Stochastic Games with Hidden States*

Yuichi Yamamoto†

First Draft: March 29, 2014
This Version: January 14, 2015

Abstract

This paper studies infinite-horizon stochastic games in which players observe noisy public information about a hidden state each period. We find that if the game is connected, the limit feasible payoff set exists and is invariant to the initial prior about the state. Building on this invariance result, we provide a recursive characterization of the equilibrium payoff set and establish the folk theorem. We also show that connectedness can be replaced with an even weaker condition, called asymptotic connectedness. Asymptotic connectedness is satisfied for generic signal distributions, if the state evolution is irreducible.

Journal of Economic Literature Classification Numbers: C72, C73.

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

*The author thanks Naoki Aizawa, Drew Fudenberg, Johannes Hörner, Atsushi Iwasaki, Michihiro Kandori, George Mailath, Takeaki Sunada, and Masatoshi Tsumagari for helpful conversations, and seminar participants at various places.

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

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. One of the crucial assumptions of these papers is that demand shocks are publicly observable *before* firms make their decisions in each period. This means that firms can perfectly adjust their price contingent on the true demand today. Unfortunately, this assumption is not always plausible, because 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. The purpose of this paper is to develop a general framework which incorporates such uncertainty, and to investigate how this 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 ω^t (booms or slumps in the Bertrand model) is given, and players have some posterior belief μ^t about the state. Players simultaneously choose actions, and then a public signal y and the next hidden state ω^{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. Our formulation accommodates a wide range of economic applications, since we allow the signal y to be informative about both the current and next states.

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 μ^t about the current state ω^t after every history. Then this posterior belief μ^t can be regarded as a common state variable of the model, and thus our model reduces to a stochastic game with *observable* states μ^t. This is a great simplification, but still the model is not as tractable as one may expect, since there are

infinitely many possible posterior beliefs so that we need to consider a stochastic game with an *infinite* state space. 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)), and accordingly the standard techniques developed in the literature do not apply to our model.[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. The positive recurrence ensures that the Markov chain is "well-behaved" in the long run; specifically, it is deeply related to *ergodicity* of the Markov chain. Loosely, ergodicity ensures 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 is invariant to the initial state, and all the existing techniques on stochastic games rely on this invariance result. In contrast, when we consider an infinite-state Markov chain, states may not be positive recurrent. Accordingly, it is well-known in the probability theory that additional assumptions are needed for an infinite-state Markov chain to be ergodic. Unfortunately, these assumptions are not satisfied in our model (see Section 5.2 for more discussions), and hence *a priori*, there is no reason to expect ergodicity of the belief evolution.

Nonetheless, we find that the invariance result of Dutta (1995) 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. This result implies that the impact of the initial belief vanishes in the long run, as long as payoffs are concerned. The proof heavily relies on the fact that the observable state in our model is players' belief; in particular, it is crucial that payoffs are linear with respect to players' belief and hence the state. We would like to stress that our proof here is substantially different from that of

[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. As will be clear from the discussion of this paper, whether such an assumption holds is a crucial question in the analysis of the infinite-state case, and they are silent on this matter. It is also worth noting that there is 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.

Dutta (1995), because we need to consider infinitely many beliefs and thus cannot assume ergodicity of the belief evolution.

Loosely speaking, our assumption, connectedness, ensures that the evolution of the *support* of the belief is well-behaved in the long run. This requirement is much weaker than assuming the belief evolution to be well-behaved, but it is enough to establish the result we want. It may be worth emphasizing that connectedness is satisfied in a wide range of applications, and can be regarded as a natural generalization of irreducibility, which is a standard assumption 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 signal structures, as long as the underlying state evolution is irreducible.

Using the above invariance result, we then establish the folk theorem for connected stochastic games. 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 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). The idea is roughly as follows. In our equilibrium, the infinite horizon is regarded as a sequence of *random blocks*, whose lengths are determined by public randomization. Specifically, at the end of each period, players refer to public randomization and decide whether the current block continues (with some probability p) or terminates (with the remaining probability $1-p$). Due to the random termination, the "effective discount factor" of each block is $p\delta$, that is, payoffs during the first block are equivalent to *payoffs for the infinite-horizon stochastic game with the discount factor* $p\delta$. Thus an equilibrium payoff vector is represented as the sum of a payoff vector for the infinite-horizon stochastic game, which can approximate any feasible payoffs, and of a continuation payoff vector from the second block. This payoff decomposition is reminiscent of that of Fudenberg, Levine, and Maskin (1994, hereafter FLM) for infinitely repeated games, in which an equilibrium payoff is the sum of a stage-game payoff in period one (which can achieve any feasible payoffs) and of a continuation payoff. Hence we can generalize the proof idea of FLM to our setup, and can establish the folk theorem.

Stochastic games are proposed by Shapley (1953). Building on earlier results

of Bewley and Kohlberg (1976) and Mertens and Neyman (1981), 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.[2]

Athey and Bagwell (2008), Escobar and Toikka (2013), and Hörner, Takahashi, and Vieille (2013) 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.[3] 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. A key assumption in their model is 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.

2 Setup

2.1 Stochastic Games with Hidden States

Let $I = \{1, \cdots, N\}$ be the set of players. At the beginning of the game, Nature chooses the state of the world ω^1 from a finite set Ω. The state may change as

[2] Independently of this paper, Renault and Ziliotto (2014) also study stochastic games with hidden states. They focus on an example of a stochastic game with thirteen states, where two of the states are absorbing. Since there are multiple absorbing states, the feasible payoff set depends on the initial prior in their example. Accordingly, their equilibrium analysis is substantially different from that of this paper (and of the literature on irreducible stochastic games).

[3] An exception is Section 6 of Hörner, Takahashi, and Vieille (2013); they consider equilibria in which some players do not reveal information and the public belief is used as a state variable, as in our model. But their analysis relies on the independent private value assumption.

time passes, and the state in period $t = 1, 2, \cdots$ is denoted by $\omega^t \in \Omega$. The state ω^t is not observable to players, and let $\mu \in \triangle\Omega$ be the common prior about ω^1.

In each period t, players move simultaneously, with player $i \in I$ choosing an action a_i from a finite set A_i. As usual, 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 ω^{t+1} which is randomly drawn from Ω. The distribution of y and ω^{t+1} depends on the current state ω^t and the current action profile $a \in A$; let $\pi^\omega(y, \tilde{\omega}|a)$ denote the probability that players observe a signal y and the next state becomes $\omega^{t+1} = \tilde{\omega}$, given $\omega^t = \omega$ and a. In this setup, a public signal y can be informative about the current state ω *and* the next state $\tilde{\omega}$, because the distribution of y may depend on ω and y may be correlated with $\tilde{\omega}$. Let $\pi_y^\omega(y|a)$ denote the marginal probability of y, and let $\pi_y^\omega(a) = (\pi_y^\omega(y|a))_{y \in Y}$ be the marginal distribution 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 ω 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 ω influences a player's expected payoff through the distribution of y. Let $g^\omega(a) = (g_i^\omega(a))_{i \in I}$ be the vector of expected payoffs. Let $\overline{g}_i = \max_{\omega,a} |2g_i^\omega(a)|$, and let $\overline{g} = \sum_{i \in I} \overline{g}_i$. Also let $\overline{\pi}$ be the minimum of $\pi^\omega(y, \tilde{\omega}|a)$ over all $(\omega, \tilde{\omega}, a, y)$ such that $\pi^\omega(y, \tilde{\omega}|a) > 0$.

Our formulation encompasses the following examples:

- *Stochastic games with observable states.* Let $Y = \Omega \times \Omega$ 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, suppose that 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 measurable with respect to the first component of y, so that stage-game payoffs are influenced by the current state but not by the next state. In this case, the signal y^{t-1} in the previous period perfectly reveals the current state ω^t, which means that players know the true state ω^t *before* they choose an action profile a^t. So this is exactly the standard stochastic games studied in the literature. In what follows, this class of games is referred to as stochastic games with observable states.

- *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 ω^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.

- *Combination of observable states and unobservable states.* Assume that ω consists of two components, ω^O and ω^U, and that the signal y^t perfectly reveals the first component of the next state, $\omega^{t+1,O}$. Then we can interpret the first component ω^O as an observable state and the second component ω^U as an unobservable state. There are many economic examples in which observable states and unobservable states coexist. For example, consider a duopoly market in which firms face uncertainty about the current demand each period, and their cost function today depends on their knowledge, know-how, or experience. The firms' experience can be regarded as an observable state variable as in Besanko, Doraszelski, Kryukov, and Satterthwaite (2010), while the uncertainty about the market demand can be described by 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 τ. 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 ω, the discount factor is δ, 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 μ. Note that for each initial prior μ, discount factor δ, and $s_{-i} \in S_{-i}^*$, player i's best reply

$s_i \in S_i^*$ exists.[4] Let $v^\omega(\delta,s) = (v_i^\omega(\delta,s))_{i\in I}$ and $v^\mu(\delta,s) = (v_i^\mu(\delta,s))_{i\in I}$.

2.2 Alternative Interpretation: Belief as a State Variable

In each period t, each player forms a belief μ^t about the current hiddent state ω^t. Since players have the same initial prior μ and the same information h^{t-1}, the posterior belief μ^t is also the same across all players. Then we can regard this belief μ^t as a common state variable; that is, our model can be interpreted as a stochastic game with *observable states* μ^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 later periods, players use Bayes' rule to update the belief, as will be explained. In each period t, player i observes the current (common) belief μ^t and then chooses an action a_i. Actions are observable, and in addition players observe a public signal y, where the probability of y conditional on an action profile a is given by $\pi_y^{\mu^t}(y|a) = \sum_{\omega\in\Omega}\mu^t(\omega)\pi_y^\omega(y|a)$. Public randomization z, which follows the uniform distribution on $[0,1]$, is also available at the end of each period. Player i's actual payoff is $u_i(a,y)$ as in the original model, so her expected stage-game payoff conditional on the posterior belief μ^t and current action profile a is $g_i^{\mu^t}(a) = \sum_{\omega\in\Omega}\mu^t(\omega)g_i^\omega(a)$.

After receiving payoffs, players compute the posterior belief μ^{t+1} about ω^{t+1} using Bayes' rule, and then go to the next period $t+1$. Specifically, given μ^t, a^t,

[4] The formal proof is as follows. Pick μ, δ, 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^∞ be the set of all functions (bounded sequences) $f : H \to \boldsymbol{R}$. 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 ω^{t+1} given the initial prior μ and the history h^t. Note that T is a mapping from l^∞ to itself, and that l^∞ with the sup norm is a complete metric space. Also T is monotonic, since $(Tf)(\mu) \le (T\tilde f)(\mu)$ for all μ if $f(\mu) \le \tilde f(\mu)$ for all μ. Moreover T is discounting, because letting $(f+c)(\mu) = f(\mu)+c$, the standard argument shows that $T(f+c)(\mu) \le (Tf)(\mu)+\delta c$ for all μ. 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}.

and y^t, the posterior belief for the next period is given by

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

for each $\tilde{\omega}$. Note that the posterior belief μ^t evolves stochastically due to the random public signal y.

In what follows, we will mostly follow this interpretation, that is, we regard our model as a stochastic game with observable states $\mu \in \triangle\Omega$. Note that the state space is $\triangle\Omega$, so we have infinitely many states. As emphasized in the introduction, it creates a new problem which does not exist in the case of finitely many states. Detailed discussions will be given in Section 5.2.

Now we give the definition of sequential equilibria in this setup. Let $\zeta : H \to \triangle\Omega$ be a belief system; i.e., $\zeta(h^t)$ represents players' posterior about ω^{t+1} after history h^t. A belief system ζ is *consistent with the initial prior* μ 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, the above requirement implies that we have to apply Bayes' rule even after someone deviates from a prescribed strategy profile. A strategy profile s is a *sequential equilibrium* in the stochastic game with the initial prior μ if there is a belief system ζ consistent with μ such that given this belief system ζ, s is sequentially rational in the stochastic game.

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 ($a_i^H = 2$), 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 ω, 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 ω^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 ω 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 does not depend too much on the underlying economic condition. If the effective price is zero, the probability of y^H is one regardless of the current state ω. We assume that the realized demand y is public information. Assume also that y and the next state $\tilde{\omega}$ are independently drawn.

Since this is the Bertrand model, 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 $a_i < a_{-i}$. If $a_i > a_{-i}$, then the opponent takes all and hence $u_i(a,y) = 0$. If $a_i = a_{-i}$, the firms share the market equally and $u_i(a,y) = \frac{a_i y}{2}$. Given ω 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 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 μ and a, it is easy to see that $g^\mu(a)$ is maximized by the 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 μ, 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 μ 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 be chosen in such a way that the sum of these values is maximized.

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 $\delta=0.7$; the horizontal axis represents the firms' belief, and the vertical axis represents their total profit. 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.

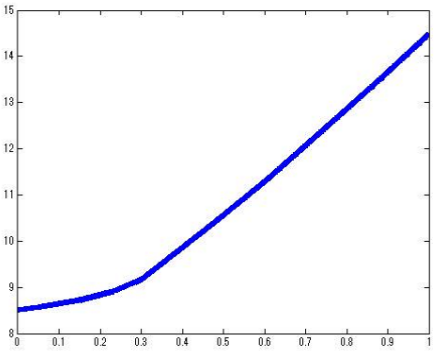

Figure 1: Value Function

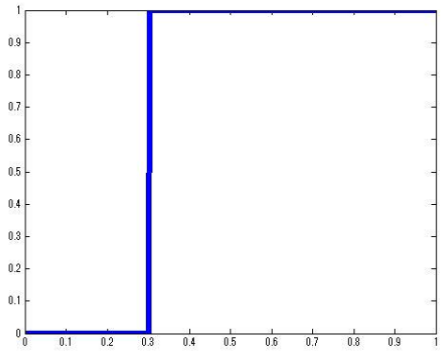

Figure 2: Optimal Policy

Figure 2 shows the optimal policy; the horizontal axis represents the firms' belief, and the vertical axis represents actions (here, 0 shows the low price a^L, while 1 shows the high price a^H). It shows that the firms can maximize the total profit in the dynamic game by a simple cut-off rule; they choose the low price a^L

when the current belief μ is less than μ^{**}, and the high price a^H otherwise, with the threshold value $\mu^{**} \approx 0.305$. This threshold value μ^{**} 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 hidden state ω than the low price, in Blackwell's sense.[5] To see this, for each a, let

$$\Pi(a) = \begin{pmatrix} \pi_y^{\omega^H}(y^H|a) & \pi_y^{\omega^H}(y^L|a) \\ \pi_y^{\omega^L}(y^H|a) & \pi_y^{\omega^L}(y^L|a) \end{pmatrix}.$$

Note that each row of the matrix is the signal distribution given a and some ω. 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) for more discussions), and in this sense, the public signal y given the low price a^L is less informative than that given the high price.

As argued, 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 is larger than the decrease in the current profit. Hence the firms should choose the high price in the optimal policy.

Note that the Nash equilibrium price a^* is never chosen in the optimal policy. The reason is that it yields no profit today, and provides no information about the state tomorrow. (Recall that the distribution of y does not depend on ω when a^* is chosen.)

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

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

us check firm i's incentive when the opponent chooses this strategy. In the punishment phase, firm i has no reason to deviate from a^*, since "playing a^* forever" is a Markovian equilibrium in this model. (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 player i deviates, firm i's current payoff is at most $g_i^{\omega^H}(a_i^L, a_{-i}^H) = 9.1$, and its continuation payoff is zero (since the firms will switch to the punishment phase). So the overall payoff by deviating to the low price 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 $\min_{\mu \in [0,1]} \frac{f(\mu)}{2} \geq 4$. Hence the above strategy profile is an equilibrium.

A couple of remarks are in order. First, the equilibrium construction here is misleadingly simple, since it relies on the existence of a Markovian equilibrium in which a^* is charged forever. In general, when the state space is infinite, the existence of Markovian equilibria is not guaranteed (see Duggan (2012) and Levy (2013)). Accordingly, it is not obvious how to punish a deviator in an equilibrium, and indeed our proof of the folk theorem is non-constructive. Specifically, we generalize the idea of self-generation of Abreu, Pearce, and Stacchetti (1990) to our setup and use it to prove the folk theorem.

Second, the solution to (1) depends on the discount factor δ. 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 for these cases are cut-off rules as in the case for $\delta = 0.7$. 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 less impact on the total profit. This property is not specific to this example; we will show in Lemma 6 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

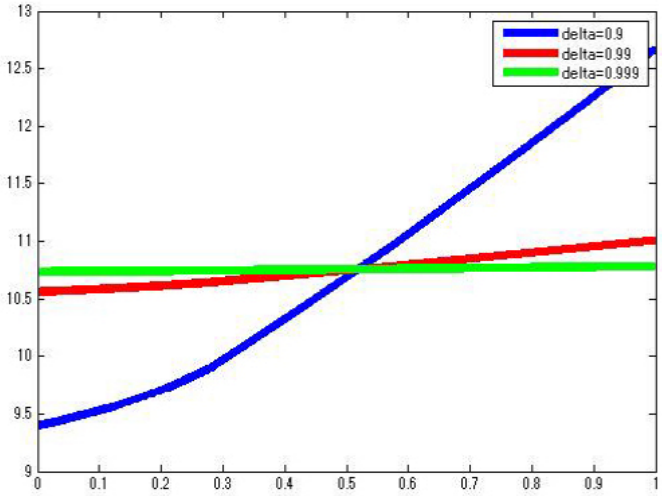

Figure 3: Value Functions for High δ

important role when we apply the self-generation technique.

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 ω is hidden information, 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 the distribution of the next state *and* the distribution of the public signal. Imposing a restriction on the signal distribution is important, since players' posterior belief, which is influenced by the public signal y, is payoff-relevant information in our model. As will be explained, connectedness can be interpreted as a condition 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 unfortunately, its definition is complex; 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 ω, $\tilde{\omega}$, a, and y such that $\pi^{\omega}_{y}(y|a) > 0$.

In words, the full support assumption implies that any state $\tilde{\omega}$ can happen tomorrow given any current state ω, 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 μ^t assigns positive probability to each state ω. Note that we do not require a full support with respect to y, so some signal y may not occur for some state ω 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 in stochastic games with observable states.

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 definition of connectedness provided here is stated in terms of beliefs. In the next subsection, we will provide the definition of connectedness in terms of primitives.

The notion of irreducibility is introduced by Dutta (1995), and it requires that each state $\tilde{\omega}$ be reachable from any state ω in finite time. Formally, $\tilde{\omega}$ is *accessible from* ω 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 ω 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.

A natural extension of irreducibility to our model is to require global accessibility of all posterior beliefs μ^t, because μ^t is an observable state in our model. A belief $\tilde{\mu} \in \triangle\Omega$ is *globally accessible* if for any initial prior μ, 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 μ and players play the action sequence $(a^1, \cdots, a^T)$. The following condition is a naive generalization of irreducibility:

$(*)$ Each belief $\tilde{\mu} \in \triangle\Omega$ is globally accessible.

Unfortunately, $(*)$ is too demanding and not satisfied in almost all applications. The problem is that there are infinitely many possible beliefs μ and thus there is no reason to expect recurrence; i.e., the posterior belief may not return to the current belief in finite time. This implies that any belief μ is not globally accessible and thus $(*)$ fails. So we need to find a condition which is weaker than $(*)$ but still parallels irreducibility of the standard model.

Our first step is to focus on the evolution of the *support* of μ^t, rather than the evolution of μ^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 that it is enough to assume global accessibility of the support of the belief μ^t, rather than global accessibility of the belief.

[6] Irreducibility of Dutta (1995) is a bit more stronger assumption than the one presented here, but for our purpose (i.e., the invariance of the feasible payoff set), this weaker requirement is enough.

Definition 3. A non-empty subset $\Omega^* \subseteq \Omega$ is *globally accessible* if there is $\pi^* > 0$ such that for any initial prior μ, there is a natural number $T \leq 4^{|\Omega|}$, an action sequence $(a^1, \cdots, a^T)$, and a belief $\tilde{\mu}$ whose support is included in Ω^* such that[7]

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

In words, $\Omega^* \subseteq \Omega$ is globally accessible if given any initial prior μ, the support of the posterior belief μ^{T+1} can be a subset of Ω^* with probability at least $\pi^* > 0$ when players play some appropriate action sequence $(a^1, \cdots, a^T)$. The parameter $\pi^* > 0$ is a lower bound of the probability of the support of the posterior being in the set Ω^*. Note that such π^* does not appear in (2), because when states are observable, the number of states is finite and thus the existence of the lower bound π^* is obvious. Here we need to explicitly assume the existence of the bound π^*, since there are infinitely many initial priors μ.

With this definition at hand, instead of $(*)$, we may assume:

$(*')$ Each subset $\Omega^* \subseteq \Omega$ is globally accessible.

Obviously $(*')$ is weaker than $(*)$, but unfortunately, it is still demanding and not satisfied in many applications. To see this, take a singleton set $\Omega^* = \{\omega\}$. For this set to be globally accessible, the posterior belief must eventually puts probability one on this state ω. However this can happen only if the signal y reveals the next state, and such an assumption is violated in most applications.

To avoid this problem, we relax $(*')$ further. The idea is to allow some sets Ω^* not to be globally accessible, and to require these sets to be *transient*. To give the definition of transience, 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 μ and players play the strategy profile s.

[7] Restricting attention to $T \leq 4^{|\Omega|}$ is without loss of generality. To see this, pick a subset $\Omega^* \subseteq \Omega$ and ω 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 5. 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 5. We can repeat this procedure to show the existence of sequences with length $T \leq 4^{|\Omega|}$ which satisfy (i) and (ii).

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 μ whose support is Ω^*, there is a natural number $T \leq 2^{|\Omega|}$ and a belief $\tilde{\mu}$ whose support is globally accessible such that[8]

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

In words, transience of Ω^* implies that if the support of the current belief is Ω^*, then regardless of future actions, the support of the posterior belief cannot stay there forever and must reach some globally accessible set with positive probability. Now we relax $(*')$ in the following way:

$(*'')$ Each subset $\Omega^* \subseteq \Omega$ is either globally accessible or transient.

As shown in Lemma 28 in Appendix C, if the support of the current belief is transient, then the support cannot return to the current one forever with positive probability. Hence, if Ω^* is transient, the time during which the support of the posterior belief stays at Ω^* is almost negligible in the long run. This implies that the existence of transient sets Ω^* does not influence the long-run behavior of the support of the posterior belief. Hence, under $(*'')$, the long-run behavior of the evolution of the support of the posterior belief is similar to the one under $(*')$, which enables us to show our key result, Lemma 6.

Our assumption, connectedness, is exactly $(*'')$, but it is stated using the posterior belief μ^t. In the next subsection, we will give the definition of connectedness based on primitives. Also we will explain that connectedness is satisfied in a wide range of examples.

4.3 Connectedness in Terms of Primitives

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

[8]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 μ whose support is Ω^* 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 μ and players play $\tilde{s}$, the support of the posterior belief never reaches a globally accessible set. The proof is standard and hence omitted.

Definition 5. 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:

(i) If the initial state is ω 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 Ω^*, 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 $t \leq T$.

The definition of globally accessibility here is equivalent to the one presented in Section 4.2. To see this, take a set Ω^* which is globally accessible in the above sense, and fix an arbitrarily initial prior μ. Note that there is at least one ω such that $\mu(\omega) \geq \frac{1}{|\Omega|}$, so pick such ω, and then pick $(a^1, \cdots, a^T)$ and $(y^1, \cdots, y^T)$ as stated in the above definition. Suppose that the initial prior is μ and players play $(a^1, \cdots, a^T)$. Then clause (i) 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 μ^{T+1} after observing this signal sequence is a subset of Ω^*, i.e., $\mu^{T+1}(\tilde{\omega}) = 0$ for all $\tilde{\omega} \notin \Omega^*$.[9] 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 ω and the action sequence $(a^1, \cdots, a^T)$. This implies that global accessibility in Definition 5 implies the one in Definition 3, by letting

$$\pi^* = \frac{1}{|\Omega|} \overline{\pi}^{4^{|\Omega|}}.$$

[9]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 $\tilde{\omega} \notin \Omega^*$.

We can also show that the converse is true, which leads to the following lemma. The formal proof can be found in Appendix B.

Lemma 1. *Definitions 3 and 5 are equivalent.*

Let $\mathscr{O}$ be the set of all globally accessible $\Omega^* \subseteq \Omega$. The set $\mathscr{O}$ is non-empty, as the whole state space $\Omega^* = \Omega$ is always globally accessible. Indeed, when $\Omega^* = \Omega$, 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 Ω 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 Ω^* 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 Ω^* 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 6. 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 $t \leq T$.

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 Ω^* at some point in the future.[10] It is obvious that the definition of transience above is equivalent to Definition 4 in the previous subsection, except that here we consider only singleton sets $\{\omega\}$.

Now we are ready to give the definition of connectedness:

[10] 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.

Definition 7. A stochastic game is *connected* if each singleton set $\{\omega\}$ is globally accessible or transient.

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 $(*'')$ in Section 4.2.

Lemma 2. *If each singleton set $\{\omega\}$ is globally accessible or transient, then any subset $\Omega^* \subseteq \Omega$ is globally accessible or transient.*

In what follows, 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:

Lemma 3. *If the state transition function has a full support, then the game is connected.*

Proof. Note that under the full support assumption, the support of the posterior belief is the whole space Ω after every history. This implies that each singleton set is transient and thus the game is connected. *Q.E.D.*

The next lemma shows that if states are observable without delay, the standard assumption in the literature, irreducibility, implies connectedness. In this sense, connectedness is a natural generalization of irreducibility.

Lemma 4. *In stochastic games with observable states, the game is connected if the state evolution is irreducible.*

Proof. Since all states are observable and the state evolution is irreducible, any singleton set $\{\omega\}$ is globally accessible, and thus the game is connected. *Q.E.D.*

Also, as the following lemma shows, irreducibility of the underlying state implies connectedness even if states are observable with delay. The proof is given in Appendix B.

Lemma 5. *In stochastic games with delayed observations, the game is connected if the state evolution is irreducible.*

The following example shows that even irreducibility is not necessary for connectedness. That is, the game can be connected even if the underlying state is not irreducible.

Example 1. Suppose that there are three states, ω_1, ω_2, and ω_3. If the current state is ω_1, the state stays at ω_1 with probability $\frac{1}{2}$ and moves to ω_2 with probability $\frac{1}{2}$. If the current state is ω_2, the state moves to ω_3 for sure. If the current state is ω_3, the state moves to ω_2 for sure. Note that this state transition is not irreducible, as there is no path from ω_2 to ω_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 ω_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 ω_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 4 and 5, we know that irreducibility of the underlying state is sufficient for connectedness, if states are observable (possibly with delay). 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 ω_2 for sure if the current state is ω_1, and vice versa. Assume that $|A_i| = 1$ for each i, and that the public signal y does not reveal the state ω, that is, $\pi_y^{\omega}(y|a) > 0$ for all ω, 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 ω, then the posterior belief puts probability one on ω in all odd periods 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.

Similarly, in the following example, the state evolution is irreducible but the game is not connected. What is different from the previous example is that the state evolution is not deterministic.

Example 3. Consider a machine with two states, ω_1 and ω_2. ω_1 is a "normal" state and ω_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 Ω, it is impossible to reach the belief μ 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 ω_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.

Remark 1. Although the game is not connected in Example 2, we can prove that the folk theorem holds in this example. A key is that the game is *asymptotically connected* here; see Section 8 for more discussions. The same argument applies to Example 3.

Remark 2. When the game is not (asymptotically) connected, Lemma 6 is not valid and the limit feasible payoff set V^{μ} may depend on μ. However, even in

this case, Proposition 2 remains true if we modify the definition of V so that $V = \bigcap_{\mu \in \triangle\Omega} V^{\mu}$. That is, if a payoff vector v is feasible and individually rational regardless of the initial prior μ, then there is a sequential equilibrium with the payoff v for δ close to one.

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 ω and the discount factor is δ, i.e., let

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

Likewise, let $V^{\mu}(\delta)$ be the set of feasible payoffs when the initial prior is μ. Note that the feasible payoff set depends on δ, as the state ω, which influences the structure of the stage game, changes over time.

Let Λ be the set of directions $\lambda \in \boldsymbol{R}^N$ with $|\lambda| = 1$. For each direction λ, we compute the "score" using the following formula:[11]

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

Roughly speaking, the score characterizes the boundary of the feasible payoff set $V^{\mu}(\delta)$ toward direction λ. For example, when λ 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 λ, let $f(\mu)$ be the corresponding score given the initial prior μ. The function f can be derived by solving the following Bellman equation:

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

where $\tilde{\mu}(\mu, a, y)$ is the belief in period two given that the initial prior is μ and players play a and observe y in period one. Note that (3) is a generalization of

[11] This maximization problem indeed has a solution. The formal proof is very similar to that provided in footnote 4, and hence omitted.

(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 δ is sufficiently large, the scores are insensitive to the initial prior. This result implies that the feasible payoff sets $V^{\mu}(\delta)$ are similar across all initial priors μ when δ is close to one. The proof can be found in Appendix B.[12]

Lemma 6. *Suppose that the game is connected. Then for each $\varepsilon > 0$, there is $\overline{\delta} \in (0,1)$ such that the following properties hold:*

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

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

(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 μ,

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

Clause (i) asserts that when δ is sufficiently large, the scores $\max_{v \in V^{\mu}(\delta)} \lambda \cdot v$ are similar across all priors μ. Also, although it is not stated in the above lemma, the proof of the lemma actually shows that the convergence of the score is $1 - \delta$.

[12] We thank Johannes Hörner for pointing out that Lemma 6 strengthens the results in the literature of partially observable Markov decision process (POMDP). It turns out that connectedness is weaker than various assumptions in the literature, including renewability of Ross (1968), reachability-detectability of Platzman (1980), and Assumption 4 and inequality (5) of Hsu, Chuang, and Arapostathis (2006). In other words, connectedness is a single condition which covers all the problems studied in these papers, 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 10 applies. (Note also that there is a minor error in Hsu, Chuang, and Arapostathis (2006); the minimum with respect to an action sequence in their Assumption 4 should be replaced with the minimum with respect to a strategy profile. Connectedness is indeed weaker than the modified Assumption 4. See Appendix D for more details.) 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. Accordingly, verifying these assumptions is not very easy; indeed, in our setup, there is a Bellman equation for each direction λ, and we need to compute the optimal policy for each problem.

That is, we can replace ε in Lemma 6 with $O(1-\delta)$. Clause (ii) ensures that for each direction λ and sufficiently large δ, there is a strategy profile s which yields a payoff vector approximating the score (i.e., the boundary of V) regardless of the initial prior μ. The existence of such a strategy profile is useful in the proof of Lemma 7.

In the next subsection, we provide a sketch of the proof of this lemma, and compares it with that of Dutta (1995).

5.2 Proof Sketch and Comparison with Dutta (1995)

Dutta (1995) considers irreducible stochastic games and shows that the limit feasible payoff set is invariant to the initial state ω. Clause (i) of Lemma 6 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 μ.

To highlight the difference, it may be helpful to explain the idea of Dutta (1995). Assume that ω is observable, and fix λ 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 ω. When players play this Markov strategy, the evolution of the state ω^t is described by a Markov chain. For simplicity, assume that the state transition function has a full support.[13] Then 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 δ close to one.[14] Hence the scores are independent of the initial state ω for δ close to one.

In our model, the score is achieved by a Markov strategy $a^* : \triangle A \to \Omega$, where the state space is $\triangle\Omega$ rather than Ω. Such a Markov strategy induces a Markov chain which governs the evolution of the belief μ^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 μ. Suppose that the Markov strategy a^* is not constant, that is, a^* induce different action profiles for some different beliefs μ and $\hat{\mu}$. Then the distribution $P(\cdot|\mu)$ of

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

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

the belief tomorrow is not continuous with respect to the current belief μ, at the point in which the Markov strategy a^* switches the action profile. This implies that the Markov chain is not Feller,[15] and thus the standard probability theory does not tell us if there is an invariant measure.[16] 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.

To illustrate the idea of our new proof, consider the coordinate direction λ 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 π 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 μ is represented by a real number between $[0,1]$. Let s^μ be the strategy profile which attains the score when the initial prior is μ, i.e., let s^μ be such that $v_i^\mu(\delta, s^\mu) = \max_{v \in V^\mu(\delta)} v_i$.

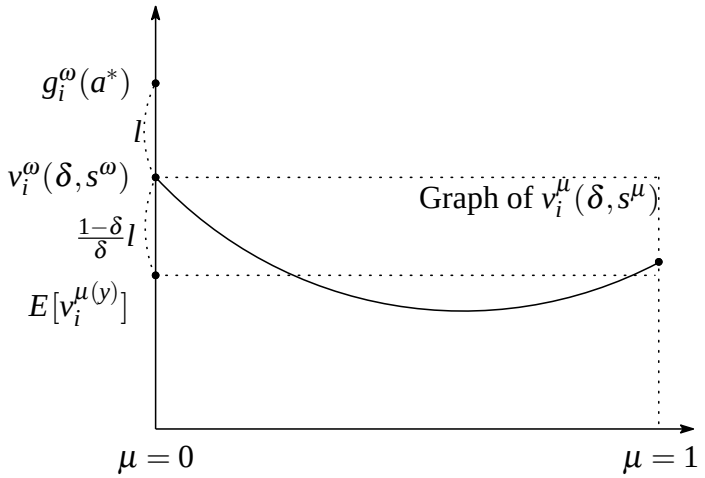

Figure 4: Score $v_i^\mu(\delta, s^\mu)$

[15] A Markov chain $P = (P(\cdot|\mu))_{\mu \in \triangle\Omega}$ is *Feller* if $P(\cdot|\mu)$ is continuous with respect to μ. 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 μ), 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.

[16] 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.")

As shown in Lemma 16 in the appendix, the score $v_i^\mu(\delta, s^\mu)$ is convex with respect to μ. (The proof relies on the fact that μ is a belief and thus player i's payoff $v_i^\mu(\delta, s)$ is linear with respect to μ 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 ω be the corresponding state. The curve in Figure 4 represents the score $v_i^\mu(\delta, s^\mu)$ for each μ; note that this is indeed a convex function and the value is maximized at $\mu = 0$. Let s^ω be the strategy profile which achieves the score at $\mu = 0$. (So the superscript ω means $\mu = 0$.) Let a^* be the action profile in period one played by s^ω. 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 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 ω and a^* is chosen in period one. Equivalently,

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

Hence the left-hand side must be of order $\frac{1-\delta}{\delta}$, that is, the expected continuation payoff $E[v_i^{\mu(y)}(\delta, s^{\mu(y)})]$ must be "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

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

To see this, note that $v_i^\omega(\delta, s^\omega)$ is the maximum score, and thus the left-hand side of (4) is non-negative for all y. Then its weighted average, $v_i^\omega(\delta, s^\omega) - E[v_i^{\mu(y)}(\delta, s^{\mu(y)})]$, is also non-negative. In particular, if there is y which does not satisfy (4), then the weighted average is also not of order $\frac{1-\delta}{\delta}$. This is a contradiction, as we know that the weighted average $v_i^\omega(\delta, s^\omega) - E[v_i^{\mu(y)}(\delta, s^{\mu(y)})]$ is of order $\frac{1-\delta}{\delta}$. Hence the result follows.

Pick an arbitrary y. Since we assume the full support of π, the corresponding posterior belief $\mu(y)$ is not close to $\mu = 0$ or $\mu = 1$; indeed we must have $\mu(y) \in$

$[\overline{\pi}, 1-\overline{\pi}]$. Consider the strategy profile $s^{\mu(y)}$, which achieves the score when the initial prior is $\mu(y)$. The dashed line in Figure 5 represents player i's payoff with this strategy profile $s^{\mu(y)}$ for each belief μ. 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 μ. 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.

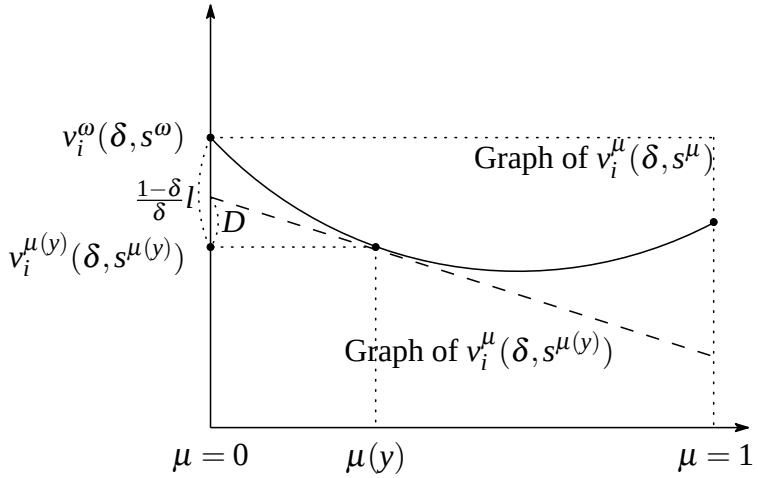

Figure 5: Payoff by $s^{\mu(y)}$

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 $v_i^{\mu(y)}(\delta, s^{\mu(y)})$ is also of order $\frac{1-\delta}{\delta}$. Since we have (4), 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 μ. 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 μ. 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 μ. This implies that the score $v_i^{\mu}(\delta, s^{\mu})$ approximates the maximal score $v_i^{\omega}(\delta, s^{\omega})$ for all μ, which proves clause (i). To establish clause (ii), use the strat-

egy profile $s^{\mu(y)}$, which approximates the maximal score $v_i^{\omega}(\delta, s^{\omega})$ regardless of the initial prior μ.

We can apply a similar argument to the case in which the dashed line is upward-sloping. Here we use the payoffs at $\mu = 1$ rather than at $\mu = 0$ to bound the slope of the dashed line.

A key idea in the above argument is that if the score for some posterior belief $\mu(y)$ approximates the maximal score $v_i^{\omega}(\delta, s^{\omega})$, then we can bound the slope of the dashed line and thus we can conclude that the score is close to the highest score for *all* beliefs μ. For this idea to work, it is crucial that the support of $\mu(y)$ is the whole state space Ω. Indeed, if we have $\mu(y) = 0$ or $\mu(y) = 1$ so that the support is not the whole state space, then we cannot bound the slope of the graph of $v_i^{\mu}(\delta, s^{\mu(y)})$, as there are arbitrarily steep lines which intersect with the curve at $\mu = 0$ or $\mu = 1$. See Figure 6.

The above discussion suggests that tracking the evolution of the *support* of the posterior belief is essential, and the notion of connectedness is motivated by this observation. Loosely, in connected stochastic games, the support of the posterior belief evolves nicely, and hence an idea similar to the above one applies even if the full support assumption does not hold. Interested readers may refer to the formal proof.

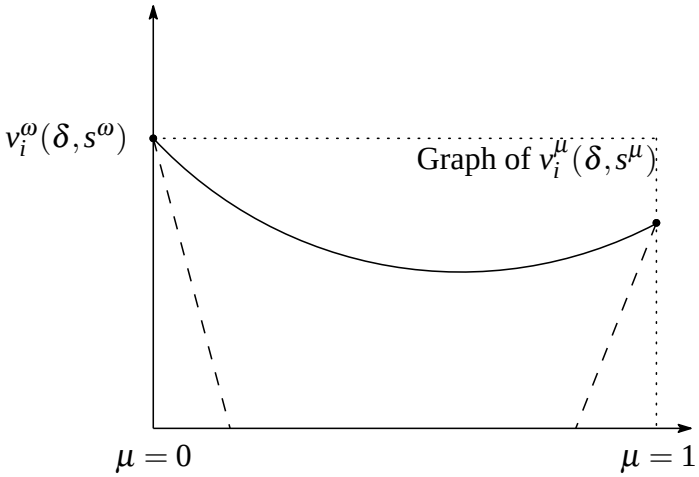

Figure 6: Without Full Support

5.3 Existence of the Limit Feasible Payoff Set

The next two lemmas establish that the "limit feasible payoff set" exists. Specifically, the first lemma shows that for each initial prior μ and direction λ, the score converges to a limiting value as δ goes to one. The second lemma shows that its convergence is uniform in λ. The proofs can be found in Appendix B.[17]

Lemma 7. *If the game is connected, then* $\lim_{\delta\to 1}\max_{v\in V^\mu(\delta)}\lambda\cdot v$ *exists for each* λ *and* μ.

Lemma 8. *For each* ε*, there is* $\overline{\delta}\in(0,1)$ *such that for each* $\lambda\in\Lambda$, $\delta\in(\overline{\delta},1)$, *and* μ,

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

Let V^μ be the set of all $v\in\boldsymbol{R}^N$ such that $\lambda\cdot v\le\lim_{\delta\to 1}\max_{v\in V^\mu(\delta)}\lambda\cdot v$ for all λ. Lemma 7 guarantees that this set is well-defined. From Lemma 6, we know that this set V^μ is independent of μ, so we denote it by V. Lemma 8 implies that this set V is indeed the "limit feasible payoff set" in that $V^\mu(\delta)$ approximates V for all μ as long as δ is close to one.

5.4 Minimax Payoffs

The *minimax payoff* to player i in the stochastic game given the initial prior μ and discount factor δ 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.[18] Let $\underline{v}_i(\delta)=\sup_{\mu\in\triangle\Omega}\underline{v}_i^\mu(\delta)$, and let $\underline{v}_i=\limsup_{\delta\to 1}\underline{v}_i^\mu$. Let V^* denote the set of feasible payoffs $v\in V$ such that $v_i\ge\underline{v}_i$

[17] Lemma 7 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.

[18] The formal proof is as follows. Pick μ and δ, and let h^t and l^∞ be as in footnote 4. For each function $f\in l^\infty$, let Tf be a function such that

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

where $\tilde\mu(h^t)$ is the posterior belief of ω^{t+1} given the initial prior μ and the history h^t. Note that T is a mapping from l^∞ to itself, and that l^∞ with the sup norm is a complete metric space. Also T is

for all i. In words, V^* is the limit set of feasible payoffs which are individually rational regardless of the initial prior μ.

In irreducible stochastic games with observable states, Dutta (1995) shows that the minimax payoff $\underline{v}_i^{\omega}(\delta)$ has a limit and its limit is invariant to the initial state ω. Unfortunately, we do not know if the same result holds when the state ω is hidden. A complication here is that the minimax payoff $\underline{v}_i^{\mu}(\delta)$ is not necessarily convex (or concave) with respect to μ, since player i maximizes the value while the opponents minimize it. Hence we cannot follow the proof technique of Lemma 6, which heavily relies on convexity of the score function.

6 Stochastic Self-Generation

The invariance result in the previous section implies that after every history h^t, the feasible 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 δ, 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

---

monotonic, because if $f(h^t) \leq \tilde{f}(h^t)$ for all h^t, then we have

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

for all α_{-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 footnote 4. 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}.

assume for simplicity that actions are perfectly observable. Let a^* be the action profile chosen by s in period one. Then the equilibrium payoff v of s must satisfy the following equation:

$$v = (1-\delta)g(a^*) + \delta w(a^*) \tag{5}$$

where $g(a^*)$ 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 from period two on is also subgame-perfect, which implies that $w(a^*)$ must be an element of 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 is chosen in such a way that it yields a payoff vector on the Pareto-efficient frontier. So (5) implies that the equilibrium payoff v is *decomposable into a Pareto-efficient payoff vector and some continuation payoff chosen from the equilibrium payoff set*, as shown in Figure 7. FLM use this structure to establish a general folk theorem.

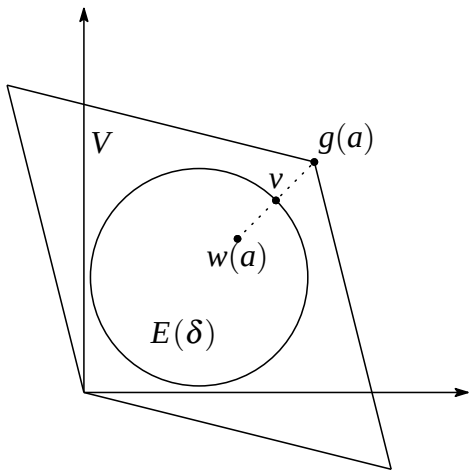

Figure 7: Payoff Decomposition

Obviously, a similar payoff decomposition is possible in our model. Let μ 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

$$v = (1-\delta)g^{\mu}(a^*) + \delta \sum_{y \in Y} \pi_y^{\mu}(y|a^*)w(a^*,y) \tag{6}$$

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 (6) is not very useful in our model. The 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.

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 as follows: Given some parameter $p \in (0,1)$, if $z^t \leq p$, then the current block continues so that period $t+1$ is the next period of the current block. On the other hand, 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 μ and sequential equilibrium s arbitrarily, and let v be the equilibrium payoff vector. 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 each random block, and does not influence the actual play by s. 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

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

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 δ), and the second is the continuation payoff from the second random block. Intuitively, a random block with the discount factor δ is payoff-equivalent to the infinite-horizon stochastic game with the "effective discount factor" $p\delta$, due to the termination probability $1-p$. Hence the term $v_i^{\mu}(p\delta, s)$ shows up in the above expression.

Now, pick p sufficiently large and then take δ 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 the equilibrium payoff vector is a weighted average of some payoff vector approximating the efficient frontier and the expected continuation payoffs, just as in Figure 7. 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 δ goes to one; hence a small variation in continuation payoffs is enough to provide appropriate incentives during the first random block. These structures are reminiscent of the payoff decomposition (5) for the standard repeated game, and we use these structures to establish the folk theorem.

Now we present a version of the self-generation theorem which decompose payoffs in the above way. When we consider the random block structure, players' incentive problem in each random block is equivalent to the one for the following *stochastic termination game*. Consider an auxiliary 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. So the length of the game is finite with probability one, but it can be arbitrarily long. 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 μ, δ, $p \in (0,1)$,

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

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

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 8. A pair (s, v) of a strategy profile and a payoff vector is *stochastically enforceable with respect to* (δ, μ, 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 9. A subset W of $\boldsymbol{R}^N$ is *stochastically self-generating with respect to* (δ, p) if for each $v \in W$ and μ, there are $s \in S$ and $w : H \to W$ such that (s, v) is stochastically enforceable with respect to (δ, μ, 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 μ, 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 μ. This comes from the fact that 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 very similar to theirs and hence omitted.

Proposition 1. *Fix δ. If W is bounded and stochastically self-generating with respect to (δ, p) for some p, then for each payoff vector $v \in W$ and initial prior μ, 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 (indeed, the minimax payoffs may depend on the initial state) and thus an average payoff in the T-period stochastic game may not be able to 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 (2013) 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 relatively easy when players use a Markov strategy, but in order to establish a folk theorem, they need to use non-Markov strategies. To deal with this problem, they consider an equilibrium in which players play a Markov strategy profile in each random block. They take the termination probability $1-p$ close to zero; then players' play is approximately Markovian, and hence the computation of payoffs by misreporting becomes easier. To sum up, in their paper, stochastic self-generation is used 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 δ simultane-

ously, while we fix p first and then take δ large enough.

7 Folk Theorem

Now we will establish the folk theorem for connected stochastic games. That is, we will show that for any payoff vector v in the interior of V^* and for any initial prior μ, there is a sequential equilibrium with the payoff v when δ is large enough. The precise statement of the folk theorem is as follows:

Definition 10. A subset W of $\boldsymbol{R}^N$ is *smooth* if it is closed and convex; it has a non-empty interior; and there is a unique unit normal for each point on the boundary of W.[19]

Proposition 2. *Suppose that the game is connected. Then for any smooth subset W of the interior of V^*, 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 initial prior μ, there is a sequential equilibrium which yields the payoff of v.*

This proposition shows that the folk theorem obtains as long as the full dimensional condition, $\dim V^* = N$, is satisfied. (If the full dimensional condition is not satisfied and $\dim V^* < N$, then there is no interior point of V^* and thus W is always empty.) It may be worth noting that the threshold of the discount factor depends on the target payoff v, but does not depend on the initial prior μ. This comes from the fact that for δ close to one, the feasible payoff set approximates the limiting set V uniformly in μ (Lemma 6).

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 W is stochastically self-generating for δ 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 A.

[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 $\boldsymbol{R}^N$.

8 Relaxing Connectedness

So far we have focused on connected stochastic games. In this section, we show that connectedness is stronger than necessary for the folk theorem; we show that *asymptotic connectedness*, which is weaker than connectedness, is sufficient to establish the folk theorem.

To illustrate the idea of asymptotic connectedness, consider Example 2 in Section 4.3, where the state ω 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 ω_1. Then the true state must be ω_1 in all odd periods, and be ω_2 in all even periods. Hence if we consider the empirical distribution of the public signals in odd periods, it should approximate π_1 with probability close to one, by the law of large numbers. Similarly, if the initial state is ω_2, the empirical distribution of the public signals in odd periods should approximate π_2. This implies that players can eventually learn the current state by aggregating the past public signals, regardless of the initial prior μ. Hence for δ close to one, the feasible payoff set should be similar across all μ, i.e., Lemma 6 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 ω_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 11. 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 μ, 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 Ω^* requires that given any initial prior μ, there is an action sequence $(a^1, \cdots, a^{T^*})$ so that the posterior belief can approximate a belief whose support is Ω^*. 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 ω, then the posterior belief in the second period will put probability close to one on ω. 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 ω_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 ω_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 distribution of the posterior belief, not primitives. However, as explained above, in some applications, checking whether each set Ω^* is asymptotically accessible is a relatively simple task. Note that Ω^* 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.

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

Definition 12. A singleton set $\{\omega\} \notin \tilde{\mathscr{O}}$ is *weakly transient* if for each pure strategy profile $s \in S^*$, there is an asymptotically accessible set $\Omega^* \in \tilde{\mathscr{O}}$, 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 $t \leq T$.

Definition 13. A stochastic game is *asymptotically connected* if each singleton set $\{\omega\}$ is asymptotically accessible or weakly 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 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.

Lemma 9. *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 ω, then $\pi_y^{\tilde{\omega}}(y|a) > 0$ for all $\tilde{\omega}$.*
- *Signals do not reveal the next state, that is, for each ω, $\tilde{\omega}$, $\hat{\omega}$, 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 $\{\pi_y^\omega(a)|\forall\omega\}$ 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 B.

As shown in Lemma 6, 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.[20] The proof can be found in Appendix B.

Lemma 10. *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 $\lambda \in \Lambda$, $\delta \in (\overline{\delta},1)$, μ, and $\tilde{\mu}$,

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

(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 μ,

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

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

[20] However, unlike Lemma 6, we do not know the rate of convergence, and in particular, we do not know if we can replace ε in the lemma with $O(1-\delta)$.

9 Conclusion

This paper considers a new class of stochastic games in which the state is hidden information. Our model is equivalent to stochastic games with observable states where players' belief is a state. Since the belief space is continuous and each belief is not necessarily positive recurrent, the analysis is significantly different from the one for stochastic games with finitely many states. We find that if the game is connected (or asymptotically connected), then the feasible payoff set is invariant to the initial belief 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 establish 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: Proof of Proposition 2

Here we prove the folk theorem for connected stochastic games. 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 δ. 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 8 for all δ. Formally, letting λ 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 δ, but if W is smooth, this set G is always in the interior of the set W for any δ 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 δ, they are in the set W, as desired.

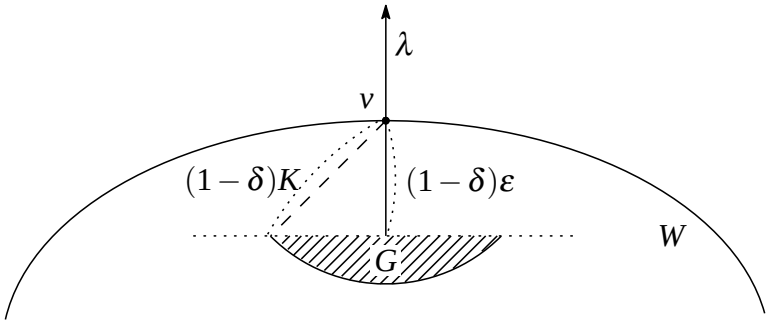

Figure 8: Continuation payoffs are chosen from G.

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, λ, $\varepsilon > 0$, $K > 0$, 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 point of W and λ is the corresponding unit normal, this set $G_{v,\lambda,\varepsilon,K,\delta}$ is exactly the set G in Figure 8.

Definition 14. 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 μ, 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 (δ, μ, 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 δ.[21] The proof is similar to Fudenberg and Yamamoto (2011b) and hence omitted.

Lemma 11. *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 μ, 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 λ 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 λ is enforceable using continuation payoffs in the set G. The proof can be found in Appendix B.

Lemma 12. *For each regular direction λ and for each $p \in (0,1)$, there is $K > 0$ such that for each μ, 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 (δ, μ, 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 λ. Assume also that p and δ are close to one. Let s be a strategy profile approximating the

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

boundary of V toward direction λ 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 8, 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 9. (In particular, clause (ii) says that continuation payoffs can be chosen from a hyperplane orthogonal to λ. This is reminiscent of the idea of "utility transfers across players" of FLM.)

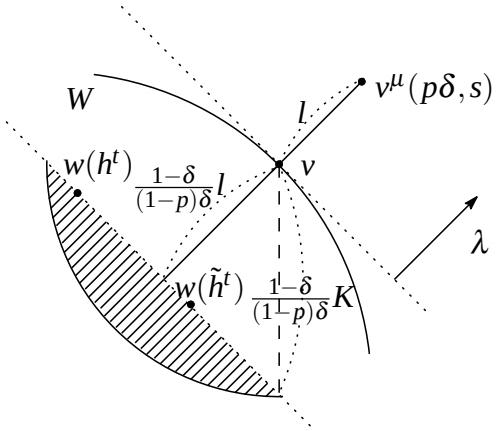

Figure 9: Continuation Payoffs for Regular Direction λ

Note that in the above lemma, the rate of convergence (the constant K in the lemma) depends on direction λ; indeed this constant K can become arbitrarily large as λ approaches a coordinate vector. However, in Lemma 15 we will extract a finite set of direction vectors so the dependence of the constant on λ 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 B.

Lemma 13. *For each $p \in (0,1)$, there is $K > 0$ such that for each μ, $s_{-i} \in S^*_{-i}$, $v \in V$, δ, 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 (δ, μ, 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 10.

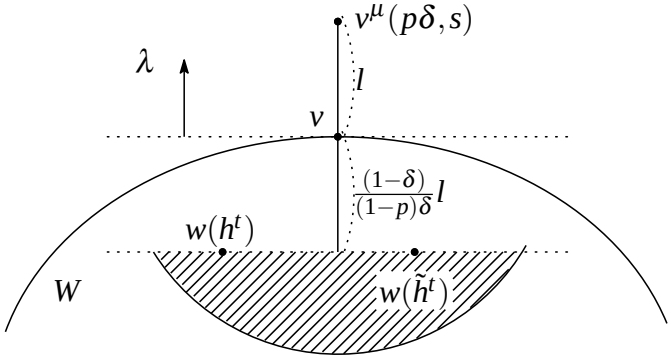

Figure 10: Continuation Payoffs for λ with $\lambda_i = 1$

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 11.

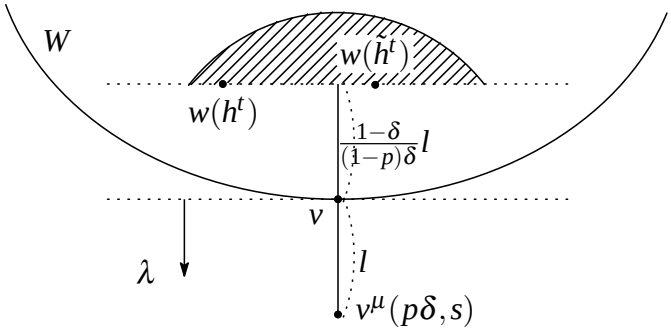

Figure 11: Continuation Payoffs for λ with $\lambda_i = -1$

To establish the above lemma, it is important that s_i is a best reply to s_{-i} given μ 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 9 through 11. The proof is found in Appendix B.

Lemma 14. *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 λ and μ, 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 μ, 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 μ, 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 B.

Lemma 15. *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 11, establishes Proposition 2.

Appendix B: Proof of Lemmas

B.1 Proof of Lemma 1

We have already explained that global accessibility in Definition 5 implies the one in Definition 3. So it is sufficient to show that the converse is true.

Let Ω^* be a globally accessible set in the sense of Definition 3. Pick $\pi^* > 0$ as stated in Definition 3, and pick ω arbitrarily. Let μ 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 Ω^* is globally accessible, we can choose an action sequence $(a^1, \cdots, a^T)$ and a belief $\tilde{\mu}$ whose support is included in Ω^* such that

$$\Pr(\mu^{T+1} = \tilde{\mu} | \mu, a^1, \cdots, a^T) \geq \pi^*. \tag{7}$$

Let $(y^1, \cdots, y^T)$ be the signal sequence which induces the posterior belief $\tilde{\mu}$ given the initial prior μ 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 (7) 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 ω and the action sequence $(a^1, \cdots, a^T)$. Obviously this sequence $(y^1, \cdots, y^T)$ satisfies clause (i) in Definition 5. Also it satisfies clause (ii) in Definition 5, since $(y^1, \cdots, y^T)$ induces the posterior belief $\tilde{\mu}$ whose support is Ω^*, given the initial prior μ whose support is the whole space Ω. Since ω can be arbitrarily chosen, the proof is completed.

B.2 Proof of Lemma 2

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 Ω^{**} is not globally accessible. In what follows, we will show that it is transient. Take a strategy $s \in S^*$ arbitrarily. Take $\Omega^* \in \mathscr{O}$, T, and $(y^1, \cdots, y^T)$, as stated in Definition 6. Suppose that the initial belief is μ with the support Ω^{**}, and that players play s. Then, since μ puts a positive probability on ω, the signal sequence $(y^1, \cdots, y^T)$ realizes with positive probability and the support of the posterior belief μ^{T+1} must be a superset Ω^{***} of Ω^*. Since Ω^* is globally accessible, so is the superset Ω^{***}. This shows that Ω^{**} is transient, as s can be arbitrary.

B.3 Proof of Lemma 5

Fix an arbitrary $\{\omega\}$ which is not globally accessible. Fix a^* and y^* such that $\pi_y^\omega(y^*|a^*) > 0$, and let Ω^* be the set of all $\tilde{\omega}$ such that $\pi^\omega(y^*, \tilde{\omega}|a^*) > 0$. It is sufficient to show that Ω^* is globally accessible; indeed, if it is the case, then $\{\omega\}$ is transient so that the game is connected.

Fix an arbitrary initial prior μ, and take an arbitrary ω^* 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 ω is positive conditional on the initial state ω^* and the action sequence $(a^1, \cdots, a^T)$, Suppose that, given the initial prior μ, 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 ω, so that with positive probability, the signal y^* realizes. Then the support of the posterior belief is Ω^*, since the signal y reveals that the current state is ω. Note that the probability of this event is at least

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

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

B.4 Proof of Lemma 6

Fix δ and λ. For each μ, let s^μ 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 μ, the score is equal to $\lambda \cdot v^{\mu}(\delta, s^{\mu})$. The following lemma shows that the score is convex with respect to μ.

Lemma 16. *$\lambda \cdot v^{\mu}(\delta, s^{\mu})$ is convex with respect to μ.*

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

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

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

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

$$\lambda \cdot v^{\omega}(\delta, s^{\omega}) \geq \lambda \cdot v^{\mu}(\delta, s^{\mu}) \tag{8}$$

for all μ. (Here the superscript ω refers to the initial prior μ which assigns probability one to the state ω.) Pick such a ω.

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 ω and players play s^{ω}. For each $(y^1, \cdots, y^T)$ with $\pi(y^1, \cdots, y^T) > 0$. let $\mu(y^1, \cdots, y^T)$ be the posterior belief μ^{T+1} given the initial state ω, the strategy profile s^{ω}, 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 $C(T) = \frac{2\bar{g}}{\bar{\pi}^T}$

Lemma 17. *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

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

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

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

This inequality, together with (8), implies that

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

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

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

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

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

Using (8) and $\lambda\cdot v^\omega(\delta,s^\omega)\geq-\overline{g}$, 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 ω and players play s^ω, 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 Ω^*. To simplify our notation, let $\mu^*=\mu(y^1,\cdots,y^T)$ and $C=\frac{C(T)}{\overline{\pi}}=\frac{2\overline{g}}{\overline{\pi}^{T+1}}$.

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

Let $s^*=s^{\mu^*}$, that is, s^* is the strategy profile which achieves the score when the initial prior is μ^*. The following lemma shows that the strategy s^* can approximate the maximal score $\lambda\cdot v^\omega(\delta,s^\omega)$ given any initial state $\tilde{\omega}$ in the set Ω^*, when δ is close to one.

Lemma 18. *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

$$\begin{aligned}\lambda \cdot v^{\mu^*}(\delta, s^*) &= \sum_{\hat{\omega} \in \Omega^*} \mu^*[\hat{\omega}] \lambda \cdot v^{\hat{\omega}}(\delta, s^*) \\ &\leq \mu^*[\tilde{\omega}] \lambda \cdot v^{\tilde{\omega}}(\delta, s^*) + \sum_{\hat{\omega} \neq \tilde{\omega}} \mu^*[\hat{\omega}] \lambda \cdot v^{\hat{\omega}}(\delta, s^{\hat{\omega}})\end{aligned}$$

Using (8),

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

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

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

By the definition of $\overline{\pi}$, we have $\mu^*[\tilde{\omega}] \geq \overline{\pi}$, and hence

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

Using Lemma 17, the result follows. *Q.E.D.*

Now, take an arbitrary μ whose support is a subset of Ω^*. The following lemma shows that when δ is close to one, the strategy s^* approximates the maximal score $\lambda \cdot v^{\omega}(\delta, s^{\omega})$ for such initial prior μ.

Lemma 19. *For each μ 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 18, the result follows. *Q.E.D.*

Since Ω^{*} is globally accessible, given any initial prior μ, 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 Ω^{*} with probability at least $\pi^{*} = \frac{1}{|\Omega|}\overline{\pi}^{4^{|\Omega|}}$.

Now let μ^{**} be such that $\mu^{**}(\omega) = \frac{1}{|\Omega|}$ for each ω. Given the initial prior μ^{**}, 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 Ω^{*}, 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 Ω^{*}, 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 Ω^{*}. Let $\mu^{(2)}$ be the posterior belief, and if its support is indeed a subset of Ω^{*}, they play the strategy profile s^{*} in the continuation game. Lemma 19 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 Ω^{*}, once again players play an action sequence which can potentially induce a posterior belief whose support is a subset of Ω^{*}. And they do the same thing over and over.

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

Lemma 20. *We have*

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

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

Suppose that the initial prior is μ^{**} 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 Ω^* for some k, players switch their continuation play to s^*, and Lemma 19 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 $\lambda \cdot g^{\omega}(a) \leq -2\overline{g}$ for all ω and a. For each t, let ρ^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

$$\lambda \cdot v^{\mu^{**}}(\delta, s^{**}) \geq (1-\delta) \sum_{t=1}^{\infty} \delta^{t-1} \left\{ \rho^{t-1}(-2\overline{g}) + (1-\rho^{t-1}) v^* \right\}. \tag{9}$$

Recall that the length of the action sequence $\boldsymbol{a}(\mu)$ is at most $4^{|\Omega|}$ for each μ. 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 μ. Hence we have

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

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

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

for each $n = 0, 1, \cdots$ and $k \in \{0, \cdots, 4^{|\Omega|} - 1\}$. Plugging this inequality into (9), we obtain

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

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

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

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

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

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

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

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

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

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 21. *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 18. Replace Ω^*, μ^*, s^*, and $\overline{\pi}$ in the proof of Lemma 18 with Ω, μ^{**}, 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 δ and $\tilde{\omega} \in \Omega$. Then from Lemma 20, 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 μ.

Lemma 22. *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 19. Replace Ω^* and s^* in the proof of Lemma 19 with Ω and s^{**}, respectively. In the last step of the proof, use Lemma 21 instead of Lemma 18. *Q.E.D.*

From Lemma 22 and

$$\lambda \cdot v^{\omega}(\delta, s^{\omega}) \geq \lambda \cdot v^{\mu}(\delta, s^{\mu}) \geq \lambda \cdot v^{\mu}(\delta, s^{**}), \tag{10}$$

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 clause (i), since $T \leq 2^{|\Omega|}$, $C = \frac{2\overline{g}}{\overline{\pi}^{T+1}}$ and $\tilde{C} = \frac{4\overline{g}}{\pi^*}$ for any δ and λ.

To prove (ii), set $s = s^{**}$. Then from Lemma 22 and (10), we obtain

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

which implies clause (ii).

B.5 Proof of Lemma 7

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

$$\left| \max_{\nu \in V^{\mu}(\overline{\delta})} \lambda \cdot \nu - \limsup_{\delta \to 1} \max_{\nu \in V^{\mu}(\delta)} \lambda \cdot \nu \right| < \frac{\varepsilon}{2}, \tag{11}$$

and such that there is s^* with

$$\left| \max_{\nu \in V^{\mu}(\overline{\delta})} \lambda \cdot \nu - \lambda \cdot v^{\tilde{\mu}}(\overline{\delta}, s^*) \right| < \frac{\varepsilon}{2} \tag{12}$$

for all $\tilde{\mu}$. (The existence of such $\overline{\delta}$ is guaranteed by Lemma 6.) It suffices to show that

$$\limsup_{\delta \to 1} \max_{\nu \in V^{\mu}(\delta)} \lambda \cdot \nu - \max_{\nu \in V^{\mu}(\delta)} \lambda \cdot \nu < \varepsilon \tag{13}$$

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 $z^\tau \leq p$ 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 μ^* and discount factor δ. Then the corresponding score is

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

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

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

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

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

Arranging,

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

Then from (11) and (12), we have

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

This implies (13), since we have

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

B.6 Proof of Lemma 8

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

$$\left| \lim_{\delta \to 1} \max_{v \in V^{\mu}(\delta)} \tilde{\lambda} \cdot v - \lim_{\delta \to 1} \max_{v \in V^{\mu}(\delta)} \lambda \cdot v \right| < \frac{\varepsilon}{3} \tag{14}$$

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

$$\left| \max_{v \in V^{\mu}(\delta)} \tilde{\lambda} \cdot v - \max_{v \in V^{\mu}(\delta)} \lambda \cdot v \right| < \frac{\varepsilon}{3} \tag{15}$$

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

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

for all $\lambda \in \Lambda^*$, $\delta \in (\overline{\delta},1)$, and μ. Plugging (14) and (15) into this, we obtain

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

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

B.7 Proof of Lemma 9

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

Take an arbitrary action sequence $(a^1,a^2,\cdots)$. Suppose that the initial prior assigns probability one on ω and that the action sequence $(a^1,a^2,\cdots)$ is played. Since signals do not reveal the current or next state, the support of the posterior belief in period t does not depend on realized signals $(y^1,\cdots,y^{t-1})$, so we denote it by $\Omega^t \subseteq \Omega$ for each t. Then, since Ω 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 $\tilde{t}-1$, that is, let $\hat{a}^k = a^{t+k-1}$ for each k. Also, let $\Omega^* = \Omega^t = \Omega^{\tilde{t}}$. By the definition, if the support of the initial prior is Ω^* and players play the sequence $(\hat{a}^1,\cdots,\hat{a}^{\tilde{t}-t})$ repeatedly, then the support of the posterior belief after period $n(\tilde{t}-t)$ is also in the set Ω^* for each natural number n.

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

posterior belief which puts probability at least $1-\varepsilon$ on Ω^*, by letting players play $(\hat{a}^1,\cdots,\hat{a}^{\tilde{t}-t})$ repeatedly.

Lemma 23. *There is $q>0$ such that for each initial prior μ, 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 Ω^* 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 μ 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 μ. 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. *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 24. *For each $\varepsilon>0$ and $q>0$, there is a natural number n and $\pi^{***}>0$ such that for each initial prior μ 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 Ω^* is the whole state space, the result holds obviously. So we assume that Ω^* is a proper subset of Ω. With an abuse of notation, let $\triangle\Omega^*$ be the set of all priors $\mu\in\triangle\Omega$ which puts probability one on Ω^*; i.e., $\triangle\Omega^*$ is the set of all μ 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 Ω^*.

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 ω 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 μ. 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

$$\Pr(y^1,\cdots,y^{\tilde{t}-t}|\overline{\mu},\hat{a}^1,\cdots,\hat{a}^{\tilde{t}-t})\geq\Pr(y^1,\cdots,y^{\tilde{t}-t}|\underline{\mu},\hat{a}^1,\cdots,\hat{a}^{\tilde{t}-t})+\tilde{\pi}. \tag{16}$$

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

$$\left(\frac{1}{1-\tilde{\pi}}\right)^n\frac{q}{1-q}\geq\frac{1-\varepsilon}{\varepsilon}. \tag{17}$$

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 μ 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 (16) 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 μ, players play $(\hat{a}^1, \cdots, \hat{a}^{\tilde{t}-t})$ and observe $(y^1, \cdots, y^{\tilde{t}-t})$.

By the definition of Ω^*, if the initial state is in the set Ω^* 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 Ω^*. (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

$$\Pr(y^1, \cdots, y^{\tilde{t}-t}, \omega^{\tilde{t}-t+1} | \omega^1, \hat{a}^1, \cdots, \hat{a}^{\tilde{t}-t}) = 0 \tag{18}$$

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

$$\begin{aligned}
&\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})}.
\end{aligned}$$

Here, the second equality comes from (18). Since $(y^1, \cdots, y^{\tilde{t}-t})$ satisfies (16), 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 Ω^* induced by the posterior belief $\hat{\mu}$ is greater than the likelihood of Ω^* induced by the initial prior μ. Note also that such a posterior belief $\hat{\mu}$ realizes with probability at least $q\tilde{\pi}$, since (16) 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 μ 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

$$\frac{\sum_{\tilde{\omega}\in\Omega^*}\mu^{\tilde{t}-t+1}(\tilde{\omega})}{\sum_{\tilde{\omega}\notin\Omega^*}\mu^{\tilde{t}-t+1}(\tilde{\omega})} \geq \frac{1}{1-\tilde{\pi}}\cdot\frac{\sum_{\tilde{\omega}\in\Omega^*}\mu(\tilde{\omega})}{\sum_{\tilde{\omega}\notin\Omega^*}\mu(\tilde{\omega})}. \tag{19}$$

Now suppose that we are in period $\tilde{t}-t+1$ and the posterior belief $\mu^{\tilde{t}-t+1}$ satisfies (19). Note that (19) 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 μ, 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 (17) 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 Ω^*, as desired. *Q.E.D.*

Fix $\varepsilon > 0$ arbitrarily. Choose q as stated in Lemma 23, and then choose n and π^{***} as stated in Lemma 24. Then the above two lemmas ensure that given any initial prior μ, 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 Ω^*. Since the bounds $|\Omega| + n(\tilde{t}-t)$ and π^{**} do not depend on the initial prior μ, this shows that Ω^* is asymptotically accessible, and hence $\{\omega\}$ is weakly transient.

B.8 Proof of Lemma 10

The proof is similar to that of Lemma 6, up to the proof of Lemma 17.

Now, since the game is asymptotically connected, $\{\omega\}$ is either asymptotically accessible or weakly transient. When it is weakly 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 ω and players play s^ω, 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 an asymptotically accessible set Ω^*. Take such T, $(y^1,\cdots,y^T)$, and Ω^*. To simplify our notation, let $\mu^* = \mu(y^1,\cdots,y^T)$ and $C = \frac{C(T)}{\overline{\pi}} = \frac{2\overline{g}}{\overline{\pi}^{T+1}}$.

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

Let $s^* = s^{\mu^*}$, that is, s^* is the strategy profile which achieves the score when the initial prior is μ^*. Then Lemmas 18 and 19 are still valid. Also, we can prove the following lemma, which extends Lemma 19; it shows that if the initial prior μ approximates some belief whose support is a subset of Ω^*, then s^* approximates the highest score.

Lemma 25. *For any $\varepsilon \in (0, \frac{1}{|\Omega|})$ and for any μ such that $\sum_{\tilde{\omega}\in\Omega^*}\mu(\tilde{\omega}) \geq 1-\varepsilon$,*

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

Proof. Pick ε and μ as stated. If the support of μ is a subset of Ω^*, the result follows from Lemma 19. So assume that the support of μ is not a subset of Ω^*. (Implicitly, this means that Ω^* is a proper subset of Ω.) Pick $\tilde{\omega} \in \Omega^*$ such that $\mu(\tilde{\omega}) \geq \frac{1}{|\Omega|}$. Note that such $\tilde{\omega}$ indeed exists, as $\varepsilon \in (0, \frac{1}{|\Omega|})$.

Let $\tilde{\mu} \in \triangle\Omega$ be such that $\tilde{\mu}(\tilde{\omega}) = \mu(\tilde{\omega}) + \sum_{\hat{\omega} \notin \Omega^*} \mu(\hat{\omega})$, $\tilde{\mu}(\hat{\omega}) = \mu(\hat{\omega})$ for all $\hat{\omega} \in \Omega^* \setminus \{\tilde{\omega}\}$, and $\tilde{\mu}(\hat{\omega}) = 0$ for all $\hat{\omega} \notin \Omega^*$. Intuitively, $\tilde{\mu}$ is chosen so that its support is a subset of Ω^* and it approximates μ (for small ε).

Let $\hat{\mu}$ be such that $\mu = \kappa\hat{\mu} + (1-\kappa)\tilde{\mu}$ for some $\kappa \in (0,1)$ and such that $\hat{\mu}(\tilde{\omega}) = 0$. In words, $\hat{\mu}$ is chosen in such a way that it is on the boundary of $\triangle\Omega$ and μ is a convex combination of $\tilde{\mu}$ and $\hat{\mu}$. By the construction, it is easy to check that $\hat{\mu} \in \triangle\Omega$. Note that κ must solve

$$\kappa\hat{\mu}(\tilde{\omega}) + (1-\kappa)\tilde{\mu}(\tilde{\omega}) = \mu(\tilde{\omega}).$$

Arranging,

$$\kappa = \frac{\tilde{\mu}(\tilde{\omega}) - \mu(\tilde{\omega})}{\tilde{\mu}(\tilde{\omega})} = \frac{\sum_{\hat{\omega} \notin \Omega^*} \mu(\hat{\omega})}{\mu(\tilde{\omega}) + \sum_{\hat{\omega} \notin \Omega^*} \mu(\hat{\omega})}.$$

Using $\sum_{\hat{\omega} \notin \Omega^*} \mu(\hat{\omega}) \leq \varepsilon$ and $\mu(\tilde{\omega}) \geq \frac{1}{|\Omega|}$, we obtain

$$\kappa < \frac{\varepsilon}{\frac{1}{|\Omega|} + \varepsilon} = \frac{\varepsilon|\Omega|}{1 + \varepsilon|\Omega|} < \varepsilon|\Omega|.$$

Since $\mu = \kappa\hat{\mu} + (1-\kappa)\tilde{\mu}$, we have

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

so that

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

Here the inequality comes from $\left|\lambda \cdot v^{\tilde{\mu}}(\delta, s^{\tilde{\mu}}) - \lambda \cdot v^{\hat{\mu}}(\delta, s^{\tilde{\mu}})\right| \leq \overline{g}$ and $\kappa < \varepsilon|\Omega|$. From Lemma 19, we know that

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

Combining these two inequalities, we obtain the result. *Q.E.D.*

Pick $\varepsilon \in (0, \frac{1}{|\Omega|})$ arbitrarily. For each asymptotically accessible set Ω^{**}, choose T and π^{**} as stated in the definition of asymptotic accessibility, and we denote them by $T(\varepsilon, \Omega^{**})$ and $\pi^{**}(\varepsilon, \Omega^{**})$. Let $\tilde{T}(\varepsilon) = \max_{\Omega^{**} \in \tilde{\mathscr{O}}} T(\varepsilon, \Omega^{**})$, and $\pi^{**}(\varepsilon) = \min_{\Omega^{**} \in \tilde{\mathscr{O}}} \pi^{**}(\varepsilon, \Omega^{**})$.

Let μ^{**} be such that $\mu^{**}(\omega) = \frac{1}{|\Omega|}$ for each ω. Since Ω^* is asymptotically accessible, for any initial prior μ, there is a natural number $T(\mu) \leq \tilde{T}(\varepsilon)$ and an action sequence $\boldsymbol{a}(\mu) = (a^1(\mu), \cdots, a^{T(\mu)}(\mu))$ 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 $\pi^{**}(\varepsilon)$. Let s^{**} be the following strategy profile:

- Let $\mu^{(1)} = \mu^{**}$.
- Players play the action sequence $\boldsymbol{a}(\mu^{(1)})$ 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 $\boldsymbol{a}(\mu^{(2)})$ 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 $\tilde{C}(\varepsilon) = \frac{4\overline{g}}{\pi^{**}(\varepsilon)}$. The following lemma is a counterpart to Lemma 20, which shows that the strategy s^{**} can approximate the maximal score when the initial prior is μ^{**}.

Lemma 26. *We have*

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

Proof. The proof is essentially the same as that of Lemma 20; we simply replace π^* and $4^{|\Omega|}$ in the proof of Lemma 20 with $\pi^{**}(\varepsilon)$ and $\tilde{T}(\varepsilon)$, and use Lemma 25 instead of Lemma 19. *Q.E.D.*

The next lemma asserts that the strategy profile s^{**} can approximate the maximal score regardless of the initial prior μ.

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

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

Proof. The proof is simply a combination of those of Lemmas 21 and 22. The only difference is to use Lemma 26 instead of Lemma 20. *Q.E.D.*

From Lemma 27 and (10), we have

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

Recall that $T \leq 2^{|\Omega|}$, $C = \frac{2\overline{g}}{\underline{\pi}^{T+1}}$, and $\tilde{C}(\varepsilon) = \frac{4\overline{g}}{\pi^{**}(\varepsilon)}$ for any δ and λ. Recall also that $\tilde{T}(\varepsilon)$ and $\pi^{**}(\varepsilon)$ do not depend on δ and λ. Hence the above inequality implies that the left-hand side can be arbitrarily small for all λ, if we take ε close to zero and then take δ close to one. This proves clause (i).

The proof of clause (ii) is similar to that of Lemma 6.

B.9 Proof of Lemma 12

Fix μ, s, λ, p, δ, 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

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

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

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

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 $z^1 : H^1 \to \boldsymbol{R}^N$ 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 λ. 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 λ, 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 μ, δ, 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 $z^{t+1} : H^{t+1} \to \boldsymbol{R}^N$ be such that

$$v_i(h^t) = v_i(h^t, a_i) + (1-p)\delta \sum_{y \in Y} \pi^{\mu(h^t)}(y | a_i, s_{-i}(h^t)) z_i^{t+1}(h^t, (a_i, s_{-i}(h^t)), y)) \tag{20}$$

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

$$\lambda \cdot z^{t+1}(h^t, (a,y)) = 0 \tag{21}$$

for all $a \in A$ and $y \in Y$. To see the meaning of (20), 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. (20) 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, (21) asserts that the bonus payment $z^{t+1}(h^{t+1})$ is on the linear hyperplane tangential to λ. Note that for each λ, we can take $\tilde{K} > 0$ such that

$$|z^{t+1}(h^t)| < \frac{1-\delta}{(1-p\delta)(1-p)\delta}\tilde{K} \tag{22}$$

for all h^t, δ, 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

$$w_i(h^t) = \tilde{v}_i + z_i^t(h^t) \tag{23}$$

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 (20) 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

$$\tilde{v}_i = v_i - \frac{1-\delta}{(1-p)\delta}(v_i^{\mu}(p\delta, s) - v_i). \tag{24}$$

This, together with (21) and (23), proves clause (ii). Also, from (23) and (24), 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 (22),

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

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).

B.10 Proof of Lemma 13

Fix p, s_{-i}, μ, δ, 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 12. Then for each h^t, let $z_j^{t+1} : H^{t+1} \to \boldsymbol{R}$ 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 12, 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 μ and discount factor $p\delta$. Hence clause (i) follows. From (24), 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.

B.11 Proof of Lemma 14

Fix W as stated, and take ε so that the ε-neighborhood of W is in the interior of V^*. Then from Lemma 8, there is p such that the ε-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.

B.12 Proof of Lemma 15

Fix W. Fix $p \in (0,1)$ and $\tilde{\varepsilon} > 0$, as stated in Lemma 14. (Here, $\tilde{\varepsilon}$ represents ε in Lemma 14.) Applying Lemmas 12 and 13 to the strategy profiles specified in Lemma 14, it follows that there is $\overline{\delta} \in (0,1)$ such that for each λ, there is $\tilde{K}_\lambda > 0$ such that for each $\delta \in (\overline{\delta}, 1)$, μ, 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 (δ, μ, 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 λ, let $K_\lambda = \frac{\tilde{K}_\lambda}{(1-p)\bar{\delta}}$. Then it follows from (ii) and (iii) that $w_{\nu,\lambda,\delta,\mu}(h^t) \in G_{\nu,\lambda,2\varepsilon,K_\lambda,\delta}$ 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 C: Consequence of Transience

As emphasized in the main text, if the support of the current belief is transient, then the support cannot return to the current one forever with positive probability. Formally, we can prove the following lemma. Let $X(\Omega^*|\mu,s)$ be the random variable X which represents the first time in which the support of the posterior belief is Ω^* given that the initial prior is μ and players play s. That is, let

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

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 Ω^* in finite time.

Lemma 28. *For any transient set Ω^*, any initial prior μ whose support is Ω^*, and any strategy profile s, the probability that the support returns to Ω^* in finite time is strictly less than one. That is,*

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

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

$$\Pr(X(\Omega^*|\mu,s) < \infty) = 1. \tag{25}$$

Pick such $s \in S^*$ and μ.

Suppose that the initial prior is μ and players play s. Since Ω^* 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 μ and s is positive.

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

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

The existence of a signal sequence which satisfies (i) and (iii) is obvious, since Ω^* 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 Ω^* before it reaches a globally accessible set. Assume that the initial prior is μ, and consider the following strategy profile $\tilde{s}$:

- Play the strategy profile s, until the support of the posterior belief returns to Ω^*.
- Once the support returns to Ω^*, then play the profile s again, until the support of the posterior belief returns to Ω^* next time.
- And so on.

By the construction, if the initial prior is μ and players play $\tilde{s}$, the support of the posterior belief cannot reach a globally accessible set. This contradicts the fact that the set Ω^* 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 Ω^{**} 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 (25) holds, if the support of the current belief is Ω^{**} and players play $s|_{(y^1, \cdots, y^T)}$ in the continuation game, then the support will reach Ω^* with probability one. That is, for all $\hat{\mu}$ whose support is Ω^{**}, 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

$$\Pr(\hat{y}^1, \cdots, \hat{y}^{\hat{T}}|\hat{\omega}, \hat{a}^1, \cdots, \hat{a}^{\hat{T}}) > 0 \tag{26}$$

and such that

$$\Pr(\hat{y}^1, \cdots, \hat{y}^{\hat{T}}, \omega^{\hat{T}+1}|\omega^1, \hat{a}^1, \cdots, \hat{a}^{\hat{T}}) = 0 \tag{27}$$

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

Pick ω arbitrarily. From (iii), the set Ω^{**} is globally accessible, and hence we can choose a natural number $\tilde{T} \leq 4^{|\Omega|}$, 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

$$\Pr(\tilde{y}^1, \cdots, \tilde{y}^{\tilde{T}}, \omega^{\tilde{T}+1} = \hat{\omega} | \omega, \tilde{a}^1, \cdots, \tilde{a}^{\tilde{T}}) > 0 \tag{28}$$

and such that

$$\Pr(\tilde{y}^1, \cdots, \tilde{y}^{\tilde{T}}, \omega^{\tilde{T}+1} | \omega^1, \tilde{a}^1, \cdots, \tilde{a}^{\tilde{T}}) = 0 \tag{29}$$

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 (26) and (27) 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 (26) and (28), we have

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

Also, from (27) and (29), 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 5 for ω. Since such $\boldsymbol{a}$ and $\boldsymbol{y}$ exist for each $\omega \in \Omega$, the set Ω^* is globally accessible. (The length of the action sequence $\boldsymbol{a}$ above may be greater than $4^{|\Omega|}$, but as discussed in footnote 7, we can always find a sequence with length $T \leq 4^{|\Omega|}$.) However this is a contradiction, as the set Ω^* is transient. *Q.E.D.*

Appendix D: Assumption 4 of Hsu, Chuang, and Arapostathis (200

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 (ω_1 and ω_2), two actions (a and $\tilde{a}$), and three signals (y_1, y_2, and y_3). If the current state is ω_1 and a is chosen, (y_1, ω_1) and (y_2, ω_2) occur with probability $\frac{1}{2}$-$\frac{1}{2}$. The same thing happens if the current state is ω_2 and $\tilde{a}$ is chosen. Otherwise, (y_3, ω_1) and (y_3, ω_2) occur with probability $\frac{1}{2}$-$\frac{1}{2}$. Intuitively, y_1 shows that the next state is ω_1 and y_2 shows that the next state is ω_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 ω_1 and $\tilde{a}$ for ω_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 ω_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.

Bewley, T., and E. Kohlberg (1976): "The Asymptotic Theory of Stochastic Games," *Mathematics of Operations Research* 1, 200-208.

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 (2013): "Truthful Equilibria in Dynamic Bayesian Games," *mimeo*.

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.

Mertens, J.F., and A. Neyman (1981): "Stochastic Games," *International Journal of Game Theory* 10, 53-66.

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.