i}^1, \cdots, a_{-i}^T)$, and the signal sequence $(y^1, \cdots, y^T)$ realizes, then the support of the belief in period $T+1$ is the set $\Omega^*$.

To prove the lemma, fix an arbitrary prior $\mu$. Pick $\omega$ such that $\mu(\omega) \geq \frac{1}{|\Omega|}$, and then choose $(a_{-i}^1, \cdots, a_{-i}^{4^{|\Omega|}})$ as in the above definition. Then for each $s_i$, choose $T$ and $(y^1, \cdots, y^T)$ as in the above definition. Let $\tilde{\mu}$ be the posterior belief in period $T+1$ given the initial prior $\mu$, player $i$'s strategy $s_i$, the opponents' play $(a_{-i}^1, \cdots, a_{-i}^T)$, and the signal sequence $(y^1, \cdots, y^T)$. From (ii), $\tilde{\mu}(\omega) = 0$ for other $\omega$. Also from (iii), $\tilde{\mu}(\omega) \geq \frac{1}{|\Omega|} \overline{\pi}^{4^{|\Omega|}}$ for all $\omega \in \Omega^{**}$.                     Q.E.D.

Now assume that the initial prior is $\overline{\mu}$ and players $-i$ play the following strategy $s_{-i}^{**}$: Let $\Delta^{**}$ denote the set of beliefs $\mu$ such that $\mu(\omega) \geq \frac{1}{|\Omega|} \overline{\pi}^{4^{|\Omega|}}$ for all $\omega \in \Omega^{**}$ and $\mu(\omega) = 0$ for other $\omega$.

- Let $\mu^{(1)} = \overline{\mu}$.

- Play the action sequence $\boldsymbol{a}_{-i}(\mu^{(1)})$ for the first $4^{|\Omega|}$ periods, unless the posterior belief reaches the set $\Delta^{**}$.

- If the posterior belief satisfies $\mu^t \in \Delta^{**}$ in some period $t \leq 4^{|\Omega|} + 1$, then stop playing $\boldsymbol{a}_{-i}(\mu^{(1)})$ and switch the play immediately to $s_{-i}^{\mu^{**}}$ in the rest of the game.

- If not, play the action sequence $\boldsymbol{a}_{-i}(\mu^{(2)})$ for the next $4^{|\Omega|}$ periods. where $\mu^{(2)}$ is the posterior belief in period $4^{|\Omega|} + 1$.

- If the posterior belief satisfies $\mu^t \in \Delta^{**}$ in some period $t \leq 2 \cdot 4^{|\Omega|} + 1$, then switch the play immediately to $s_{-i}^{\mu^{**}}$ in the rest of the game.

- And so on.

Let $v_i^{\overline{\mu}}(s_{-i}^{**})$ be player $i$'s payoff when the initial prior is $\overline{\mu}$, the opponents play $s_{-i}^{**}$, and player $i$ plays a best reply. Note that after the switch to $s_{-i}^{\mu^{**}}$, player $i$'s continuation payoff is equal to $v_i^{\mu}(s_{-i}^{\mu^{**}})$ where $\mu$ is the posterior belief when the switch happens. Since this switch eventually happens with probability one, the overall payoff $v_i^{\overline{\mu}}(s_{-i}^{**})$ cannot be significantly greater than the value $v_i^{**} = \max_{\mu \in \Delta^{**}} v_i^{\mu}(s_{-i}^{\mu^{**}})$. Formally, we have the following lemma.

72

**Lemma 36.** $v_i^{\overline{\mu}}(s_{-i}^{**}) \le v_i^{**} + (1 - \delta^{4^{|\Omega|}})2\overline{g}$.

*Proof.* The proof is very similar to that of Lemma 25. Let $\rho^t$ denote the probability that the switch to $s_{-i}^{\mu^{**}}$ does not happen until the end of period $t$. Let $\rho^0 = 1$. Then we have

$$v_i^{\overline{\mu}}(s_{-i}^{**}) \le \sum_{t=1}^{\infty} \delta^{t-1} \left\{ \rho^{t-1}\overline{g} + (1 - \rho^{t-1})v_i^{**} \right\}.$$

Recall that the length of the action sequence $\boldsymbol{a}_{-i}(\mu)$ is at most $4^{|\Omega|}$ for each $\mu$, and that the probability that the action sequence $\boldsymbol{a}_{-i}(\mu)$ does not induce the switch to $s^*$ is at most $1 - \pi^*$. Hence we have

$$\rho^{n4^{|\Omega|}+k} \le (1 - \pi^*)^n$$

for each $n = 0, 1, \cdots$ and $k \in \{0, \cdots, 4^{|\Omega|} - 1\}$. This inequality, together with $\overline{g} \ge v_i^{**}$, implies that

$$\rho^{n4^{|\Omega|}+k}\overline{g} + (1 - \rho^{n4^{|\Omega|}+k})v_i^{**} \le (1 - \pi^*)^n\overline{g} + \{1 - (1 - \pi^*)^n\}v_i^{**}$$

for each $n = 0, 1, \cdots$ and $k \in \{0, \cdots, 4^{|\Omega|} - 1\}$. Plugging this inequality into the first one, we obtain

$$v_i^{\overline{\mu}}(s_{-i}^{**}) \le (1 - \delta) \sum_{n=1}^{\infty} \sum_{k=1}^{4^{|\Omega|}} \delta^{(n-1)4^{|\Omega|}+k-1} \left[ \begin{array}{c} (1 - \pi^*)^{n-1}\overline{g} \\ +\{1 - (1 - \pi^*)^{n-1}\}v_i^{**} \end{array} \right].$$

Then as in the proof of Lemma 25, the standard algebra shows

$$v_i^{\overline{\mu}}(s_{-i}^{**}) \le \frac{(1 - \delta^{4^{|\Omega|}})\overline{g}}{1 - (1 - \pi^*)\delta^{4^{|\Omega|}}} + \frac{\delta^{4^{|\Omega|}}\pi^*}{1 - (1 - \pi^*)\delta^{4^{|\Omega|}}}v_i^{**}.$$

Since $\frac{\delta^{4^{|\Omega|}}\pi^*}{1-(1-\pi^*)\delta^{4^{|\Omega|}}} = 1 - \frac{1-\delta^{4^{|\Omega|}}}{1-(1-\pi^*)\delta^{4^{|\Omega|}}}$, we have

$$v_i^{\overline{\mu}}(s_{-i}^{**}) \le v_i^{**} + \frac{1 - \delta^{4^{|\Omega|}}}{1 - (1 - \pi^*)\delta^{4^{|\Omega|}}}(\overline{g} - v_i^{**}).$$

Since $1 - (1 - \pi^*)\delta^{4^{|\Omega|}} > 1 - (1 - \pi^*) = \pi^*$ and $v_i^{**} \ge -\overline{g}$, the result follows.

<div align="right">*Q.E.D.*</div>

From Lemma 36 and the fact that $v_i^{\overline{\mu}}(s_{-i}^{**})$ is at least the minimax payoff $v_i^{\overline{\mu}}(s_{-i}^{\overline{\mu}})$, we have

$$v_i^{\overline{\mu}}(s_{-i}^{\overline{\mu}}) \le v_i^{**} + (1 - \delta^{4^{|\Omega|}})2\overline{g}.$$

This, together with Lemma 33, implies that

$$\overline{v}_i(s_{-i}^{\mu^*}) \le v_i^{**} + (1 - \delta^{4^{|\Omega|}})2\overline{g} + \frac{1 - \delta^T}{\delta^T}C' + (1 - \delta)\tilde{C}'.$$

On the other hand, by the definition, we know that $v_i^{**} \le \overline{v}_i(s_{-i}^{\mu^*})$. Hence

$$\left| \overline{v}_i(s_{-i}^{\mu^*}) - v_i^{**} \right| \le \frac{1 - \delta^T}{\delta^T}C' + (1 - \delta)\tilde{C}' + (1 - \delta^{4^{|\Omega|}})2\overline{g}.$$

Let $\hat{\mu}$ be a solution to $\max_{\mu \in \Delta^{**}} v_i^{\mu}(s_{-i}^{\mu^{**}})$. The inequality here says that player $i$'s best possible payoff given the initial prior $\hat{\mu}$ and the opponents' strategy $s_{-i}^{\mu^{**}}$ is close to the value $\overline{v}_i(s_{-i}^{\mu^*})$. That is,

$$\left| \overline{v}_i(s_{-i}^{\mu^*}) - v_i^{\hat{\mu}}(s_{-i}^{\mu^{**}}) \right| \le \frac{1 - \delta^T}{\delta^T}C' + (1 - \delta)\tilde{C}' + (1 - \delta^{4^{|\Omega|}})2\overline{g}.$$

The following lemma shows that the same result holds even if we replace the initial prior $\hat{\mu}$ with $\tilde{\omega} \in \Omega^{**}$. Let $C''' = \frac{C'|\Omega|}{\pi^{4^{|\Omega|}}}$, $\tilde{C}''' = \frac{\tilde{C}'|\Omega|}{\pi^{4^{|\Omega|}}}$, and $\hat{C} = \frac{2\overline{g}|\Omega|}{\pi^{4^{|\Omega|}}}$

**Lemma 37.** *For any $\tilde{\omega} \in \Omega^{**}$,*

$$\left| \overline{v}_i(s_{-i}^{\mu^*}) - v_i^{\tilde{\omega}}(s_{-i}^{\mu^{**}}) \right| \le \frac{1 - \delta^T}{\delta^T}C''' + (1 - \delta)\tilde{C}''' + (1 - \delta^{4^{|\Omega|}})\hat{C}.$$

*Proof.* The proof is very similar to that of Lemma 31. Replace $v_i^{\tilde{\omega}}(s_{-i}^{\mu(h^T|\mu_\eta^*)})$ in Lemma 31 with $v_i^{\tilde{\omega}}(s_{-i}^{\mu^{**}})$, $v_i^{\mu(h^T|\omega)}(s_{-i}^{\mu(h^T|\mu_\eta^*)})$ with $v_i^{\hat{\mu}}(s_{-i}^{\mu^{**}})$. $\qquad$ Q.E.D.

The next lemma shows that the same result holds for all initial prior $\mu$ whose support is $\Omega^{**}$. The proof is identical with that of Lemma 24 and hence omitted.

**Lemma 38.** *For any $\mu$ with $\mu(\tilde{\omega}) = 0$ with $\tilde{\omega} \notin \Omega^{**}$,*

$$\left| \overline{v}_i(s_{-i}^{\mu^*}) - v_i^{\mu}(s_{-i}^{\mu^{**}}) \right| \le \frac{1 - \delta^T}{\delta^T}C''' + (1 - \delta)\tilde{C}''' + (1 - \delta^{4^{|\Omega|}})\hat{C}.$$

Letting $\mu = \mu^{**}$, the above lemma implies that the minimax payoff given the initial prior $\mu^{**}$ is close to $\bar{v}_i(s_{-i}^{\mu^*})$. (More precisely, the difference between the two values is of order $1-\delta$.) Then from Lemma 34, we can conclude that the minimal minimax payoff $v_i^\mu(s_{-i}^\mu)$ is also close to $\bar{v}_i(s_{-i}^{\mu^*})$. This, together with Lemma 33, implies clause (i).

To prove clause (ii), let $\mu^{***} = (\frac{1}{|\Omega|}, \cdots, \frac{1}{|\Omega|})$. The discussion above shows that the difference between $\bar{v}_i(s_{-i}^{\mu^*})$ and the minimax payoff $v_i^{\mu^{***}}(s_{-i}^{\mu^{***}})$ is of order $1-\delta$. Then as in the proof of Lemma 37, we can show that the difference between $\bar{v}_i(s_{-i}^{\mu^*})$ and the payoff $v_i^\omega(s_{-i}^{\mu^{***}})$ is of order $1-\delta$ for all $\omega$, and that the difference between $\bar{v}_i(s_{-i}^{\mu^*})$ and the payoff $v_i^\omega(s^{\mu^{***}})$ is of order $1-\delta$ for all $\omega$. (The derivation of the latter result is identical with the one for the derivation of (25).) Then as in Lemma 38, we can prove that both $v_i^\mu(s_{-i}^{\mu^{***}})$ and $v_i^\mu(s^{\mu^{***}})$ are close to $\bar{v}_i(s_{-i}^{\mu^*})$ for each $\mu$. Since the minimax payoff $v_i^\mu(s_{-i}^\mu)$ is close to $\bar{v}_i(s_{-i}^{\mu^*})$ for each $\mu$, clause (ii) holds by setting $s = s^{\mu^{***}}$.

## E.10   Proof of Lemma 13

Very similar to that of Lemma 9. Here we use Lemma 12 instead of Lemma 8.

## E.11   Proof of Lemma 14

Take an arbitrary singleton set $\{\omega\}$ which is not asymptotically accessible. It is sufficient to show that this set $\{\omega\}$ is asymptotically transient.

Take an arbitrary pure strategy $s \in S^*$. Suppose that the initial prior assigns probability one on $\omega$ and that the strategy $s$ is played. Take an arbitrary history $h^{2^{|\Omega|}+1}$ with length $2^{|\Omega|}+1$ which can happen with positive probability. Given this history $h^{2^{|\Omega|}+1}$ and the initial state $\omega$, we can compute the posterior belief $\mu^t$ for each period $t \in \{1, \cdots, 2^{|\Omega|+1}+1\}$. Let $\Omega^t$ be the support of the belief $\mu^t$ in period $t$. Since $\Omega$ is finite, there must be $t$ and $\tilde{t} \neq t$ such that $\Omega^t = \Omega^{\tilde{t}}$. Pick such $t$ and $\tilde{t}$, and without loss of generality, assume that $t \leq 2^{|\Omega|}$ and $\tilde{t} \leq 2^{|\Omega|+1}$. Let $(\hat{a}^1, \cdots, \hat{a}^{\tilde{t}-t})$ be the action sequence chosen from period $t$ to period $\tilde{t}-1$, according to the history $h^{2^{|\Omega|}+1}$. Also, let $\Omega^* = \Omega^t = \Omega^{\tilde{t}}$. Since signals do not reveal the state, if the support of the initial prior is $\Omega^*$ and players play the sequence $(\hat{a}^1, \cdots, \hat{a}^{\tilde{t}-t})$ repeatedly, then the support of the posterior belief after pe-

riod $n(\tilde{t} - t)$ is also in the set $\Omega^*$ for each natural number $n$, regardless of the realized signal sequence.

Recall that our goal is to prove asymptotic transience of $\{\omega\}$. For this, it is sufficient to show that the set $\Omega^*$ is asymptotically accessible; that is, given any initial prior $\mu$, there is an action sequence and a signal sequence so that the posterior belief puts probability at least $1 - \varepsilon$ on $\Omega^*$. To prove this, the following two lemmas are useful. The first lemma shows that there is $q > 0$ such that given any initial prior $\mu$, there is an action sequence and a signal sequence so that the posterior belief puts probability at least $q$ on $\Omega^*$. Then the second lemma shows that from such a belief (i.e., a belief which puts probability at least $q$ on $\Omega^*$), we can reach some posterior belief which puts probability at least $1 - \varepsilon$ on $\Omega^*$, by letting players play $(\hat{a}^1, \cdots, \hat{a}^{\tilde{t}-t})$ repeatedly.

**Lemma 39.** *There is $q > 0$ such that for each initial prior $\mu$, there is a natural number $T \leq |\Omega|$ and an action sequence $(\tilde{a}^1, \cdots, \tilde{a}^T)$,*

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

*for some $\tilde{\mu}$ with $\sum_{\tilde{\omega} \in \Omega^*} \tilde{\mu}(\tilde{\omega}) \geq q$.*

*Proof.* Since the state evolution is irreducible, for each initial state $\hat{\omega} \in \Omega$, there is a natural number $T \leq |\Omega|$, an action sequence $(\tilde{a}^1, \cdots, \tilde{a}^T)$, and a signal sequence $(y^1, \cdots, y^T)$ such that

$$\sum_{\omega^{T+1} \in \Omega^*} \Pr(y^1, \cdots, y^T, \omega^{T+1}|\hat{\omega}, \tilde{a}^1, \cdots, \tilde{a}^T) \geq \overline{\pi}^T.$$

That is, if the initial state is $\hat{\omega}$ and players play $(\tilde{a}^1, \cdots, \tilde{a}^T)$, then the state in period $T + 1$ can be in the set $\Omega^*$ with positive probability. Let

$$q(\hat{\omega}) = \frac{\sum_{\omega^{T+1} \in \Omega^*} \Pr(y^1, \cdots, y^T, \omega^{T+1}|\hat{\omega}, \tilde{a}^1, \cdots, \tilde{a}^T)}{|\Omega| \max_{\tilde{\omega} \in \Omega} \Pr(y^1, \cdots, y^T|\tilde{\omega}, \tilde{a}^1, \cdots, \tilde{a}^T)}.$$

Here we write $q(\hat{\omega})$, because the right-hand side depends on the choice of $(\tilde{a}^1, \cdots, \tilde{a}^T)$ and $(y^1, \cdots, y^T)$, which in turn depends on the choice of $\hat{\omega}$. By the definition, $q(\hat{\omega}) > 0$ for each $\hat{\omega}$. Let $q = \min_{\hat{\omega} \in \Omega} q(\hat{\omega}) > 0$.

In what follows, we will show that this $q$ satisfies the property stated in the lemma. Pick $\mu$ arbitrarily, and then pick $\hat{\omega}$ with $\mu(\hat{\omega}) \geq \frac{1}{|\Omega|}$ arbitrarily. Choose

$T$, $(\tilde{a}^1, \cdots, \tilde{a}^T)$, and $(y^1, \cdots, y^T)$ as stated above. Let $\tilde{\mu}$ be the posterior belief after $(\tilde{a}^1, \cdots, \tilde{a}^T)$ and $(y^1, \cdots, y^T)$ given the initial prior $\mu$. Then

$$
\begin{aligned}
\sum_{\tilde{\omega} \in \Omega^*} \tilde{\mu}(\tilde{\omega}) &= \frac{\sum_{\omega^1 \in \Omega} \sum_{\omega^{T+1} \in \Omega^*} \mu(\omega^1) \Pr(y^1, \cdots, y^T, \omega^{T+1} | \omega^1, \tilde{a}^1, \cdots, \tilde{a}^T)}{\sum_{\omega^1 \in \Omega} \mu(\omega^1) \Pr(y^1, \cdots, y^T | \omega^1, \tilde{a}^1, \cdots, \tilde{a}^T)} \\
&\geq \frac{\sum_{\omega^{T+1} \in \Omega^*} \mu(\hat{\omega}) \Pr(y^1, \cdots, y^T, \omega^{T+1} | \hat{\omega}, \tilde{a}^1, \cdots, \tilde{a}^T)}{\sum_{\omega^1 \in \Omega} \mu(\omega^1) \Pr(y^1, \cdots, y^T | \omega^1, \tilde{a}^1, \cdots, \tilde{a}^T)} \\
&\geq q(\hat{\omega}) \geq q.
\end{aligned}
$$

Also, the above belief $\tilde{\mu}$ realizes with probability

$$
\Pr(y^1, \cdots, y^T | \mu, \tilde{a}^1, \cdots, \tilde{a}^T) \geq \mu(\hat{\omega}) \Pr(y^1, \cdots, y^T | \hat{\omega}, \tilde{a}^1, \cdots, \tilde{a}^T) \geq \frac{\overline{\pi}^T}{|\Omega|} \geq \frac{\overline{\pi}^{|\Omega|}}{|\Omega|},
$$

as desired. $\hfill$ *Q.E.D.*

To simplify the notation, let $\hat{\boldsymbol{a}}(n)$ be the action sequence which consists of $n$ cycles of $(\hat{a}^1, \cdots, \hat{a}^{\tilde{t}-t})$, that is,

$$
\hat{\boldsymbol{a}}(n) = (\hat{a}^1, \cdots, \hat{a}^{\tilde{t}-t}, \hat{a}^1, \cdots, \hat{a}^{\tilde{t}-t}, \cdots, \hat{a}^1, \cdots, \hat{a}^{\tilde{t}-t}).
$$

**Lemma 40.** *For each $\varepsilon > 0$ and $q > 0$, there is a natural number $n$ and $\pi^{**} > 0$ such that for each initial prior $\mu$ with $\sum_{\tilde{\omega} \in \Omega^*} \mu(\tilde{\omega}) \geq q$, there is some belief $\tilde{\mu}$ with $\sum_{\tilde{\omega} \in \Omega^*} \tilde{\mu}(\tilde{\omega}) \geq 1 - \varepsilon$ such that*

$$
\Pr(\mu^{n(\tilde{t}-t)+1} = \tilde{\mu} | \mu, \hat{\boldsymbol{a}}(n)) > \pi^{**}.
$$

*Proof.* If $\Omega^*$ is the whole state space, the result holds obviously. So we assume that $\Omega^*$ is a proper subset of $\Omega$. With an abuse of notation, let $\triangle \Omega^*$ be the set of all priors $\mu \in \triangle \Omega$ which puts probability one on $\Omega^*$; i.e., $\triangle \Omega^*$ is the set of all $\mu$ such that $\sum_{\tilde{\omega} \in \Omega^*} \mu(\tilde{\omega}) = 1$. Likewise, let $\triangle(\Omega \setminus \Omega^*)$ be the set of all priors $\mu \in \triangle \Omega$ which puts probability zero on $\Omega^*$.

Let $\Pr(\cdot | \omega, \hat{a}^1, \cdots, \hat{a}^{\tilde{t}-t})$ be the probability distribution of $(y^1, \cdots, y^{\tilde{t}-t})$ induced by the initial state $\omega$ and the action sequence $(\hat{a}^1, \cdots, \hat{a}^{\tilde{t}-t})$. Similarly, let $\Pr(\cdot | \mu, \hat{a}^1, \cdots, \hat{a}^{\tilde{t}-t})$ be the distribution when the initial prior is $\mu$. Since the signal distributions $\{\pi_Y^\omega(a) | \omega \in \Omega\}$ are linearly independent for $a = \hat{a}^1$, the distributions $\{\Pr(\cdot | \omega, \hat{a}^1, \cdots, \hat{a}^{\tilde{t}-t}) | \omega \in \Omega\}$ are also linearly independent. Hence the convex hull of $\{\Pr(\cdot | \omega, \hat{a}^1, \cdots, \hat{a}^{\tilde{t}-t}) | \omega \in \Omega^*\}$ and that of $\{\Pr(\cdot | \omega, \hat{a}^1, \cdots, \hat{a}^{\tilde{t}-t} | \omega \notin \Omega^*\}$

do not intersect. Let $\kappa > 0$ be the distance between these convex hulls. Then for each $\overline{\mu} \in \triangle \Omega^*$ and $\underline{\mu} \in \triangle(\Omega \setminus \Omega^*)$,

$$\left| \Pr(\cdot | \overline{\mu}, \hat{a}^1, \cdots, \hat{a}^{\tilde{t}-t}) - \Pr(\cdot | \underline{\mu}, \hat{a}^1, \cdots, \hat{a}^{\tilde{t}-t}) \right| \geq \kappa.$$

This implies that there is $\tilde{\pi} \in (0,1)$ such that for each $\overline{\mu} \in \triangle \Omega^*$ and $\underline{\mu} \in \triangle(\Omega \setminus \Omega^*)$, there is a signal sequence $(y^1, \cdots, y^{\tilde{t}-t})$ such that

$$\Pr(y^1, \cdots, y^{\tilde{t}-t} | \overline{\mu}, \hat{a}^1, \cdots, \hat{a}^{\tilde{t}-t}) \geq \Pr(y^1, \cdots, y^{\tilde{t}-t} | \underline{\mu}, \hat{a}^1, \cdots, \hat{a}^{\tilde{t}-t}) + \tilde{\pi}. \quad (26)$$

Pick such a number $\tilde{\pi} \in (0,1)$.

Choose $\varepsilon > 0$ and $q > 0$ arbitrarily, and let $n$ be a natural number such that

$$\left( \frac{1}{1-\tilde{\pi}} \right)^n \frac{q}{1-q} \geq \frac{1-\varepsilon}{\varepsilon}. \quad (27)$$

Since $\tilde{\pi} \in (0,1)$, the existence of $n$ is guaranteed. In what follows, we will show that this $n$ and $\pi^{**} = (q\tilde{\pi})^n$ satisfy the condition stated in the lemma.

Pick $\mu$ such that $\sum_{\tilde{\omega} \in \Omega^*} \mu(\tilde{\omega}) \geq q$, and let $\overline{\mu} \in \triangle \Omega^*$ be such that $\overline{\mu}(\omega) = \frac{\mu(\omega)}{\sum_{\tilde{\omega} \in \Omega^*} \mu(\tilde{\omega})}$ for each $\omega \in \Omega^*$ and $\overline{\mu}(\omega) = 0$ for $\omega \notin \Omega^*$. Likewise, let $\underline{\mu} \in \triangle(\Omega \setminus \Omega^*)$ be such that $\underline{\mu}(\omega) = \frac{\mu(\omega)}{\sum_{\tilde{\omega} \notin \Omega^*} \mu(\tilde{\omega})}$ for each $\omega \notin \Omega^*$ and $\underline{\mu}(\omega) = 0$ for $\omega \in \Omega^*$. Choose $(y^1, \cdots, y^{\tilde{t}-t})$ so that (26) holds for this $\overline{\mu}$ and $\underline{\mu}$. Let $\hat{\mu}$ be the posterior belief in period $\tilde{t} - t + 1$ when the initial prior is $\mu$, players play $(\hat{a}^1, \cdots, \hat{a}^{\tilde{t}-t})$ and observe $(y^1, \cdots, y^{\tilde{t}-t})$.

By the definition of $\Omega^*$, if the initial state is in the set $\Omega^*$ and players play $(\hat{a}^1, \cdots, \hat{a}^{\tilde{t}-t})$, then the state in period $\tilde{t} - t + 1$ must be in the set $\Omega^*$. (To see this, suppose not and the state in period $\tilde{t} - t + 1$ can be $\tilde{\omega} \notin \Omega^*$ with positive probability. Then, since signals do not reveal the current or next state, the set $\Omega^{\tilde{t}}$ must contain $\tilde{\omega}$, implying that $\Omega^{\tilde{t}} \neq \Omega^*$. This is a contradiction.) That is, we must have

$$\Pr(y^1, \cdots, y^{\tilde{t}-t}, \omega^{\tilde{t}-t+1} | \omega^1, \hat{a}^1, \cdots, \hat{a}^{\tilde{t}-t}) = 0 \quad (28)$$

for all $\omega^1 \in \Omega^*$ and $\omega^{\tilde{t}-t+1} \notin \Omega^*$. Then we have

$$
\frac{\sum_{\tilde{\omega} \in \Omega^*} \hat{\mu}(\tilde{\omega})}{\sum_{\tilde{\omega} \notin \Omega^*} \hat{\mu}(\tilde{\omega})}
$$

$$
= \frac{\sum_{\omega^1 \in \Omega} \sum_{\omega^{\tilde{t}-t+1} \in \Omega^*} \mu(\omega^1) \Pr(y^1, \cdots, y^{\tilde{t}-t}, \omega^{\tilde{t}-t+1} | \omega^1, \hat{a}^1, \cdots, \hat{a}^{\tilde{t}-t})}{\sum_{\omega^1 \in \Omega} \sum_{\omega^{\tilde{t}-t+1} \notin \Omega^*} \mu(\omega^1) \Pr(y^1, \cdots, y^{\tilde{t}-t}, \omega^{\tilde{t}-t+1} | \omega^1, \hat{a}^1, \cdots, \hat{a}^{\tilde{t}-t})}
$$

$$
\geq \frac{\sum_{\omega^1 \in \Omega^*} \sum_{\omega^{\tilde{t}-t+1} \in \Omega^*} \mu(\omega^1) \Pr(y^1, \cdots, y^{\tilde{t}-t}, \omega^{\tilde{t}-t+1} | \omega^1, \hat{a}^1, \cdots, \hat{a}^{\tilde{t}-t})}{\sum_{\omega^1 \in \Omega} \sum_{\omega^{\tilde{t}-t+1} \notin \Omega^*} \mu(\omega^1) \Pr(y^1, \cdots, y^{\tilde{t}-t}, \omega^{\tilde{t}-t+1} | \omega^1, \hat{a}^1, \cdots, \hat{a}^{\tilde{t}-t})}
$$

$$
= \frac{\sum_{\omega^1 \in \Omega^*} \sum_{\omega^{\tilde{t}-t+1} \in \Omega} \mu(\omega^1) \Pr(y^1, \cdots, y^{\tilde{t}-t}, \omega^{\tilde{t}-t+1} | \omega^1, \hat{a}^1, \cdots, \hat{a}^{\tilde{t}-t})}{\sum_{\omega^1 \notin \Omega^*} \sum_{\omega^{\tilde{t}-t+1} \notin \Omega^*} \mu(\omega^1) \Pr(y^1, \cdots, y^{\tilde{t}-t}, \omega^{\tilde{t}-t+1} | \omega^1, \hat{a}^1, \cdots, \hat{a}^{\tilde{t}-t})}
$$

$$
\geq \frac{\sum_{\omega^1 \in \Omega^*} \sum_{\omega^{\tilde{t}-t+1} \in \Omega} \mu(\omega^1) \Pr(y^1, \cdots, y^{\tilde{t}-t}, \omega^{\tilde{t}-t+1} | \omega^1, \hat{a}^1, \cdots, \hat{a}^{\tilde{t}-t})}{\sum_{\omega^1 \notin \Omega^*} \sum_{\omega^{\tilde{t}-t+1} \in \Omega} \mu(\omega^1) \Pr(y^1, \cdots, y^{\tilde{t}-t}, \omega^{\tilde{t}-t+1} | \omega^1, \hat{a}^1, \cdots, \hat{a}^{\tilde{t}-t})}
$$

$$
= \frac{\sum_{\omega^1 \in \Omega^*} \mu(\omega^1) \Pr(y^1, \cdots, y^{\tilde{t}-t} | \omega^1, \hat{a}^1, \cdots, \hat{a}^{\tilde{t}-t})}{\sum_{\omega^1 \notin \Omega^*} \mu(\omega^1) \Pr(y^1, \cdots, y^{\tilde{t}-t} | \omega^1, \hat{a}^1, \cdots, \hat{a}^{\tilde{t}-t})}
$$

$$
= \frac{\Pr(y^1, \cdots, y^{\tilde{t}-t} | \overline{\mu}, \hat{a}^1, \cdots, \hat{a}^{\tilde{t}-t}) \sum_{\tilde{\omega} \in \Omega^*} \mu(\tilde{\omega})}{\Pr(y^1, \cdots, y^{\tilde{t}-t} | \underline{\mu}, \hat{a}^1, \cdots, \hat{a}^{\tilde{t}-t}) \sum_{\tilde{\omega} \notin \Omega^*} \mu(\tilde{\omega})}.
$$

Here, the second equality comes from (28). Since $(y^1, \cdots, y^{\tilde{t}-t})$ satisfies (26), we have

$$
\frac{\Pr(y^1, \cdots, y^{\tilde{t}-t} | \overline{\mu}, \hat{a}^1, \cdots, \hat{a}^{\tilde{t}-t})}{\Pr(y^1, \cdots, y^{\tilde{t}-t} | \underline{\mu}, \hat{a}^1, \cdots, \hat{a}^{\tilde{t}-t})} \geq \frac{1}{1 - \tilde{\pi}}.
$$

Applying this inequality to the previous one, we obtain

$$
\frac{\sum_{\tilde{\omega} \in \Omega^*} \hat{\mu}(\tilde{\omega})}{\sum_{\tilde{\omega} \notin \Omega^*} \hat{\mu}(\tilde{\omega})} \geq \frac{1}{1 - \tilde{\pi}} \cdot \frac{\sum_{\tilde{\omega} \in \Omega^*} \mu(\tilde{\omega})}{\sum_{\tilde{\omega} \notin \Omega^*} \mu(\tilde{\omega})}.
$$

Since $\frac{1}{1-\tilde{\pi}} > 1$, this inequality implies that the likelihood of $\Omega^*$ induced by the posterior belief $\hat{\mu}$ is greater than the likelihood of $\Omega^*$ induced by the initial prior $\mu$. Note also that such a posterior belief $\hat{\mu}$ realizes with probability at least $q\tilde{\pi}$, since (26) implies

$$
\Pr(y^1, \cdots, y^{\tilde{t}-t} | \mu, \hat{a}^1, \cdots, \hat{a}^{\tilde{t}-t}) \geq q \Pr(y^1, \cdots, y^{\tilde{t}-t} | \overline{\mu}, \hat{a}^1, \cdots, \hat{a}^{\tilde{t}-t}) \geq q\tilde{\pi}.
$$

In sum, we have shown that given any initial prior $\mu$ with $\sum_{\tilde{\omega} \in \Omega^*} \mu(\tilde{\omega}) \geq q$, if players play the action sequence $(\hat{a}^1, \cdots, \hat{a}^{\tilde{t}-t})$, then with probability at least $q\tilde{\pi}$, the posterior belief $\mu^{\tilde{t}-t+1}$ satisfies

$$
\frac{\sum_{\tilde{\omega} \in \Omega^*} \mu^{\tilde{t}-t+1}(\tilde{\omega})}{\sum_{\tilde{\omega} \notin \Omega^*} \mu^{\tilde{t}-t+1}(\tilde{\omega})} \geq \frac{1}{1 - \tilde{\pi}} \cdot \frac{\sum_{\tilde{\omega} \in \Omega^*} \mu(\tilde{\omega})}{\sum_{\tilde{\omega} \notin \Omega^*} \mu(\tilde{\omega})}. \tag{29}
$$

Now suppose that we are in period $\tilde{t} - t + 1$ and the posterior belief $\mu^{\tilde{t}-t+1}$ satisfies (29). Note that (29) implies $\sum_{\omega \in \Omega^*} \mu^{\tilde{t}-t+1}(\omega) \geq q$, and hence we can apply the above argument once again; that is, we can show that if players play the action sequence $(\hat{a}^1, \cdots, \hat{a}^{\tilde{t}-t})$ again for the next $\tilde{t} - t$ periods, then with probability at least $q\tilde{\pi}$, the posterior belief $\mu^{2(\tilde{t}-t)+1}$ satisfies

$$\frac{\sum_{\tilde{\omega} \in \Omega^*} \mu^{2(\tilde{t}-t)+1}(\tilde{\omega})}{\sum_{\tilde{\omega} \notin \Omega^*} \mu^{2(\tilde{t}-t)+1}(\tilde{\omega})} \geq \frac{1}{1-\tilde{\pi}} \cdot \frac{\sum_{\tilde{\omega} \in \Omega^*} \hat{\mu}(\tilde{\omega})}{\sum_{\tilde{\omega} \notin \Omega^*} \hat{\mu}(\tilde{\omega})} \geq \left(\frac{1}{1-\tilde{\pi}}\right)^2 \cdot \frac{\sum_{\tilde{\omega} \in \Omega^*} \mu(\tilde{\omega})}{\sum_{\tilde{\omega} \notin \Omega^*} \mu(\tilde{\omega})}.$$

Iterating this argument, we can prove the following statement: Given that the initial prior is $\mu$, if players play the action sequence $\hat{\boldsymbol{a}}(n)$, then with probability at least $\pi^{**} = (q\tilde{\pi})^n$, the posterior belief $\mu^{n(\tilde{t}-t)+1}$ satisfies

$$\frac{\sum_{\tilde{\omega} \in \Omega^*} \mu^{n(\tilde{t}-t)+1}(\tilde{\omega})}{\sum_{\tilde{\omega} \notin \Omega^*} \mu^{n(\tilde{t}-t)+1}(\tilde{\omega})} \geq \left(\frac{1}{1-\tilde{\pi}}\right)^n \cdot \frac{\sum_{\tilde{\omega} \in \Omega^*} \mu(\tilde{\omega})}{\sum_{\tilde{\omega} \notin \Omega^*} \mu(\tilde{\omega})}.$$

Plugging (27) and $\frac{\sum_{\tilde{\omega} \in \Omega^*} \mu(\tilde{\omega})}{\sum_{\tilde{\omega} \notin \Omega^*} \mu(\tilde{\omega})} \geq \frac{q}{1-q}$ into this,

$$\frac{\sum_{\tilde{\omega} \in \Omega^*} \mu^{n(\tilde{t}-t)+1}(\tilde{\omega})}{\sum_{\tilde{\omega} \notin \Omega^*} \mu^{n(\tilde{t}-t)+1}(\tilde{\omega})} \geq \frac{1-\varepsilon}{\varepsilon}.$$

This implies that the posterior belief puts probability at least $1 - \varepsilon$ on $\Omega^*$, as desired. *Q.E.D.*

Fix $\varepsilon > 0$ arbitrarily. Choose $q$ as stated in Lemma 39, and then choose $n$ and $\pi^{**}$ as stated in Lemma 40. Then the above two lemmas ensure that given any initial prior $\mu$, there is an action sequence with length $T^* \leq |\Omega| + n(\tilde{t} - t)$ such that with probability at least $\pi^* = \frac{\overline{\pi}^{|\Omega|}\pi^{**}}{|\Omega|}$, the posterior belief puts probability at least $1 - \varepsilon$ on $\Omega^*$. Since the bounds $|\Omega| + n(\tilde{t} - t)$ and $\pi^{**}$ do not depend on the initial prior $\mu$, this shows that $\Omega^*$ is asymptotically accessible, and hence $\{\omega\}$ is asymptotically transient.

## E.12  Proof of Lemma 15

Fix $\delta$ and $\lambda$. Let $s^{\mu}$ and $\omega$ be as in the proof of Lemma 5. We first prove the following lemma, which says that if two initial priors $\mu$ and $\tilde{\mu}$ are close, then the corresponding scores are also close.

**Lemma 41.** *For any* $\varepsilon \in (0, \frac{1}{|\Omega|})$, $\mu$, *and* $\tilde{\mu}$ *with* $\max_{\tilde{\omega} \in \Omega} |\mu(\tilde{\omega}) - \tilde{\mu}(\tilde{\omega})| \leq \varepsilon$,

$$\left| \lambda \cdot v^\mu(\delta, s^\mu) - \lambda \cdot v^{\tilde{\mu}}(\delta, s^{\tilde{\mu}}) \right| \leq \varepsilon \bar{g} |\Omega|.$$

*Proof.* Without loss of generality, assume that $\lambda \cdot v^\mu(\delta, s^\mu) \geq \lambda \cdot v^{\tilde{\mu}}(\delta, s^{\tilde{\mu}})$. Then

$$
\begin{aligned}
\left| \lambda \cdot v^\mu(\delta, s^\mu) - \lambda \cdot v^{\tilde{\mu}}(\delta, s^{\tilde{\mu}}) \right| &\leq \left| \lambda \cdot v^\mu(\delta, s^\mu) - \lambda \cdot v^{\tilde{\mu}}(\delta, s^\mu) \right| \\
&= \left| \sum_{\tilde{\omega} \in \Omega} \mu(\tilde{\omega}) \lambda \cdot v^{\tilde{\omega}}(\delta, s^\mu) - \sum_{\tilde{\omega} \in \Omega} \tilde{\mu}(\tilde{\omega}) \lambda \cdot v^{\tilde{\omega}}(\delta, s^\mu) \right| \\
&\leq \sum_{\tilde{\omega} \in \Omega} \lambda \cdot v^{\tilde{\omega}}(\delta, s^{\tilde{\omega}}) |\mu(\tilde{\omega}) - \tilde{\mu}(\tilde{\omega})|.
\end{aligned}
$$

Since $\lambda \cdot v^{\tilde{\omega}}(\delta, s^{\tilde{\omega}}) \leq \bar{g}$ and $|\mu(\tilde{\omega}) - \tilde{\mu}(\tilde{\omega})| \leq \varepsilon$, the result follows. $\qquad$ *Q.E.D.*

Since there are only finitely many subsets $\Omega^* \subset \Omega$, there is $\tilde{\pi}^* > 0$ such that for each asymptotically transient $\Omega^*$, $\tilde{\pi}^*$ satisfies the condition stated in the definition of asymptotic transience. Pick such $\tilde{\pi}^* > 0$.

Pick $\varepsilon \in (0, \frac{1}{|\Omega|})$ arbitrarily. Then there is a natural number $T$ and $\pi^* > 0$ such that for each asymptotically accessible $\Omega^*$, $T$ and $\pi^*$ satisfy the condition stated in the definition of asymptotically accessibility, and such that for each asymptotically transient $\Omega^*$, $T$ satisfies the condition stated in the definition of asymptotic transience. Pick such $T$ and $\pi^* > 0$.

Since the game is asymptotically connected, $\{\omega\}$ is either asymptotically accessible or asymptotically transient. When it is asymptotically transient, there is an asymptotically accessible set $\Omega^*$, a natural number $T^* \leq T$, and a signal sequence $(y^1, \cdots, y^{T^*})$ such that if the initial state is $\omega$ and players play $s^\omega$, then the signal sequence $(y^1, \cdots, y^{T^*})$ appears with positive probability and the resulting posterior belief $\mu(y^1, \cdots, y^{T^*})$ satisfies $\sum_{\tilde{\omega} \in \Omega^*} \mu(y^1, \cdots, y^{T^*})[\tilde{\omega}] \geq 1 - \varepsilon$ and $\mu(y^1, \cdots, y^{T^*})[\tilde{\omega}] \geq \tilde{\pi}^*$ for all $\tilde{\omega} \in \Omega^*$. Take such $\Omega^*$, $T^*$, and $(y^1, \cdots, y^{T^*})$. Let $\pi(y^1, \cdots, y^{T^*})$ be the probability that the signal sequence $(y^1, \cdots, y^{T^*})$ happens given the initial state $\omega$ and the strategy profile $s^\omega$, and let $\mu^* = \mu(y^1, \cdots, y^{T^*})$ be the resulting posterior belief. Since $(y^1, \cdots, y^{T^*})$ is induced by the initial state $\omega$ and the pure strategy $s^\omega$, we have $\pi(y^1, \cdots, y^{T^*}) \geq \bar{\pi}^{T^*}$.

When $\{\omega\}$ is asymptotically accessible, let $\Omega^* = \{\omega\}$, $T^* = 0$, and $\mu^* \in \triangle \Omega$ with $\mu^*(\omega) = 1$.

Let $C(T) = \frac{2\overline{g}}{\pi^{T*}}$ and $C(T) = \frac{2\overline{g}}{\pi^{\overline{T}}}$. Since Lemma 22 holds, we have

$$\left| \lambda \cdot v^\omega(\delta, s^\omega) - \lambda \cdot v^{\mu^*}(\delta, s^{\mu^*}) \right| \leq \frac{1-\delta^{T^*}}{\delta^{T^*}} C(T^*) \leq \frac{1-\delta^T}{\delta^T} C(T). \qquad (30)$$

That is, the score with the initial prior $\mu^*$ is close to the maximal score, when $\delta$ is close to one. (Recall that $T$ and $C(T)$ depends on $\varepsilon$ but not on $\delta$.)

Note that the belief $\mu^*$ approximates some belief whose support is $\Omega^*$, that is, $\sum_{\tilde{\omega} \in \Omega^*} \mu^*[\tilde{\omega}] \geq 1 - \varepsilon$. Note also that $\mu^*[\tilde{\omega}] \geq \tilde{\pi}^*$ for all $\tilde{\omega} \in \Omega^*$. Hence there is a belief $\tilde{\mu}^*$ whose support is $\Omega^*$ such that $\tilde{\mu}^*[\tilde{\omega}] \geq \tilde{\pi}^*$ for all $\tilde{\omega} \in \Omega^*$, and such that $\tilde{\mu}^*$ is $\varepsilon$-close to $\mu^*$ in that $\max_{\tilde{\omega} \in \Omega} |\mu^*(\tilde{\omega}) - \tilde{\mu}^*(\tilde{\omega})| \leq \varepsilon$. Lemma 41 implies that these two beliefs $\mu^*$ and $\tilde{\mu}^*$ induce similar scores, that is,

$$\left| \lambda \cdot v^{\mu^*}(\delta, s^{\mu^*}) - \lambda \cdot v^{\tilde{\mu}^*}(\delta, s^{\tilde{\mu}^*}) \right| \leq \varepsilon \overline{g} |\Omega|.$$

Plugging this into (30), we obtain the following lemma, which says that the score with the initial prior $\tilde{\mu}^*$ is close to the maximal score. Let $s^* = s^{\tilde{\mu}^*}$.

**Lemma 42.** *We have*

$$\left| \lambda \cdot v^\omega(\delta, s^\omega) - \lambda \cdot v^{\tilde{\mu}^*}(\delta, s^*) \right| \leq \frac{1-\delta^T}{\delta^T} C(T) + \varepsilon \overline{g} |\Omega|.$$

Recall that the support of $\tilde{\mu}^*$ is $\Omega^*$. The next lemma shows that the strategy profile $s^*$ approximates the maximal score even if the initial state is $\tilde{\omega} \in \Omega^*$. Let $C = \frac{C(T)}{\tilde{\pi}^*}$.

**Lemma 43.** *For each $\tilde{\omega} \in \Omega^*$,*

$$\left| \lambda \cdot v^\omega(\delta, s^\omega) - \lambda \cdot v^{\tilde{\omega}}(\delta, s^*) \right| \leq \frac{1-\delta^T}{\delta^T} C + \frac{\varepsilon \overline{g} |\Omega|}{\tilde{\pi}^*}.$$

*Proof.* Very similar to that of Lemma 23. Specifically, replace $\lambda \cdot v^{\mu^*}(\delta, s^*)$ in the proof of Lemma 23 with $\lambda \cdot v^{\tilde{\mu}^*}(\delta, s^*)$. Use Lemma 42 instead of Lemma 22. *Q.E.D.*

The next lemma also shows that the strategy profile $s^*$ approximates the maximal score for any initial state is $\mu$ whose support is $\Omega^*$. The proof is very similar to that of Lemma 24 and hence omitted.

**Lemma 44.** *For each $\mu$ such that $\mu(\tilde{\omega}) = 0$ for all $\tilde{\omega} \notin \Omega^*$,*

$$|\lambda \cdot v^\omega(\delta, s^\omega) - \lambda \cdot v^\mu(\delta, s^*)| \leq \frac{1 - \delta^T}{\delta^T} C + \frac{\varepsilon \overline{g} |\Omega|}{\tilde{\pi}^*}.$$

The above lemma, together with Lemma 41, implies the following lemma: It says that the strategy profile $s^*$ approximates the maximal score for any initial prior $\mu$ which approximates some belief whose support is $\Omega^*$.

**Lemma 45.** *For each $\mu$ such that $\sum_{\tilde{\omega} \in \Omega^*} \mu(\tilde{\omega}) \geq 1 - \varepsilon$,*

$$|\lambda \cdot v^\omega(\delta, s^\omega) - \lambda \cdot v^\mu(\delta, s^*)| \leq \frac{1 - \delta^T}{\delta^T} C + \frac{2\varepsilon \overline{g} |\Omega|}{\tilde{\pi}^*}.$$

Let $\mu^{**}$ be such that $\mu^{**}(\omega) = \frac{1}{|\Omega|}$ for each $\omega$. Since $\Omega^*$ is asymptotically accessible, for any initial prior $\mu$, there is a natural number $T(\mu) \leq T$ and an action sequence $a(\mu) = (a^1(\mu), \cdots, a^{T(\mu)}(\mu))$ such that the probability that the posterior belief $\mu^{T(\mu)+1}$ satisfies $\sum_{\tilde{\omega} \in \Omega^*} \mu^{T(\mu)+1}(\tilde{\omega}) \geq 1 - \varepsilon$ is at least $\pi^*$. Let $s^{**}$ be the following strategy profile:

- Let $\mu^{(1)} = \mu^{**}$.

- Players play the action sequence $a(\mu^{(1)})$ for the first $T(\mu^{(1)})$ periods.

- If the posterior belief $\mu^{(2)}$ satisfies $\sum_{\tilde{\omega} \in \Omega^*} \mu^{(2)}(\tilde{\omega}) \geq 1 - \varepsilon$, then players play $s^*$ in the continuation game.

- If not, players play $a(\mu^{(2)})$ for the next $T(\mu^{(2)})$ periods.

- If the posterior belief $\mu^{(3)}$ satisfies $\sum_{\tilde{\omega} \in \Omega^*} \mu^{(3)}(\tilde{\omega}) \geq 1 - \varepsilon$, then players play $s^*$ in the continuation game.

- And so on.

Let $\tilde{C} = \frac{4\overline{g}}{\pi^*}$. The following lemma is a counterpart to Lemma 25, which shows that the strategy $s^{**}$ can approximate the maximal score when the initial prior is $\mu^{**}$.

**Lemma 46.** *We have*

$$\left| \lambda \cdot v^\omega(\delta, s^\omega) - \lambda \cdot v^{\mu^{**}}(\delta, s^{**}) \right| \leq \frac{1 - \delta^T}{\delta^T} C + (1 - \delta^T)\tilde{C} + \frac{2\varepsilon \overline{g} |\Omega|}{\tilde{\pi}^*}.$$

*Proof.* The proof is essentially the same as that of Lemma 25; we simply replace $4^{|\Omega|}$ in the proof of Lemma 25 with $T$, and use Lemma 45 instead of Lemma 24. *Q.E.D.*

The next lemma extends the above one; it shows that the same result holds regardless of the initial prior $\mu$.

**Lemma 47.** *For all $\mu \in \triangle\Omega$,*

$$|\lambda \cdot v^\omega(\delta, s^\omega) - \lambda \cdot v^\mu(\delta, s^{**})| \leq \frac{1-\delta^T}{\delta^T}C|\Omega| + (1-\delta^T)\tilde{C}|\Omega| + \frac{2\varepsilon\overline{g}|\Omega|^2}{\tilde{\pi}^*}.$$

*Proof.* The proof is simply a combination of those of Lemmas 26 and 27. The only difference is to use Lemma 46 instead of Lemma 25. *Q.E.D.*

From Lemma 47 and (11), we have

$$|\lambda \cdot v^\omega(\delta, s^\omega) - \lambda \cdot v^\mu(\delta, s^\mu)| \leq \frac{1-\delta^T}{\delta^T}C|\Omega| + (1-\delta^T)\tilde{C}|\Omega| + \frac{2\varepsilon\overline{g}|\Omega|^2}{\tilde{\pi}^*}.$$

Note that $T$ and $\pi^*$ depend on $\varepsilon$ but not on $\delta$ and $\lambda$. This in turn implies $C$ and $\tilde{C}$ depend on $\varepsilon$ but not on $\delta$ and $\lambda$. Note also that $\tilde{\pi}^*$ does not depend on $\varepsilon$, $\delta$, or $\lambda$. Hence the above inequality implies that the left-hand side can be arbitrarily small for all $\lambda$, if we take $\varepsilon$ close to zero and then take $\delta$ close to one. This proves clause (i).

The proof of clause (ii) is similar to that of Lemma 5.

## E.13   Proof of Lemma 17

Fix $\mu$, $s$, $\lambda$, $p$, $\delta$, and $v$ as stated. Consider the game $\Gamma(\mu, \delta, p, \tilde{w})$ with a function $\tilde{w}$ such that $\tilde{w}_i(h^t) = \tilde{v}_i$ for all $i$ and $h^t$; i.e., we consider the case in which players' payoffs when the game terminates are constant. Let

$$\tilde{v}_i = \frac{v_i - (1-\delta)E\left[\sum_{t=1}^\infty (\delta p)^{t-1} g_i^{\omega^t}(a^t) \Big| \mu, s\right]}{(1-p)\sum_{t=1}^\infty p^{t-1}\delta^t}.$$

Intuitively, we choose $\tilde{v}_i$ in such a way that player $i$'s expected payoff in the game $\Gamma(\mu, \delta, p, \tilde{w})$ given the strategy profile $s$ is equal to $v_i$; indeed, we have

$$(1-\delta)E\left[\sum_{t=1}^\infty (p\delta)^{t-1} g_i^{\omega^t}(a^t) \Big| \mu, s\right] + (1-p)\sum_{t=1}^\infty p^{t-1}\delta^t\tilde{v}_i = v_i.$$

Also, let $v_i(a_i)$ denote player $i$'s payoff when she makes a one-shot deviation to $a_i$ in period one.

Now let $z^1 : H^1 \to \mathbf{R}^N$ be such that

$$v_i = v_i(a_i) + (1-p)\delta \sum_{y \in Y} \pi^\mu(y|a_i, s_{-i}(h^0)) z_i^1(a_i, s_{-i}(h^0), y)$$

for all $a_i \in A_i$, and

$$\lambda \cdot z^1(a, y) = 0$$

for all $a \in A$ and $y \in Y$. The first equation implies that player $i$ is indifferent over all actions $a_i$ in the first period of the game $\Gamma(\mu, \delta, p, \tilde{w})$, if she can obtain an additional "bonus payment" $z_i^1(a, y)$ when the game terminates at the end of period one. The second equation implies that the bonus payment vector $z^1(a, y)$ is on the linear hyperplane tangential to $\lambda$. The existence of such $z^1$ comes from the fact that actions are perfectly observable. The proof is very similar to that of Lemmas 5.3 and 5.4 of FLM and hence omitted. Note also that for each $\lambda$, we can take $\tilde{K} > 0$ such that $|z^1(a, y)| < \frac{1-\delta}{(1-p\delta)(1-p)\delta} \tilde{K}$ for all $\mu$, $\delta$, and $p$; this is because $|v_i - v_i(a_i)|$ is at most $\frac{1-\delta}{1-p\delta} \overline{g}$. (Note that the term $(1-p) \sum_{t=1}^\infty p^{t-1} \delta^t \tilde{v}_i$ appears both in $v_i$ and $v_i(a_i)$, so that it cancels out when we compute the difference $v_i - v_i(a_i)$.)

Similarly, for each $h^t$, we will specify a function $z^{t+1}$ to provide appropriate incentives in period $t+1$ of the game $\Gamma(\mu, \delta, p, \tilde{w})$. For each history $h^t$, let $v_i(h^t)$ denote player $i$'s continuation payoff after $h^t$ in the game $\Gamma(\mu, \delta, p, \tilde{w})$ when players play $s|_{h^t}$. Also let $v_i(h^t, a_i)$ denote her continuation payoff when she makes a one-shot deviation to $a_i$ in the next period. Let $\mu(h^t)$ represent the posterior belief after $h^t$. Let $z^{t+1} : H^{t+1} \to \mathbf{R}^N$ be such that

$$v_i(h^t) = v_i(h^t, a_i) + (1-p)\delta \sum_{y \in Y} \pi^{\mu(h^t)}(y|a_i, s_{-i}(h^t)) z_i^{t+1}(h^t, (a_i, s_{-i}(h^t), y)) \tag{31}$$

for all $h^t$ and $a_i \in A_i$, and

$$\lambda \cdot z^{t+1}(h^t, (a, y)) = 0 \tag{32}$$

for all $a \in A$ and $y \in Y$. To see the meaning of (31), suppose that now we are in period $t+1$ of the game $\Gamma(\mu, \delta, p, \tilde{w})$ and that the past history was $h^t$. (31) implies

that player $i$ is indifferent over all actions in the current period if she can obtain a bonus payment $z_i^{t+1}(h^{t+1})$ when the game terminates at the end of period $t+1$. Also, (32) asserts that the bonus payment $z^{t+1}(h^{t+1})$ is on the linear hyperplane tangential to $\lambda$. Note that for each $\lambda$, we can take $\tilde{K} > 0$ such that

$$|z^{t+1}(h^t)| < \frac{1-\delta}{(1-p\delta)(1-p)\delta}\tilde{K} \tag{33}$$

for all $h^t$, $\delta$, and $p$. Here we can choose $\tilde{K}$ uniformly in $h^t$, since actions are observable and there are only finitely many pure action profiles.

Now we construct the continuation payoff function $w$. Let $w : H \to \boldsymbol{R}$ be such that

$$w_i(h^t) = \tilde{v}_i + z_i^t(h^t) \tag{34}$$

for each $i$, $t \geq 1$, and $h^t$. That is, we consider the continuation payoff function $w$ which gives the constant value $\tilde{v}_i$ and the bonus payment $z_i^t(h^t)$ to player $i$ when the game terminates at the end of period $t$. This continuation payoff function makes player $i$ indifferent over all actions in each period $t$, regardless of the past history. (Note that $z_i^t(h^t)$ does not influence player $i$'s incentive in earlier periods $\tilde{t} < t$, since (31) implies that the expected value of $z_i^t(h^t)$ conditional on $h^{t-1}$ is equal to zero for all $h^{t-1}$ as long as player $i$ does not deviate in period $t$.) Also, the resulting payoff vector is $v$. Therefore clause (i) follows.

By the definition of $\tilde{v}_i$, we have

$$\tilde{v}_i = v_i - \frac{1-\delta}{(1-p)\delta}(v_i^\mu(p\delta,s) - v_i). \tag{35}$$

This, together with (32) and (34), proves clause (ii). Also, from (34) and (35), we have

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

Then from (33),

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

Since $v \in V$, we have $|v^\omega(p\delta,s) - v| \leq \overline{g}$. Hence, by letting $K > \frac{\tilde{K}}{1-p} + \overline{g}$, we have clause (iii).

## E.14 Proof of Lemma 18

Fix $p$, $s_{-i}$, $\mu$, $\delta$, $v$, and $s_i$ as stated. For each $j \neq i$, let $\tilde{v}_j$, $v_j(h^t)$, and $v_j(h^t, a_j)$ be as in the proof of Lemma 17. Then for each $h^t$, let $z_j^{t+1} : H^{t+1} \to \mathbf{R}$ be such that

$$v_j(h^t) = v_j(h^t, a_j) + (1-p)\delta \sum_{y \in Y} \pi^\mu(y|a_j, s_{-j}(h^t)) z_j^{t+1}(h^t, (a_j, s_{-j}(h^t)), y))$$

for all $h^t$ and $a_j \in A_j$, and let

$$w_j(h^t) = \tilde{v}_j + z_j^t(h^t).$$

Then player $j$'s incentive constraints are satisfied.

Now, consider player $i$'s continuation payoff $w_i$. Let $\tilde{v}_i$ be as in the proof of Lemma 17, and let

$$w_i(h^t) = \tilde{v}_i$$

for all $h^t$. With this constant continuation payoff, player $i$'s incentive constraints are satisfied because $s_i$ is a best reply to $s_{-i}$ given initial prior $\mu$ and discount factor $p\delta$. Hence clause (i) follows. From (35), we have

$$w_i(h^t) = v_i - \frac{1-\delta}{(1-p)\delta}(v_i^\mu(p\delta, s) - v_i),$$

which proves clause (ii) follows. Also, letting $K > \overline{g}_i$, we have $K > |v_i^\mu(p\delta, s) - v_i|$ so that clause (iii) follows.

## E.15 Proof of Lemma 19

Fix $W$ as stated, and take $\varepsilon$ so that the $\varepsilon$-neighborhood of $W$ is in the interior of $V^*$. Then from Lemma 10, there is $p$ such that the $\varepsilon$-neighborhood of $W$ is included in the feasible payoff set $V^\mu(p)$ and such that $v_i - \varepsilon > \underline{v}_i(p)$ for all $i$ and $v \in W$. Then clauses (i) through (iii) hold.

## E.16 Proof of Lemma 20

Fix $W$. Fix $p \in (0, 1)$ and $\tilde{\varepsilon} > 0$, as stated in Lemma 19. (Here, $\tilde{\varepsilon}$ represents $\varepsilon$ in Lemma 19.) Applying Lemmas 17 and 18 to the strategy profiles specified in Lemma 19, it follows that there is $\overline{\delta} \in (0, 1)$ such that for each $\lambda$, there is $\tilde{K}_\lambda > 0$ such that for each $\delta \in (\overline{\delta}, 1)$, $\mu$, and $v \in W$, there is a strategy profile $s_{v,\lambda,\delta,\mu}$ and a function $w_{v,\lambda,\delta,\mu}$ such that

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

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

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

Set $\varepsilon = \frac{\tilde{\varepsilon}}{2(1-p)}$, and for each $\lambda$, let $K_\lambda = \frac{\tilde{K}_\lambda}{(1-p)\overline{\delta}}$. Then it follows from (ii) and (iii) that $w_{v,\lambda,\delta,\mu}(h^t) \in G_{v,\lambda,2\varepsilon,K_\lambda,\delta}$ for all $t$ and $h^t$. The rest of the proof is the same as that of Lemma 8 of Fudenberg and Yamamoto (2011b).

# Appendix F: Existence of Maximizers

**Lemma 48.** *For each initial prior $\mu$, discount factor $\delta$, and $s_{-i} \in S_{-i}^*$, player $i$'s best reply $s_i \in S_i^*$ exists.*

*Proof.* The formal proof is as follows. Pick $\mu$, $\delta$, and $s_{-i} \in S_{-i}^*$. With an abuse of notation, let $h^t = (a^\tau, y^\tau)_{\tau=1}^t$ denote a history with length $t$ without information about public randomization, and let $l^\infty$ be the set of all functions (bounded sequences) $f : H \to \mathbf{R}$. For each function $f \in l^\infty$, let $Tf$ be a function such that

$$(Tf)(h^t) = \max_{a_i \in A_i}\left[(1-\delta)g_i^{\tilde{\mu}(h^t)}(a_i, s_{-i}(h^t)) + \delta \sum_{a_{-i} \in A_{-i}} \sum_{y \in Y} s_{-i}(h^t)[a_{-i}]\pi_Y^{\tilde{\mu}(h^t)}(y|a)f(h^t, a, y)\right]$$

where $\tilde{\mu}(h^t)$ is the posterior belief of $\omega^{t+1}$ given the initial prior $\mu$ and the history $h^t$. Note that $T$ is a mapping from $l^\infty$ to itself, and that $l^\infty$ with the sup norm is a complete metric space. Also $T$ is monotonic, since $(Tf)(\mu) \leq (T\tilde{f})(\mu)$ for all $\mu$ if $f(\mu) \leq \tilde{f}(\mu)$ for all $\mu$. Moreover $T$ is discounting, because letting $(f+c)(\mu) = f(\mu) + c$, the standard argument shows that $T(f+c)(\mu) \leq (Tf)(\mu) + \delta c$ for all $\mu$. Then from Blackwell's theorem, the operator $T$ is a contraction mapping and thus has a unique fixed point $f^*$. The corresponding action sequence is a best reply to $s_{-i}$. *Q.E.D.*

**Lemma 49.** $\max_{v \in V^\mu(\delta)} \lambda \cdot v$ *has a solution.*

*Proof.* Identical with that of the previous lemma. *Q.E.D.*

**Lemma 50.** *There is $s_{-i}$ which solves $\min_{s_{-i} \in S_{-i}^*} \max_{s_i \in S_i^*} v_i^\mu(\delta, s)$.*

*Proof.* The formal proof is as follows. Pick $\mu$ and $\delta$, and let $h^t$ and $l^\infty$ be as in the proof of Lemma 48. For each function $f \in l^\infty$, let $Tf$ be a function such that

$$(Tf)(h^t) = \min_{\alpha_{-i} \in \times_{j \neq i} \triangle A_j} \max_{a_i \in A_i} \left[ (1-\delta)g_i^{\tilde{\mu}(h^t)}(a_i, \alpha_{-i}) + \delta \sum_{a_{-i} \in A_{-i}} \sum_{y \in Y} \alpha_{-i}(a_{-i}) \pi_Y^{\tilde{\mu}(h^t)}(y|a) f(h^t, a, y) \right]$$

where $\tilde{\mu}(h^t)$ is the posterior belief of $\omega^{t+1}$ given the initial prior $\mu$ and the history $h^t$. Note that $T$ is a mapping from $l^\infty$ to itself, and that $l^\infty$ with the sup norm is a complete metric space. Also $T$ is monotonic, because if $f(h^t) \leq \tilde{f}(h^t)$ for all $h^t$, then we have

$$(Tf)(h^t) \leq \max_{a_i \in A_i} \left[ (1-\delta)g_i^{\tilde{\mu}(h^t)}(a_i, \alpha_{-i}) + \delta \sum_{a_{-i} \in A_{-i}} \sum_{y \in Y} \alpha_{-i}(a_{-i}) \pi_Y^{\tilde{\mu}(h^t)}(y|a) f(h^t, a, y) \right]$$

$$\leq \max_{a_i \in A_i} \left[ (1-\delta)g_i^{\tilde{\mu}(h^t)}(a_i, \alpha_{-i}) + \delta \sum_{a_{-i} \in A_{-i}} \sum_{y \in Y} \alpha_{-i}(a_{-i}) \pi_Y^{\tilde{\mu}(h^t)}(y|a) \tilde{f}(h^t, a, y) \right]$$

for all $\alpha_{-i}$ and $h^t$, which implies $(Tf)(h^t) \leq (T\tilde{f})(h^t)$ for all $h^t$. Moreover, $T$ is discounting as in the proof of Lemma 48. Then from Blackwell's theorem, the operator $T$ is a contraction mapping and thus has a unique fixed point $f^*$. The corresponding action sequence is the minimizer $s_{-i}$. *Q.E.D.*

## Appendix G: Comparison with Dutta (1995)

Dutta (1995) considers irreducible stochastic games and shows that the limit feasible payoff set is invariant to the initial state $\omega$. Clause (i) of Lemma 5 extends his result to stochastic games with hidden states. As argued in the introduction, our proof technique is substantially different from that of Dutta (1995), since we need to deal with technical complications arising from infinitely many states $\mu$.

To highlight the difference, it may be helpful to explain the idea of Dutta (1995). Assume that $\omega$ is observable, and fix $\lambda$ arbitrarily. From the principle of optimality, the score must be attained by a Markov strategy $a^* : \Omega \to A$, where $a^*(\omega)$ denotes the action profile when the current state is $\omega$. When players play this Markov strategy, the evolution of the state $\omega^t$ is described by a Markov chain. For simplicity, assume that the state transition function has a full support.[22] Then

---

[22] As Dutta (1995) shows, the full support assumption here can be replaced with irreducibility.

it is immediate that the Markov chain is ergodic, i.e., the time-average of the state converges to the unique invariant measure regardless of the initial state. Ergodicity guarantees that the initial state does not influence players' time-average payoffs, which approximate their discounted payoffs for $\delta$ close to one.[23] Hence the scores are independent of the initial state $\omega$ for $\delta$ close to one.

In our model, the score is achieved by a Markov strategy $a^* : \triangle\Omega \to A$, where the state space is $\triangle\Omega$ rather than $\Omega$. Such a Markov strategy induces a Markov chain which governs the evolution of the belief $\mu^t$, so let $P(\cdot|\mu) \in \triangle(\triangle\Omega)$ be the distribution of the posterior belief $\tilde{\mu}$ in the next period given the current belief $\mu$. Suppose that the Markov strategy $a^*$ is not constant, that is, $a^*$ induce different action profiles for some different beliefs $\mu$ and $\hat{\mu}$. Then the distribution $P(\cdot|\mu)$ of the belief tomorrow is not continuous with respect to the current belief $\mu$, at the point in which the Markov strategy $a^*$ switches the action profile. This implies that the Markov chain is not Feller,[24] and thus the standard probability theory does not tell us if there is an invariant measure.[25] Hence, *a priori* there is no reason to expect ergodicity, and thus the proof idea of Dutta (1995), which relies on ergodicity, is not applicable.

## Appendix H: Assumption 4 of Hsu, Chuang, and Arapostathis (2006

Hsu, Chuang, and Arapostathis (2006) claims that their Assumption 4 implies their Assumption 2. However it is incorrect, as the following example shows.

Suppose that there is one player, two states ($\omega_1$ and $\omega_2$), two actions ($a$ and $\tilde{a}$), and three signals ($y_1$, $y_2$, and $y_3$). If the current state is $\omega_1$ and $a$ is chosen, $(y_1, \omega_1)$ and $(y_2, \omega_2)$ occur with probability $\frac{1}{2}$-$\frac{1}{2}$. The same thing happens if the

---

[23]This result follows from Abel's theorem.

[24]A Markov chain $P = (P(\cdot|\mu))_{\mu \in \triangle\Omega}$ is *Feller* if $P(\cdot|\mu)$ is continuous with respect to $\mu$. It is well-known that a Markov chain $P$ has an invariant measure if it is Feller and the state space is compact (Theorem 3.1.1 of Da Prato and Zabczyk (1996)). On the other hand, if a Markov chain is not Feller (even if $P(\cdot|\mu)$ is discontinuous only at finitely many $\mu$), it may not have an invariant measure. For example, suppose that the state space is $[0, 1]$, $\mu^{t+1} = \frac{\mu^t}{2}$ if $\mu^t \in (0, 1]$, and $\mu^{t+1} = 1$ if $\mu^t = 0$. We have discontinuity only at $\mu = 0$, but there is no invariant measure.

[25]Even if the Markov chain has an invariant measure, there may be more than one invariant measures; indeed, the *Doeblin condition*, which is often assumed for the uniqueness of invariant measure, is not satisfied here. See Doob (1953) for more details. (The condition is stated as "Hypothesis D.")

current state is $\omega_2$ and $\tilde{a}$ is chosen. Otherwise, $(y_3, \omega_1)$ and $(y_3, \omega_2)$ occur with probability $\frac{1}{2}$-$\frac{1}{2}$. Intuitively, $y_1$ shows that the next state is $\omega_1$ and $y_2$ shows that the next state is $\omega_2$, while $y_3$ is not informative about the next state. And as long as the action matches the current state (i.e., $a$ for $\omega_1$ and $\tilde{a}$ for $\omega_2$), the signal $y_3$ never happens so that the state is revealed each period. A stage-game payoff is $0$ if the current signal is $y_1$ or $y_2$, and $-1$ if $y_3$.

Suppose that the initial prior puts probability one on $\omega_1$. The optimal policy asks to choose $a$ in period one and any period $t$ with $y^{t-1} = y_1$, and asks to choose $\tilde{a}$ in any period $t$ with $y^{t-1} = y_2$. If this optimal policy is used, then it is easy to verify that the support of the posterior is always a singleton set and thus their Assumption 2 fails. On the other hand, their Assumption 4 holds by letting $k_0 = 2$. This shows that Assumption 4 does not imply Assumption 2.

To fix this problem, the minimum with respect to an action sequence in Assumption 4 should be replaced with the minimum with respect to a strategy. The modified version of Assumption 4 is more demanding than connectedness in this paper.

# 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 (2008): "Collusion with Persistent Cost Shocks," *Econometrica* 76, 493-540.

Bagwell, K., and R.W. Staiger (1997): "Collusion over the Business Cycle," *RAND Journal of Economics* 28, 82-106.

Besanko, D., U. Doraszelski, Y. Kryukov, and M. Satterthwaite (2010): "Learning by Doing, Organizational Forgetting, and Industry Dynamics," *Econometrica* 78, 453-508.

Da Prato, G., and J. Zabczyk (1996): *Ergodicity for Infinite Dimensional Systems*, Cambridge University Press.

Doob, J.L. (1953): *Stochastic Processes*, John Wiley & Sons, Inc.; Chapman & Hall, Limited.

Duggan, J. (2012): "Noisy Stochastic Games," *Econometrica* 80, 2017-2045.

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

Dutta, P., and K. Sundaram (1998): "The Equilibrium Existence Problem in General Markovian Games," in M. Majumdar (ed), *Organizations with Incomplete Information*, Cambridge University Press.

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.

Haltiwanger, J., and J.E. Harrington Jr. (1991): "The Impact of Cyclical Demand Movements on Collusive Behavior," *RAND Journal of Economics* 22, 89-106.

Hao, D., A. Iwasaki, M. Yokoo, Y.J. Joe, M. Kandori, and I. Obara (2012): "Single Agent Optimal Pricing under Market Uncertainty by Using POMDP," International Joint Agent Workshop and Symposium 2012.

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 (2011): "Recursive Methods in Discounted Stochastic Games II: Infinite State Space," *mimeo*.

Hörner, J., S. Takahashi, and N. Vieille (2015): "Truthful Equilibria in Dynamic Bayesian Games," forthcoming in *Econometrica*.

Hsu, S.-P., D.-M. Chuang, and A. Arapostathis (2006): "On the Existence of Stationary Optimal Policies for Partially Observed MDPs under the Long-Run Average Cost Criterion," *Systems and Control Letters* 55, 165-173.

Kandori, M. (1991): "Correlated Demand Shocks and Price Wars During Booms," *Review of Economic Studies* 58, 171-180.

Kandori, M. (1992): "The Use of Information in Repeated Games with Imperfect Monitoring," *Review of Economic Studies* 59, 581-594.

Levy, Y. (2013): "Discounted Stochastic Games with No Stationary Nash Equilibrium: Two Examples," *Econometrica* 81, 1973-2007.

Platzman, L.K. (1980): "Optimal Infinite-Horizon Undiscounted Control of Finite Probabilistic Systems," *SIAM Journal on Control and Optimization* 18, 362-380.

Renault, J., and B. Ziliotto (2014): "Hidden Stochastic Games and Limit Equilibrium Payoffs," *mimeo*.

Rosenberg, D., E. Solan, and N. Vieille (2002): "Blackwell Optimality in Markov Decision Processes with Partial Observation," *Annals of Statistics* 30, 1178-1193.

Ross, S.M. (1968): "Arbitrary State Markovian Decision Processes," *Annals of Mathematical Statistics* 6, 2118-2122.

Rotemberg, J., and G. Saloner (1986): "A Supergame-Theoretic Model of Price Wars during Booms," *American Economic Review* 76, 390-407.

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.