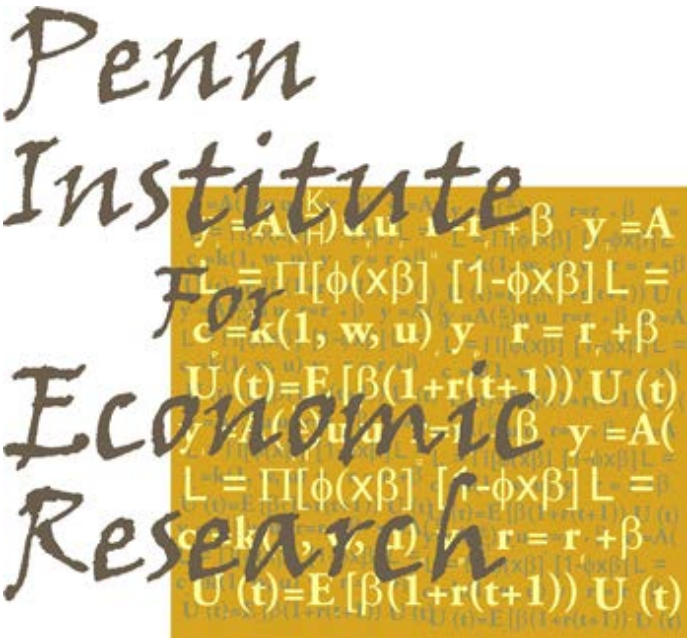

Penn Institute for Economic Research
Department of Economics
University of Pennsylvania
3718 Locust Walk
Philadelphia, PA 19104-6297
pier@econ.upenn.edu
http://economics.sas.upenn.edu/pier

PIER Working Paper 16-012

BY

Yuichi Yamamoto

"Stochastic Games with Hidden States"

http://ssrn.com/abstract=2563612

Stochastic Games with Hidden States*
Third Version

Yuichi Yamamoto†

First Draft: March 29, 2014
This Version: April 26, 2016

Abstract

This paper studies infinite-horizon stochastic games in which players observe payoffs and noisy public information about a hidden state each period. We find that, very generally, the feasible and individually rational payoff set is invariant to the initial prior about the state in the limit as the discount factor goes to one. This result ensures that players can punish or reward the opponents via continuation payoffs in a flexible way. Then we prove the folk theorem, assuming that public randomization is available. The proof is constructive, and introduces the idea of random blocks to design an effective punishment mechanism.

Journal of Economic Literature Classification Numbers: C72, C73.

Keywords: stochastic game, hidden state, uniform connectedness, robust connectedness, random blocks, folk theorem.

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

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

Contents

1 Introduction

When agents have a long-run relationship, underlying economic conditions may change over time. A leading example is a repeated Bertrand competition with stochastic demand shocks. Rotemberg and Saloner (1986) explore optimal collusive pricing when random demand shocks are i.i.d. each period. Haltiwanger and Harrington (1991), Kandori (1991), and Bagwell and Staiger (1997) further extend the analysis to the case in which demand fluctuations are cyclic or persistent. A common assumption of these papers is that demand shocks are publicly observable *before* firms make their decisions in each period. This means that in their model, firms can perfectly adjust their price contingent on the true demand today. However, in the real world, firms often face uncertainty about the market demand when they make decisions. Firms may be able to learn the current demand shock through their sales *after* they make decisions; but then in the next period, a new demand shock arrives, and hence they still face uncertainty about the true demand. When such uncertainty exists, equilibrium strategies considered in the existing work are no longer equilibria, and players may want to "experiment" to obtain better information about the hidden state. This paper investigates whether and how players can maintain cooperation in this new environment.

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. The signal y can be informative about both the current and next states, which ensures that our formulation accommodates a wide range of economic applications, including games with delayed observations and a combination of observed and unobserved states.

Since we assume that actions are perfectly observable, players have no private information, and hence after every history, different players have the same posterior belief μ^t about the current state ω^t. Then this posterior belief μ^t can be regarded as a common state variable, and 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 would like: Since there are infinitely many possible posterior beliefs, we need to consider a stochastic game with *infinite* states. This is in a sharp contrast with past work which assumes *finite* states (Dutta (1995), Fudenberg and Yamamoto (2011b), and Hörner, Sugaya, Takahashi, and Vieille (2011)).[1]

In general, the analysis of stochastic games is different from that of repeated games, because the action today influences the distribution of the future states, which in turn influences the stage-game payoffs in the future. For the finite-state case, past work shows that this effect vanishes for patient players, under a mild condition. Formally, if states are *communicating* in that players can move the state from any state to any other state, then the feasible payoff set is invariant to the initial state in the limit as the discount factor goes to one. This invariance result ensures that even if someone deviates today and influences the distribution of the state tomorrow, it does not change the feasible payoff set in the continuation game from tomorrow; so players can choose the continuation payoff in a flexible way, regardless of the action today. This property allows us to discipline players' play via intertemporal incentives as in repeated games.

Why is the feasible payoff set invariant for the finite-state case? To see this, consider the welfare-maximizing payoff vector in the feasible payoff set, and suppose that players play a strategy profile which achieves this payoff. Without loss of generality, we can assume that it is a Markov strategy so that the state follows a Markov process. When states are finite, a Markov process is *ergodic* under a mild assumption; indeed, under the communicating state assumption, the Markov processes here is ergodic. Ergodicity ensures that the initial state cannot influence the state in a distant future. This immediately implies that the welfare-maximizing payoff is invariant to the initial state, since patient players care only about payoffs in a distant future. A similar argument shows that the entire feasible payoff set is also invariant to the initial prior.

On the other hand, when states are infinite, a Markov process is not ergodic

[1]For the infinite-state case, the existence of Markov perfect equilibria is extensively studied. See recent work by Duggan (2012) and Levy (2013), and an excellent survey by Dutta and Sundaram (1998). In contrast to this literature, we consider general non-Markovian equilibria. Hörner, Takahashi, and Vieille (2011) consider non-Markovian equilibria in infinite states, but they assume that the limit equilibrium payoff set is invariant to the initial state. That is, they directly assume a sort of ergodicity and do not investigate when it is the case.

in many cases. This is essentially because states are not *positive recurrent* in the sense that the state may not return to the current state forever. While there are some sufficient conditions for ergodicity of infinite-state Markov chains (e.g. *Doeblin condition*, see Doob (1953)) , these conditions are not satisfied in our setup, as in the context of partially observable Markov decision process (POMDP).[2] Accordingly, *a priori*, we cannot rule out the possibility that the invariance result does not hold. This implies that the action today may have a long-lasting effect on future feasible payoffs, which potentially makes it difficult to discipline players' play.

The first finding of this paper is that, despite the potential lack of ergodicity, the above invariance result still holds in our setup. Specifically, we show that if the game satisfies a new property called *uniform connectedness*, then the feasible payoff set is invariant to the initial belief for patient players. We also show that the minimax payoff for patient players is invariant to the initial belief under a similar assumption called *robust connectedness*. Our proof is substantially different from that in the literature, since the techniques which refer to ergodic theorems are not directly applicable due to infinite states.

Our assumption, uniform connectedness, is a condition about how the *support* of the belief evolves over time. Its formal definition is given in Section 3, but roughly, it requires that players can jointly drive the support of the belief from any set Ω^* to any other set $\tilde{\Omega}^*$, except the case in which the set $\tilde{\Omega}^*$ is "transient" in the sense that the support cannot stay at $\tilde{\Omega}^*$ forever. (Here, Ω^* and $\tilde{\Omega}^*$ denote subsets of the whole state space Ω.) Note that uniform connectedness is *not* a condition on the evolution of the belief itself, and thus it does not imply ergodicity of the belief evolution. Also it does not imply ergodicity of the support either, because the exact probability distribution of the support tomorrow depends not only on the current support but on the fine details of the current belief. Nonetheless, this condition is sufficient for our invariance result. A key step in the proof is to find a uniform bound on the variability of feasible payoffs over beliefs with the same support. In turns out that this bound is close to zero, which implies that the feasible payoff set is almost determined by the support of the belief. Hence what is essential is how the support changes over time, which suggests that uniform connectedness is useful to obtain the invariance result. It may be worth noting

[2]This is explained in the introduction of Rosenberg, Solan, and Vieille (2002), for example.

that when states are observable and communicating, uniform connectedness is always satisfied; so uniform connectedness allows us to develop a general theory which subsumes the existing models as a special case.

Uniform connectedness is satisfied in a wide range of applications with hidden states. Two examples are presented in Section 4. The first example is a repeated Bertrand model in which a hidden demand shock (state) follows a Markov process. As noted, experimentation naturally arises in our hidden-state model; for example, in this Bertrand model, firms may choose prices which do not maximize the expected profit today, in order to obtain better information about the hidden state. We illustrate how such experimentation is incorporated with efficient equilibria. While the value of information is temporary, we find that even patient firms do not stop experimentation. The second example is a novel natural resource management problem. In this example, the state is the number of fish living in the gulf. The state may increase or decrease over time, due to natural increase or overfishing. Since the fishermen (players) cannot directly count the number of fish in the gulf, this is one of the examples in which the belief about the hidden state plays an important role in applications. The state evolution is more complicated than in the Bertrand model, because actions (whether to fish or not) directly influence the state tomorrow. We show that community-based institutions are helpful to manage local environmental resource competition.

In addition to that, we show in Appendix B that the invariance result holds under a condition even weaker than uniform connectedness, called *asymptotic uniform connectedness*. Asymptotic uniform connectedness is satisfied for generic hidden-state games, as long as the underlying states are communicating. This means that the payoff invariance result (almost always) holds as long as the evolution of the hidden state satisfies the standard assumption in the literature.

Section 6 presents the second main result of the paper, the folk theorem; assuming uniform connectedness and robust connectedness, we show that any feasible and individually rational payoffs are achieved by sequential equilibria for patient players. Note that in our environment, even the existence of a trivial equilibrium (such as a Markov perfect equilibrium) is an open question, due to infinite states.[3] This also suggests that we cannot construct a simple trigger strategy in

[3]Levy (2013) presents an example of infinite-state stochastic games which does not have a Markov perfect equilibrium. When the belief is the state, a Markov perfect equilibrium does not

general, as it uses a trivial equilibrium for a punishment. Our proof of the folk theorem develops a general method to find an equilibrium in such an environment. Indeed, our proof is *constructive*, and directly describes how players reward or punish the opponents in our equilibrium.

The main challenge in the proof is to figure out an effective punishment mechanism in our model. In the standard repeated-game model, Fudenberg and Maskin (1986) consider a simple equilibrium in which a deviator will be minimaxed for T periods and then those who minimaxed will be rewarded. Promising a reward after the minimax play is important, because the minimax profile itself is not an equilibrium and players would be reluctant to minimax without such a reward. Unfortunately, this "T-period punishment mechanism" does not directly extend to our environment. To see this, suppose that we fix δ first and then take sufficiently large T. In this case, δ^T approaches zero, which implies that players do not care about payoffs after the minimax play. So even if we promise a reward after the minimax play, it cannot provide appropriate incentives. What if we take sufficiently large T first and then take $\delta \to 1$, as in Fudenberg and Maskin (1986)? In this case, due to the potential failure of belief ergodicity, for any fixed T, the average payoff over T periods can be quite different from the infinite-horizon game payoff, and hence quite different from the minimax payoff. Hence the minimax play over T periods may not be an effective punishment.

To solve this problem, we introduce the idea of *random blocks*, whose lengths are randomly determined by public randomization. Specifically, at the end of each period, public randomization determines whether the current random block continues or terminates with probability p and $1-p$. This random block is payoff-equivalent to *the infinite-horizon game with the discount factor* $p\delta$, due to the termination probability $1-p$. Hence if players play the minimax strategy during the random block, the expected payoff during the block is exactly the minimax payoff with the discount factor $p\delta$. Now, suppose that we first fix sufficiently large p, and then take $\delta \to 1$. Then although the expected block length $\frac{1}{1-p}$ is quite long, the discount factor δ is arbitrarily close to one, so players still care about the payoffs after the random block. Hence a promised reward after the

exist in the bargaining model (Fudenberg, Levine, and Tirole (1985)) and in the reputation model (Mailath and Samuelson (2006)). While their results do not directly apply to our setup, they suggest that a similar problem may arise in our model.

minimax play provides appropriate incentives.[4]

Shapley (1953) proposes the framework of stochastic games. 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 his result to games with public monitoring. All these papers assume that the state of the world is publicly observable at the beginning of each period.[5]

Athey and Bagwell (2008), Escobar and Toikka (2013), and Hörner, Takahashi, and Vieille (2015) consider repeated Bayesian games in which the state changes as time goes and players have private information about the current state each period. They look at equilibria in which players report their private information truthfully, which means that the state is perfectly revealed before they choose actions each period.[6] In contrast, in this paper, players have only limited information about the true state and the state is not perfectly revealed.

Wiseman (2005), Fudenberg and Yamamoto (2010), Fudenberg and Yamamoto (2011a), and Wiseman (2012) study repeated games with unknown states. They all assume that the state of the world is fixed at the beginning of the game and does not change over time. Since the state influences the distribution of a public signal each period, players can (almost) perfectly learn the true state by aggregating all the past public signals. In contrast, in our model, the state changes as time goes and thus players never learn the true state perfectly.

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

[4] Interestingly, some papers on macroeconomics (such as Arellano (2008)) assume that punishment occurs in a random block; we thank Juan Pablo Xandri for pointing this out. Our analysis is different from theirs because random blocks endogenously arise in equilibrium.

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

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

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. 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}. 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.

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 $\overline{g}_i = \max_{\omega, a} |2 g_i^\omega(a)|$, and let $\overline{g} = \sum_{i \in I} \overline{g}_i$. Also let $\overline{\pi}$ be the minimum of $\pi^\omega(y, \tilde{\omega}|a)$ over all $(\omega, \tilde{\omega}, a, y)$ such that $\pi^\omega(y, \tilde{\omega}|a) > 0$.

Our formulation encompasses the following examples:

- *Stochastic games with observable states.* Let $Y = \Omega \times \Omega$ and suppose that $\pi^\omega(y, \tilde{\omega}|a) = 0$ for $y = (y_1, y_2)$ such that $y_1 \neq \omega$ or $y_2 \neq \tilde{\omega}$. That is, the first component of the signal y reveals the current state and the second component reveals the next state. Suppose also that $u_i(a, y)$ does not depend on the second component y_2, so that stage-game payoffs are influenced by the current state only. Since the signal in the previous period perfectly reveals the current state, players know the state ω^t *before* they move. This is exactly the standard stochastic games studied in the literature.

- *Stochastic games with delayed observations.* 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. So players observe the state *after* they move.

- *Observable 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 com-

ponent of the next state, ω_O^{t+1}. Then we can interpret ω_O as an observable state and ω_U as an unobservable state. One of the examples which fits this formulation is a duopoly market in which firms face uncertainty about the demand, and their cost function depends on their knowledge, know-how, or experience. The firms' experience can be described as an observable state variable as in Besanko, Doraszelski, Kryukov, and Satterthwaite (2010), and the uncertainty about the market demand as an unobservable state.

In the infinite-horizon stochastic game, players have a common discount factor $\delta \in (0,1)$. Let $(\omega^\tau, a^\tau, y^\tau)$ be the state, the action profile, and the public signal in period τ. Then the history up to period $t \geq 1$ is denoted by $h^t = (a^\tau, y^\tau)_{\tau=1}^t$. 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 a mapping $s_i : H \to \triangle A_i$ Let S_i be the set of all strategies for player i, and let $S = \times_{i \in I} S_i$. 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}, player i's best reply s_i exists; see Appendix E for the proof. 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 hidden 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 reduces to 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 period $t \geq 2$, players use Bayes' rule to update the belief; given μ^{t-1}, a^{t-1}, and y^{t-1}, let

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

for each $\tilde{\omega}$. Given this (common) belief μ^t, players chooses actions a^t, and then observe a signal y according to the distribution $\pi_Y^{\mu^t}(y|a) = \sum_{\omega\in\Omega}\mu^t(\omega)\pi_Y^{\omega}(y|a)$. Player i's expected stage-game payoff given μ^t and a^t is $g_i^{\mu^t}(a^t) = \sum_{\omega\in\Omega}\mu^t(\omega)g_i^{\omega}(a^t)$.

Now we give the definition of sequential equilibria. Let $\zeta : H \to \triangle\Omega$ be a belief system; i.e., $\zeta(h^t)$ is the posterior about 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, given the initial prior μ, a consistent belief is unique at each information set which is reachable by some strategy. (So essentially there is a unique belief system ζ consistent with μ.) A strategy profile s is a *sequential equilibrium* in the stochastic game with the initial prior μ if s is sequentially rational given the belief system ζ consistent with μ.

3 Uniformly Connected Stochastic Games

As discussed in the introduction, stochastic games can be different from infinitely repeated games in general, because the action today may influence the distribution of the future states and hence the future stage-game payoffs. To avoid a difficulty arising from this structure, most of the existing papers assume that players can move the state from any state to any other state (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, we need to identify an appropriate condition which parallels the assumption in the standard model. We find that one of such conditions is *uniform connectedness*, which imposes a restriction on how the *support* of the posterior belief evolves over time.

3.1 Full Support Assumption

Uniform connectedness is satisfied in a wide range of examples, but its definition is a bit complex, Hence it would be desirable to have a simple sufficient condition for uniform 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 $\pi_Y^{\omega}(y|a) > 0$.

In words, the full support assumption holds if 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 uniform connectedness, and it rules out many economic applications. For example, the full support assumption is never satisfied for stochastic games with observable states; so the standard model is ruled out. Also, the full support assumption is never satisfied if the action and/or the signal today has a huge impact on the state evolution so that some state $\tilde{\omega}$ cannot happen tomorrow conditional on some (a,y). One of such examples is the natural resource management problem in Section 4.2; in this example, if the fishermen catch too much fish today, the state (the number of fish in the gulf) cannot be the highest state tomorrow because natural increase is slow. To study such economic examples, uniform connectedness, which is more general than the full support assumption, is a useful concept.

3.2 Uniform Connectedness

In this subsection, we introduce the idea of uniform connectedness, which relaxes the full support assumption. Also we illustrate how uniform connectedness is related to the common assumption in the literature. While uniform connectedness is more general than the full support assumption, its definition is somewhat complicated; so those who are less interested in generalization of the full support assumption may skip this subsection. It should not cause any problem in later sections, except Section 4.2 (which shows that uniform connectedness is satisfied in the natural resource management problem) and Section 5.3 (which explains how to prove Proposition 4 under uniform connectedness).

As noted, for games with observable states, most of the existing papers assume that there is a path from any state to any other state. 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

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

where $\Pr(\omega^{T+1} = \tilde{\omega} | \omega, a^1, \cdots, a^T)$ denotes the probability of the state in period $T+1$ being $\tilde{\omega}$ 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 ω. The *communicating state* assumption in the literature requires all states ω to be globally accessible (Dutta (1995), Fudenberg and Yamamoto (2011b), and Hörner, Sugaya, Takahashi, and Vieille (2011)).

Since the belief μ is the state variable in our model, a natural extension of the above assumption is to assume that there be a path from any belief μ to any other belief $\tilde{\mu}$. Unfortunately, this approach does not work, because such a condition is too demanding and never satisfied. A major 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.[7]

Instead, we focus on the evolution of the *support* of the belief, rather than the evolution of the belief itself. Advantage of this approach is that the set of possible supports is finite, and thus the recurrence problem stated above is not an issue. Of course, the support of the belief is only coarse information about the belief, so imposing a condition on the evolution of the support is much weaker than imposing a condition on the evolution of the belief. However, as it will turn out, this is precisely the condition we need for Proposition 4.

To illustrate the idea of uniform connectedness, suppose that there are three states ($\Omega = \{\omega_1, \omega_2, \omega_3\}$) so that there are seven possible supports Ω_1, $\cdots$, Ω_7. Figure 1 shows how the support of the belief changes over time. For each arrow, there is an action profile which lets the support move along the arrow with positive probability. For example, there is an action profile which moves the support from Ω_1 to Ω_2 with positive probability. Each thick arrow is a move which must happen with positive probability *regardless of* the action profile. The thick forked arrow from Ω_6 means that the support must move to either Ω_2 or Ω_3 with positive probability regardless of the action profile, but its destination may depend on the action profile. Note that the evolution of the support described in the picture is

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

well-defined, because if two initial priors μ and $\tilde{\mu}$ have the same support, then after every history h^t, the corresponding posterior beliefs $\mu(h^t)$ and $\tilde{\mu}(h^t)$ have the same support.

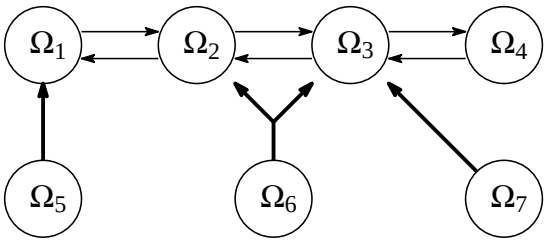

Figure 1: Connectedness

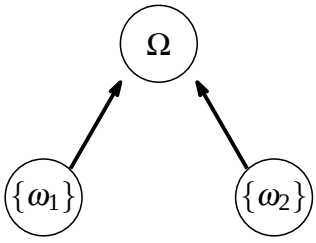

Figure 2: Full Support

In this example, the support Ω_1 is *globally accessible* in the sense that there is a path to Ω_1 from any current support; for example, the support can move from Ω_7 to Ω_1 through Ω_3 and Ω_2. (Formally, global accessibility is more general than this because it requires only that there be a path to Ω_1 or a subset of Ω_1. Details will be given later.) Likewise, Ω_2, Ω_3, and Ω_4 are globally accessible. So the four supports Ω_1, Ω_2, Ω_3, and Ω_4 are "connected" in the sense that the support can go back and forth within these supports.

The support Ω_5 is not globally accessible, because it is not accessible from Ω_1. However, this support Ω_5 is *uniformly transient* in the sense that if the current support is Ω_5, then *regardless of players' play*, the support cannot stay there forever and must move to other globally accessible set (in this case Ω_1) with positive probability, due to the thick arrow. Similarly, the supports Ω_6 and Ω_7 are uniformly transient, as the support must move to globally accessible sets Ω_2 or Ω_3, depending on the chosen action profile. An important consequence of uniform transience is that if the current support is uniformly transient, the support will eventually move to a globally accessible set with probability one, regardless of players' play. Uniform connectedness requires each support $\Omega^* \subseteq \Omega$ to be globally accessible or uniformly transient; in other words, it requires that all "essential" supports be globally accessible. Figure 1 is one of the examples which satisfy uniform connectedness.

Uniform connectedness is weaker than the full support assumption. To see this, suppose that there are two states, ω_1 and ω_2, and that the full support assumption holds. Figure 2 shows how the support changes in this situation. We

have two thick arrows to the whole state space Ω, because the full support assumption ensures that the support of the posterior belief must be Ω regardless of the current support. The set Ω is globally accessible because there is a path from any support. Also the sets $\{\omega_1\}$ and $\{\omega_2\}$ are uniformly transient, because the support must move to the globally accessible set Ω regardless of players' actions. Note that the same result holds even if there are more than two states; the whole state space Ω is globally accessible, and all proper subsets $\Omega^* \subset \Omega$ are uniformly transient. Hence the full support assumption implies uniform connectedness.

Now we provide the formal definition of global accessibility and uniform transience. Let $\Pr(\mu^{T+1} = \tilde{\mu}|\mu, s)$ denote the probability of the posterior belief in period $T+1$ being $\tilde{\mu}$ given that the initial prior is μ and players play the strategy profile s. Similarly, let $\Pr(\mu^{T+1} = \tilde{\mu}|\mu, a^1, \cdots, a^T)$ denote the probability given that players play the action sequence $(a^1, \cdots, a^T)$ in the first T periods.

Definition 2. A non-empty subset $\Omega^* \subseteq \Omega$ is *globally accessible* if there is $\pi^* > 0$ such that for any initial prior μ, there is a natural number $T \leq 4^{|\Omega|}$, an action sequence $(a^1, \cdots, a^T)$, and a belief $\tilde{\mu}$ whose support is included in Ω^* such that

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

Global accessibility of Ω^* requires that given any initial prior μ, players can move the support of the posterior belief to Ω^* or its subset with probability at least $\pi^* > 0$, by choosing some appropriate action sequence which may depend on μ.[8] A couple of remarks are in order. First, the above condition requires that there be a lower bound $\pi^* > 0$ on the probability of the posterior belief reaching $\tilde{\mu}$, while (1) does not. The reason why π^* does not show up in (1) is that when states are observable, possible initial states are finite and thus a lower bound $\pi^* > 0$ always exists. On the other hand, we need to explicitly assume the existence of the bound π^* here, since there are infinitely many initial priors μ. Note that there is π^* which works for all globally accessible sets Ω^*, since there are only finitely many supports.

[8] Replacing the action sequence $(a^1, \cdots, a^T)$ in the definition with a strategy profile s does not weaken the condition; that is, as long as there is a strategy profile which satisfies the condition stated in the definition, we can find an action sequence which satisfies the same condition. On the other hand, the strategy profile s in Definition 3 cannot be replaced with an action sequence $(a^1, \cdots, a^T)$.

The existence of the lower bound $\pi^* > 0$ ensures that in the infinite-horizon game, players can eventually move the support to Ω^* (or its subset) with probability one. To see this, suppose that Ω^* is globally accessible with $T = 1$. That is, if players choose appropriate actions, each period the support moves to Ω^* with probability at least π^*. Hence the probability that the support cannot reach Ω^* until period t is at most $(1-\pi^*)^{t-1}$, which converges to zero as t goes to infinity. This property plays a crucial role in the proof of Proposition 4.

Second, globally accessibility of Ω^* does not require the support of the posterior to be exactly equal to Ω^*; it requires only that the support of the posterior to be a subset of Ω^*. So global accessibility is a weaker condition than the one we discussed using Figure 1. Due to this property, it is obvious that the whole state space $\Omega^* = \Omega$ is globally accessible for any game. Also if a subset Ω^* is globally accessible, then so is any superset $\tilde{\Omega}^* \supseteq \Omega^*$; this is because global accessibility ensures the existence of a path to Ω^*, which is a subset of $\tilde{\Omega}^*$.

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

Next, we give the formal definition of uniform transience. Like global accessibility, the restriction $T \leq 2^{|\Omega|}$ is without loss of generality.

Definition 3. A subset $\Omega^* \subseteq \Omega$ is *uniformly transient* if it is not globally accessible and for any pure strategy profile s and for any μ whose support is Ω^*, there is a natural number $T \leq 2^{|\Omega|}$ and a belief $\tilde{\mu}$ whose support is globally accessible such that $\Pr(\mu^{T+1} = \tilde{\mu} | \mu, s) > 0$.

In words, uniform transience of Ω^* requires that if the support of the current belief is Ω^*, then *regardless of future actions*, the support of the posterior belief must reach some globally accessible set with positive probability at some point. In the definition above, there is no lower bound $\pi^* > 0$ on the probability of the support reaching to a globally accessible set. However, as Proposition D2 in Appendix D shows, when the game is uniformly connected, we can find a lower bound $\pi^* > 0$, that is, the support must reach a globally accessible set with probability at least $\pi^* > 0$ regardless of the initial prior μ. Also the probability of the support returning to the uniformly transient set Ω^* is bounded away from one, as Proposition D3 shows. Taken together, we can conclude that the probability of

the support being uniformly transient in period T approximates zero when T is large enough. So when we consider the long-run evolution of the support of the belief, uniformly transient supports are "not essential" in the sense that the time during which the support stays at these sets is almost negligible.

As noted, a superset of a globally accessible set is globally accessible. Using this property, we can show that a superset of a uniformly transient set is globally accessible or uniformly transient. These results are useful when we check uniform connectedness in applications, so we record them as a proposition. The proof is given in Appendix C.

Proposition 1. *A superset of a globally accessible set is globally accessible. Also, a superset of a uniformly transient set is globally accessible or uniformly transient.*

Our assumption, uniform connectedness, requires that each subset Ω^* be either globally accessible or uniformly transient. This ensures that regardless of the initial prior and regardless of future actions, the support of the posterior belief must eventually reach some globally accessible set and move within these globally accessible sets.

Definition 4. A stochastic game is *uniformly connected* if each subset $\Omega^* \subseteq \Omega$ is globally accessible or uniformly transient.

As argued, uniform connectedness is weaker than the full support condition. This generalization is important, because the full support assumption is not satisfied when actions or signals directly influence the state so that some state cannot happen tomorrow conditional on some (a,y). In Section 4.2, we consider one of such examples, the natural resource management problem, and show that it satisfies uniform connectedness.

For stochastic games with observable states, uniform connectedness reduces to a simple condition. Specifically, as the following proposition shows, uniform connectedness is equivalent to assuming that each state ω be globally accessible or uniformly transient; here, a state ω is *uniformly transient* if for any pure strategy profile s, there is a natural number T and a globally accessible state $\tilde{\omega}$ so that $\Pr(\omega^{T+1} = \tilde{\omega}|\omega, s) > 0$. This result shows that uniform connectedness is weaker than the communicating state assumption of Dutta (1995), which requires each

state ω to be globally accessible. So our model subsumes the ones studied in the literature as a special case. Also the proposition shows that the same result holds even for stochastic games with delayed observations. The proof can be found in Appendix C.

Proposition 2. *In stochastic games with observable states, the game is uniformly connected if and only if each state ω is globally accessible or uniformly transient. Similarly, in stochastic games with delayed observations, the game is uniformly connected if and only if each state ω is globally accessible or uniformly transient.*

Unfortunately, the above result does not hold when the state ω is not observable. Indeed, as will be shown in Appendix B, there is an example in which each state is globally accessible but nonetheless the game is not uniformly connected. So the if part of the above proposition does not extend to the hidden-state case. On the other hand, as the next proposition shows, the only if part remains true even if the state ω is not observable. The proof can be found in Appendix C.

Proposition 3. *The game is uniformly connected only if each state ω is globally accessible or uniformly transient.*

This proposition gives a necessary condition for uniform connectedness; if there is a state ω which is neither globally accessible nor uniformly transient, the uniformly connectedness is never satisfied. So for example, uniform connectedness is never satisfied if there are multiple absorbing states.[9]

The definition of uniform connectedness here is stated using the posterior belief μ^t. In Appendix D, we give an equivalent definition of uniform connectedness based on primitives. Using this definition, one can check if a given game is uniformly connected or not in finitely many steps.

[9]To confirm this, suppose that two states ω_1 and ω_2 are absorbing. Then the set $\{\omega_1\}$ is not globally accessible, because the support cannot move from $\{\omega_2\}$ to $\{\omega_1\}$. Also it is not uniformly transient because the support cannot move to any other set. Hence uniform connectedness does not hold, regardless of the signal structure.

4 Examples

4.1 Bertrand Competition with Hidden Demand Shocks

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 the 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 is not very sensitive to the underlying economic condition. If the effective price is zero, the probability of y^H is one regardless of the current state ω. We assume that the realized demand y is public information. Assume also that y and the next state $\tilde{\omega}$ are independently drawn.

This is the Bertrand model, so a firm with a lower price takes the whole market share. Accordingly, firm i's current profit is $u_i(a,y) = a_i y$ if $a_i < a_{-i}$, and $u_i(a,y) = 0$ if $a_i > a_{-i}$. If $a_i = a_{-i}$, the firms share the market equally so that $u_i(a,y) = \frac{a_i y}{2}$. 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 high price a^H if $\mu \geq \frac{44}{117} \approx 0.376$, and by the low price a^L if $\mu \leq \frac{44}{117}$.

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

$$f(\mu) = \max_{a\in A} \left[(1-\delta)g^\mu(a) + \delta \sum_{y\in Y} \pi_Y^\mu(y|a) f(\tilde{\mu}(\mu,a,y)) \right] \tag{2}$$

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, (2) says that the total profit $f(\mu)$ consists of today's profit $g^\mu(a)$ and the expectation of the future profits $f(\tilde{\mu}(\mu,a,y))$, and that the current action should maximize it.

For each discount factor $\delta \in (0,1)$, we can derive an approximate solution to (2) by value function iteration with a discretized belief space. Figure 3 shows the value function f for $\delta = 0.7$. As one can see, the value function f is upward sloping, which means that the total profit becomes larger when the initial prior becomes more optimistic.

Figure 4 shows the optimal policy; in the vertical axis, 0 means the low price a^L, while 1 means the high price a^H. It shows that the optimal policy is a simple cut-off rule, that is, the optimal action is the low price a^L when the current belief μ is less than μ^{**}, and is 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, $\frac{44}{117} \approx$

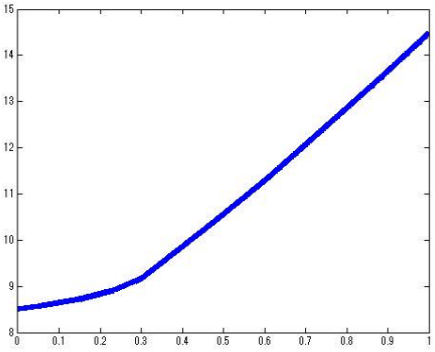

Figure 3: Value Function
x-axis: belief μ. y-axis: payoffs.

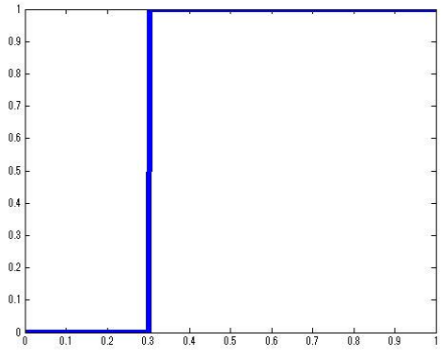

Figure 4: Optimal Policy
x-axis: belief μ. y-axis: actions.

0.376. So when the current belief is $\mu \in (\mu^*, \frac{44}{117})$, the firms choose the high price although it 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.[10] To see this, for each a, let $\Pi(a)$ denote the two-by-two matrix with rows $(\pi_Y^\omega(y^H|a), \pi_Y^\omega(y^L|a))$ for each 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)), and in this sense, the public signal y given the low price a^L is less informative than that given the high price.

So 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 $\mu \in (\mu^*, \frac{44}{117})$, this effect dominates the decrease in the current profit. Hence the optimal policy chooses the high price.

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,

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

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

It is worth emphasizing that the firms do "experiments" in this efficient equilibrium. As argued, when the current belief is $\mu \in (\mu^*, \frac{44}{117})$, the firms choose the high price a^H in order to obtain better information, although it does not maximize the current expected payoff.

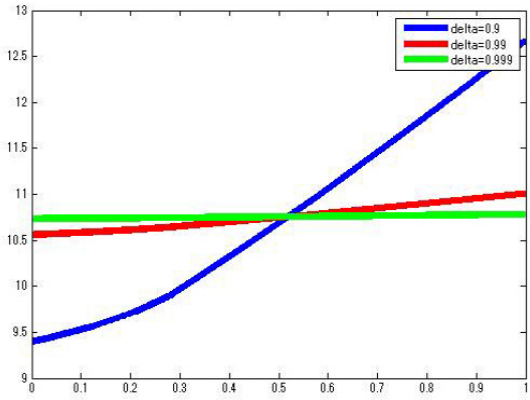

Figure 5: Value Functions for High δ
x-axis: belief μ. y-axis: payoffs.

Of course, the solution to (2) depends on the discount factor δ. Figure 5 illustrates how the value function changes when the firms become more patient; it gives the value functions for $\delta = 0.9$, $\delta = 0.99$, and $\delta = 0.999$. The optimal policies are still cut-off rules, and the cut-off value is $\mu = 0.285$ for $\delta = 0.9$, $\mu = 0.276$ for $\delta = 0.99$, and $\mu = 0.275$ for $\delta = 0.999$. Note that the cut-off value is

decreasing in the discount factor; that is, patient firms experiment more frequently. This result comes from the fact that patient firms care more about future profits, and thus information about the hidden state tomorrow is more valuable.

One may think that this result is counter-intuitive, because the impact of the current hidden state on future states is temporary. Indeed, the hidden state ω follows an ergodic Markov process, and the state tomorrow has only a negligible impact on the state in a distant future. This implies that the information about the hidden state has only a negligible impact on average payoffs for patient firms. Nonetheless, the numerical solution above shows that patient firms experiment more frequently. To reconcile this issue, note that the impact of the stage-game payoff today on average payoffs is also vanishing for patient firms, at a rate $1-\delta$. Since the optimal policy is determined by the trade-off between the stage-game payoff today and the value of information, we need to investigate which effect is more persistent when δ approaches one. Now, it is easy to see that the latter effect dominates the former; indeed, if we consider the *unnormalized payoff* rather than the average payoff, the impact of the future payoffs increases in δ, while the impact of the stage-game payoff does not change. Hence patient firms values the information about the hidden state more, and they experiment more frequently.

As Figure 5 shows, when the firms become patient, the value function becomes almost flat. That is, the firms' initial prior has almost no impact on the total profit. This property is not specific to this example; we show in Proposition 4 that under uniform connectedness, the feasible payoff set is invariant to the initial prior in the limit as the discount factor goes to one. As discussed in the introduction, this invariance result plays a crucial role in order to obtain the folk theorem.

4.2 Natural Resource Management

In the Bertrand example in the last subsection, we have seen that the firms can attain efficiency by the trigger strategy in which a deviator will be punished by the Markov equilibrium "forever a^*." However, such a trigger strategy may not work in other games, because the existence of Markov equilibria is not guaranteed when the state space is infinite (see Duggan (2012) and Levy (2013)) and it is less clear how to punish a deviator. In this subsection, we consider an example in which a Markov equilibrium may not exist, and explain that efficiency is still achievable

even in such a case. Also we explain how to check uniform connectedness in a given model.

Suppose that two fishermen live near a gulf. The state of the world is the number of fish in the gulf, which is denoted by $\omega \in \{0, \cdots, K\}$ where K is the maximal capacity. The fishermen cannot directly observe the number of fish, ω, so their beliefs about ω impact their incentives in a crucial way.

Each period, each fisherman decides whether to "Fish" (F) or "Do Not Fish" (N); so fisherman i's action set is $A_i = \{F, N\}$. Let $y_i \in Y_i = \{0, 1, 2\}$ denote the amount of fish caught by fisherman i, and let $\pi_Y^\omega(y|a)$ denote the probability of the outcome $y = (y_1, y_2)$ given the current state ω and the current action profile a. We assume that if fisherman i chooses N, then he cannot catch anything so that $y_i = 0$. That is, $\pi_Y^\omega(y|a) = 0$ if there is i with $a_i = N$ and $y_i > 0$. We also assume that the fishermen cannot catch more than the number of fish in the gulf, so $\pi_Y^\omega(y|a) = 0$ for ω, a, and y such that $\omega < y_1 + y_2$. We assume $\pi_Y^\omega(y|a) > 0$ for all other cases, so the signal y does not reveal the hidden state ω.

Fisherman i's utility in each stage game is 0 if he chooses N, and is $y_i - c$ if he chooses F. Here $c > 0$ denotes the cost of choosing F, which involves effort cost, fuel cost for a fishing vessel, and so on. We assume that $c < \sum_{y \in Y} \pi_Y^\omega(y|F, a_{-i}) y_i$ for some ω and a_{-i}, that is, the cost is not too high and the fishermen can earn positive profits by choosing F, at least for some state ω and the opponents' action a_{-i}. If this assumption does not hold, no one fishes in any equilibrium.

Over time, the number of fish may increase or decrease due to natural increase or overfishing. Specifically, we assume that the number of fish in period $t+1$ is determined by the following formula:

$$\omega^{t+1} = \omega^t - (y_1^t + y_2^t) + \varepsilon^t. \tag{3}$$

In words, the number of fish tomorrow is equal to the number of fish in the gulf today minus the amount of fish caught today, plus a random variable $\varepsilon^t \in \{-1, 0, 1\}$, which captures natural increase or decrease of fish. Intuitively, $\varepsilon = 1$ implies that some fish had an offspring or new fish came to the gulf from the open sea. Similarly, $\varepsilon = -1$ implies that some fish died out or left the gulf. Let $\Pr(\cdot|\omega, a, y)$ denote the probability distribution of ε given the current ω, a, and y. We assume that the state ω^{t+1} is always in the state space $\Omega = \{0, \cdots, K\}$, that is, $\Pr(\varepsilon = -1|\omega, a, y) = 0$ if $\omega - y_1 - y_2 = 0$ and $\Pr(\varepsilon = 1|\omega, a, y) = 0$ if $\omega - y_1 - y_2 = K$.

We also assume that $\Pr(\varepsilon = 1|\omega, a, y) = 0$ if $\omega - y_1 - y_2 = 0$ and $a \neq (N, N)$, that is, if the state reaches the lowest ($\omega - y_1 - y_2 = 0$) and at least one of the fishermen try to catch ($a \neq (N, N)$), there will be no natural increase. This captures the idea that there is a critical biomass level below which the growth rate drops rapidly. (This assumption simplifies the computation of the minimax payoff, but it is not essential; even if this assumption is dropped, all the results remain valid, as will be explained.) We assume $\Pr(\varepsilon|\omega, a, y) > 0$ for all other cases.

This model can be interpreted as a dynamic version of "tragedy of commons." The fish in the gulf is public good, and overfishing may result in resource depletion. Competition for natural resources like this is quite common in the real world, due to growing populations, economic integration, and resource-intensive patterns of consumption. For example, each year Russian and Japanese officials discuss salmon fishing within 200 nautical miles of the Russian coast, and set Japan's salmon catch quota. (The quota for 2015 is 1310 tons for chum salmon, 503 tons for sockeye, 103 tons for pink, 41 tons for silver, and 5 tons for king.) Often times, it is argued that community-based institutions are helpful to manage local environmental resource competition. Our goal here is to provide its theoretical foundation.

As in the Bertrand model in the previous subsection, the optimal resource management plan can be characterized by solving the Bellman equation (2). That is, the fishermen can maximize the sum of their infinite-horizon game payoffs by playing the optimal policy which solves (2). In general, this optimal policy is not incentive compatible, and the fishermen may want to deviate from it in order to maximize their own payoffs. So the question is whether there is an equilibrium in which the fishermen indeed play this optimal policy on the equilibrium path.

The equilibrium analysis here is more complicated than that for the Bertrand model, because a Markov equilibrium may not exist in this example and the trigger strategy cannot be used. Nonetheless it is still possible to construct an approximately efficient equilibrium. A key is that this example satisfies uniform connectedness: Uniform connectedness ensures that the feasible payoff set is invariant to the initial prior for patient players (Proposition 4), which in turn implies that the folk theorem obtains in this example (Proposition 7). Hence as long as players are patient enough, there is an equilibrium which approximates the efficient payoff. The proof of the folk theorem is constructive, so we can describe

how players punish a deviator.

In what follows, we show that this game indeed satisfies uniform connectedness, and hence the feasible payoff set is invariant to the initial prior. Note that this example does not satisfy the full support assumption, because the state ω cannot jump from 0 to K in one period. So introducing the notion of uniform connectedness is crucial to study this example.

We first show that $\Omega^* = \{0\}$ is globally accessible, that is, we show that given any initial prior μ, players can move the support to $\{0\}$. Pick an arbitrary initial prior μ. Suppose that the fishermen do not fish for the first K periods, and then they both fish and observe $y = (1,1)$ for the next $K-1$ periods. We claim that the support of the posterior after this history is $\{0\}$, as desired. To see this, note that if $y = (1,1)$ is observed, the state tomorrow will decrease at least by one, because the fishermen caught more than the possible natural increase $\varepsilon = 1$. In particular, if the current state is $\omega = 2$, the state must decrease by two because we must have $\varepsilon = 0$ for that period. Hence, even if the current state is the highest level $\omega = K$, if $y = (1,1)$ is observed for $K-1$ consecutive periods, the state must reach the lowest level $\omega = 0$. This shows that the support after the above history is indeed $\{0\}$. Not fishing for the first K periods ensures that the state when they start to fish can be indeed the highest state with positive probability.

Global accessibility of $\Omega^* = \{0\}$ also requires that the probability of the above history is bounded from zero uniformly in μ. Pick an initial prior μ, and let μ^{K+1} be the posterior belief in period $K+1$ given that no one fishes in the past. Due to the natural increase, this posterior μ^{K+1} puts probability at least $\overline{\pi}^K$ to the highest state K, that is, $\mu^{K+1}(K) \geq \overline{\pi}^K$. (Recall that $\overline{\pi}$ is the maximum of π, see Section 2.1.) Let $\pi(k,\tilde{k})$ denote the probability that $y = (1,1)$ is observed and the state tomorrow is $\omega = \tilde{k}$ given the current state $\omega = k$ and the action profile (F,F). Then the probability of the fishermen observing $y = (1,1)$ for the $K-1$ consecutive periods is

$$\mu^{K+1}(K)\pi(K,K-1)\pi(K-1,K-2)\cdots\pi(3,2)\pi(2,0).$$

Since $\pi(2,0) \geq \overline{\pi}$ and $\pi(k,k-1) \geq \overline{\pi}$ for each $k \geq 3$, this probability is at least $\overline{\pi}^{2K-1}$. This bound does not depend on the initial prior μ, and hence the set $\Omega^* = \{0\}$ is indeed globally accessible.[11]

[11]To obtain this uniform bound, it is essential that the fishermen do not fish for the first K

We also claim that any other set $\Omega^* \neq \{0\}$ is either globally accessible or uniformly transient. First of all, Proposition 1 ensures that any superset of $\{\omega\}$ is globally accessible. So pick Ω^* which does not contain $\omega = 0$, and pick an arbitrary belief μ with support Ω^*. Since there is the possibility of natural decrease ($\varepsilon = -1$), regardless of the history, the posterior belief after $K-1$ periods must put positive probability on $\omega = 0$, implying that its support is globally accessible. This shows that Ω^* is uniformly transient (or globally accessible, if it satisfies the condition for global accessibility).

Hence this example satisfies uniform connectedness, and the limit feasible payoff set is independent of the initial prior. Also, the limit minimax payoff is invariant to the initial prior. To see this, note first that player i's minimax payoff is at least 0, as he can obtain at least 0 by not fishing each period. On the other hand, if the opponent fishes every period to punish player i, then regardless of player i's play and regardless of the initial prior μ, the state must reach the lowest state $\omega = 0$ in finite time and stays there forever. Accordingly player i's average payoff is at most 0 in the limit as $\delta \to 1$, regardless of the initial prior μ. Taken together, the limit minimax payoff payoff set is 0 regardless of the initial prior. Then as Proposition 7 shows, the folk theorem holds, and there is an equilibrium in which the fishermen earn the efficient payoff.

In Section 6, we describe how to construct such an (approximately) efficient equilibrium. Here we highlight key features of our equilibrium strategy:

- Players follow the optimal policy which solves (2) almost all periods, until someone deviates.

- If player i unilaterally deviates, go to the punishment phase for player i. Here player i is minimaxed for a while (by playing F each period) and after that, players play the strategy profile which achieves some payoff vector $w(i)$ forever. The payoff $w(i)$ is an "intermediate payoff," which Pareto-

periods. To see this, suppose that the initial prior μ puts probability η on the highest state K. Suppose that the fishermen do not play (N,N) and they simply play (F,F) for the first $K-1$ periods. If they observe $y = (1,1)$ in all periods, then the support of the posterior reaches $\{0\}$. However, the probability of such a signal sequence is

$$\mu(K)\pi(K,K-1)\pi(K-1,K-2)\cdots\pi(3,2)\pi(2,0),$$

which converges to zero as $\mu(K) = \eta$ goes to zero.

dominates the minimax payoff $(0,0)$ but is Pareto-dominated by the efficient payoff characterized by (2).

A couple of remarks are in order. First, the optimal policy to (2) depends on the parameters K, c, and $\pi^\omega(y,\tilde{\omega}|a)$. So if we want to have more detailed descriptions of the optimal policy, we need to make more assumptions on these parameters.

Second, the computation of the minimax payoff above heavily relies on the assumption that $\varepsilon = 0$ when the state is the lowest and someone fishes. If this assumption does not hold, the limit minimax payoff may be greater than 0. Nonetheless, even in such a case, we can still show that the limit minimax payoff is invariant to the initial prior, because the game is *robustly connected*. See Appendix A for more discussions.

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 $V^\omega(\delta) = \mathrm{co}\{v^\omega(\delta,s)|s \in S\}$ where co$B$ denotes the convex hull of the set B. 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 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:

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

Note that this maximization problem indeed has a solution; see Appendix E for the proof. 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 score given the initial prior μ. The

function f can be derived by solving the following Bellman equation:

$$f(\mu) = \max_{a \in A} \left[(1-\delta)\lambda \cdot g^{\mu}(a) + \delta \sum_{y \in Y} \pi_Y^{\mu}(y|a) f(\tilde{\mu}(\mu,a,y)) \right] \tag{4}$$

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 (4) is a generalization of (2), which characterizes the best possible profit in the stochastic Bertrand model; indeed, when $\lambda = (\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}})$, (4) reduces to (2).

In Section 4.1, 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 proposition generalizes this observation; it shows that under uniform connectedness, if δ is sufficiently large, the scores do not depend on the initial prior. So the feasible payoff sets $V^{\mu}(\delta)$ are similar across all initial priors μ when δ is close to one.[12] Note that the same result holds under the full support assumption, because uniform connectedness is weaker than the full support assumption.

Proposition 4. *Suppose that the game is uniformly connected. Then for each $\varepsilon > 0$, there is $\overline{\delta} \in (0,1)$ such that for any $\lambda \in \Lambda$, $\delta \in (\overline{\delta}, 1)$, μ, and $\tilde{\mu}$,*

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

Although it is not stated in the proposition, in the proof we show that the score converges at the rate of $1-\delta$. That is, we can replace ε in Proposition 4 with $O(1-\delta)$.

This proposition extends the invariance result of Dutta (1995) to the hidden-state case. The proof technique is different, because his proof essentially relies on ergodic theorems, which are not applicable to our model due to infinite states. In

[12] We thank Johannes Hörner for pointing out that Proposition 4 strengthens the results on the time-average dynamic programming equation in the POMDP model. It turns out that uniform connectedness is weaker than sufficient conditions found in the literature, including renewability of Ross (1968), reachability-detectability of Platzman (1980), and Assumption 4 of Hsu, Chuang, and Arapostathis (2006). (There is a minor error in Hsu, Chuang, and Arapostathis (2006); see Appendix F for more details.) Indeed, the natural resource management problem in this paper does not satisfy any assumptions above, but it satisfies uniform connectedness. Similarly, Examples B1 and B2 in Appendix B satisfies asymptotic uniform connectedness but not the assumptions in the literature.

the next two subsections, we explain the proof idea of Proposition 4. The formal proof is given in Appendix C.

Now we define the "limit feasible payoff set." The following proposition shows that for each initial prior μ and direction λ, the score converges to a limiting value as δ goes to one. From Proposition 4 above, this limit is independent of μ. The proof can be found in Appendix C.[13]

Proposition 5. *Pick an arbitrary direction λ. If the game is uniformly connected, then the limit $\lim_{\delta\to 1}\max_{v\in V^\mu(\delta)}\lambda\cdot v$ exists for each μ and is independent of μ.*

Let V^μ be the set of all $v\in \boldsymbol{R}^N$ such that $\lambda\cdot v\leq \lim_{\delta\to 1}\max_{v\in V^\mu(\delta)}\lambda\cdot v$ for all λ. Proposition 5 above guarantees that this set V^μ is independent of μ, so we denote it by V. This set V is the limit feasible payoff set, in the sense that $V^\mu(\delta)$ approximates V for all μ when δ is close to one. Indeed, the following proposition shows that the convergence of the score function $\lambda\cdot v^\mu(\delta,s^\mu)$ is uniform in λ and μ. The proof can be found in Appendix C.

Proposition 6. *If the game is uniformly connected, then for each $\varepsilon>0$, there is $\overline{\delta}\in(0,1)$ such that for each $\lambda\in\Lambda$, $\delta\in(\overline{\delta},1)$, and μ,*

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

5.2 Proof Sketch under the Full Support Assumption

In this subsection, we illustrate the proof idea of Proposition 4 under the full support assumption. 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 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 $v_i^\mu(\delta,s^\mu)=\max_{v\in V^\mu(\delta)}v_i$.

As shown in Lemma C1 in Appendix C, the score $v_i^\mu(\delta,s^\mu)$ is convex with respect to μ. (The proof relies on the fact that player i's payoff $v_i^\mu(\delta,s)$ is linear in a belief μ for a given s.) This implies that the score must be maximized by $\mu=0$

[13]Theorem 2 of Rosenberg, Solan, and Vieille (2002) show the existence of the limit of the score, but for completeness, we provide a (simple and new) proof. We thank Johannes Hörner for pointing this out.

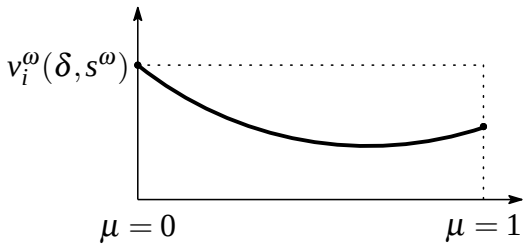

Figure 6: Convex Score Function $v_i^{\mu}(\delta, s^{\mu})$

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

The rest of the proof sketch is divided into two steps. In the first step, we explain that if there is an interior belief $\mu^* \in (0,1)$ whose score approximates the maximal score, then the score function is almost flat so that the score for *every* belief μ approximates the maximal score. The proof relies on simple geometric observation and convexity of the score function. Thanks to this result, in order to establish Proposition 4, we do not need to compute the score for each belief separately; instead, it is sufficient to find an interior belief μ^* whose score approximates the maximal score.

Then in the second step, we explain that there indeed exists such an interior belief μ^*. The full support assumption plays an important role in this step; in particular, we use the fact that the posterior belief is an interior belief after every history.

5.2.1 Step 1: Almost Flat Convex Curve

To simplify the argument, suppose that there is an interior belief $\mu^* \in (0,1)$ whose score not only approximates but exactly equals the maximal score. That is, we assume that $v_i^{\mu^*}(\delta, s^{\mu^*}) = v_i^{\omega}(\delta, s^{\omega})$ for some interior belief $\mu^* \in (0,1)$. Under this assumption, we show that the score function $v_i^{\mu}(\delta, s^{\mu})$ is flat so that the score $v_i^{\mu}(\delta, s^{\mu})$ is equal to the maximal score $v_i^{\omega}(\delta, s^{\omega})$ for all $\mu \in [0,1]$.

We prove by contradiction, so suppose that the score function is not flat and

hence there is a belief $\tilde{\mu} \in [0,1]$ whose score $v_i^{\tilde{\mu}}(\delta, s^{\tilde{\mu}})$ is strictly less than the maximal score $v_i^{\omega}(\delta, s^{\omega})$. Suppose that $\tilde{\mu} < \mu^*$. Then since the score at μ^* is equal to the maximal value, the convex score function must look like Figure 7, and the score for $\mu > \mu^*$ must exceed the maximal score, which is a contradiction. Similarly, if $\tilde{\mu} > \mu^*$, then the score for $\mu < \mu^*$ must exceed the maximal score, as Figure 8 shows. In any case this is a contradiction, and hence the score function must be flat.

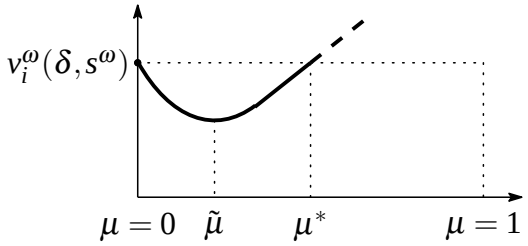

Figure 7: Case with $\tilde{\mu} < \mu^*$

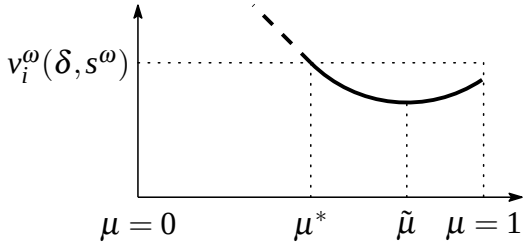

Figure 8: Case with $\tilde{\mu} > \mu^*$

Note that, for the above argument to work, it is important that μ^* is an interior belief. Indeed, if not and we have the boundary belief $\mu^* = 1$, we can draw a non-flat convex score function which attains the maximal score at both $\mu = 0$ and μ^*, as Figure 9 shows. So in order to conclude that the score function is flat, we need to show that there is an *interior belief* μ^* whose score is equal to the maximal score.

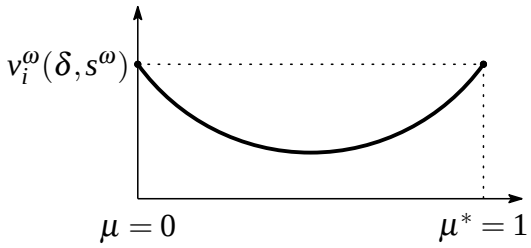

Figure 9: Convex Score Function When $\mu^* = 1$

The result above easily extends to the case in which the score for μ^* approximates (but is not equal to) the maximal score. In that case, we can prove that the score function is almost flat and the score for *every* belief μ approximates the maximal score. (Formally, to obtain such a result, we also need to assume that μ^* be not too close to the boundary points 0 or 1. See Lemma C2 in Appendix

C for more details. Looking ahead, the belief μ^* found in Step 2 satisfies this requirement, because the full support assumption implies $\mu^* \in [\overline{\pi}, 1-\overline{\pi}]$, where $\overline{\pi}$ is the minimum of the π function.)

5.2.2 Step 2: Existence of μ^* Approximating the Maximal Score

Now we show that there indeed exists an interior belief μ^* whose score approximates the maximal score. Recall that the score function is maximized at the belief $\mu = 0$, and s^ω is the strategy profile which achieves this maximal score. Let a^* be the action profile induced by s^ω in period one. Let $\mu(y)$ be the posterior belief in period two when the initial prior is $\mu = 0$ and the outcome in period one is (a^*, y). Pick an arbitrary signal y^* which appears in period one with positive probability, and let $\mu^* = \mu(y^*)$. We show that this μ^* satisfies the desired property; that is, the score for μ^* approximates the maximal score. Note that under the full support assumption, the belief μ^* is indeed an interior belief.

Since the score $v_i^\omega(\delta, s^\omega)$ satisfies (4), we must 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,

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

For simplicity, assume that ω and s^ω do not depend on δ. (Lemma C3 in Appendix C shows that the result easily extends to the case in which they depend on δ.) Then the above equation implies that

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

That is, the expected continuation payoff $E[v_i^{\mu(y)}(\delta, s^{\mu(y)})]$ approximates 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 $v_i^{\mu(y)}(\delta, s^{\mu(y)})$ approximates the maximal score so that

$$v_i^\omega(\delta, s^\omega) - v_i^{\mu(y)}(\delta, s^{\mu(y)}) = O\left(\frac{1-\delta}{\delta}\right). \tag{6}$$

To see this, note that

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

Since $v_i^{\omega}(\delta, s^{\omega})$ is the maximum score, the term in the curly brackets in the right-hand side is non-negative for all y. Thus, if there is y such that the term in the curly brackets is not of order $\frac{1-\delta}{\delta}$, then the right-hand side is not of order $\frac{1-\delta}{\delta}$. However this contradicts (5), and hence (6) holds for all y. This in particular means that (6) holds for $y = y^*$, so that the score for the belief $\mu^* = \mu(y^*)$ approximates the maximal score, as desired.

5.3 Proof Sketch under Uniform Connectedness

Under the full support assumption, the support of the posterior belief is the whole state space Ω after every history. In Step 2 above, this property is used to show that the posterior belief $\mu^* = \mu(y^*)$ in period two is an interior belief. This is important, because if not and the posterior belief μ^* is a boundary belief, the result from Step 1 does not apply, as we have discussed at the end of Step 1.

In general, when we consider uniformly connected stochastic games, the full support assumption may not hold, and hence the support of the posterior may not reach the whole state space Ω. Nonetheless we can show that the same result holds under uniform connectedness. In what follows, we explain how to extend the proof idea in the previous subsection to uniformly connected stochastic games. The argument is a bit complicated, so those who are not interested in technical details may skip this subsection.

The proof consists of three steps: The first two steps are similar to the ones in the previous subsection, while the last step is new. We consider an arbitrary finite state space Ω, so a belief μ is not represented by a single number. With an abuse of notation, for each subset $\Omega^* \subseteq \Omega$, let $\triangle\Omega^*$ denote the set of all beliefs whose support is Ω^* or its subset. That is, $\mu \in \triangle\Omega^*$ puts probability zero on each state $\tilde{\omega} \notin \Omega^*$. Also $\mu \in \triangle\Omega^*$ is called a *relative interior belief* if it assigns positive probability to each state $\tilde{\omega} \in \Omega^*$.

In the first step of the proof, we show that for each subset $\Omega^* \subseteq \Omega$, if there is a relative interior belief $\mu^* \in \triangle\Omega^*$ whose score approximates the maximal score,[14]

[14]Again, we need that this belief μ^* be not too close to the boundary of $\triangle\Omega^*$. See Lemma C2

then the score function is almost flat over the space $\triangle\Omega^*$ so that the score for *every* belief $\mu \in \triangle\Omega^*$ approximates the maximal score. In other words, if there is a belief μ whose score approximates the maximal score, then so does the score for *every* belief with the same support. The proof technique is quite similar to Step 1 in the previous subsection.

In the second step, we show that there is a globally accessible set Ω^* such that the score for *every* belief $\mu \in \triangle\Omega^*$ approximates the maximal score. Thanks to the result of the first step, it is sufficient to show that there is a relative interior belief $\mu^* \in \triangle\Omega^*$ whose score approximates the maximal score. The proof idea is as follows. Let ω be the state which achieves the maximal score. Uniform connectedness ensures that the set $\{\omega\}$ is either globally accessible or uniformly transient. If it is globally accessible, the result immediately follows by setting $\Omega^* = \{\omega\}$. So we focus on the case in which $\{\omega\}$ is uniformly transient. Suppose that the initial state is ω and players play the optimal policy s^ω. In Step 2 in the previous subsection, we have shown that after every on-path history h^1 with length one, the continuation payoff approximates the maximal score. Using a similar technique, we can show that after every on-path history h^t with length t, the continuation payoff approximates the maximal score. Now, the uniform transience of $\{\omega\}$ ensures that there be an on-path history h^t such that the support of the corresponding posterior belief $\mu(h^t)$ is globally accessible. Let $\mu^* = \mu(h^t)$ denote this posterior belief, and let Ω^* denote its support. This Ω^* satisfies the desired condition. Indeed, it is globally accessible, and the score for the relative interior belief μ^*, which is equal to the continuation payoff after the history h^t, approximates the maximal score.

Take Ω^* as in the second step. If this set Ω^* is the whole state space Ω, the result in the second step implies Proposition 4. However, in general, Ω^* may be smaller than Ω. To deal with this problem, in the last step, we show that the score for *every* belief μ approximates the maximal score. The proof idea is as follows. Take an arbitrary initial prior μ, and consider the following strategy profile $\tilde{s}^\mu$: Players try to move the support of the posterior to Ω^* or its subset, and once it happens (say period t), they switch the play immediately and use the optimal policy s^{μ^t} in the rest of the game. Since Ω^* is globally accessible, the support indeed reaches Ω^* in finite time with probability one (if players choose actions

in Appendix C for more details.

appropriately). So for patient players, waiting time until the switch is almost negligible. This implies that the payoff by this strategy profile $\tilde{s}^{\mu}$ is approximated by the expected continuation payoff after the switch to s^{μ^t}, and from the result in the second step, we know that this continuation payoff approximates the maximal score. Hence the payoff by the strategy $\tilde{s}^{\mu}$ approximates the maximal score. This implies the result, because the payoff by the optimal policy s^{μ} must be higher than the one by $\tilde{s}^{\mu}$, and hence even closer to the maximal score.

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

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

As will be shown in Appendix E, the minimizer s_{-i} indeed exists, and it is possibly a mixed strategy. Since players have no private information, the standard argument shows that sequential equilibrium payoffs must be at least the minimax payoff.

For some economic applications, the minimax payoff can be easily computed. For example, the limit minimax payoff is achieved by "forever a^*_{-i}" in the Bertrand model in Section 4.1, and by "forever fish" in the natural resource management problem in Section 4.2. However, there are other examples in which computation of the minimax payoff is complicated, and one may wonder if there is a general condition which guarantees invariance of the limit minimax payoff.

For stochastic games with observable states, irreducibility of Fudenberg and Yamamoto (2011b) ensures that the limit minimax payoff is invariant to the initial state ω. Irreducibility requires that players $-i$ can move the state from any state to any other state *regardless of* player i's play. Formally, $\tilde{\omega}$ is *robustly accessible despite i* if for each ω, there is a (possibly mixed) action sequence $(\alpha^1_{-i}, \cdots, \alpha^{|\Omega|}_{-i})$ such that for any player i's strategy s_i, there is a natural number $T \leq |\Omega|$ such that $\Pr(\omega^{T+1} = \tilde{\omega} | \omega, s_i, \alpha^1_{-i}, \cdots, \alpha^T_{-i}) > 0$. Irreducibility requires each state ω to be robustly accessible despite i for each i.

This paper extends the above invariance result to the hidden-state model. The assumption we make is *robust connectedness*, which requires that players $-i$ can drive the support of the belief from any set Ω^* to any other set $\tilde{\Omega}^*$ regardless

of player i's play, except the case in which $\tilde{\Omega}^*$ is transient in some sense. This assumption is satisfied in a wide range of applications; for example, the game is robustly connected if the full support assumption holds. Robust connectedness is closely related to uniform connectedness, but neither implies the other. See Appendix A for the formal definition of robust connectedness.

As will be shown in Propositions A1 and A2 in Appendix A, if the game is robustly connected, the limit minimax payoff $\underline{v}_i^\mu = \lim_{\delta\to 1} \underline{v}_i^\mu(\delta)$ exists and is invariant to the initial prior μ. We denote it by $\underline{v}_i$, and let V^* denote the limit set of feasible and individually rational payoffs; that is, V^* is the set of all feasible payoffs $v \in V$ such that $v_i \geq \underline{v}_i$ for all i.

5.5 Proof Sketch of Invariance of the Minimax Payoff

The proof of invariance of the minimax payoff is quite different from that of the feasible payoff set. A new complication here is that the minimax payoff $\underline{v}_i^\mu(\delta)$ is not necessarily convex (or concave) with respect to μ, since player i maximizes the value while the opponents minimize it. Hence the technique developed in Step 1 in Section 5.2 does not directly apply to the minimax payoff $\underline{v}_i^\mu(\delta)$.

In what follows, we briefly describe how to prove invariance of the minimax payoffs, under the full support assumption. The argument is a bit complicated, so again those who are not interested in technical details may skip this subsection.

Pick δ close to one, and let s_{-i}^μ denote the minimax strategy for the initial prior μ. Let $v_i^{\tilde{\mu}}(s_{-i}^\mu) = \max_{s_i \in S_i} v_i^{\tilde{\mu}}(s_i, s_{-i}^\mu)$, that is, let $v_i^{\tilde{\mu}}(s_{-i}^\mu)$ denote player i's best payoff against the minimax strategy s_{-i}^μ when the initial prior is $\tilde{\mu}$. When $\tilde{\mu} = \mu$, the payoff $v_i^{\tilde{\mu}}(s_{-i}^\mu)$ is the minimax payoff for the belief μ. On the other hand when $\tilde{\mu} \neq \mu$, this payoff is not the minimax payoff for any belief. Like the score function in Section 5.2, we can show that for a given minimax strategy s_{-i}^μ, player i's payoff $v_i^{\tilde{\mu}}(s_{-i}^\mu)$ is convex with respect to the belief $\tilde{\mu}$. Note that different beliefs μ induce different minimax strategies s_{-i}^μ, and hence different convex curves $v_i^{\tilde{\mu}}(s_{-i}^\mu)$. See Figure 10. In what follows, we work on the set of these convex curves.

For simplicity, assume that there is $(\mu^*, \tilde{\mu}^*)$ which maximizes $v_i^{\tilde{\mu}}(s_{-i}^\mu)$, that is, $v_i^{\tilde{\mu}^*}(s_{-i}^{\mu^*}) \geq v_i^{\tilde{\mu}}(s_{-i}^\mu)$ for all $(\mu, \tilde{\mu})$. The belief $\tilde{\mu}^*$ must put probability one on some state ω, since $v_i^{\tilde{\mu}}(s_{-i}^{\mu^*})$ is convex and hence maximized when $\tilde{\mu}$ is an extreme point

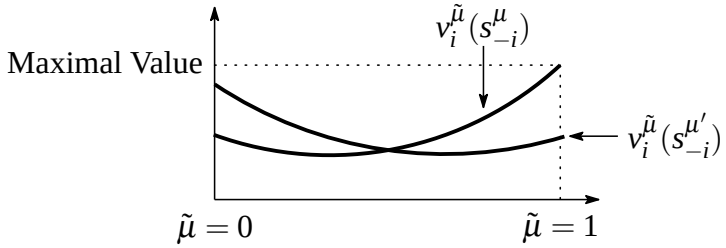

Figure 10: Convex Curves Induced By s_{-i}^{μ} and $s_{-i}^{\mu'}$

of $\triangle\Omega$. So we denote this belief by ω instead of $\tilde{\mu}^*$. In what follows, we call $v_i^{\omega}(s_{-i}^{\mu^*})$ *the maximal value*. Intuitively, the maximal value is the upper bound of the set of the convex curves $v_i^{\tilde{\mu}}(s_{-i}^{\mu})$, as shown in Figure 10.

In the first step of the proof, we show that for a given minimax strategy s_{-i}^{μ}, if the corresponding curve $v_i^{\tilde{\mu}}(s_{-i}^{\mu})$ approximates the maximal score for some interior belief $\tilde{\mu}$, then the curve is almost flat and approximates the maximal score for *all* beliefs $\tilde{\mu}$. The proof technique is quite similar to Step 1 in Section 5.2.

In the second step of the proof, we show that there is a belief μ^{**} whose minimax payoff approximates the maximal value. The proof idea is as follows. Suppose that the initial state is ω and the opponents play $s_{-i}^{\mu^*}$. Suppose that player i takes a best reply so that her payoff is the maximal value. Pick an arbitrary on-path history h^1 with length one, and let $\mu(h^1|\omega)$ be the posterior after h^1. Then as in Step 2 in Section 5.2, we can show that the continuation payoff $v_i^{\mu(h^1|\omega)}(s_{-i}^{\mu^*}|_{h^1})$ after the history h^1 approximates the maximal value. Now, note that the minimax strategy is Markov, and hence the continuation strategy $s_{-i}^{\mu^*}|_{h^1}$ is also the minimax strategy; that is, we have $s_{-i}^{\mu^*}|_{h^1} = s_{-i}^{\mu^{**}}$ for some belief μ^{**}. So the above result implies that the payoff $v_i^{\mu(h^1|\omega)}(s_{-i}^{\mu^{**}})$ approximates the maximal value. Then since $\mu(h^1|\omega)$ is an interior belief under the full support assumption, the result in the first step ensures that the convex curve $v_i^{\tilde{\mu}}(s_{-i}^{\mu^{**}})$ is almost flat and approximates the maximal score for all beliefs $\tilde{\mu}$. In particular, when $\tilde{\mu} = \mu^{**}$, this implies that the minimax payoff $v_i^{\mu^{**}}(s_{-i}^{\mu^{**}})$ approximates the maximal value, as desired.

In the third step, we show that the minimax payoff for any interior belief μ approximates the maximal score. Suppose not so that there is an interior belief μ whose minimax is much lower than the maximal value. Pick such a μ. Then from the result in the first step, player i's payoff $v_i^{\tilde{\mu}}(s_{-i}^{\mu})$ must be much lower than

the maximal value for *all* interior beliefs $\tilde{\mu}$; otherwise the convex curve $v_i^{\tilde{\mu}}(s_{-i}^{\mu})$ approximates the maximal value for all beliefs $\tilde{\mu}$ and so does the minimax payoff $v_i^{\mu}(s_{-i}^{\mu})$. Now, pick μ^{**} as in the second step, and suppose that it is the initial prior. Suppose that the opponents play the following strategy $\tilde{s}_{-i}$: Play some action a_{-i} in period one, and then in the rest of the game play the minimax strategy s_{-i}^{μ} for the belief μ above. Suppose that player i takes a best reply. Then by the definition of s_{-i}^{μ}, player i's continuation payoff from period two on is much lower than the maximal value, and so is her overall payoff. However this is a contradiction, because this payoff must be at least the minimax payoff for μ^{**}, which approximates the maximal value by the definition of μ^{**}.

In the last step, we show that the minimax payoff for any boundary belief μ approximates the maximal score. Suppose that the initial prior is a boundary belief μ, and that the opponents play the minimax strategy s_{-i}^{μ}. Note that the posterior belief in period two is an interior belief, since we impose the full support assumption. Thus the result in the third step ensures that player i's continuation payoff from period two on, which is the minimax payoff for the posterior belief, approximates the maximal value. This in turn implies that player i's overall payoff approximates the maximal value, and hence the result.

6 Folk Theorem

In this section, we show that the folk theorem holds if the limit of the feasible and individually rational payoff set is invariant to the initial prior. When there is a Markov equilibrium, the efficient payoff can be easily sustained by a trigger strategy, just as explained in Section 4.1. However, since the state is infinite in our model, a Markov equilibrium may not exist in general, and a trigger strategy may not work. Here we illustrate how to design an effective punishment mechanism for such a case. Throughout this section, we assume that public randomization z, which follows the uniform distribution $U[0,1]$, is available.

6.1 Punishment over Random Blocks

In this section, we consider an equilibrium in which a deviator will be punished by the minimax strategy. Since the minimax strategy does not constitute an equi-

librium, we cannot ask players to play the minimax strategy forever; players must stop playing the minimax strategy at some point and after that we need to reward those who actually played the minimax strategy.

Now the question is when players should stop the minimax play. As discussed in the introduction, playing the minimax strategy for T periods is not an effective punishment. Indeed, when we take T as given and then take $\delta \to 1$, due to the lack of the belief ergodicity, the average payoff over the T-period block can be quite different from the infinite-horizon game payoff and hence quite different from the minimax payoff. This is in a sharp contrast with Dutta (1995) and Hörner, Sugaya, Takahashi, and Vieille (2011), who consider the case with finite states. Since they assume finite states, the minimax strategy induces an ergodic state evolution, and thus the average payoff over the T-period block approximates the infinite-horizon game payoff in the limit as $\delta \to 1$.

To solve this problem, this paper considers an equilibrium with *random blocks*. Unlike the T-period block, the length of the random block is not fixed and is determined by public randomization $z \in [0,1]$. Specifically, at the end of each period t, players determine whether to continue the current block or not in the following way: Given some parameter $p \in (0,1)$, if $z^t \leq p$, the current block continues so that period $t+1$ is still included in the current random block. On the other hand, if $z^t > p$, then the current block terminates. So the random block terminates with probability $1-p$ each period.

A key is that the random block is payoff-equivalent to the infinite-horizon game with the discount factor $p\delta$, due to the random termination probability $1-p$. Hence given the current belief μ, player i's average payoff during the block cannot exceed the minimax payoff $\underline{v}_i^{\mu}(p\delta)$ if the opponents use the minimax strategy for the initial prior μ and the discount factor $p\delta$ (not δ) during the block. In particular, this payoff approximates the limit minimax payoff $\underline{v}_i$ when both p and δ are close to one. (Note that taking p close to one implies that the expected duration of the block is long.) In this sense, the opponents can indeed punish player i by playing the minimax strategy over the random block.

In the proof of the folk theorem, we pick p close to one, and then take $\delta \to 1$. This implies that although the random block is long in expectation, players puts a higher weight on the continuation payoff after the block than the payoff during the current block; so a small variation in continuation payoffs is enough to provide

appropriate incentives during the random block. This ensures that a small amount of reward after the block is enough to provide incentives to play the minimax strategy.

The idea of random blocks is useful in other parts of the proof of the folk theorem, too. For example, it ensures that the payoff on the equilibrium path does not change much after every history. See the proof in Section 6.3 for more details.

Independently of this paper, Hörner, Takahashi, and Vieille (2015) also propose the idea of random blocks, which they call "random switching." However, their model and motivation are quite different from ours. They study repeated adverse-selection games in which players report their private information every period. In their model, a player's incentive to disclose her information depends on the impact of her report on her flow payoffs until the effect of the initial state vanishes. Measuring this impact is difficult in general, but it becomes tractable when the equilibrium strategy has the random switching property. That is, they use random blocks in order to measure payoffs by misreporting. In contrast, in this paper, the random blocks ensure that playing the minimax strategy over the block indeed approximate the minimax payoff. Another difference between the two papers is the order of limits. They take the limits of p and δ simultaneously, while we fix p first and then take δ large enough. Lastly, it may be worth emphasizing that our proof of the folk theorem is constructive, while their proof is not. So our proof directly illustrates how the random blocks are used in equilibrium.

6.2 Folk Theorem for Stochastic Games with Hidden States

Now we establish the folk theorem, assuming that the feasible and individually rational payoff set is invariant to the initial prior in the limit as $\delta \to 1$. This invariance assumption ensures that the feasible and individually rational payoff set in any continuation game (possibly with a posterior belief quite different from the initial belief) is exactly the same as the one for the initial game. Hence, while a player's deviation may influence the distribution of the posterior belief tomorrow, it does not change future feasible payoffs, which bounds a player's incentive to deviate.

As shown by Proposition 4 and discussed in Section 5.4, the invariance assumption is satisfied as long as uniform connectedness and robust connectedness

hold. So the following proposition ensures that the folk theorem holds under these assumptions (or simply the full support assumption). It may be worth emphasizing that this proposition encompasses the folk theorem by Dutta (1995) as a special case in which the state is observable each period.

Proposition 7. *Suppose that the feasible and individually rational payoff set is invariant to the initial prior in the limit as $\delta \to 1$, and that the limit payoff set V^* is full dimensional. Assume also that public randomization is available. Then for any interior point $v \in V^*$, there is $\underline{\delta} \in (0,1)$ such that for any $\delta \in (\underline{\delta}, 1)$ and for any initial prior μ, there is a sequential equilibrium with the payoff v.*

In addition to the invariance assumption, the proposition requires the full dimensional assumption. This assumption allows us to construct a player-specific punishment mechanism; that is, it ensures that we can punish player i (decrease player i's payoff) while not doing so to all other players. Note that this assumption is common in the literature, for example, Fudenberg and Maskin (1986) prove the folk theorem for repeated games with observable actions and the full dimensional assumption.

Fudenberg and Maskin (1986) also show that the full dimensional assumption is dispensable if there are only two players and the minimax strategies are pure actions. The reason is that player-specific punishments are not necessary in such a case; they consider an equilibrium in which players mutually minimax each other over T periods after any deviation, and find that it effectively disciplines players' incentives. Unfortunately, this idea does not directly extend to our setup, since a player's incentive to deviate from the mutual minimax play can be quite large in stochastic games; this is so especially because the payoff by the mutual minimax play is not necessarily invariant to the initial prior, unlike the minimax payoff. To avoid this problem, we consider player-specific punishments even for the two-player case, which requires the full dimensional assumption.

The proof of the proposition is constructive, and combines the idea of random blocks with the player-specific punishments of Fudenberg and Maskin (1986). In the next subsection, we prove this proposition assuming that the minimax strategies are pure strategies. Then we briefly discuss how to extend the proof to the case with mixed minimax strategies. The formal proof for games with mixed minimax strategies will be given in Appendix C.

6.3 Equilibrium with Pure Minimax Strategies

Take an interior point $v \in V^*$. We construct a sequential equilibrium with the payoff v when δ is close to one. For now, assume that the minimax strategies are pure strategies. Under this assumption, the equilibrium has a quite simple form, as will be presented below. Note that the natural resource management problem in Section 4.2 satisfies this assumption. To simplify the notation, we also assume that there are only two players, but this is not essential; the proof easily extends to the case with more than two players.

Pick payoff vectors $w(1)$ and $w(2)$ from the interior of the limit payoff set V^* such that the following two conditions hold:

(i) $w(i)$ is Pareto-dominated by the target payoff v, i.e., $w_i(i) < v_i$ for each i.

(ii) Each player i prefers $w(j)$ over $w(i)$, i.e., $w_i(i) < w_i(j)$ for each i and $j \neq i$.

The full dimensional condition ensures that such $w(1)$ and $w(2)$ exist. See Figure 11 to see how to choose these payoffs $w(i)$. In this figure, the payoffs are normalized so that the limit minimax payoff vector is $\underline{v} = (\underline{v}_1, \underline{v}_2) = (0,0)$.

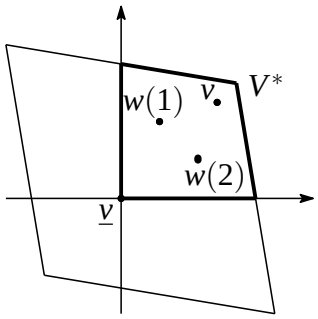

Figure 11: Payoffs $w(1)$ and $w(2)$

Looking ahead, the payoffs $w(1)$ and $w(2)$ are regarded as "long-run payoffs in the punishment phase." That is, when player i has deviated and players start to punish her, their payoff in the continuation game will be approximately $w(i)$ in our equilibrium. (We use player-specific punishments, so the payoff depends on the identity of the deviator.) Property (i) above implies that each player i prefers the cooperative payoff v over the punishment payoff, so no one wants to stop

cooperation. Property (ii) ensures that the punishment is a credible threat. To see this, suppose that player $j \neq i$ has deviated and players are about to punish her. Then player i is indeed willing to punish player j, because if player i does not, then now player i will be punished and the long-run payoff changes from $w(j)$ to $w(i)$, which lowers player i's payoff.

When $p \in (0,1)$ is close to one, the payoff vectors v, $w(1)$, and $w(2)$ are all included in the interior of the feasible payoff set $V^{\mu}(p)$ with the discount factor p for each belief μ. (Here we use Proposition 6, which ensures that the score converges to the limit uniformly in λ.) Also, when $p \in (0,1)$ is close to one, we have $\sup_{\mu \in \triangle\Omega} \underline{v}_i^{\mu}(p) < w_i(i)$ for each i, because the minimax payoff $\underline{v}_i^{\mu}(p)$ converges to the limit minimax payoff for $p \to 1$. Pick p so that these conditions are satisfied. By continuity, if the discount factor δ is close to one, then the payoff vectors v, $w(1)$, and $w(2)$ are all included in the interior of the feasible payoff set $V^{\mu}(p\delta)$ with the discount factor $p\delta$.

In our equilibrium, players are in the "regular (cooperative) phase" or in the "punishment phase for player 1" or in the "punishment phase for player 2." In the regular phase, the infinite horizon is regarded as a series of random blocks, and in each random block, players play a pure strategy profile which exactly achieves the target payoff v as the average payoff during the block. To be precise, suppose that we are currently in the initial period of a random block, and let μ be the current belief. If there is a pure strategy profile s which achieves the payoff v given the discount factor $p\delta$ and the belief μ, (that is, $v^{\mu}(p\delta, s) = v$), then players use this strategy until the block ends. If there is no such a pure strategy profile, players use public randomization to generate v. That is, players choose one of the extreme points of $V^{\mu}(p\delta)$ via public randomization at the beginning of the block, and then play the corresponding pure strategy until the block ends.

When players follow the strategy profile above, the average payoff in the entire game is also equal to v, because it is a weighted average of payoffs over all blocks. Of course, the average payoff in the current block can be quite different from v once the public randomization (which determines one of the extreme points) realizes. However, if we take δ close to one, players do not care much about the payoffs in the current block, and what matters is the payoffs in the continuation game from the second block. Hence the average payoff in the entire game is still close to v, even after public randomization realizes. This property is due to the

random block structure, and will play an important role when we check incentive conditions.

Players stay at the regular phase as long as no one deviates. However, once someone (say, player i) deviates unilaterally, then they will switch to the punishment phase for player i immediately. In the punishment phase for player i, the infinite horizon is regarded as a sequence of random blocks, like in the regular phase. In the first K blocks, the opponent (player $j \neq i$) minimaxes player i, and player i chooses a best reply. Specifically, in each block, letting μ be the belief at the beginning of the block, the opponent plays the minimax strategy for the belief μ and the discount factor $p\delta$ until the current block ends. On the other hand, player i maximizes her payoff during these K consecutive blocks. (Note that player i's play here can be possibly different from maximizing the block payoff in each block separately, since her action today influences the posterior belief in the next block, which in turn influences the payoff in the next block. Note that player i's payoff in each block cannot exceed $\sup_{\mu\in\triangle} \underline{v}_i^{\mu}(p\delta)$, since the opponent minimaxes player i.) After the K blocks, players change their play to achieve the post-minimax payoff $w(i)$; that is, in each random block, players play a pure strategy profile s which exactly achieves $w(i)$ as the average payoff in the block (i.e., $v^{\mu}(p\delta, s) = w(i)$ where μ is the current belief). If such s does not exist, players use public randomization to generate $w(i)$, just as in the regular phase. The parameter K will be specified later.

If no one deviates from the above play, players stay at the punishment phase for player i forever. Likewise, even if player i deviates in the first K random blocks, it is ignored and players continue the play. If player i deviates after the first K blocks (i.e., if she deviates from the post-minimax play) then players restart the punishment phase for player i immediately; from the next period, the opponent starts to minimax player i. If the opponent (player $j \neq i$) deviates in any period of the punishment phase for player i, players switch to the punishment phase for player j, in order to punish player j. See Figure 12.

Now, choose K such that

$$-\overline{g} - \frac{1}{1-p}\overline{g} + \frac{K-1}{1-p} w_i(i) > \overline{g} + \frac{K}{1-p} \sup_{\mu\in\triangle\Omega} \underline{v}_i^{\mu}(p) \tag{7}$$

for each i. Note that (7) indeed holds for sufficiently large K, as $\sup_{\mu\in\triangle\Omega} \underline{v}_i^{\mu}(p) < w_i(i)$. To interpret (7), suppose that we are now in the punishment phase for player

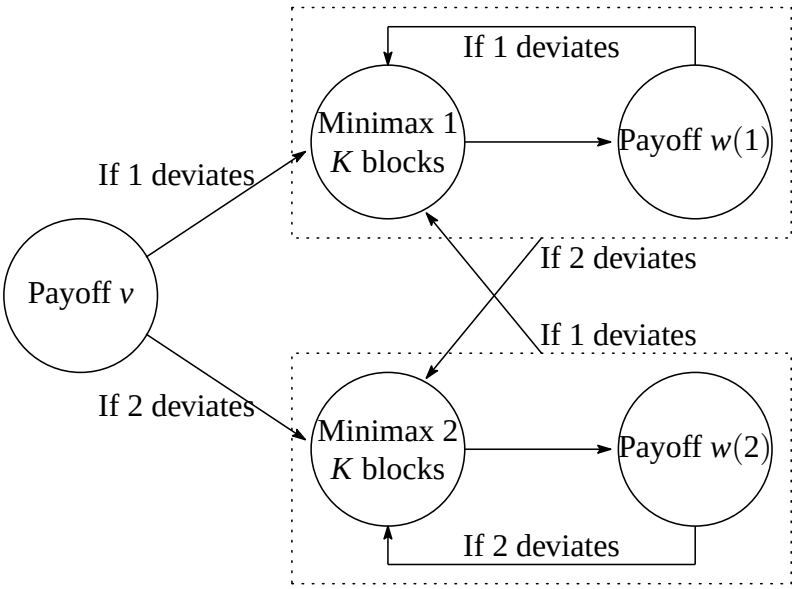

Figure 12: Equilibrium strategy

i, in particular a period in which players play the strategy profile with payoff $w(i)$. (7) ensures that player i's deviation today is not profitable for δ close to one. To see why, suppose that player i deviates today. Then her stage-game payoff today is at most $\overline{g}$, and then she will be minimaxed for the next K random blocks. Since the expected length of each block is $\frac{1}{1-p}$, the (unnormalized) expected payoff during the minimax phase is at most $\frac{K}{1-p}\sup_{\mu\in\triangle\Omega}\underline{v}_i^{\mu}(p)$ when $\delta\to 1$. So the right-hand side of (7) is an upper bound on player i's unnormalized payoff until the minimax play ends, when she deviates. The payoffs after the minimax play do not show up in (7), because the corresponding continuation payoff is $w_i(i)$ regardless of whether or not player i deviates today.

On the other hand, the left-hand side of (7) is a lower bound of player i's payoff when she does not deviate. The first term $-\overline{g}$ is the lower bound of her stage-game payoff today. The next term $-\frac{1}{1-p}\overline{g}$ is the lower bound of her expected unnormalized payoff from tomorrow until the end of the block which involves tomorrow. (Since tomorrow is not necessarily the first period of the block, the average payoff can be different from $w_i(i)$ and thus we use the lower bound $-\overline{g}$ here. Note also that whether tomorrow is the first period of the block or not does not influence the expected length until the end of the block.) The last term

$\frac{K-1}{1-p} w_i(i)$ is her expected unnormalized payoff for the next $K-1$ blocks.

We claim that the strategy profile above is indeed an equilibrium for sufficiently large δ. First of all, consider player i's incentive in the regular phase, in which players play the pure strategy profile with the payoff v each block. If she does not deviate, her average payoff is approximately v_i, because she does not care much about the payoff during the current block when δ is close to one. On the other hand, if she deviates, her average payoff is approximately $w_i(i)$. Since $v_i > w_i(i)$, any deviation today is not profitable.

Next, consider player i's incentive in player j's punishment phase. If she does not deviate, her average payoff is approximately $w_i(j)$. On the other hand, if she deviates, her average payoff is approximately $w_i(i)$. Since $w_i(j) > w_i(i)$, any deviation today is not profitable.

Now, consider player i's incentive in player i's punishment phase. During the minimax play, player i's play is optimal and hence any deviation is not profitable. Also, when players play the pure strategy with the payoff $w(i)$, (7) ensures that the gain by deviating is less than the future loss. Hence any deviation is not profitable. Overall, each player i has no profitable deviation after every history, and hence the above strategy profile is an equilibrium.

As noted, it is easy to extend the above equilibrium strategy to the case with more than two players. Indeed, the same construction works if we choose the payoff vectors $\{w(i)\}_{i \in I}$ such that $w_i(i) < v_i$ for each i and $w_i(i) < w_i(j)$ for each i and $j \neq i$.

When the minimax strategies are mixed strategies, we need to modify the above equilibrium construction and make player i indifferent over all actions when she minimaxes player $j \neq i$. To do so, we follow the idea of Fudenberg and Maskin (1986); we perturb $w_i(j)$, player i's continuation payoff after the minimax play, to make her indifferent. The amount of the perturbation depends on player i's actions during the minimax play. Roughly, we increase the continuation payoff $w_i(j)$ if player i chose actions yielding low payoffs during the minimax play, and we decrease it if she chose actions yielding high payoffs. This perturbation makes player i indifferent over all actions during the minimax play. Also, the amount of the perturbation can be arbitrarily small when δ approaches one, so all other incentive constraints still hold. See Appendix C for the formal proof.

Remark 1. When the game is not (asymptotically) uniformly connected, the limit

payoff set V^{μ} may depend on μ. However, even in this case, the following result holds. Let $\underline{v}_i^{\mu} = \limsup_{\delta \to 1} \underline{v}_i^{\mu}(\delta)$, and $V^{*\mu}$ be the set of $v \in V^{\mu}$ such that $v_i \geq \underline{v}_i^{\mu}$ for all i, and let V^* be the intersection of $V^{*\mu}$ over all μ. Then for any interior point v of V^* and any initial prior μ, there is a sequential equilibrium with the payoff v when δ is large enough. The proof is exactly the same as Proposition 7.

Remark 2. Note that the proof above does not rely on the fact that the belief is the state. So Proposition 7 is valid for general stochastic games with infinite states, as long as the invariance condition holds and the limit payoff set has the interior.

7 Concluding Remarks

This paper considers a new class of stochastic games in which the state is hidden information. We find that, very generally, the feasible and individually rational payoff set is invariant to the initial belief in the limit as the discount factor goes to one. Then we introduce the idea of random blocks and prove the folk theorem.

Proposition 4 shows that uniform connectedness ensures invariance of the limit feasible payoff set. In Appendix B, we show that uniform connectedness can be relaxed further; the invariance result in Proposition 4 holds even if uniform connectedness is replaced with a weaker condition, called *asymptotic uniform connectedness*. Asymptotic uniform connectedness is deeply related to state learning, and is satisfied in quite general environments. For example, asymptotic uniform connectedness holds if each state ω is globally accessible or uniformly transient, and if different hidden states induce different signal distributions (more formally, for each fixed action profile a, the signal distributions $\{(\pi_Y^{\omega}(y|a))_{y \in Y} | \omega \in \Omega\}$ are linearly independent). Note that the latter condition is satisfied for generic signal structures as long as the signal space is large enough so that $|Y| \geq |\Omega|$. So the payoff invariance result "almost always" holds if the evolution of the hidden state ω satisfies the communicating state assumption. See Proposition B1 in Appendix B for more details.

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 im-

plies 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: Minimax Payoffs and Robust Connectedness

In Section 5.4, we have briefly explained the idea of robust connectedness, which ensures invariance of the minimax payoffs. Here we present the formal definition of robust connectedness and the related results.

We first introduce the notion of robust accessibility.

Definition A1. A non-empty subset $\Omega^* \subseteq \Omega$ is *robustly accessible despite player i* if there is $\pi^* > 0$ such that for any initial prior μ, there is an action sequence $(\alpha_{-i}^1, \cdots, \alpha_{-i}^{4^{|\Omega|}})$ such that for any strategy s_i, there is a natural number $T \leq 4^{|\Omega|}$ and a belief $\tilde{\mu}$ with support Ω^* such that[15]

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

Robust accessibility of Ω^* requires that the opponents can drive the support of the belief to Ω^* regardless of player i's play. Robust accessibility is different from global accessibility in two respects. First, we consider a mixed action sequence $(\alpha_{-i}^1, \cdots, \alpha_{-i}^{4^{|\Omega|}})$, rather than a pure action sequence.[16] Second, the support of the resulting belief $\tilde{\mu}$ must be precisely equal to Ω^*; on the other hand, global accessibility requires only that the support of the posterior belief $\tilde{\mu}$ be a subset of Ω^*. Note that if Ω^* is robustly accessible despite i for some i, it is globally accessible.

Next, we define transience.

Definition A2. A subset $\Omega^* \subseteq \Omega$ is *transient given player i* if it is not robustly accessible despite i and there is $\pi^* > 0$ such that for any μ whose support is Ω^*, there is player i's action sequence $(\alpha_i^1, \cdots, \alpha_i^{4^{|\Omega|}})$ such that for any strategy s_{-i} of the opponents, there is a natural number $T \leq 4^{|\Omega|}$ and a belief $\tilde{\mu}$ whose support is robustly accessible despite i such that

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

[15] Like global accessibility, restricting attention to $T \leq 4^{|\Omega|}$ is without loss of generality. To see this, note that there is an equivalent definition of robust accessibility, as discussed in the proof of Lemma C7. Suppose that for some strategy s_i, there is no $T \leq 4^{|\Omega|}$ such that the condition stated there is not satisfied; then we can find a strategy $\tilde{s}_i$ such that the condition stated there is not satisfied for every natural number T.

[16] Replacing the action sequence $(\alpha_{-i}^1, \cdots, \alpha_{-i}^{4^{|\Omega|}})$ in the definition with a strategy s_{-i} does not relax the condition at all.

Transience is different from uniform transience in many aspects. First, the support of the posterior belief must eventually reach a robustly accessible set, rather than a globally accessible set. Second, transience requires only that the support reaches a robustly accessible set when player i plays a particular action sequence $(\alpha_i^1, \cdots, \alpha_i^{4^{|\Omega|}})$. On the other hand, uniform transience requires that the support must reach a globally accessible set regardless of players' play. Due to this property, transience of Ω^* does not necessarily imply uniform transience of Ω^*, and accordingly robust connectedness does not necessarily imply uniform connectedness. Third, transience requires that the probability that the support of the belief reaches a robust accessible set is at least π^* regardless of the initial prior with support Ω^*; accordingly, we have the restriction $T \leq 4^{|\Omega|}$ as in the definition of global accessibility.

We now introduce one more condition, called the support merging condition.

Definition A3. The game satisfies the *support merging condition* if for each state ω and for each pure strategy profile s, there is a natural number $T \leq 4^{|\Omega|}$ and a history h^T such that $\Pr(h^T|\omega, s) > 0$ and such that the support of the posterior belief induced by the initial state ω and the history h^T is the same as the one induced by the initial prior $\mu = (\frac{1}{|\Omega|}, \cdots, \frac{1}{|\Omega|})$ and the history h^T.

The support merging condition ensures that two posterior beliefs induced by different initial priors ω and $\mu = (\frac{1}{|\Omega|}, \cdots, \frac{1}{|\Omega|})$ must eventually have the same support with positive probability, regardless of the play. Note that this condition is trivially satisfied in many examples; for example, under the full support assumption, the support of the posterior belief is Ω regardless of the initial belief, and hence the support merging condition holds.

Now we are ready to state the definition of robust connectedness. It requires each set Ω^* to be robustly accessible or transient, in addition to the support merging condition.

Definition A4. The game is *robustly connected* if the support merging condition holds and if for each i, each non-empty subset $\Omega^* \subseteq \Omega$ is robustly accessible despite i or transient given i.

As discussed in Section 5.4, when the game is robustly connected and δ is sufficiently large, the minimax payoffs are similar across all initial priors μ. Formally, we have the following proposition. The proof is given in Appendix C.

Proposition A1. *Suppose that the game is robustly connected. Then for each $\varepsilon > 0$, there is $\overline{\delta} \in (0,1)$ such that $|\underline{v}_i^{\mu}(\delta) - \underline{v}_i^{\tilde{\mu}}(\delta)| < \varepsilon$ for any $\delta \in (\overline{\delta}, 1)$, μ, and $\tilde{\mu}$.*

The next proposition shows that the limit minimax payoff exists. See Appendix C for the proof.

Proposition A2. *If the game is robustly connected, then the limit $\lim_{\delta \to 1} \underline{v}_i^{\mu}(\delta)$ exists for any μ, and is independent of μ.*

Like uniform connectedness, robust connectedness is weaker than the full support assumption. Indeed, if the full support assumption holds, the whole state space Ω is robustly accessible despite i and any other set Ω^* is transient. Also, as argued, the support merging condition holds. Hence the full support assumption implies robust connectedness.

The following proposition shows that the result similar to Proposition 2 holds for stochastic games with observable states. It also shows that the result similar to Proposition 3 holds. The proof is similar to those of Propositions 2 and 3 and hence omitted. A state ω is *transient given player i* if there is player i's action sequence $(\alpha_i^1, \cdots, \alpha_i^{|\Omega|})$ such that if the initial state is ω, with positive probability, the state reaches a state which is robustly accessible despite i within $|\Omega|$ periods, regardless of the opponents' strategy s_{-i}.

Proposition A3. *The game is robustly connected only if for each i, each state ω is robustly accessible despite i or transient given i. In particular, for stochastic games with observable states, the game is robustly connected if and only if for each i, each state ω is robustly accessible despite i or transient given i.*

Unfortunately, the second result in Proposition 2 does not extend, that is, for stochastic games with delayed observations, assuming each state to be robustly accessible or transient is not sufficient for robust connectedness. The main problem is that robust accessibility is much more demanding than global accessibility, in the sense that it requires that the support of the posterior must move to Ω^* regardless of player i's play. Such a condition is never satisfied if player i's action has a significant impact on the support tomorrow.

For example, suppose that there are two players, and there are three states, ω_A, ω_B, and ω_C. Each player has three actions, A, B, and C. Assume that the state is

observed with delay, so $Y = \Omega$ and the signal today is equal to the current state with probability one. Suppose that the state tomorrow is determined by the action profile today, and a player's action increases the probability of the corresponding state by $\frac{1}{2}$. For example, if both players choose A, then the state tomorrow is ω_A for sure. If one player chooses A and the opponent chooses B, then ω_A and ω_B are equally likely. So regardless of the opponent's play, if a player chooses A, then ω_A will appear with probability at least $\frac{1}{2}$. This shows that each state is robustly accessible despite i for each i. Unfortunately, robust connectedness is not satisfied in this example. Indeed, any set Ω^* is neither robustly accessible nor transient. For example, any set Ω^* which does not include some state ω is not robustly accessible despite 1, because if player 1 always chooses the action corresponding to ω each period, the posterior must put probability at least $\frac{1}{2}$ on ω. Also the whole set Ω is not robustly accessible, because in any period, the posterior puts probability zero on some state ω. Since there is no robustly accessible set, any set cannot be transient either.

Note, however, that robust connectedness is just a sufficient condition for invariance of the limit minimax payoff. The following proposition shows, for stochastic games with delayed observations, the limit minimax payoff is invariant to the initial prior in general. The proof relies the fact that there are only finitely many possible posterior beliefs for games with observation delays; see Appendix C.

Proposition A4. *Consider stochastic games with delayed observations, and suppose that for each i, each state ω is robustly accessible despite i or transient given i. Then for each $\varepsilon > 0$, there is $\overline{\delta} \in (0,1)$ such that $|\underline{v}_i^{\mu}(\delta) - \underline{v}_i^{\tilde{\mu}}(\delta)| < \varepsilon$ for any $\delta \in (\overline{\delta}, 1)$, μ, and $\tilde{\mu}$.*

This proposition suggests that robust connectedness is much stronger than necessary for invariance of the limit minimax payoff. Yet, robust connectedness is a useful concept, because it is satisfied in many economic examples. Consider the natural resource management problem in Section 4.2, and assume that there is natural increase even if the state is the lowest and someone fishes; that is, $\Pr(\varepsilon = 1|\omega, a, y) > 0$ when $\omega - y_1 - y_2 = 0$ and $a \neq (N, N)$. Now the limit minimax payoff is not equal to zero, and the minimax strategy can take a complicated form. Nonetheless we can show that the limit minimax payoff is invariant to the

initial prior, because the game is robustly connected. In what follows, we show that robust connectedness is indeed satisfied.

We first show that $\Omega^* = \Omega$ is robustly accessible for each player i. To see this, pick an arbitrary initial prior μ, and pick an arbitrary strategy profile s. Suppose that $y = (0,0)$ is observed for the first K periods. (This history happens with probability at least $\overline{\pi}^K$, regardless of (μ, s).) Then due to the possibility of natural increase and decrease, the support of the posterior must be the whole state space, Ω. This shows that $\Omega^* = \Omega$ is robustly accessible despite i for each i.

Also, any other set $\Omega^* \neq \Omega$ is either robustly accessible despite i or transient given i. To see this, pick an arbitrary initial prior μ, and pick an arbitrary strategy profile s. Suppose that $y = (0,0)$ is observed for the first K periods. Then as in the case above, the support of the posterior moves to the whole state space Ω, which is robustly accessible. Hence all these sets $\Omega^* \neq \Omega$ are transient (or robustly accessible, if they satisfy the relevant condition.)

Finally, the support merging condition holds for the same reason: Regardless of the initial prior and the strategy profile, if $y = (0,0)$ is observed for the first K periods, the support will be Ω. Hence the game is robustly connected.

Appendix B: Relaxing Uniform Connectedness

Proposition 4 shows that uniform connectedness ensures invariance of the feaseible payoff set. Here we show that it is indeed possible to replace uniform connectedness with a weaker condition, called *asymptotic uniform connectedness*.

Before we describe the idea of asymptotic uniform connectedness, it is useful to understand when uniform connectedness is not satisfied and why we want to relax it. We present two examples in which uniform connectedness is not satisfied; in these examples, the communicating state assumption of Dutta (1995) is satisfied but nonetheless uniform connectedness does not hold. These examples show that assuming global accessibility of each ω does not necessarily imply uniform connectedness.

Example B1. 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 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 uniformly transient.

In the next example, the state evolution is not deterministic.

Example B2. 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. Uniform connectedness is not satisfied in this example, since $\{\omega_2\}$ is neither globally accessible nor uniformly 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 uniformly 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 $\{\omega_2\}$.

Now we illustrate the idea of asymptotic uniform connectedness, Consider Example B1, and 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., Proposition 4 should remain valid in this example, even though the game is not uniformly 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 invariance of the feasible payoff set. Formally, asymptotic accessibility is defined as follows:

Definition B1. 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 $\Pr(\mu^{T^*+1} = \tilde{\mu} | \mu, a^1, \cdots, a^{T^*}) \geq \pi^*$ 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 B1. 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 B2 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 $\{\omega_2\}$.

Note that Ω^* is asymptotically accessible whenever it is globally accessible. Hence the whole state space $\Omega^* = \Omega$ is always asymptotically accessible. Next,

we give the definition of asymptotic uniform transience.

Definition B2. A singleton set $\{\omega\}$ is *asymptotically uniformly transient* if it is not asymptotically accessible and there is $\tilde{\pi}^* > 0$ such that for any $\varepsilon > 0$, there is a natural number T such that for each pure strategy profile s, there is an asymptotically accessible set Ω^*, a natural number $T^* \leq T$, and a belief $\tilde{\mu}$ such that $\Pr(\mu^{T^*+1} = \tilde{\mu} | \omega, s) > 0$, $\sum_{\tilde{\omega} \in \Omega^*} \tilde{\mu}(\tilde{\omega}) \geq 1 - \varepsilon$, and $\tilde{\mu}(\tilde{\omega}) \geq \tilde{\pi}^*$ for all $\tilde{\omega} \in \Omega^*$.

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

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

Definition B3. A stochastic game is *asymptotically uniformly connected* if each singleton set $\{\omega\}$ is asymptotically accessible or asymptotically uniformly transient.

Asymptotic uniform connectedness is weaker than uniform connectedness. Indeed, Examples B1 and B2 satisfy asymptotic uniform connectedness but do not satisfy uniform connectedness.

Unfortunately, checking asymptotic uniform connectedness in a given example is often a daunting task, because we need to compute how the posterior belief looks like in a distant future. However, the following proposition provides a simple sufficient condition for asymptotic uniform connectedness:

Proposition B1. *The game is asymptotically uniformly connected if each state ω is globally accessible or uniformly transient, and for each action profile a and each proper subset $\Omega^* \subset \Omega$,*

$$co\{\pi_Y^\omega(a) | \omega \in \Omega^*\} \cap co\{\pi_Y^\omega(a) | \omega \notin \Omega^*\} = \emptyset.$$

[17] Asymptotic uniform transience requires $\tilde{\mu}(\tilde{\omega}) \geq \tilde{\pi}^*$, that is, the posterior belief $\tilde{\mu}$ is not close to the boundary of $\triangle\Omega^*$. We can show that this condition is automatically satisfied in the definition of uniform transience, as in Lemma C7.

In words, the game is asymptotically uniformly connected if each state ω is globally accessible or uniformly transient, and if players can statistically distinguish whether the current state ω is in the set Ω^* or not through the public signal y. Loosely, the latter condition ensures that players can eventually learn the current support after a long time at least for some history, which implies asymptotic accessibility of some sets Ω^*. See Appendix C for the formal proof.

Note that the second condition in the above proposition is satisfied if the signal distributions $\{\pi_Y^\omega(a)|\omega \in \Omega\}$ are linearly independent for each a. Note also that linear independence is satisfied for generic signal structures as long as the signal space is large enough so that $|Y| \geq |\Omega|$. So asymptotic uniform connectedness generically holds as long as the evolution of the hidden state ω satisfies the communicating state assumption and the signal space is large enough, as discussed in Section 7.

As shown in Proposition 4, uniform connectedness ensures invariance of the feasible payoff set. The following proposition shows that the invariance result remains valid as long as the game is asymptotically uniformly connected.[18] The proof can be found in Appendix C.

Proposition B2. *If the game is asymptotically uniformly connected, then for each $\varepsilon > 0$, there is $\overline{\delta} \in (0,1)$ such that for any $\lambda \in \Lambda$, $\delta \in (\overline{\delta},1)$, μ, and $\tilde{\mu}$,*

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

In the same spirit, we can find a condition which is weaker than robust connectedness but still ensures invariance of the limit minimax payoff. The idea is quite similar to asymptotic connectedness and hence the details are omitted.

Appendix C: Proofs

C.1 Proof of Proposition 1

It is obvious that any superset of a globally accessible set is globally accessible. So it is sufficient to show that any superset of a uniformly transient set is globally accessible or uniformly transient.

[18] However, unlike Proposition 4, we do not know the rate of convergence, and in particular, we do not know if we can replace ε in the proposition with $O(1-\delta)$.

Let Ω^* be a uniformly transient set, and take a superset $\tilde{\Omega}^*$. Suppose that $\tilde{\Omega}^*$ is not globally accessible. In what follows, we show that it is uniformly transient. Take a strategy profile s arbitrarily. Since Ω^* is uniformly transient, there is T and $(y^1,\cdots,y^T)$ such that if the support of the initial prior is Ω^* and players play s, the signal sequence $(y^1,\cdots,y^T)$ appears with positive probability and the support of the posterior belief μ^{T+1} is globally accessible. Pick such T and $(y^1,\cdots,y^T)$. Now, suppose that the support of the initial prior is $\tilde{\Omega}^*$ and players play s. Then since $\tilde{\Omega}^*$ is a superset of Ω^*, the signal sequence $(y^1,\cdots,y^T)$ realizes with positive probability and the support of the posterior belief $\tilde{\mu}^{T+1}$ is a superset of the support of μ^{T+1}. Since the support of μ^{T+1} is globally accessible, so is the superset. This shows that $\tilde{\Omega}^*$ is uniformly transient, as s can be arbitrary.

C.2 Proof of Proposition 2

Consider stochastic games with observable states. For the if part, it is obvious that a singleton set $\{\omega\}$ with globally accessible ω is globally accessible, and other sets Ω^* are uniformly transient. The only if part follows from Proposition 3.

Next, consider stochastic games with delayed observations. Again the only if part follows from Lemma 3, so we focus on the if part. We first prove that if ω is uniformly transient, then the set $\{\omega\}$ is uniformly transient. To prove this, take a uniformly transient state ω, and take an arbitrary pure strategy profile s. Since ω is uniformly transient, there must be a history h^{t-1} such that if the initial state is ω and players play s, the history h^{t-1} realizes with positive probability and the posterior puts positive probability on some globally accessible state ω^* Pick such h^{t-1} and ω^*. Let h^t be the history such that the history until period $t-1$ is h^{t-1}, and then players played $s(h^{t-1})$ and observed $y=\omega^*$ in period t. By the definition, this history h^t happens with positive probability given the initial state ω and the strategy profile s. Now, let Ω^* be the support of the posterior belief after h^t. To prove that $\{\omega\}$ is uniformly transient, it is sufficient to show that this set Ω^* is globally accessible, because it ensures that the support must move from $\{\omega\}$ to a globally accessible set regardless of players' play s. (For $\{\omega\}$ to be uniformly transient, we also need to show that $\{\omega\}$ is not globally accessible, but it follows from the fact that ω is not globally accessible.)

To prove that Ω^* is globally accessible. pick an arbitrary prior μ, and pick $\tilde{\omega}$

such that $\mu(\tilde{\omega}) \geq \frac{1}{|\Omega|}$. Since ω^* is globally accessible, there is an action sequence $(a^1, \cdots, a^T)$ which moves the state from $\tilde{\omega}$ to ω^* with positive probability. Pick such an action sequence, and pick a signal sequence $(y^1, \cdots, y^T)$ which happens when the state moves from $\tilde{\omega}$ to ω^*. Now, suppose that the initial prior is μ and players play $(a^1, \cdots, a^T, s(h^{t-1}))$. Then by the definition, with positive probability, players observe the signal sequence $(y^1, \cdots, y^T)$ during the first T periods and then the signal $y^{T+1} = \omega^*$ in period $T+1$. Obviously the support of the posterior after such a history is Ω^*, so this shows that the support can move to Ω^* from any initial prior. Also the probability of this move is at least $\mu(\tilde{\omega})\overline{\pi}^{T+1} \geq \frac{\overline{\pi}^{T+1}}{|\Omega|}$ for all initial prior μ. Hence Ω^* is globally accessible, as desired.

So far we have shown that $\{\omega\}$ is uniformly transient if ω is uniformly transient. To complete the proof of the if part, we show that when ω is globally accessible, $\{\omega\}$ is globally accessible or uniformly transient. So fix an arbitrary $\{\omega\}$ such that ω is globally accessible yet $\{\omega\}$ is not globally accessible. It is sufficient to show that $\{\omega\}$ is uniformly transient. To do so, fix arbitrary a^* and y^* such that $\pi_Y^{\omega}(y^*|a^*) > 0$, and let Ω^* be the set of all $\tilde{\omega}$ such that $\pi^{\omega}(y^*, \tilde{\omega}|a^*) > 0$. Then just as in the previous previous paragraph, we can show that Ω^* is globally accessible, which implies that $\{\omega\}$ is uniformly transient.

C.3 Proof of Proposition 3

For each state ω, let $\Omega(\omega)$ denote the set of all states reachable from the state ω. That is, $\Omega(\omega)$ is the set of all states $\tilde{\omega}$ such that there is a natural number $T \geq 1$ and an action sequence $(a^1, \cdots, a^T)$ such that the probability of the state in period $T+1$ being $\tilde{\omega}$ is positive given the initial state ω and the action sequence $(a^1, \cdots, a^T)$.

The proof consists of three steps. In the first step, we show that the game is uniformly connected only if $\Omega(\omega) \cap \Omega(\tilde{\omega}) \neq \emptyset$ for all ω and $\tilde{\omega}$. In the second step, we show that the condition considered in the first step (i.e., $\Omega(\omega) \cap \Omega(\tilde{\omega}) \neq \emptyset$ for all ω and $\tilde{\omega}$) holds if and only if there is a globally accessible state ω. This and the result in the first step imply that the game is uniformly connected only if there is a globally accessible state ω. Then in the last step, we show that the game is uniformly connected only if each state ω is globally accessible or uniformly transient.

C.3.1 Step 1: Uniformly Connected Only If $\Omega(\omega) \cap \Omega(\tilde{\omega}) \neq \emptyset$

Here we show that the game is uniformly connected only if $\Omega(\omega) \cap \Omega(\tilde{\omega}) \neq \emptyset$ for all ω and $\tilde{\omega}$. It is equivalent to show that if $\Omega(\omega) \cap \Omega(\tilde{\omega}) = \emptyset$ for some ω and $\tilde{\omega}$, then the game is not uniformly connected.

So suppose that $\Omega(\omega) \cap \Omega(\tilde{\omega}) = \emptyset$ for ω and $\tilde{\omega}$. Take an arbitrary state $\hat{\omega} \in \Omega(\omega)$. To prove that the game is not uniformly connected, it is sufficient to show that the singleton set $\{\hat{\omega}\}$ is not globally accessible or uniformly transient.

We first show that the set $\{\hat{\omega}\}$ is not globally accessible. More generally, we show that any set $\Omega^* \subseteq \Omega(\omega)$ is not globally accessible. Pick $\Omega^* \subseteq \Omega(\omega)$ arbitrarily. Then $\Omega^* \cap \Omega(\tilde{\omega}) = \emptyset$, and hence there is no action sequence which moves the state from $\tilde{\omega}$ to some state in the set Ω^* with positive probability. This means that if the initial prior puts probability one on $\tilde{\omega}$, then regardless of the past history, the posterior belief never puts positive probability on any state in the set Ω^*, and thus the support of the posterior belief is never included in the set Ω^*. Hence the set Ω^* is not globally accessible, as desired.

Next, we show that the set $\{\hat{\omega}\}$ is not uniformly transient. Note first that $\hat{\omega} \in \Omega(\omega)$ implies $\Omega(\hat{\omega}) \subseteq \Omega(\omega)$. That is, if $\hat{\omega}$ is accessible from ω, then any state accessible from $\hat{\omega}$ is accessible from ω. So if the initial state is $\hat{\omega}$, then in any future period, the state must be included in the set $\Omega(\omega)$ regardless of players' play. This implies that if the initial prior puts probability one on $\hat{\omega}$, then regardless of the players' play, the support of the posterior belief is always included in the set $\Omega(\omega)$; this implies that the support never reaches a globally accessible set, because we have seen in the previous paragraph that any set $\Omega^* \subseteq \Omega(\omega)$ is not globally accessible. Hence $\{\omega\}$ is not uniformly transient, as desired.

C.3.2 Step 2: Uniformly Connected Only If There is Globally Accessible ω

Here we show that $\Omega(\omega) \cap \Omega(\tilde{\omega}) \neq \emptyset$ for all ω and $\tilde{\omega}$ if and only if there is a globally accessible state ω. This and the result in the previous step implies that the game is uniformly connected only if there is a globally accessible state ω.

The if part simply follows from the fact that if ω is globally accessible, then $\omega \in \Omega(\tilde{\omega})$ for all $\tilde{\omega}$. So we prove the only if part. That is, we show that if $\Omega(\omega) \cap \Omega(\tilde{\omega}) \neq \emptyset$ for all ω and $\tilde{\omega}$, then there is a globally accessible state ω. So assume that $\Omega(\omega) \cap \Omega(\tilde{\omega}) \neq \emptyset$ for all ω and $\tilde{\omega}$.

Since the state space is finite, the states can be labeled as $\omega_1, \omega_2, \cdots, \omega_K$. Pick $\omega^* \in \Omega(\omega_1) \cap \Omega(\omega_2)$ arbitrarily; possibly we have $\omega^* = \omega_1$ or $\omega^* = \omega_2$. By the definition, ω^* is accessible from ω_1 and ω_2.

Now pick $\omega^{**} \in \Omega(\omega^*) \cap \Omega(\omega_3)$. By the definition, this state ω^{**} is accessible from ω_3. Also, since ω^{**} is accessible from ω^* which is accessible from ω_1 and ω_2, ω^{**} is accessible from ω_1 and ω_2. So this state ω^{**} is accessible from ω_1, ω_2, and ω_3. Repeating this process, we can eventually find a state which is accessible from all states ω. This state is globally accessible, as desired.

C.3.3 Step 3: Uniformly Connected Only If Each ω is Globally Accessible or Uniformly Transient

Now we prove that the game is uniformly connected only if each state ω is globally accessible or uniformly transient. It is equivalent to show that if there is a state ω which is not globally accessible or uniformly transient, then the game is not uniformly connected.

We prove this by contradiction, so suppose that the state ω^* is not globally accessible or uniformly transient, and that the game is uniformly connected. Since ω^* is not globally accessible or uniformly transient, there is a strategy profile s such that if the initial state is ω^*, the state never reaches a globally accessible state. Pick such a strategy profile s, and let Ω^* be the set of states accessible from ω^* with positive probability given the strategy profile s. That is, Ω^* is the set of states which can happen with positive probability in some period $t \geq 2$ if the initial state is ω and the strategy profile is s. (Note that Ω^* is different from $\Omega(\omega^*)$, as the strategy profile s is given here.) By the definition of s, any state in Ω^* is not globally accessible.

Since the game is uniformly connected, the singleton set $\{\omega^*\}$ must be either globally accessible or uniformly transient. It cannot be globally accessible, because ω^* is not globally accessible and hence there is some state ω such that ω^* is not accessible from ω; if the initial prior puts probability one on such ω, then regardless of the play, the posterior never puts positive probability on ω^*. So the singleton set $\{\omega^*\}$ must be uniformly transient. This requires that if the initial prior puts probability one on ω^* and players play the profile s, then the support of the posterior must eventually reach some globally accessible set. By the definition

of Ω^*, given the initial prior ω^* and the profile s, the support of the posterior must be included in Ω^*. This implies that there is a globally accessible set $\tilde{\Omega}^* \subseteq \Omega^*$.

However, this is a contradiction, because any set $\tilde{\Omega}^* \subseteq \Omega^*$ cannot be globally accessible. To see this, recall that the game is uniformly connected, and then as shown in Step 2, there must be a globally accessible state, say ω^{**}. Then $\Omega^* \cap \Omega(\omega^{**}) = \emptyset$, that is, any state in Ω^* is not accessible from ω^{**}. Indeed if not and some state $\omega \in \Omega^*$ is accessible from ω^{**}, then the state ω is globally accessible, which contradicts with the fact that any state in Ω^* is not globally accessible. Now, if the initial prior puts probability one on ω^{**}, then regardless of the play, the posterior belief never puts positive probability on any state in the set Ω^*, and hence the support of the posterior belief is never included in the set Ω^*. This shows that any subset $\tilde{\Omega}^* \subseteq \Omega^*$ is not globally accessible, which is a contradiction.

C.4 Proof of Proposition 4

Fix δ and the direction λ. For each μ, let s^μ be a pure-strategy profile which solves $\max_{s \in S} \lambda \cdot v(\delta, s)$. That is, s^μ is the profile which achieves the score given the initial prior μ. For each initial prior μ, the score is denoted by $\lambda \cdot v^\mu(\delta, s^\mu)$. The following lemma shows that the score is convex with respect to μ.

Lemma C1. *$\lambda \cdot v^\mu(\delta, s^\mu)$ 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. That is, given δ and λ, the score $\lambda \cdot v^\mu(\delta, s^\mu)$ is maximized when the belief μ puts probability one on some state ω. Pick such ω. In what follows, the score for this ω is called the *maximal score*.

The rest of the proof consists of three steps. In the first step, we show that for each set Ω^*, if there is a relative interior belief $\mu \in \triangle\Omega^*$ whose score approximates the maximal score, then the score function is almost flat over the space $\triangle\Omega^*$ and the score for *every* belief $\mu \in \triangle\Omega^*$ approximates the maximal score. This result generalizes the one presented in Step 1 in Section 5.2.

In the second step, we show that there is a globally accessible set Ω^* such that the score for any belief with support Ω^* approximates the maximal score. The result in the first step ensures that it suffices to find a belief μ^* whose support Ω^* is globally accessible and whose score approximates the maximal score. To find such μ, we follow the technique presented in Step 2 in Section 5.2.

In the last step, we show that the score for any belief μ (here the support of μ is not necessarily Ω^*) approximates the maximal score. To get the proof idea, pick an arbitrary initial belief μ, and consider the following strategy $\tilde{s}^\mu$: Players try to move the support of the belief to Ω^*, and once it happens (say in period t), they switch their play to the optimal policy s^{μ^t} in the continuation game. Note that players can indeed move the support to Ω^* in finite time with probability one, because Ω^* is globally accessible. Hence for patient players, waiting time is almost negligible. Note also that the continuation payoff after the switch to s^{μ^t} approximates the maximal score, as shown in the second step. Hence for δ close to one, this strategy profile $\tilde{s}^\mu$ approximates the maximal score. This implies the desired result, because the score for μ is even higher and thus closer to the maximal score.

C.4.1 Step 1: Almost Flat Convex Curve

Here, we show that if there is a belief μ whose score approximates the maximal score, then the score for *every* belief $\tilde{\mu}$ with the same support as μ approximates the maximal score. Specifically, we prove the following lemma.

Lemma C2. *Pick an arbitrary belief μ. Let Ω^* denote its support and let $p = \min_{\tilde{\omega}\in\Omega^*} \mu(\tilde{\omega})$, which measures the distance from μ to the boundary of $\triangle\Omega^*$. Then for each $\tilde{\mu} \in \triangle\Omega^*$,*

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

To interpret this lemma, let $\mu = (\frac{1}{|\Omega|}, \cdots, \frac{1}{|\Omega|})$, and suppose that

$$|\lambda \cdot v^{\omega}(\delta, s^{\omega}) - \lambda \cdot v^{\mu}(\delta, s^{\mu})| < \varepsilon \tag{9}$$

where ε is a positive number close to zero. That is, assume that the score for the belief $\mu = (\frac{1}{|\Omega|}, \cdots, \frac{1}{|\Omega|})$ approximates the maximal score $\lambda \cdot v^{\omega}(\delta, s^{\omega})$. Then the above lemma implies that

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

for all $\tilde{\mu} \in \triangle\Omega$. Since ε is close to zero, this implies that the score $\lambda \cdot v^{\tilde{\mu}}(\delta, s^{\tilde{\mu}})$ approximates the maximal score for *every* belief $\tilde{\mu}$.

A similar argument applies even if the support of μ is not the whole state space. For example, let μ be the uniform distribution over some set $\tilde{\Omega} \subset \Omega$ (so the support of μ is $\tilde{\Omega}$) and suppose that (9) holds for this μ. Then the above lemma implies that

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

for all $\tilde{\mu} \in \triangle\tilde{\Omega}$. That is, the score for every belief $\tilde{\mu}$ with the same support as μ approximates the maximal score.

Note that, for the above argument to work, the belief μ should not be too close to the boundary of $\triangle\Omega^*$, where Ω^* is the support of μ. Indeed, if μ is close to the boundary of $\triangle\Omega^*$, the parameter p approximates zero so that the bound $\frac{|\lambda \cdot v^{\omega}(\delta, s^{\omega}) - \lambda \cdot v^{\mu}(\delta, s^{\mu})|}{p}$ can be arbitrarily large. So the more precise interpretation of the above lemma is: If there is a belief μ such that the score for μ approximates the maximal score and such that μ is not too close to the boundary of $\triangle\Omega^*$, then the score for *every* belief $\tilde{\mu}$ with support Ω^* approximates the maximal score.

The proof idea of the above lemma is already presented in Section 5.2. The formal proof below is more complicated because now we need to consider the case in which the score for the belief μ approximates but is not equal to the maximal score.

Proof. Pick an arbitrary belief μ and let Ω^* be the support of μ. Pick $\tilde{\omega} \in \Omega^*$

arbitrarily. Then we have

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

Applying (8) to the above inequality, we obtain

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

Arranging,

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

Dividing both sides by $\mu(\tilde{\omega})$,

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

Since $\lambda \cdot v^{\omega}(\delta, s^{\omega}) - \lambda \cdot v^{\mu}(\delta, s^{\mu}) > 0$, we can apply $\mu(\tilde{\omega}) \geq p = \min_{\tilde{\omega} \in \Omega^*} \mu(\tilde{\omega})$ to the above inequality and obtain

$$\lambda \cdot v^{\omega}(\delta, s^{\omega}) - \lambda \cdot v^{\tilde{\omega}}(\delta, s^{\mu}) \leq \frac{\lambda \cdot v^{\omega}(\delta, s^{\omega}) - \lambda \cdot v^{\mu}(\delta, s^{\mu})}{p}. \tag{10}$$

Pick an arbitrary belief $\tilde{\mu} \in \triangle\Omega^*$. Recall that (10) holds for each $\tilde{\omega} \in \Omega^*$. Multiplying both sides of (10) by $\tilde{\mu}(\tilde{\omega})$ and summing over all $\tilde{\omega} \in \Omega^*$,

$$\lambda \cdot v^{\omega}(\delta, s^{\omega}) - \lambda \cdot v^{\tilde{\mu}}(\delta, s^{\mu}) \leq \frac{\lambda \cdot v^{\omega}(\delta, s^{\omega}) - \lambda \cdot v^{\mu}(\delta, s^{\mu})}{p}.$$

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

$$\lambda \cdot v^{\omega}(\delta, s^{\omega}) - \lambda \cdot v^{\tilde{\mu}}(\delta, s^{\tilde{\mu}}) \leq \frac{\lambda \cdot v^{\omega}(\delta, s^{\omega}) - \lambda \cdot v^{\mu}(\delta, s^{\mu})}{p}.$$

Taking the absolute values of both sides, we obtain the result. *Q.E.D.*

C.4.2 Step 2: Bound on the Scores for All Beliefs with Support Ω^*

Here, we show that there is a globally accessible set Ω^* such that the score for any belief $\mu \in \triangle\Omega^*$ approximates the maximal score. More precisely, we prove the following lemma:

Lemma C3. *There is a globally accessible set $\Omega^* \subseteq \Omega$ such that for all $\mu \in \triangle\Omega^*$,*

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

The proof idea is as follows. Since the game is uniformly connected, $\{\omega\}$ is globally accessible or uniformly transient. If it is globally accessible, let $\Omega^* = \{\omega\}$. This set Ω^* satisfies the desired property, because the set $\triangle\Omega^*$ contains only the belief $\mu = \omega$, and the score for this belief is exactly equal to the maximal score.

Now, consider the case in which $\{\omega\}$ is uniformly transient. Suppose that the initial state is ω and the optimal policy s^{ω} is played. Since $\{\omega\}$ is uniformly transient, there is a natural number $T \leq 2^{|\Omega|}$ and a history h^T such that the history h^T appears with positive probability and the support of the posterior belief after the history h^T is globally accessible. Take such T and h^T. Let μ^* denote the posterior belief after this history h^T and let Ω^* denote its support. By the definition, Ω^* is globally accessible. Using the technique similar to the one in Section 5.2, we can show that the continuation payoff after this history h^T approximates the maximal score, and so does the score for this belief μ^*. Then Lemma C2 implies that the score for any belief $\mu \in \triangle\Omega^*$ approximates the maximal score, as desired.

Proof. First, consider the case in which $\{\omega\}$ is globally accessible. Let $\Omega^* = \{\omega\}$. Then this set Ω^* satisfies the desired property, because $\triangle\Omega^*$ contains only the belief $\mu = \omega$, and the score for this belief is exactly equal to the maximal score.

Next, consider the case in which $\{\omega\}$ is uniformly transient. Take T, h^T, μ^*, and Ω^* as stated above. By the definition, the support of μ^* is Ω^*. Also, since μ^* is induced by the initial state ω, we have $\mu^*(\tilde{\omega}) \geq \overline{\pi}^T$ for each $\tilde{\omega} \in \Omega^*$. (Recall that $\overline{\pi}$ is the minimum of the π function.) In this sense, μ^* is not too close to the boundary of $\triangle\Omega^*$.

For each history $\tilde{h}^T$, let $\mu(\tilde{h}^T)$ denote the posterior belief given the initial state ω and the history $\tilde{h}^T$. From the principle of optimality, we have

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

Using (8), $\mu(h^T) = \mu^*$, and $(1-\delta)\sum_{t=1}^{T}\delta^{t-1}E[\lambda \cdot g^{\omega^t}(a^t)|\omega^1 = \omega, s^\omega] \leq (1-\delta^T)\overline{g}$, we obtain

$$\begin{aligned}\lambda \cdot v^\omega(\delta, s^\omega) \leq & (1-\delta^T)\overline{g} + \delta^T \Pr(h^T|\omega, s^\omega)\lambda \cdot v^{\mu^*}(\delta, s^{\mu^*}) \\ & + \delta^T(1 - \Pr(h^T|\omega, s^\omega))\lambda \cdot v^\omega(\delta, s^\omega).\end{aligned}$$

Arranging, we have

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

Note that $\Pr(h^T|\omega, s^\omega) \geq \overline{\pi}^T$, because it is the probability given the initial state ω and the pure strategy s^ω. Hence we have

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

Since (8) ensures that the left-hand side is non-negative, taking the absolute values of both sides and using $\lambda \cdot v^\omega(\delta, s^\omega) \geq -\overline{g}$,

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

That is, the score for the belief μ^* approximates the maximal score if δ is close to one. As noted, we have $\mu^*(\tilde{\omega}) \geq \overline{\pi}^T$ for each $\tilde{\omega} \in \Omega^*$. Then applying Lemma C2 to the inequality above, we obtain

$$|\lambda \cdot v^\omega(\delta, s^\omega) - \lambda \cdot v^\mu(\delta, s^\mu)| \leq \frac{(1-\delta^T)2\overline{g}}{\delta^T \overline{\pi}^{2T}}$$

for each $\mu \in \triangle\Omega^*$. This implies the desired inequality, since $T \leq 2^{|\Omega|}$. *Q.E.D.*

Pick Ω^* as in the above lemma. If $\Omega^* = \Omega$ (this is the case, for example, when the full support assumption holds) then the lemma implies Proposition 4. However, often times Ω^* is smaller than Ω and thus the above lemma does not bound the score for $\mu \notin \triangle\Omega^*$. In Step 3, we explain how to complete the proof for such a case.

C.4.3 Step 3: Bound on the Scores for All Beliefs μ

We begin with providing a preliminary lemma. In the definition of global accessibility, the action sequence which moves the support to a globally accessible set Ω^* depends on the current belief. The following lemma shows that such a belief-dependent action sequence can be replaced with a belief-independent sequence if we allow mixed actions. That is, if players mix all actions equally each period, then the support will reach Ω^* regardless of the current belief. Note that π^* in the lemma can be possibly different from the one in the definition of global accessibility.

Lemma C4. *Let Ω^* be a globally accessible set. Suppose that players randomize all actions equally each period. Then there is $\pi^* > 0$ such that given any initial prior μ, there is a natural number $T \leq 4^{|\Omega|}$ such that the support of the posterior belief at the beginning of period $T+1$ is a subset of Ω^* with probability at least π^*.*

Proof. Take $\pi^* > 0$ as stated in the definition of global accessibility of Ω^*. Take an arbitrary initial prior μ, and take an action sequence $(a^1, \cdots, a^T)$ as stated in the definition of global accessibility of Ω^*.

Suppose that players mix all actions each period. Then the action sequence $(a^1, \cdots, a^T)$ realizes with probability $\frac{1}{|A|^T}$, and it moves the support of the posterior to a subset of Ω^* with probability at least π^*. Hence, in sum, playing mixed actions each period moves the support to a subset of Ω^* with probability at least $\frac{1}{|A|^T} \cdot \pi^*$. This probability is bounded from zero for all μ, and hence the proof is completed. *Q.E.D.*

Pick Ω^* as in Step 2, so that it is globally accessible. Take $\pi^* > 0$ as stated in Lemma C4. This means that if players mix all actions each period, the support will move to Ω^* or its subset within $4^{|\Omega|}$ periods with probability at least π^* regardless of the initial prior.

Pick an initial prior μ, and suppose that players play the following strategy profile $\tilde{s}^\mu$:

- Players randomize all actions equally likely, until the support of the posterior belief becomes a subset of Ω^*.

- Once the support of the posterior belief becomes a subset of Ω^* in some period t, players play s^{μ^t} in the rest of the game. (They do not change the play after that.)

That is, players wait until the support of the belief reaches Ω^*, and once it happens, they switch the play to the optimal policy s^{μ^t} in the continuation game. Lemma C3 guarantees that the continuation play after the switch to s^{μ^t} approximates the maximal score $\lambda \cdot v^{\omega}(\delta, s^{\omega})$. Also, Lemma C4 ensures that this switch occurs with probability one and waiting time is almost negligible for patient players. Hence the payoff by this strategy profile $\tilde{s}^{\mu}$ approximates the maximal score. Formally, we have the following lemma.

Lemma C5. *For each μ,*

$$|\lambda \cdot v^{\omega}(\delta, s^{\omega}) - \lambda \cdot v^{\mu}(\delta, \tilde{s}^{\mu})| \leq \frac{(1-\delta^{2^{|\Omega|}})2\overline{g}}{\delta^{2^{|\Omega|}}\underline{\pi}^{4^{|\Omega|}}} + \frac{(1-\delta^{4^{|\Omega|}})3\overline{g}}{\pi^*}.$$

Proof. Pick an arbitrary belief μ. If $\frac{(1-\delta^{2^{|\Omega|}})2\overline{g}}{\delta^{2^{|\Omega|}}\underline{\pi}^{4^{|\Omega|}}} \geq \overline{g}$, then the result obviously holds because we have $|\lambda \cdot v^{\omega}(\delta, s^{\omega}) - \lambda \cdot v^{\mu}(\delta, \tilde{s}^{\mu})| \leq \overline{g}$. So in what follows, we assume that $\frac{(1-\delta^{2^{|\Omega|}})2\overline{g}}{\delta^{2^{|\Omega|}}\underline{\pi}^{4^{|\Omega|}}} < \overline{g}$.

Suppose that the initial prior is μ and players play the strategy profile $\tilde{s}^{\mu}$. Let $\Pr(h^t|\mu, \tilde{s}^{\mu})$ be the probability of h^t given the initial prior μ and the strategy profile $\tilde{s}^{\mu}$, and let $\mu^{t+1}(h^t|\mu, \tilde{s}^{\mu})$ denote the posterior belief in period $t+1$ given this history h^t. Let H^{*t} be the set of histories h^t such that $t+1$ is the first period at which the support of the posterior belief μ^{t+1} is in the set Ω^*. Intuitively, H^{*t} is the set of histories h^t such that players will switch their play to $s^{\mu^{t+1}}$ from period $t+1$ on, according to $\tilde{s}^{\mu}$.

Note that the payoff $v^{\mu}(\delta, \tilde{s}^{\mu})$ by the strategy profile $\tilde{s}^{\mu}$ can be represented as the sum of the two terms: The expected payoffs before the switch to s^{μ^t} occurs, and the payoffs after the switch. That is, we have

$$\begin{aligned}\lambda \cdot v^{\mu}(\delta, \tilde{s}^{\mu}) = &\sum_{t=1}^{\infty} \left(1 - \sum_{\tilde{t}=0}^{t-1} \sum_{h^{\tilde{t}} \in H^{*\tilde{t}}} \Pr(h^{\tilde{t}}|\mu, \tilde{s}^{\mu})\right) (1-\delta)\delta^{t-1} E\left[\lambda \cdot g^{\omega^t}(a^t)|\mu, \tilde{s}^{\mu}\right] \\ &+ \sum_{t=0}^{\infty} \sum_{h^t \in H^{*t}} \Pr(h^t|\mu, \tilde{s}^{\mu})\delta^t \lambda \cdot v^{\mu^{t+1}(h^t|\mu, \tilde{s}^{\mu})}(\delta, s^{\mu^{t+1}(h^t|\mu, \tilde{s}^{\mu})})\end{aligned}$$

where the expectation operator is taken conditional on that the switch has not happened yet. Note that the term $1-\sum_{\tilde{t}=0}^{t-1}\sum_{h^{\tilde{t}}\in H^{*\tilde{t}}}\Pr(h^{\tilde{t}}|\mu,\tilde{s}^{\mu})$ is the probability that players still randomize all actions in period t because the switch has not happened by then. To simplify the notation, let ρ^t denote this probability. From Lemma C3, we know that

$$\lambda\cdot v^{\mu^{t+1}(h^t|\mu,\tilde{s}^{\mu})}(\delta,s^{\mu^{t+1}(h^t|\mu,\tilde{s}^{\mu})})\geq v^*$$

for each $h^t\in H^{*t}$, where $v^*=\lambda\cdot v^{\omega}(\delta,s^{\omega})-\frac{(1-\delta^{2^{|\Omega|}})2\overline{g}}{\delta^{2^{|\Omega|}}\underline{\pi}^{4^{|\Omega|}}}$. Applying this and $\lambda\cdot g^{\omega^t}(a^t)\geq -2\overline{g}$ to the above equation, we obtain

$$\lambda\cdot v^{\mu}(\delta,\tilde{s}^{\mu})\geq\sum_{t=1}^{\infty}\rho^t(1-\delta)\delta^{t-1}(-2\overline{g})+\sum_{t=0}^{\infty}\sum_{h^t\in H^{*t}}\Pr(h^t|\mu,\tilde{s}^{\mu})\delta^t v^*.$$

Using $\sum_{t=0}^{\infty}\sum_{h^t\in H^{*t}}\Pr(h^t|\mu,\tilde{s}^{\mu})\delta^t=\sum_{t=1}^{\infty}(1-\delta)\delta^{t-1}\sum_{\tilde{t}=0}^{t-1}\sum_{h^{\tilde{t}}\in H^{*\tilde{t}}}\Pr(h^{\tilde{t}}|\mu,\tilde{s}^{\mu})=\sum_{t=1}^{\infty}(1-\delta)\delta^{t-1}(1-\rho^t)$, we obtain

$$\lambda\cdot v^{\mu}(\delta,\tilde{s}^{\mu})\geq(1-\delta)\sum_{t=1}^{\infty}\delta^{t-1}\left\{\rho^t(-2\overline{g})+(1-\rho^t)v^*\right\}. \tag{11}$$

According to Lemma C4, the probability that the support reaches Ω^* within $4^{|\Omega|}$ periods is at least $\underline{\pi}^*$. This implies that the probability that players still randomize all actions in period $4^{|\Omega|}+1$ is at most $1-\underline{\pi}^*$. Similarly, for each natural number n, the probability that players still randomize all actions in period $n4^{|\Omega|}+1$ is at most $(1-\underline{\pi}^*)^n$, that is, $\rho^{n4^{|\Omega|}+1}\leq(1-\underline{\pi}^*)^n$. Then since ρ^t is weakly decreasing in t, we obtain

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

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

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

for each $n=0,1,\cdots$ and $k\in\{1,\cdots,4^{|\Omega|}\}$. Plugging this inequality into (11), we obtain

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

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

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

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

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

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

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

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

$$\begin{aligned}
& \lambda \cdot v^{\omega}(\delta, s^{\omega}) - \lambda \cdot v^{\mu}(\delta, \tilde{s}^{\mu}) \\
& \leq \frac{(1-\delta^{4^{|\Omega|}})2\overline{g}}{1-(1-\pi^*)\delta^{4^{|\Omega|}}} + \frac{\delta^{4^{|\Omega|}}\pi^*(1-\delta^{2^{|\Omega|}})2\overline{g}}{\{1-(1-\pi^*)\delta^{4^{|\Omega|}}\}\delta^{2^{|\Omega|}}\overline{\pi}^{4^{|\Omega|}}} + \frac{(1-\delta^{4^{|\Omega|}})\overline{g}}{1-(1-\pi^*)\delta^{4^{|\Omega|}}} \\
& \leq \frac{(1-\delta^{4^{|\Omega|}})3\overline{g}}{1-(1-\pi^*)} + \frac{\pi^*(1-\delta^{2^{|\Omega|}})2\overline{g}}{\{1-(1-\pi^*)\}\delta^{2^{|\Omega|}}\overline{\pi}^{4^{|\Omega|}}} \\
& = \frac{(1-\delta^{4^{|\Omega|}})3\overline{g}}{\pi^*} + \frac{(1-\delta^{2^{|\Omega|}})2\overline{g}}{\delta^{2^{|\Omega|}}\overline{\pi}^{4^{|\Omega|}}}
\end{aligned}$$

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

Note that

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

that is, the score for μ is at least $\lambda \cdot v^{\mu}(\delta, \tilde{s}^{\mu})$ (this is because $\tilde{s}^{\mu}$ is not the optimal policy) and is at most the maximal score. Then from Lemma C5, we have

$$\begin{aligned} |\lambda \cdot v^{\omega}(\delta, s^{\omega}) - \lambda \cdot v^{\mu}(\delta, s^{\mu})| &\leq |\lambda \cdot v^{\omega}(\delta, s^{\omega}) - \lambda \cdot v^{\mu}(\delta, \tilde{s}^{\mu})| \\ &\leq \frac{(1-\delta^{2^{|\Omega|}})2\overline{g}}{\delta^{2^{|\Omega|}}\underline{\pi}^{4^{|\Omega|}}} + \frac{(1-\delta^{4^{|\Omega|}})3\overline{g}}{\pi^{*}}, \end{aligned}$$

as desired.

C.5 Proof of Proposition 5

In this proof, we show only the existence of the limit of the score. Independence of the limit score follows from Proposition 4.

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

$$\left| \max_{v \in V^{\mu}(\overline{\delta})} \lambda \cdot v - \limsup_{\delta \to 1} \max_{v \in V^{\mu}(\delta)} \lambda \cdot v \right| < \frac{\varepsilon}{2} \tag{12}$$

and such that

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

for each $\tilde{\mu}$. Note that Proposition 4 guarantees that such $\overline{\delta}$ exists.

For each $\tilde{\mu}$, let $s^{\tilde{\mu}}$ be a pure strategy profile which achieves the score given $\tilde{\mu}$ and $\overline{\delta}$. That is, $\lambda \cdot v^{\tilde{\mu}}(\overline{\delta}, s^{\tilde{\mu}}) = \max_{v \in V^{\tilde{\mu}}(\overline{\delta})} \lambda \cdot v$. In what follows, we show that

$$\lambda \cdot v^{\mu}(\delta, s^{\mu}) > \limsup_{\tilde{\delta} \to 1} \max_{v \in V^{\mu}(\tilde{\delta})} \lambda \cdot v - \varepsilon \tag{14}$$

for each $\delta \in (\overline{\delta}, 1)$. That is, we show that when the true discount factor is δ, the optimal policy s^{μ} for the discount factor $\overline{\delta}$ can achieve a payoff better than the limit superior of the score. Since the optimal policy s^{μ} for the discount factor $\overline{\delta}$ is not necessarily the optimal policy for δ, the score for δ is greater than $\lambda \cdot v^{\mu}(\delta, s^{\mu})$, and hence (14) ensures that the score for δ is greater than the limit superior of the score. Since this is true for all $\delta \in (\overline{\delta}, 1)$, the limit superior is the limit, as desired.

So pick an arbitrary $\delta \in (\overline{\delta}, 1)$, and compute $\lambda \cdot v^{\mu}(\delta, s^{\mu})$, the payoff by the optimal policy s^{μ} for the discount factor $\overline{\delta}$ when the true discount factor is δ. To

evaluate this payoff, we regard the infinite horizon as a series of random blocks, as in Section 6. The termination probability is $1-p$, where $p = \frac{\underline{\delta}}{\delta}$. Then, since s^μ is Markov, playing s^μ in the infinite-horizon game is payoff-equivalent to playing the following strategy profile:

- During the first random block, play s^μ.
- During the kth random block, play s^{μ^k} where μ^k is the belief in the initial period of the kth block.

Given this interpretation, the payoff $\lambda \cdot v^\mu(\delta, s^\mu)$ is represented as the sum of the random block payoffs, that is,

$$\lambda \cdot v^\mu(\delta, s^\mu) = (1-\delta) \sum_{k=1}^{\infty} \left(\frac{\delta(1-p)}{1-p\delta} \right)^{k-1} E\left[\frac{\lambda \cdot v^{\mu^k}(p\delta, s^{\mu^k})}{1-p\delta} \middle| \mu, s^\mu \right] \quad (15)$$

The meaning of the right-hand side is as follows. The term $1-\delta$ is needed because we want to convert unnormalized payoffs to average payoffs. The term $\frac{v^{\mu^k}(p\delta, s^{\mu^k})}{1-p\delta}$ is the expected unnormalized payoff during the kth random block, evaluated at the beginning of that block. Indeed, given μ^k, we have

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

The term $\left(\frac{\delta(1-p)}{1-p\delta} \right)^{k-1}$ is the "expected discount factor" due to earlier $k-1$ random blocks. To see this, consider the case in which $k=2$. The expected discount factor due to the first random block is

$$(1-p)\delta + (1-p)p\delta^2 + (1-p)p^2\delta^3 + \cdots = \frac{\delta(1-p)}{1-p\delta},$$

Here the first term comes from the fact that the length of the first block is one with probability $1-p$, in which case the payoff in the second block is discounted by δ. The second term from the fact that the length of the first block is two with probability $(1-p)p$, in which case the payoff in the second block is discounted by δ^2.

Since $p\delta = \underline{\delta}$, we have $\lambda \cdot v^{\tilde{\mu}}(p\delta, s^{\tilde{\mu}}) = \max_{v \in V^{\tilde{\mu}}(\underline{\delta})} \lambda \cdot v$ for each $\tilde{\mu}$. Hence (13) implies that

$$\lambda \cdot v^{\tilde{\mu}}(p\delta, s^{\tilde{\mu}}) > \lambda \cdot v^\mu(\underline{\delta}, s^\mu) - \frac{\varepsilon}{2}$$

for each $\tilde{\mu}$. Plugging this into (15), we have

$$\begin{aligned}\lambda \cdot v^{\mu}(\delta, s^{\mu}) &> (1-\delta) \sum_{k=1}^{\infty} \left(\frac{\delta(1-p)}{1-p\delta} \right)^{k-1} \left(\frac{\lambda \cdot v^{\mu}(\overline{\delta}, s^{\mu})}{1-p\delta} - \frac{\varepsilon}{2(1-p\delta)} \right) \\ &= \lambda \cdot v^{\mu}(\overline{\delta}, s^{\mu}) - \frac{\varepsilon}{2}.\end{aligned}$$

Then using (12), we obtain (14).

We would like to emphasize that this proof does not assume public randomization. Indeed, random blocks are useful for computing the payoff by the strategy s^{μ}, but the strategy s^{μ} itself does not use public randomization.

C.6 Proof of Proposition 6

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 $\times_{i \in I}[-\overline{g}, \overline{g}]$ for all δ and μ. Hence, for each λ, there is an open set $U_{\lambda} \subset \Lambda$ containing λ such that

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

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

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

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

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

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

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

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

C.7 Proof of Proposition 7 with Mixed Minimax Strategies

Here we explain how to extend the proof provided in Section 6.3 to the case in which the minimax strategies are mixed strategies. As explained, the only thing we need to do is to perturb the continuation payoff $w_i(j)$ so that player i is indifferent over all actions in each period during the minimax play.

We first explain how to perturb the payoff, and then explain why it makes player i indifferent. For each μ and a, take a real number $R_i(\mu,a)$ such that $g_i^{\mu}(a)+R_i(\mu,a)=0$. Intuitively, in the one-shot game with the belief μ, if player i receives the bonus payment $R_i(\mu,a)$ in addition to the stage-game payoff, she will be indifferent over all action profiles and her payoff will be zero. Suppose that we are now in the punishment phase for player $j \neq i$, and that the minimax play over K blocks is done. For each $k \in \{1,\cdots,K\}$. let $(\mu^{(k)},a^{(k)})$ denote the belief and the action profile in the last period of the kth block of the minimax play. Then the perturbed continuation payoff is defined as

$$w_i(j)+(1-\delta)\sum_{k=1}^{K}\frac{(1-p\delta)^{K-k}}{\{\delta(1-p)\}^{K-k+1}}R_i(\mu^{(k)},a^{(k)}).$$

That is, the continuation payoff is now the original value $w_i(j)$ plus the K perturbation terms $R_i(\mu^{(1)},a^{(1)})$, $\cdots$, $R_i(\mu^{(K)},a^{(K)})$, each of which is multiplied by the coefficient $(1-\delta)\frac{(1-p\delta)^{K-k}}{\{\delta(1-p)\}^{K-k+1}}$.

We now verify that player i is indifferent over all actions during the minimax play. First, suppose that we are now in the last block of the minimax play, and consider player i's incentive during this block. For simplicity, call the current period $t=1$, and let (μ^1,a^1) denote the belief and the action profile today. Also let (μ^t,a^t) denote the belief and the action profile in the tth period from now on. Note that the action in the last block can influence the perturbation term $R_i(\mu^{(K)},a^{(K)})$ but not on the term $R_i(\mu^{(k)},a^{(k)})$ for each $k<K$, because the past history is already given. Hence in what follows, we focus on the term $\frac{1-\delta}{\delta(1-p)}R_i(\mu^{(K)},a^{(K)})$ only and ignore the other perturbation terms.

In general, the action profile today influences the stage-game payoff today, and the future stage-game payoffs through the belief tomorrow. However, we claim that these impacts are offset by the perturbation term $\frac{1-\delta}{\delta(1-p)}R_i(\mu^{(K)},a^{(K)})$. Note that, if we ignore the perturbation terms $R_i(\mu^{(k)},a^{(k)})$ for each $k<K$, player i's

unnormalized payoff in the entire game from now on is

$$\sum_{t=1}^{\infty}(p\delta)^{t-1}E[g_i^{\mu^t}(a^t)]+\sum_{t=1}^{\infty}(1-p)p^{t-1}\delta^t\frac{1}{1-\delta}\left(w_i(j)+\frac{(1-\delta)E[R_i(\mu^t,a^t)]}{\delta(1-p)}\right)$$

where the expectation is taken with respect to μ^t and a^t, conditional on that the block does not terminate until period t. Note that the first term is the expected payoff in the current block, while the second term is the continuation payoff after the block. $(1-p)p^{t-1}$ in the second term is the probability of period t being the last period of the block, in which case the perturbation term is $R_i(\mu^t,a^t)$ and the continuation payoff is discounted by δ^t. The term $\frac{1}{1-\delta}$ in the second term is needed in order to convert the average payoff to the unnormalized payoff. The above payoff can be rewritten as

$$\sum_{t=1}^{\infty}(p\delta)^{t-1}E[g_i^{\mu^t}(a^t)+R_i(\mu^t,a^t)]+\frac{\delta(1-p)}{(1-\delta)(1-p\delta)}w_i(j).$$

Since $g_i^{\mu}(a)+R_i(\mu,a)=0$, the actions and the beliefs during the current block cannot influence this payoff at all. Hence player i is indifferent over all actions in each period during the block.

A similar argument applies to other blocks. For example, player i is indifferent in each period during the $(K-1)$st block, due to the perturbation term $\frac{(1-\delta)(1-p\delta)}{\{\delta(1-p)\}^2}R_i(\mu^{(K-1)},a^{(K-1)})$. The coefficient here is different from the one in the previous case, because one more random block remains before player i receives the perturbation term. Indeed, the "expected discount factor" due to the next random block is

$$\delta(1-p)+\delta^2p(1-p)+\delta^3p^2(1-p)+\cdots=\frac{\delta(1-p)}{1-p\delta}.$$

Here the first term on the left-hand side comes from the fact that the length of the next block is one with probability $1-p$, in which case discounting due to the next block is δ. Similarly, the second term comes from the fact that the length of the next block is two with probability $p(1-p)$, in which case discounting due to the next block is δ^2. If we multiply the coefficient $\frac{(1-\delta)(1-p\delta)}{\{\delta(1-p)\}^2}$ and the expected discount factor $\frac{\delta(1-p)}{1-p\delta}$, it will be $\frac{1-\delta}{\delta(1-p)}$, which is exactly equal to the coefficient in the previous case.

So player i is indifferent in all periods during the minimax play. Note also that the perturbed payoff converges to the original payoff $w_i(j)$ in the limit as $\delta \to 1$, because the perturbation terms are of order $1-\delta$. Hence for sufficiently large δ, the perturbed payoff vector is in the feasible payoff set, and all other incentive constraints are still satisfied.

C.8 Proof of Proposition A1

Before we go to the main proof, we provide two preliminary lemmas. The first lemma is about the support merging condition. Recall that under the support merging condition, two posterior beliefs induced by different initial priors ω and $\mu = (\frac{1}{|\Omega|}, \cdots, \frac{1}{|\Omega|})$ must eventually have the same support with positive probability, given any pure strategy profile s. The lemma shows that the same result holds for any μ with $\mu(\omega) > 0$ and for any mixed strategy profile s. Also without loss of generality, we can assume that the probability is uniformly bounded away from zero.

Lemma C6. *Suppose that the support merging condition holds. Then for each ω, for each μ with $\mu(\omega) > 0$, and for each (possibly mixed) strategy profile s, there is a natural number $T \leq 4^{|\Omega|}$ and a history h^T such that $\Pr(h^T|\omega, s) > (\frac{|\overline{\pi}|}{|A|})^T$ and such that the support of the posterior belief induced by the initial state ω and the history h^T is identical with the one induced by the initial prior μ and the history h^T.*

Proof. Take ω, μ, and s as stated. Take a pure strategy profile $\tilde{s}$ such that for each t and h^t, $\tilde{s}(h^t)$ chooses a pure action profile which is chosen with probability at least $\frac{1}{|A|}$ by $s(h^t)$.

Since the support merging condition holds, there is a natural number $T \leq 4^{|\Omega|}$ and a history h^T such that $\Pr(h^T|\omega, \tilde{s}) > 0$ and such that the support of the posterior belief induced by the initial state ω and the history h^T is identical with the one induced by the initial prior $\tilde{\mu} = (\frac{1}{|\Omega|}, \cdots, \frac{1}{|\Omega|})$ and the history h^T. We show that T and h^T here satisfies the desired properties.

Note that $\Pr(h^T|\omega, \tilde{s}) \geq \overline{\pi}^T$, because of the definition of $\overline{\pi}$. This implies that $\Pr(h^T|\omega, s) \geq (\frac{\overline{\pi}}{|A|})^{4^{|\Omega|}}$, since each period the action profile by s coincides with the one by $\tilde{s}$ with probability at least $\frac{1}{|A|}$. Also, since $\mu(\omega) > 0$, the support of

the belief induced by (ω, h^T) must included in the support induced by (μ, h^T), which must be included in the support induced by $(\tilde{\mu}, h^T)$. Since the first and last supports are the same, all three must be the same, implying that the support of the belief induced by (ω, h^T) is identical with the support induced by (μ, h^T), as desired. *Q.E.D.*

The next lemma is about robust accessibility. Recall that if Ω^* is robustly accessible, then for any initial prior μ, there is an action sequence $(\alpha_{-i}^1, \cdots, \alpha_{-i}^T)$ such that for any strategy s_i, the support reaches Ω^* with positive probability. The lemma ensures that we do not need to use such a belief-dependent action sequence; it is sufficient to use the action sequence such that all pure actions are mixed equally each period. The lemma also shows that without loss of generality, we can assume that the posterior belief when the support reaches Ω^* is not too close to the boundary of the belief space $\triangle\Omega^*$.

Lemma C7. *Suppose that Ω^* is robustly accessible despite i. Then there is $\pi^* > 0$ such that if the opponents mix all actions equally likely each period, then for any initial prior μ and for any strategy s_i, there is a natural number $T \leq 4^{|\Omega|}$ and a belief $\tilde{\mu} \in \triangle\Omega^*$ such that the posterior belief μ^{T+1} equals $\tilde{\mu}$ with probability at least π^* and such that $\tilde{\mu}(\omega) \geq \frac{1}{|\Omega|}\pi^{4^{|\Omega|}}$ for all $\omega \in \Omega^*$.*

Proof. We first show that Ω^* is robustly accessible only if the following condition holds:[19] For each state $\omega \in \Omega$ and for any s_i, there is a natural number $T \leq 4^{|\Omega|}$ and a pure action sequence $(a_{-i}^1, \cdots, a_{-i}^T)$, and a signal sequence $(y^1, \cdots, y^T)$ such that the following properties are satisfied:

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

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

[19] We can also show that the converse is true, so that Ω^* is robustly accessible if and only if the condition stated here is satisfied. Indeed, if the condition here is satisfied, then the condition stated in the definition of robust accessibility is satisfied by the action sequence $(\alpha_{-i}^1, \cdots, \alpha_{-i}^{4^{|\Omega|}})$ which mix all pure actions equally each period.

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

To see this, suppose not so that there is ω and s_i such that any action sequence and any signal sequence cannot satisfy (i) through (iii) simultaneously. Pick such ω and s_i. We will show that Ω^* is not robustly accessible.

Pick a small $\varepsilon > 0$ and let μ be such that $\mu(\omega) > 1-\varepsilon$ and and $\mu(\tilde{\omega}) > 0$ for all $\tilde{\omega}$. That is, consider μ which puts probability at least $1-\varepsilon$ on ω. Then by the definition of ω and s_i, the probability that the support reaches Ω^* given the initial prior μ and the strategy s_i is less than ε. Since this is true for any small $\varepsilon > 0$, the probability of the support reaching Ω^* must approach zero as $\varepsilon \to 0$, and hence Ω^* cannot be robustly accessible, as desired.

Now we prove the lemma. Fix an arbitrary prior μ, and pick ω such that $\mu(\omega) \geq \frac{1}{|\Omega|}$. Then for each s_i, choose T, $(a_{-i}^1, \cdots, a_{-i}^T)$, and $(y^1, \cdots, y^T)$ as stated in the above condition. (i) ensures that if the initial prior is μ, player i plays s_i, and the opponents mix all actions equally, the action sequence $(a_{-i}^1, \cdots, a_{-i}^T)$ and the signal sequence $(a_{-i}^1, \cdots, a_{-i}^T)$ are observed with probability at least $\mu(\omega)(\frac{\overline{\pi}}{|A|^T})^T \geq \frac{1}{|\Omega|}(\frac{\overline{\pi}}{|A|^T})^{4^{|\Omega|}}$. Let $\tilde{\mu}$ be the posterior belief in period $T+1$ in this case. From (iii), $\tilde{\mu}(\omega) \geq \frac{1}{|\Omega|}\overline{\pi}^{4^{|\Omega|}}$ for all $\omega \in \Omega^*$. From (ii), $\tilde{\mu}(\omega) = 0$ for other ω. *Q.E.D.*

Now we prove the proposition. Fix δ and i. In what follows, "robustly accessible" means "robustly accessible despite i," and "transient" means "transient given i."

For a given strategy s_{-i} and a prior $\tilde{\mu}$, let $v_i^{\tilde{\mu}}(s_{-i})$ denote player i's best possible payoff; that is, let $v_i^{\tilde{\mu}}(s_{-i}) = \max_{s_i \in S_i} v_i^{\tilde{\mu}}(\delta, s_i, s_{-i})$. Like Lemma C1, we can show that this payoff $v_i^{\tilde{\mu}}(s_{-i})$ is convex with respect to $\tilde{\mu}$.

Let s^μ denote the minimax strategy profile given the initial prior μ. Pick an arbitrary μ and pick the minimax strategy s_{-i}^μ. Then the payoff $v_i^{\tilde{\mu}}(s_{-i}^\mu)$ is convex with respect to $\tilde{\mu}$. In what follows, when we say *the convex curve* $v_i^{\tilde{\mu}}(s_{-i}^\mu)$ or *the convex curve induced by* s_{-i}^μ, it refers to the convex function $v_i^{\tilde{\mu}}(s_{-i}^\mu)$ whose domain is restricted to $\tilde{\mu} \in \triangle(\text{supp}\mu)$. So when $\text{supp}\mu = \Omega$, the convex curve represents player i's payoff $v_i^{\tilde{\mu}}(s_{-i}^\mu)$ for each initial prior $\tilde{\mu} \in \triangle\Omega$. On the other hand, when $\text{supp}\mu = \{\omega\}$, the convex curve is simply a scalar $v_i^\omega(s_{-i}^\mu)$. Note that

$v_i^{\tilde{\mu}}(s_{-i}^{\mu})$ denotes the minimax payoff when $\tilde{\mu} = \mu$, but when $\tilde{\mu} \neq \mu$, it is not the minimax payoff for any initial prior.

For each belief μ, let

$$\overline{v}_i(s_{-i}^{\mu}) = \max_{\tilde{\mu} \in \triangle(\text{supp}\mu)} v_i^{\tilde{\mu}}(s_{-i}^{\mu}),$$

that is, $\overline{v}_i(s_{-i}^{\mu})$ is the highest payoff achieved by the convex curve induced by s_{-i}^{μ}. Note that different initial priors μ induce different minimax strategies s_{-i}^{μ}, and hence different convex functions, and hence different highest payoffs $\overline{v}_i(s_{-i}^{\mu})$. Now, choose μ^* so that the corresponding highest payoff $\overline{v}_i(s_{-i}^{\mu^*})$ approximates the supremum of the highest payoffs over all beliefs μ; that is, choose μ^* such that

$$\left| \overline{v}_i(s_{-i}^{\mu^*}) - \sup_{\mu \in \triangle\Omega} \overline{v}_i(s_{-i}^{\mu}) \right| < 1 - \delta. \tag{18}$$

We call $\overline{v}_i(s_{-i}^{\mu^*})$ *the maximal value*, because it approximates $\sup_{\mu \in \triangle\Omega} \overline{v}_i(s_{-i}^{\mu})$, which is greater than any payoff achieved by any convex curves. The way we choose μ^* is essentially the same as in Section 5.5, but here we allow the possibility that $\sup_{\mu \in \triangle\Omega} \overline{v}_i(s_{-i}^{\mu})$ is actually the supremum, not the max.

Since $v_i^{\tilde{\mu}}(s_{-i}^{\mu^*})$ is convex, it is maximized when $\tilde{\mu}$ is an extreme point. Let $\omega \in \text{supp}\mu^*$ denote this extreme point, that is, $v_i^{\omega}(s_{-i}^{\mu^*}) \geq v_i^{\tilde{\mu}}(s_{-i}^{\mu^*})$ for all $\tilde{\mu} \in \triangle(\text{supp}\mu^*)$. In general, the maximal value $\overline{v}_i(s_{-i}^{\mu^*}) = v_i^{\omega}(s_{-i}^{\mu^*})$ is *not* the minimax payoff for any initial prior, because the state ω can be different from the belief μ^*.

The rest of the proof consists of four steps. In the first step, we show that given the opponents' strategy s_{-i}^{μ}, if the corresponding convex curve $v_i^{\tilde{\mu}}(s_{-i}^{\mu})$ approximates the maximal value for some relative interior belief $\tilde{\mu} \in \triangle(\text{supp}\mu)$, then the curve is almost flat and approximates the maximal value for *every* belief $\tilde{\mu} \in \triangle(\text{supp}\mu)$. This proof technique is very similar to the one for Step 1 in Section 5.2.

In the second step, we show that there is a belief μ^{**} whose minimax payoff approximates the maximal value. The proof idea is similar to Lemma C3 in the proof of Proposition 4, which extends Step 2 in Section 5.2. However, the argument here is more complicated, because the maximal value is not the minimax payoff for any belief. (In contrast, in the proof of Proposition 4, the maximal score is the score for some state ω.) To deal with this problem, we use the support merging condition (in particular Lemma C6).

In the third step, we show that for any belief μ whose support is robustly accessible, the corresponding minimax payoff approximates the maximal value. This part is new compared to the proof of Proposition 4, and the idea is as follows. Suppose not so that there is a belief μ such that its support is robustly accessible and the corresponding minimax payoff is much lower than the maximal value. Pick such μ, and let Ω^* denote its support. Then the result from the first step ensures that the convex curve $v_i^{\tilde{\mu}}(s_{-i}^{\mu})$ must be much lower than the maximal value uniformly in all $\tilde{\mu}$ with support Ω^*; otherwise, the result from the first step implies that the convex curve $v_i^{\tilde{\mu}}(s_{-i}^{\mu})$ approximates the maximal value for all $\tilde{\mu} \in \triangle\Omega^*$, and hence the minimax payoff for μ approximates the maximal value, which is a contradiction. Now, take μ^{**} as in the second step, and suppose that the initial prior is μ^{**}. Suppose also that players $-i$ play the following strategy $\tilde{s}_{-i}$: Mix all actions each period until the support of the posterior reaches Ω^*, and once it happens, play the minimax strategy s_{-i}^{μ} for the belief μ above. Suppose that player i takes a best reply. Then by the definition of s_{-i}^{μ}, player i's continuation payoff after the switch to s_{-i}^{μ} is much lower than the maximal value regardless of the belief $\tilde{\mu} \in \triangle\Omega^*$ at the time of the switch; this implies that her overall payoff $v_i^{\mu^{**}}(\tilde{s}_{-i})$ is also much lower than the maximal value, because the switch must happen with probability one from Lemma C7. This is a contradiction, because $\tilde{s}_{-i}$ is not the minimax strategy $s_{-i}^{\mu^{**}}$ and the payoff $v_i^{\mu^{**}}(\tilde{s}_{-i})$ must be at least the minimax payoff for μ^{**}, which approximates the maximal value.

In the last step, we show that for any belief μ whose support is transient, the minimax payoff for the belief μ approximates the maximal value. To see the idea, suppose that the initial prior is μ whose support is transient, and that the opponents play the minimax strategy s_{-i}^{μ}. Suppose that player i chooses the following strategy $\tilde{s}_i$: Mix all actions each period until the support of the posterior becomes robustly accessible, and once it happens, play a best reply in the continuation game. Then player i's payoff approximates the maximal value, because (since the game is robustly connected) the switch happens with probability one and the continuation payoff after the switch approximates the maximal value, as shown in the third step. This implies the result, because player i's minimax payoff is better than this payoff, and hence even closer to the maximal value.

C.8.1 Step 1: Almost Flat Convex Curve

The following lemma is a counterpart to Lemma C2, which bounds the convex curve $v_i^{\tilde{\mu}}(s_{-i}^{\mu})$ uniformly in the belief $\tilde{\mu}$.

Lemma C8. *Take two beliefs μ and $\tilde{\mu}$ with the same support. Let Ω^* denote this support, and let $p = \min_{\tilde{\omega}\in\Omega^*} \tilde{\mu}(\tilde{\omega})$, which measures the distance from $\tilde{\mu}$ to the boundary of $\triangle\Omega^*$. Then for each $\hat{\mu} \in \triangle\Omega^*$,*

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

To interpret this lemma, take an arbitrary μ and let Ω^* denote its support. Then the lemma ensures that if the convex curve $v_i^{\tilde{\mu}}(s_{-i}^{\mu})$ approximates the maximal value for some relative interior belief $\tilde{\mu} \in \triangle\Omega^*$, then the curve is almost flat and the payoff $v_i^{\hat{\mu}}(s_{-i}^{\mu})$ approximates the maximal value for any initial beliefs $\hat{\mu} \in \triangle\Omega^*$. For example, suppose that $\tilde{\mu}$ is the uniform distribution over Ω^* and satisfies

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

where ε is a positive number close to zero. Then $p = \frac{1}{|\Omega^*|}$ and hence the above lemma implies that

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

for all $\hat{\mu} \in \triangle\Omega^*$. Since ε is close to zero, this ensures that the payoff $v_i^{\hat{\mu}}(s_{-i}^{\mu})$ approximates the maximal value for *every* initial belief $\hat{\mu} \in \triangle\Omega^*$. Like Lemma C2, the belief $\tilde{\mu}$ above should not be too close to the boundary of $\triangle\Omega^*$; otherwise the parameter p is close to zero so that the right-hand side of the inequality becomes arbitrarily large.

The proof of the lemma is similar to that of Lemma C2.

Proof. Pick μ, $\tilde{\mu}$, Ω^*, and p as stated. Let s_i be player i's best reply against s_{-i}^{μ} given the initial prior $\tilde{\mu}$. Pick an arbitrary $\tilde{\omega} \in \Omega^*$. Note that

$$v_i^{\tilde{\mu}}(s_{-i}^{\mu}) = \sum_{\hat{\omega}\in\Omega^*} \tilde{\mu}(\hat{\omega}) v_i^{\hat{\omega}}(\delta, s_i, s_{-i}^{\mu}).$$

Then using $v_i^{\hat{\omega}}(\delta, s_i, s_{-i}^{\mu}) \leq \overline{v}_i(s_{-i}^{\mu^*}) + (1-\delta)$ for each $\hat{\omega} \neq \tilde{\omega}$, we obtain

$$v_i^{\tilde{\mu}}(s_{-i}^{\mu}) \leq \tilde{\mu}(\tilde{\omega}) v_i^{\tilde{\omega}}(\delta, s_i, s_{-i}^{\mu}) + (1-\tilde{\mu}(\tilde{\omega}))\{\overline{v}_i(s_{-i}^{\mu^*}) + (1-\delta)\}.$$

Arranging,

$$\tilde{\mu}(\tilde{\omega}) \left\{ \overline{v}_i(s_{-i}^{\mu^*}) + (1-\delta) - v_i^{\tilde{\omega}}(\delta, s_i, s_{-i}^{\mu}) \right\} \leq \overline{v}_i(s_{-i}^{\mu^*}) + (1-\delta) - v_i^{\tilde{\mu}}(s_{-i}^{\mu}).$$

Since the left-hand side is non-negative, taking the absolute values of both sides and dividing them by $\tilde{\mu}(\tilde{\omega})$,

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

Since $\tilde{\mu}[\tilde{\omega}] \geq p$, we have

$$\left| \overline{v}_i(s_{-i}^{\mu^*}) + (1-\delta) - v_i^{\tilde{\omega}}(\delta, s_i, s_{-i}^{\mu}) \right| \leq \frac{\left| \overline{v}_i(s_{-i}^{\mu^*}) + (1-\delta) - v_i^{\tilde{\mu}}(s_{-i}^{\mu}) \right|}{p}. \tag{19}$$

Now, pick an arbitrary $\hat{\mu} \in \triangle\Omega^*$. Note that (19) holds for each $\tilde{\omega} \in \Omega^*$. So multiply both sides of (19) by $\hat{\mu}(\tilde{\omega})$ and summing over all $\tilde{\omega} \in \Omega^*$,

$$\sum_{\tilde{\omega} \in \Omega^*} \hat{\mu}(\tilde{\omega}) \left| \overline{v}_i(s_{-i}^{\mu^*}) + (1-\delta) - v_i^{\tilde{\omega}}(\delta, s_i, s_{-i}^{\mu}) \right| \leq \frac{\left| \overline{v}_i(s_{-i}^{\mu^*}) + (1-\delta) - v_i^{\tilde{\mu}}(s_{-i}^{\mu}) \right|}{p}. \tag{20}$$

Then we have

$$\begin{aligned}
\left| \overline{v}_i(s_{-i}^{\mu^*}) + (1-\delta) - v_i^{\hat{\mu}}(s_{-i}^{\mu}) \right| &\leq \left| \overline{v}_i(s_{-i}^{\mu^*}) + (1-\delta) - v_i^{\hat{\mu}}(\delta, s_i, s_{-i}^{\mu}) \right| \\
&= \left| \sum_{\tilde{\omega} \in \Omega^*} \hat{\mu}(\tilde{\omega}) \left\{ \overline{v}_i(s_{-i}^{\mu^*}) + (1-\delta) - v_i^{\tilde{\omega}}(\delta, s_i, s_{-i}^{\mu}) \right\} \right| \\
&= \sum_{\tilde{\omega} \in \Omega^*} \hat{\mu}(\tilde{\omega}) \left| \overline{v}_i(s_{-i}^{\mu^*}) + (1-\delta) - v_i^{\tilde{\omega}}(\delta, s_i, s_{-i}^{\mu}) \right| \\
&\leq \frac{\left| \overline{v}_i(s_{-i}^{\mu^*}) + (1-\delta) - v_i^{\tilde{\mu}}(s_{-i}^{\mu}) \right|}{p}.
\end{aligned}$$

Here the first inequality follows from the fact that s_i is not a best reply given $\hat{\mu}$, and the last inequality follows from (20). *Q.E.D.*

C.8.2 Step 2: Existence of μ^{**}

For any history h^T with length T, let $\mu(h^T|\mu^*)$ be the posterior after history h^T conditional on the initial prior μ^*, and let $\mu(h^T|\omega)$ be the posterior conditional on the initial state ω. Let s_i^* be player i's best reply against $s_{-i}^{\mu^*}$ given the initial state ω. Since $\mu^*(\omega) > 0$, Lemma C6 ensures that there is a natural number $T \leq 4^{|\Omega|}$ and a history h^T such that $\Pr(h^T|\omega, s_i^*, s_{-i}^{\mu}) > (\frac{\overline{\pi}}{|A|})^T$ and such that the two posterior beliefs $\mu(h^T|\omega)$ and $\mu(h^T|\mu^*)$ have the same support. Pick such T and h^T.

Suppose that the initial state is ω, and that players play $(s_i^*, s_{-i}^{\mu^*})$. This implies that player i's payoff achieves the maximal value. Now, suppose that the play proceeds until the end of period T and the past history was h^T. By the definition of h^T, such a history indeed realizes with probability at least $(\frac{\overline{\pi}}{|A|})^T$. The following lemma shows that player i's continuation payoff $v_i^{\mu(h^T|\omega)}(s_{-i}^{\mu^*}|_{h^T})$ after this history approximates the maximal value. The proof is similar to Step 2 in Section 5.2.

Lemma C9. *We have*

$$\left| v_i^{\omega}(s_{-i}^{\mu^*}) + (1-\delta) - v_i^{\mu(h^T|\omega)}(s_{-i}^{\mu^*}|_{h^T}) \right| \leq \frac{(1-\delta^{4^{|\Omega|}})2\overline{g}|A|^{4^{|\Omega|}}}{\delta^{4^{|\Omega|}}\overline{\pi}^{4^{|\Omega|}}} + \frac{(1-\delta)|A|^{4^{|\Omega|}}}{\overline{\pi}^{4^{|\Omega|}}}.$$

Proof. Note that

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

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

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

Note that for each on-path history $\tilde{h}^T$, we have

$$v_i^{\mu(\tilde{h}^T|\omega)}(s_{-i}^{\mu^*}|_{\tilde{h}^T}) = v_i^{\mu(\tilde{h}^T|\omega)}(s_{-i}^{\mu(\tilde{h}^T|\mu^*)}) \leq \overline{v}_i(s_{-i}^{\mu(\tilde{h}^T|\mu^*)}) \leq v_i^{\omega}(s_{-i}^{\mu^*}) + (1-\delta). \quad (21)$$

Here the equality follows from the fact that the minimax strategy is Markov and $s_{-i}^{\mu^*}|_{\tilde{h}^T} = s_{-i}^{\mu(\tilde{h}^T|\mu^*)}$. The first inequality uses $\mu^*(\omega) > 0$, which ensures that the

support induced by ω and $\tilde{h}^T$ is a subset of the support induced by μ^* and $\tilde{h}^T$. The last inequality follows from (18). Plugging (21) into the above inequality,

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

Since (21) holds for $\tilde{h}^T = h^T$ and $\Pr(h^T|\omega, s_i^*, s_{-i}^{\mu^*}) \geq (\frac{\overline{\pi}}{|A|})^T$,

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

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

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

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

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

From (18), the left-hand side is positive. Thus taking the absolute value of the left-hand side and using $v_i^{\omega}(s_{-i}^{\mu^*}) \geq -\overline{g}$. we obtain

$$\left| v_i^{\omega}(s_{-i}^{\mu^*}) + (1-\delta) - v_i^{\mu(h^T|\omega)}(s_{-i}^{\mu^*}|_{h^T}) \right| \leq \frac{|A|^T(1-\delta^T)2\overline{g}}{\delta^T \overline{\pi}^T} + (1-\delta)\left(\frac{|A|}{\overline{\pi}}\right)^T.$$

Then the result follows because $T \leq 4^{|\Omega|}$. *Q.E.D.*

Let $\mu^{**} = \mu(h^T|\mu^*)$. Since the minimax strategy is Markov, $s_{-i}^{\mu^*}|_{h^T} = s_{-i}^{\mu^{**}}$ and hence the above lemma implies that

$$\left| v_i^{\omega}(s_{-i}^{\mu^*}) + (1-\delta) - v_i^{\mu(h^T|\omega)}(s_{-i}^{\mu^{**}}) \right| \leq \frac{(1-\delta^{4^{|\Omega|}})2\overline{g}|A|^{4^{|\Omega|}}}{\delta^{4^{|\Omega|}}\overline{\pi}^{4^{|\Omega|}}} + \frac{(1-\delta)|A|^{4^{|\Omega|}}}{\overline{\pi}^{4^{|\Omega|}}}.$$

That is, the convex curve $v_i^{\tilde{\mu}}(s_{-i}^{\mu^{**}})$ approximates the maximal score for some belief $\tilde{\mu} = \mu(h^T|\omega) \in \triangle(\text{supp}\mu^{**})$. (Here, $\mu(h^T|\omega) \in \triangle(\text{supp}\mu^{**})$ is ensured by Lemma C6.) Note that $\mu(h^T|\omega)[\tilde{\omega}] \geq \overline{\pi}^T \geq \overline{\pi}^{4^{|\Omega|}}$, because $\mu(h^T|\omega)$ is the posterior induced by the initial belief which puts probability one on ω. This implies that the distance from $\tilde{\mu} = \mu(h^T|\omega)$ to the boundary of $\triangle(\text{supp}\mu^{**})$ is at least $\overline{\pi}^{4^{|\Omega|}}$, and then Lemma C8 ensures that

$$\left| v_i^{\omega}(s_{-i}^{\mu^*}) + (1-\delta) - v_i^{\hat{\mu}}(s_{-i}^{\mu^{**}}) \right| \leq \frac{(1-\delta^{4^{|\Omega|}})2\overline{g}|A|^{4^{|\Omega|}}}{\delta^{4^{|\Omega|}}\overline{\pi}^{(4^{|\Omega|}+4^{|\Omega|})}} + \frac{(1-\delta)|A|^{4^{|\Omega|}}}{\overline{\pi}^{(4^{|\Omega|}+4^{|\Omega|})}}$$

for all $\hat{\mu} \in \triangle(\text{supp}\mu^{**})$. That is, the convex curve induced by $s_{-i}^{\mu^{**}}$ is almost flat and approximates the maximal score for all beliefs $\hat{\mu} \in \triangle(\text{supp}\mu^{**})$. In particular, by letting $\hat{\mu} = \mu^{**}$, we have

$$\left| v_i^{\omega}(s_{-i}^{\mu^*}) + (1-\delta) - v_i^{\mu^{**}}(s_{-i}^{\mu^{**}}) \right| \leq \frac{(1-\delta^{4^{|\Omega|}})2\overline{g}|A|^{4^{|\Omega|}}}{\delta^{4^{|\Omega|}}\overline{\pi}^{(4^{|\Omega|}+4^{|\Omega|})}} + \frac{(1-\delta)|A|^{4^{|\Omega|}}}{\overline{\pi}^{(4^{|\Omega|}+4^{|\Omega|})}}, \quad (22)$$

that is, the minimax payoff for the belief μ^{**} approximates the maximal value.

C.8.3 Step 3: Minimax Payoffs when the Support is Robustly Accessible

In this step, we show that for any belief μ whose support is robustly accessible, the minimax payoff for μ approximates the maximal value.

For a given belief μ, let Δ^{μ} denote the set of beliefs $\tilde{\mu} \in \triangle(\text{supp}\mu)$ such that $\tilde{\mu}(\tilde{\omega}) \geq \frac{1}{|\Omega|}\overline{\pi}^{4^{|\Omega|}}$ for all $\tilde{\omega} \in \text{supp}\mu$. Intuitively, Δ^{μ} is the set of all beliefs $\tilde{\mu}$ with the same support as μ, except the ones which are too close to the boundary of $\triangle(\text{supp}\mu)$.

Now, assume that the initial prior is μ^{**}. Pick a belief μ whose support is robustly accessible, and suppose that the opponents play the following strategy $\tilde{s}_{-i}^{\mu}$:

- The opponents mix all actions equally likely each period, until the posterior belief becomes an element of Δ^{μ}.
- If the posterior belief becomes an element of Δ^{μ} in some period, then they play the minimax strategy s_{-i}^{μ} in the rest of the game. (They do not change the play after that.)

Intuitively, the opponents wait until the belief reaches Δ^μ, and once it happens, they switch the play to the minimax strategy s^μ_{-i}. From Lemma C7, the switch actually happens in finite time with probability one regardless of player i's play, so player i's payoff against the above strategy is approximated by the expected continuation payoff after the switch. Since the belief $\tilde{\mu}$ at the time of the switch is always in the set Δ^μ, this continuation payoff is at most

$$K_i^\mu = \max_{\tilde{\mu} \in \Delta^\mu} v_i^{\tilde{\mu}}(s^\mu_{-i}).$$

Hence player i's payoff against the above strategy $\tilde{s}^\mu_{-i}$ is approximately at most K_i^μ. Formally, we have the following lemma:

Lemma C10. *For each belief μ whose support is robustly accessible,*

$$v_i^{\mu^{**}}(\tilde{s}^\mu_{-i}) \leq K_i^\mu + \frac{(1-\delta^{4^{|\Omega|}})2\overline{g}}{\pi^*}.$$

Proof. The proof is very similar to that of Lemma C5. Pick a belief μ whose support is robustly accessible. Suppose that the initial prior is μ^{**}, the opponents play $\tilde{s}^\mu_{-i}$, and player i plays a best reply. Let ρ^t denote the probability that players $-i$ still randomize actions in period t. Then as in the proof of Lemma C5, we have

$$v_i^{\mu^{**}}(\tilde{s}^\mu_{-i}) \leq \sum_{t=1}^{\infty} \delta^{t-1} \left\{ \rho^t \overline{g} + (1-\rho^t) K_i^\mu \right\},$$

because the stage-game payoff before the switch to s^μ_{-i} is bounded from above by $\overline{g}$, and the continuation payoff after the switch is bounded from above by $K_i^\mu = \max_{\tilde{\mu} \in \Delta^\mu} v_i^{\tilde{\mu}}(s^\mu_{-i})$.

As in the proof of Lemma C5, we have

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

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

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

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

$$v_i^{\mu^{**}}(\tilde{s}^\mu_{-i}) \leq (1-\delta) \sum_{n=1}^{\infty} \sum_{k=1}^{4^{|\Omega|}} \delta^{(n-1)4^{|\Omega|}+k-1} \left[\begin{array}{l} (1-\pi^*)^{n-1}\overline{g} \\ +\{1-(1-\pi^*)^{n-1}\} K_i^\mu \end{array} \right].$$

Then as in the proof of Lemma C5, the standard algebra shows

$$v_i^{\mu^{**}}(\tilde{s}_{-i}^{\mu}) \leq \frac{(1-\delta^{4^{|\Omega|}})\overline{g}}{1-(1-\pi^*)\delta^{4^{|\Omega|}}} + \frac{\delta^{4^{|\Omega|}}\pi^* K_i^{\mu}}{1-(1-\pi^*)\delta^{4^{|\Omega|}}}.$$

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

$$v_i^{\mu^{**}}(\tilde{s}_{-i}^{\mu}) \leq K_i^{\mu} + \frac{(1-\delta^{4^{|\Omega|}})(\overline{g}-K_i^{\mu})}{1-(1-\pi^*)\delta^{4^{|\Omega|}}}.$$

Since $1-(1-\pi^*)\delta^{4^{|\Omega|}} > 1-(1-\pi^*) = \pi^*$ and $K_i^{\mu} \geq -\overline{g}$, the result follows. *Q.E.D.*

For now, ignore the term $\frac{(1-\delta^{4^{|\Omega|}})2\overline{g}}{\pi^*}$, as it approximates zero when $\delta \to 1$. Then the above lemma shows that the payoff K_i^{μ} is at least $v_i^{\mu^{**}}(\tilde{s}_{-i}^{\mu})$, which must be at least the minimax payoff $v_i^{\mu^{**}}(s_{-i}^{\mu^{**}})$ due to the fact that $\tilde{s}_{-i}^{\mu}$ is not necessarily the minimax strategy. On the other hand, the payoff K_i^{μ} cannot exceed the maximal value. Hence the payoff K_i^{μ} is between the minimax payoff $v_i^{\mu^{**}}(s_{-i}^{\mu^{**}})$ and the maximal value. Now, from Step 2, we already know that these two bounds are close each other; hence the payoff $K_i^{\mu} = \max_{\tilde{\mu}\in\Delta^{\mu}} v_i^{\tilde{\mu}}(s_{-i}^{\mu})$ approximates the maximal value. That is, the convex curve induced by s_{-i}^{μ} approximates the maximal value for some belief $\tilde{\mu} \in \Delta^{\mu}$. Then Lemma C8 ensures that the convex curve is almost flat and that the payoff $v_i^{\hat{\mu}}(s_{-i}^{\mu})$ approximates the maximal value for all beliefs $\tilde{\mu} \in \triangle(\mathrm{supp}\mu)$. When $\hat{\mu} = \mu$, it ensures that the minimax payoff $v_i^{\mu}(s_{-i}^{\mu})$ approximates the maximal value. Formally, we have the following lemma.

Lemma C11. *For each belief μ whose support is robustly accessible,*

$$\begin{aligned}&\left|\overline{v}_i(s_{-i}^{\mu^*}) + (1-\delta) - v_i^{\mu}(s_{-i}^{\mu})\right| \\ &\leq \frac{(1-\delta^{4^{|\Omega|}})2\overline{g}|\Omega|}{\pi^*\overline{\pi}^{4^{|\Omega|}}} + \frac{(1-\delta^{4^{|\Omega|}})2\overline{g}|A|^{4^{|\Omega|}}|\Omega|}{\delta^{4^{|\Omega|}}\overline{\pi}^{(4^{|\Omega|}+4^{|\Omega|}+4^{|\Omega|})}} + \frac{(1-\delta)|A|^{4^{|\Omega|}}|\Omega|}{\overline{\pi}^{(4^{|\Omega|}+4^{|\Omega|}+4^{|\Omega|})}}.\end{aligned}$$

Proof. Pick an arbitrary belief μ whose support is robustly accessible. From Lemma C10 and the fact that $v_i^{\mu^{**}}(\tilde{s}_{-i}^{\mu})$ is at least the minimax payoff $v_i^{\mu^{**}}(s_{-i}^{\mu^{**}})$, we have

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

This, together with (22), implies that

$$\overline{v}_i(s_{-i}^{\mu^*}) + (1-\delta) \leq K_i^\mu + \frac{(1-\delta^{4^{|\Omega|}})2\overline{g}}{\pi^*} + \frac{(1-\delta^{4^{|\Omega|}})2\overline{g}|A|^{4^{|\Omega|}}}{\delta^{4^{|\Omega|}}\overline{\pi}^{(4^{|\Omega|}+4^{|\Omega|})}} + \frac{(1-\delta)|A|^{4^{|\Omega|}}}{\overline{\pi}^{(4^{|\Omega|}+4^{|\Omega|})}}.$$

Since $K_i^\mu \leq \overline{v}_i(s_{-i}^{\mu^*}) + (1-\delta)$,

$$\left|\overline{v}_i(s_{-i}^{\mu^*}) + (1-\delta) - K_i^\mu\right| \leq \frac{(1-\delta^{4^{|\Omega|}})2\overline{g}}{\pi^*} + \frac{(1-\delta^{4^{|\Omega|}})2\overline{g}|A|^{4^{|\Omega|}}}{\delta^{4^{|\Omega|}}\overline{\pi}^{(4^{|\Omega|}+4^{|\Omega|})}} + \frac{(1-\delta)|A|^{4^{|\Omega|}}}{\overline{\pi}^{(4^{|\Omega|}+4^{|\Omega|})}}.$$

Then by the definition of K_i^μ, there is $\tilde{\mu} \in \Delta^\mu$ such that

$$\left|\overline{v}_i(s_{-i}^{\mu^*}) + (1-\delta) - v_i^{\tilde{\mu}}(s_{-i}^{\mu})\right| \leq \frac{(1-\delta^{4^{|\Omega|}})2\overline{g}}{\pi^*} + \frac{(1-\delta^{4^{|\Omega|}})2\overline{g}|A|^{4^{|\Omega|}}}{\delta^{4^{|\Omega|}}\overline{\pi}^{(4^{|\Omega|}+4^{|\Omega|})}} + \frac{(1-\delta)|A|^{4^{|\Omega|}}}{\overline{\pi}^{(4^{|\Omega|}+4^{|\Omega|})}}.$$

Then using Lemma C8 and $\tilde{\mu}(\tilde{\omega}) \geq \frac{1}{|\Omega|}\overline{\pi}^{4^{|\Omega|}}$ for each $\tilde{\omega} \in \text{supp}\mu$,

$$\begin{aligned}
&\left|\overline{v}_i(s_{-i}^{\mu^*}) + (1-\delta) - v_i^{\tilde{\mu}}(s_{-i}^{\mu})\right| \\
&\leq \frac{(1-\delta^{4^{|\Omega|}})2\overline{g}|\Omega|}{\pi^*\overline{\pi}^{4^{|\Omega|}}} + \frac{(1-\delta^{4^{|\Omega|}})2\overline{g}|A|^{4^{|\Omega|}}|\Omega|}{\delta^{4^{|\Omega|}}\overline{\pi}^{(4^{|\Omega|}+4^{|\Omega|}+4^{|\Omega|})}} + \frac{(1-\delta)|A|^{4^{|\Omega|}}|\Omega|}{\overline{\pi}^{(4^{|\Omega|}+4^{|\Omega|}+4^{|\Omega|})}}.
\end{aligned}$$

for all $\tilde{\mu} \in \triangle(\text{supp}\mu)$. By setting $\tilde{\mu} = \mu$, we obtain the result. *Q.E.D.*

C.8.4 Step 4: Minimax Payoffs when the Support is Transient

The previous step shows that the minimax payoff for any belief whose support is robustly accessible approximates the maximal value. Now we show that the minimax payoff for any belief whose support is transient approximates the maximal value.

So pick an arbitrary belief μ whose support is transient. Suppose that the initial prior is μ and the opponents use the minimax strategy s_{-i}^μ. Suppose that player i plays the following strategy $\tilde{s}_i^\mu$:

- Player i mixes all actions equally likely each period, until the support of the posterior belief becomes robustly accessible.
- If the support of the posterior belief becomes robustly accessible, then play a best reply in the rest of the game.

Intuitively, player i waits until the support of the posterior belief becomes robustly accessible, and once it happens, she plays a best reply to the opponents' continuation strategy $s_{-i}^{\mu^t}$, where μ^t is the belief when the switch happens. (Here the opponents' continuation strategy is the minimax strategy $s_{-i}^{\mu^t}$, since the strategy s_{-i}^{μ} is Markov and induces the minimax strategy in every continuation game.) Note that player i's continuation payoff after the switch is exactly equal to the minimax payoff $v_i^{\mu^t}(s_{-i}^{\mu^t})$. Lemma C11 ensures that this continuation payoff approximates the maximal value, regardless of the belief μ^t at the time of the switch. Then since the switch must happen in finite time with probability one, player i's payoff by playing the above strategy $\tilde{s}_i^{\mu}$ also approximates the maximal value. Precisely, we have the following lemma:

Lemma C12. *For any μ whose support is transient,*

$$\begin{aligned}&\left|\overline{v}_i(s_{-i}^{\mu^*})+(1-\delta)-v_i^{\mu}(\delta,\tilde{s}_i^{\mu},s_{-i}^{\mu})\right| \\ &\leq \frac{(1-\delta^{4^{|\Omega|}})4\overline{g}|\Omega|}{\pi^*\overline{\pi}^{4^{|\Omega|}}}+\frac{(1-\delta^{4^{|\Omega|}})2\overline{g}|A|^{4^{|\Omega|}}|\Omega|}{\delta^{4^{|\Omega|}}\overline{\pi}^{(4^{|\Omega|}+4^{|\Omega|}+4^{|\Omega|})}}+\frac{(1-\delta)|A|^{4^{|\Omega|}}|\Omega|}{\overline{\pi}^{(4^{|\Omega|}+4^{|\Omega|}+4^{|\Omega|})}}.\end{aligned}$$

Proof. The proof is very similar to that of Lemma C10 and hence omitted. *Q.E.D.*

Note that the strategy $\tilde{s}_i^{\mu}$ is not a best reply against s_{-i}^{μ}, and hence we have

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

Then from the lemma above, we can conclude that the minimax payoff for any belief μ whose support is transient approximates the maximal payoff, as desired.

C.9 Proof of Proposition A2

In this proof, we show only the existence of the limit of the minimax payoff. Independence of the minimax payoff follows from Proposition A1.

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

$$\left|\underline{v}_i^{\mu}(\overline{\delta}) - \liminf_{\delta\to 1}\underline{v}_i^{\mu}(\delta)\right| < \frac{\varepsilon}{2} \tag{23}$$

and such that

$$\left|\underline{v}_i^{\mu}(\overline{\delta}) - \underline{v}_i^{\tilde{\mu}}(\overline{\delta})\right| < \frac{\varepsilon}{2} \tag{24}$$

for each $\tilde{\mu}$. Note that Proposition A1 guarantees that such $\overline{\delta}$ exists.

For each $\tilde{\mu}$, let $s_{-i}^{\tilde{\mu}}$ be the minimax strategy given $\tilde{\mu}$ and $\overline{\delta}$. In what follows, we show that

$$\max_{s_i \in S_i} v_i^{\mu}(\delta, s_i, s_{-i}^{\mu}) < \liminf_{\delta \to 1} \underline{v}_i^{\mu}(\delta) + \varepsilon \tag{25}$$

for each $\delta \in (\overline{\delta}, 1)$. That is, we show that when the true discount factor is δ, player i's best payoff against the minimax strategy for the discount factor $\overline{\delta}$ is worse than the limit inferior of the minimax payoff. Since the minimax strategy for the discount factor $\overline{\delta}$ is not necessarily the minimax strategy for δ, the minimax payoff for δ is less than $\max_{s_i \in S_i} v_i^{\mu}(\delta, s_i, s_{-i}^{\mu})$. Hence (25) ensures that the minimax payoff for δ is worse than the limit inferior of the minimax payoff. Since this is true for all $\delta \in (\overline{\delta}, 1)$, the limit inferior is the limit, as desired.

So pick an arbitrary $\delta \in (\overline{\delta}, 1)$, and compute $\max_{s_i \in S_i} v_i^{\mu}(\delta, s_i, s_{-i}^{\mu})$, player i's best payoff against the minimax strategy for the discount factor $\overline{\delta}$. To evaluate this payoff, we regard the infinite horizon as a series of random blocks, as in Section 6. The termination probability is $1-p$, where $p = \frac{\overline{\delta}}{\delta}$. Then, since s_{-i}^{μ} is Markov, playing s_{-i}^{μ} in the infinite-horizon game is the same as playing the following strategy profile:

- During the first random block, play s_{-i}^{μ}.
- During the kth random block, play $s_{-i}^{\mu^k}$ where μ^k is the belief in the initial period of the kth block.

Then as in the proof of Proposition 5, the payoff $\max_{s_i \in S_i} v_i^{\mu}(\delta, s_i, s_{-i}^{\mu})$ is represented as the sum of the random block payoffs, that is,

$$\max_{s_i \in S_i} v_i^{\mu}(\delta, s_i, s_{-i}^{\mu}) = (1-\delta) \sum_{k=1}^{\infty} \left(\frac{\delta(1-p)}{1-p\delta} \right)^{k-1} E\left[\frac{v_i^{\mu^k}(p\delta, s_i^{\mu^k}, s_{-i}^{\mu^k})}{1-p\delta} \middle| \mu, s_i^{\mu^1}, s_{-i}^{\mu} \right]$$

where $s_i^{\mu^k}$ is the optimal (Markov) strategy in the continuation game from the kth block with belief μ^k. Note that $s_i^{\mu^k}$ may not maximize the payoff during the kth block, because player i needs to take into account the fact that her action during the kth block influences μ^{k+1} and hence the payoffs after the kth block. But in

any case, we have $v_i^{\mu^k}(p\delta, s_i^{\mu^k}, s_{-i}^{\mu^k}) \leq \underline{v}_i^{\mu^k}(\overline{\delta})$ because $s_{-i}^{\mu^k}$ is the minimax strategy with discount factor $p\delta = \overline{\delta}$. Hence

$$\max_{s_i \in S_i} v_i^{\mu}(\delta, s_i, s_{-i}^{\mu}) \leq (1-\delta) \sum_{k=1}^{\infty} \left(\frac{\delta(1-p)}{1-p\delta} \right)^{k-1} E\left[\frac{\underline{v}_i^{\mu^k}(\overline{\delta})}{1-p\delta} \middle| \mu, s_i^{\mu^1}, s_{-i}^{\mu} \right]$$

Using (24),

$$\begin{aligned} \max_{s_i \in S_i} v_i^{\mu}(\delta, s_i, s_{-i}^{\mu}) &< (1-\delta) \sum_{k=1}^{\infty} \left(\frac{\delta(1-p)}{1-p\delta} \right)^{k-1} \left(\frac{\underline{v}_i^{\mu}(\overline{\delta})}{1-p\delta} + \frac{\varepsilon}{2(1-p\delta)} \right) \\ &= \underline{v}_i^{\mu}(\overline{\delta}) + \frac{\varepsilon}{2} \end{aligned}$$

Then using (23), we obtain (25).

Like the proof of Proposition 5, this proof does not assume public randomization. Indeed, random blocks are useful for computing the payoff by the strategy s_{-i}^{μ}, but the strategy s_{-i}^{μ} itself does not use public randomization.

C.10 Proof of Proposition A4

The proof technique is quite similar to that of Proposition A1, so here we present only the outline of the proof. Fix δ and i. Let $v_i^{\mu}(s_{-i})$ denote player i's best payoff against s_{-i} conditional on the initial prior μ, just as in the proof of Proposition A1. Let $\overline{v}_i$ be the supremum of the minimax payoffs $\underline{v}_i^{\mu}(\delta)$ over all μ. In what follows, we call it the *maximal value* and show that the minimax payoff for any belief μ approximates the maximal value. Pick μ^* so that the minimax payoff $\underline{v}_i^{\mu^*}(\delta)$ for this belief μ^* approximates the maximal value.

Let $\mu(\omega, a)$ denote the posterior belief given that in the last period, the hidden state was ω and players chose a. Pick an arbitrary robustly accessible state ω. Suppose that the initial prior is μ^* and that the opponents use the following strategy $\tilde{s}_{-i}^{\omega}$:

- Mix all actions a_{-i} equally, until they observe $y = \omega$.
- Once it happens (say in period t), then from the next period $t+1$, they play the minimax strategy $s_{-i}^{\mu^{t+1}} = s_{-i}^{\mu(\omega, a^t)}$.

That is, the opponents wait until the signal y reveals that the state today was ω, and once it happens, play the minimax strategy in the rest of the game. Suppose that player i takes a best reply. Since ω is robustly accessible, the switch happens in finite time with probability one, and thus player i's payoff is approximately her expected continuation payoff after the switch. Since the opponents mix all actions until the switch occurs, her expected continuation payoff is at most

$$K_i^\omega = \max_{a_i \in A_i} \sum_{a_{-i} \in A_{-i}} \frac{1}{|A_{-i}|} \underline{v}_i^{\mu(\omega,a)}(\delta).$$

Hence her overall payoff $v_i^{\mu^*}(\tilde{s}_{-i}^\omega)$ is approximately at most K_i^ω; the formal proof is very similar to that of Lemma C10 and hence omitted.

Now, since $\tilde{s}_{-i}^\omega$ is not the minimax strategy $s_{-i}^{\mu^*}$, player i's payoff $v_i^{\mu^*}(\tilde{s}_{-i}^\omega)$ must be at least the minimax payoff $\underline{v}_i^{\mu^*}(\delta)$, which is approximated by $\bar{v}_i$. Hence the above result ensures that K_i^ω is approximately at least $\bar{v}_i$. On the other hand, by the definition, we have $K_i^\omega \leq \bar{v}_i$. Taken together, K_i^ω must approximate the maximal value $\bar{v}_i$.

Let a_i^ω be the maximizer which achieves K_i^ω. Recall that in the definition of K_i^ω, we take the expected value with respect to a_{-i} assuming that a_{-i} is uniformly distributed over A_{-i}. We have shown that this expected value K_i^ω approximates the maximal value $\bar{v}_i$. Now we claim that the same result holds even if we do not take the expectation with respect to a_{-i}, that is, $\underline{v}_i^{\mu(\omega,a_i^\omega,a_{-i})}(\delta)$ approximates the maximal value $\bar{v}_i$ regardless of a_{-i}. The proof technique is quite similar to Step 2 in Section 5.2 and hence omitted. Note that the result so far is true for all robustly accessible states ω. So $\underline{v}_i^{\mu(\omega,a_i^\omega,a_{-i})}(\delta)$ approximates the maximal value $\bar{v}_i$ for any a_{-i} and any globally accessible state ω.

Now we show that the minimax payoff for any belief μ approximates the maximal value. Pick an arbitrary belief μ, and suppose that the opponents play the minimax strategy s_{-i}^μ. Suppose that player i plays the following strategy s_i:

- Mix all actions a_i equally, until there is some globally accessible state ω and time t such that $a_i^t = a_i^\omega$ and $y^t = \omega$.
- Once it happens, then from the next period $t+1$, she plays a best reply.

Since each state ω is robustly accessible or transient, the switch happens in finite time with probability one. Also, player i's continuation payoff after the switch

is $\underline{v}_i^{\mu(\omega,a_i^\omega,a_{-i})}(\delta)$ for some a_{-i} and some robustly accessible ω, which approximates the maximal value. Hence player i's overall payoff by s_i approximates the maximal value, which ensures that the minimax payoff approximates the maximal value.

C.11 Proof of Proposition B1

We begin with a preliminary lemma: It shows that for each initial state ω and pure strategy profile s, there is a pure strategy s^* such that if the initial state is ω and players play s^*, the support which arises at any on-path history is the one which arises in the first $2^{|\Omega|}+1$ periods when players played s. Let $\Omega(\omega,h^t)$ denote the support of the posterior given the initial state ω and the history h^t.

Lemma C13. *For each state ω and each pure strategy profile s, there is a pure strategy profile s^* such that for any history h^t with $\Pr(h^t|\omega,s^*)>0$, there is a natural number $\tilde{t}\leq 2^{|\Omega|}$ and $\tilde{h}^{\tilde{t}}$ such that $\Pr(\tilde{h}^{\tilde{t}}|\omega,s)>0$ and $\Omega(\omega,h^t)=\Omega(\omega,\tilde{h}^{\tilde{t}})$.*

Proof. Pick ω and s as stated. We focus on s^* such that players' action today depends only on the current support, that is, $s^*(h^t)=s^*(\tilde{h}^{\tilde{t}})$ if $\Omega(\omega,h^t)=\Omega(\omega,\tilde{h}^{\tilde{t}})$. So we denote the action given the support Ω^* by $s^*(\Omega^*)$. For each support Ω^*, let h^t be the earliest on-path history with $\Omega(\omega,h^t)=\Omega^*$ when players play s. That is, choose h^t such that $\Pr(h^t|\omega,s)>0$, $\Omega(\omega,h^t)=\Omega^*$, and $\Omega(\omega,\tilde{h}^{\tilde{t}})\neq\Omega^*$ for all $\tilde{h}^{\tilde{t}}$ with $\tilde{t}<t$. (When such h^t does not exist, let $h^t=h^0$.) Then set $s^*(\Omega^*)=s(h^t)$. It is easy to check that this strategy profile s^* satisfies the desired property. *Q.E.D.*

Now we prove Proposition B1. Pick an arbitrary singleton set $\{\omega\}$ which is not asymptotically accessible. It is sufficient to show that this set $\{\omega\}$ is asymptotically uniformly transient. (Like Proposition 1, we can show that a superset of an asymptotically accessible set is asymptotically accessible, and a superset of an asymptotically uniformly transient set is asymptotically accessible or asymptotically uniformly transient.) In particular, it is sufficient to show that if the initial state is ω, given any pure strategy profile, the support reaches an asymptotically accessible set within $2^{|\Omega|}+1$ periods.

So pick an arbitrary pure strategy profile s. Choose s^* as in the above lemma. Let $\mathscr{O}$ be the set of supports Ω^* which arise with positive probability when the initial state is ω and players play s^*. In what follows, we show that there is an

asymptotically accessible support $\Omega^* \in \mathscr{O}$; this implies that $\{\omega\}$ is asymptotically uniformly transient, because such a support Ω^* realizes with positive probability within $2^{|\Omega|}+1$ periods when the initial state is ω and players play s.

If $\Omega \in \mathscr{O}$, then the result immediately holds by setting $\Omega^* = \Omega$. So in what follows, we assume $\Omega \notin \mathscr{O}$. We prove the existence of an asymptotically accessible set $\Omega^* \in \mathscr{O}$ in two steps. In the first step, we show that there is $q > 0$ and $\tilde{\Omega}^* \in \mathscr{O}$ such that given any initial prior μ, players can move the belief to the one which puts probability at least q on the set $\tilde{\Omega}^*$. Then in the second step, we show that from such a belief (i.e., a belief which puts probability at least q on Ω^*), players can move the belief to the one which puts probability at least $1-\varepsilon$ on some $\Omega^* \in \mathscr{O}$. Taken together, it turns out that for any initial prior μ, players can move the belief to the one which puts probability at least $1-\varepsilon$ on the set $\Omega^* \in \mathscr{O}$, which implies asymptotic accessibility of Ω^*.

The following lemma corresponds to the first step of the proof. It shows that from any initial belief, players can move the belief to the one which puts probability at least q on the set $\tilde{\Omega}^*$.

Lemma C14. *There is $q > 0$ and a set $\tilde{\Omega}^* \in \mathscr{O}$ such that for each initial prior μ, there is a natural number $T \leq |\Omega|$, an action sequence $(a^1, \cdots, a^T)$, and a history h^T such that $\Pr(h^T | \mu, a^1, \cdots, a^T) \geq \frac{\overline{\pi}^{|\Omega|}}{|\Omega|}$ and $\sum_{\tilde{\omega} \in \Omega^*} \tilde{\mu}(\tilde{\omega}) \geq q$, where $\tilde{\mu}$ is the posterior given the initial prior μ and the history h^T.*

Proof. We first show that there is $\tilde{\Omega}^* \in \mathscr{O}$ which contains at least one globally accessible state $\tilde{\omega}$. Suppose not so that all states in any set $\Omega^* \in \mathscr{O}$ are uniformly transient. Suppose that the initial state is ω^* and players play s^*. Then the support of the posterior is always an element of $\mathscr{O}$, and thus in each period t, regardless of the past history h^t, the posterior puts probability zero on any globally accessible state ω. This is a contradiction, because the standard argument shows that the probability of the state in period t being uniformly transient converges to zero as $t \to \infty$.

So there is $\tilde{\Omega}^* \in \mathscr{O}$ which contains at least one globally accessible state $\tilde{\omega}$. Pick such $\tilde{\Omega}^*$ and $\tilde{\omega}$. Global accessibility of $\tilde{\omega}$ ensures that for each initial state $\hat{\omega} \in \Omega$, there is a natural number $T \leq |\Omega|$, an action sequence $(a^1, \cdots, a^T)$, and a signal sequence $(y^1, \cdots, y^T)$ such that

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

That is, if the initial state is $\hat{\omega}$ and players play $(a^1,\cdots,a^T)$, then the state in period $T+1$ can be in the set Ω^* with positive probability. For each $\hat{\omega}$, choose such $(a^1,\cdots,a^T)$ and $(y^1,\cdots,y^T)$, and let

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

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 show that this q and the set $\tilde{\Omega}^*$ above satisfy 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, $(a^1,\cdots,a^T)$, and $(y^1,\cdots,y^T)$ as stated above. Let $\tilde{\mu}$ be the posterior belief after $(a^1,\cdots,a^T)$ and $(y^1,\cdots,y^T)$ given the initial prior μ. Then

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

This implies that the posterior $\tilde{\mu}$ puts probability at least q on $\tilde{\Omega}^*$, since $\tilde{\omega}\in\tilde{\Omega}^*$. Also, the above belief $\tilde{\mu}$ realizes with probability

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

as desired. *Q.E.D.*

Choose $\tilde{\Omega}^*\in\mathscr{O}$ as in the above lemma. Let $\tilde{s}^*$ be the continuation strategy of s^* given that the current support is $\tilde{\Omega}^*$, that is, let $\tilde{s}^*=s^*|_{h^t}$ where h^t is chosen such that $\Pr(h^t|\omega^*,s^*)>0$ and $\Omega(\omega^*,h^t)=\tilde{\Omega}^*$. (If such h^t is not unique, pick one arbitrarily.) By the definition, if the initial support is $\tilde{\Omega}^*$ and players play $\tilde{s}^*$, the posterior is an element of $\mathscr{O}$ after every history.

The following lemma corresponds to the second step of the proof. It shows that if the initial prior puts probability at least q on the set $\tilde{\Omega}^*$ and players play $\tilde{s}^*$, then with some probability π^{**}, players learn the support from the realized signals and the posterior puts $1-\varepsilon$ on some set $\Omega^*\in\mathscr{O}$.

Lemma C15. *For each $\varepsilon>0$ and $q>0$, there is a natural number T, a set $\Omega^*\in\mathscr{O}$, and $\pi^{**}>0$ such that for each initial prior μ with $\sum_{\tilde{\omega}\in\tilde{\Omega}^*}\mu(\tilde{\omega})\geq q$, there is a history h^T such that $\Pr(h^T|\mu,\tilde{s}^*)>\pi^{**}$ and the posterior $\tilde{\mu}$ given the initial prior μ and the history h^T satisfies $\sum_{\tilde{\omega}\in\Omega^*}\tilde{\mu}(\tilde{\omega})\geq 1-\varepsilon$.*

Proof. Recall that $\Omega \notin \mathscr{O}$, so any $\Omega^* \in \mathscr{O}$ is a proper subset of Ω. By the assumption, given any $\Omega^* \in \mathscr{O}$ and a, the convex hull of $\{\pi_Y^\omega(a)|\omega \in \Omega^*\}$ and that of $\{\pi_Y^\omega(a)|\omega \notin \Omega^*\}$ do not intersect. Let $\kappa(\Omega^*, a) > 0$ be the distance between these two convex hulls, i.e.,

$$\left\| \pi_Y^{\overline{\mu}}(a) - \pi_Y^{\underline{\mu}}(a) \right\| \geq \kappa(\Omega^*, a)$$

for each $\overline{\mu} \in \triangle\tilde{\Omega}^*$ and $\underline{\mu} \in \triangle(\Omega \setminus \tilde{\Omega}^*)$. (Here $\|\cdot\|$ denotes the sup norm.) Let $\kappa > 0$ be the minimum of $\kappa(\Omega^*, a)$ over all $\Omega^* \in \mathscr{O}$ and $a \in A$.

Pick an initial prior μ as stated, that is, μ puts probability at least q on $\tilde{\Omega}^*$. Let $\Omega^1 = \tilde{\Omega}^*$, and let $\overline{\mu}$ be the marginal distribution on Ω^1, that is, $\overline{\mu}(\tilde{\omega}) = \frac{\mu(\tilde{\omega})}{\sum_{\hat{\omega} \in \Omega^1} \mu(\hat{\omega})}$ for each $\tilde{\omega} \in \Omega^1$ and $\overline{\mu}(\tilde{\omega}) = 0$ for other $\tilde{\omega}$. Likewise, let $\underline{\mu}$ be the marginal distribution on $\Omega \setminus \Omega^1$, that is, $\underline{\mu}(\tilde{\omega}) = \frac{\mu(\tilde{\omega})}{\sum_{\hat{\omega} \notin \Omega^1} \mu(\hat{\omega})}$ for each $\tilde{\omega} \notin \Omega^1$ and $\underline{\mu}(\tilde{\omega}) = 0$ for other $\tilde{\omega}$. Let a denote the action profile chosen in period one by $\tilde{s}^*$. Then by the definition of κ, there is a signal y such that

$$\pi_Y^{\overline{\mu}}(y|a) \geq \pi_Y^{\underline{\mu}}(y|a) + \kappa. \tag{26}$$

Intuitively, (26) implies that the signal y is more likely if the initial state is in the set Ω^1. Hence the posterior belief must put higher weight on the event that the initial state was in Ω^1. To be more precise, let μ^2 be the posterior belief in period two given the initial prior μ, the action profile a, and the signal y. Also, let Ω^2 be the support of the posterior in period two given the same history but the initial prior was $\overline{\mu}$ rather than μ. Intuitively, the state in period two must be in Ω^2 if the initial state was in Ω^1. Then we have $\sum_{\tilde{\omega} \in \Omega^2} \mu^2(\tilde{\omega}) > \sum_{\tilde{\omega} \in \Omega^1} \mu(\tilde{\omega})$ because the signal y indicates that the initial state was in Ω^1.

Formally, this result can be verified as follows. By the definition, if the initial state is in the set $\tilde{\Omega}^*$ and players play a and observe y, then the state in period two must be in the set Ω^2. That is, we must have

$$\pi^{\tilde{\omega}}(y, \hat{\omega}|a) = 0 \tag{27}$$

for all $\tilde{\omega} \in \Omega^1$ and $\hat{\omega} \notin \Omega^2$. Then we have

$$\begin{aligned}
\frac{\sum_{\tilde{\omega}\in\Omega^2}\mu^2(\tilde{\omega})}{\sum_{\tilde{\omega}\notin\Omega^2}\mu^2(\tilde{\omega})} &= \frac{\sum_{\tilde{\omega}\in\Omega}\sum_{\hat{\omega}\in\Omega^2}\mu(\tilde{\omega})\pi^{\tilde{\omega}}(y,\hat{\omega}|a)}{\sum_{\tilde{\omega}\in\Omega}\sum_{\hat{\omega}\notin\Omega^2}\mu(\tilde{\omega})\pi^{\tilde{\omega}}(y,\hat{\omega}|a)} \\
&= \frac{\sum_{\tilde{\omega}\in\Omega}\sum_{\hat{\omega}\in\Omega}\mu(\tilde{\omega})\pi^{\tilde{\omega}}(y,\hat{\omega}|a)}{\sum_{\tilde{\omega}\notin\Omega^1}\sum_{\hat{\omega}\notin\Omega^2}\mu(\tilde{\omega})\pi^{\tilde{\omega}}(y,\hat{\omega}|a)} \\
&\geq \frac{\sum_{\tilde{\omega}\in\Omega^1}\sum_{\hat{\omega}\in\Omega}\mu(\tilde{\omega})\pi^{\tilde{\omega}}(y,\hat{\omega}|a)}{\sum_{\tilde{\omega}\notin\Omega^1}\sum_{\hat{\omega}\in\Omega}\mu(\tilde{\omega})\pi^{\tilde{\omega}}(y,\hat{\omega}|a)} \\
&= \frac{\pi_Y^{\overline{\mu}}(y|a)\sum_{\tilde{\omega}\in\Omega^1}\mu(\tilde{\omega})}{\pi_Y^{\underline{\mu}}(y|a)\sum_{\tilde{\omega}\notin\Omega^1}\mu(\tilde{\omega})} \\
&\geq \frac{1}{1-\kappa}\cdot\frac{\sum_{\tilde{\omega}\in\Omega^1}\mu(\tilde{\omega})}{\sum_{\tilde{\omega}\notin\Omega^1}\mu(\tilde{\omega})}.
\end{aligned}$$

Here, the second equality comes from (27), and the last inequality from (26). Since $\frac{1}{1-\kappa} > 1$, this implies that the likelihood of Ω^2 induced by the posterior belief μ^2 is greater than the likelihood of Ω^1 induced by the initial prior μ, as desired. Note also that such a posterior belief μ^2 realizes with probability at least $q\kappa$, since (26) implies

$$\pi_Y^{\mu}(y|a) \geq q\pi_Y^{\overline{\mu}}(y|a) \geq q\kappa.$$

We apply a similar argument to the posterior belief in period three: Assume that period one is over and the outcome is as above, so the belief in period two is μ^2. Let $\overline{\mu}^2$ be the marginal distribution of μ^2 on Ω^2, and let $\underline{\mu}^2$ be the marginal distribution on $\Omega \setminus \Omega^2$. Let a^2 be the action profile chosen in period two by $\tilde{s}^*$ after the signal y in period one. Then choose a signal y^2 so that $\pi_Y^{\overline{\mu}^2}(y^2|a^2) \geq \pi_Y^{\underline{\mu}^2}(y^2|a^2) + \kappa$, and let μ^3 be the posterior belief in period three after observing y^2 in period two. Then as above, we can show that

$$\frac{\sum_{\tilde{\omega}\in\Omega^3}\mu^3(\tilde{\omega})}{\sum_{\tilde{\omega}\notin\Omega^3}\mu^3(\tilde{\omega})} \geq \frac{1}{1-\kappa}\cdot\frac{\sum_{\tilde{\omega}\in\Omega^2}\mu^2(\tilde{\omega})}{\sum_{\tilde{\omega}\notin\Omega^2}\mu^2(\tilde{\omega})} \geq \left(\frac{1}{1-\kappa}\right)^2\frac{\sum_{\tilde{\omega}\in\Omega}\mu(\tilde{\omega})}{\sum_{\tilde{\omega}\notin\Omega}\mu(\tilde{\omega})}$$

where Ω^3 is the support of the posterior if the initial support was Ω^1 and players play $\tilde{s}^*$ and observe the signal y and then y^2. The probability of this signal is again at least $q\kappa$.

Iterating this argument, we can prove that for any natural number T, there is a signal sequence $(y^1, \cdots, y^T)$ and a set Ω^{T+1} such that if players play the profile

$\tilde{s}^*$, the signal sequence realizes with probability at least $\pi^{**} = (q\kappa)^T$, and the posterior belief μ^{T+1} satisfies

$$\frac{\sum_{\tilde{\omega}\in\Omega^{T+1}}\mu^{T+1}(\tilde{\omega})}{\sum_{\tilde{\omega}\notin\Omega^{T+1}}\mu^{T+1}(\tilde{\omega})} \geq \left(\frac{1}{1-\kappa}\right)^T \cdot \frac{\sum_{\tilde{\omega}\in\Omega^1}\mu(\tilde{\omega})}{\sum_{\tilde{\omega}\notin\Omega^1}\mu(\tilde{\omega})} \geq \left(\frac{1}{1-\kappa}\right)^T \frac{q}{1-q}.$$

Note that the set Ω^{T+1} is an element of $\mathscr{O}$, by the construction.

Now, choose $\varepsilon > 0$ and $q > 0$ arbitrarily, and then pick T large enough that $(\frac{1}{1-\kappa})^T \frac{q}{1-q} \geq \frac{1-\varepsilon}{\varepsilon}$. Then the above posterior belief μ^{T+1} puts probability at least $1-\varepsilon$ on $\Omega^{T+1} \in \mathscr{O}$. So by letting $\Omega^* = \Omega^{T+1}$, the result holds. *Q.E.D.*

Fix $\varepsilon > 0$ arbitrarily. Choose q and $\tilde{\Omega}^*$ as stated in Lemma C14, and then choose Ω^*, T, and π^{**} as stated in Lemma C15. Then the above two lemmas ensure that given any initial prior μ, there is an action sequence with length $T^* \leq |\Omega| + T$ 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 $|\Omega|+T$ and π^{**} do not depend on the initial prior μ, this shows that Ω^* is asymptotically accessible. Then $\{\omega\}$ is asymptotically uniformly transient, as $\Omega^* \in \mathscr{O}$.

C.12 Proof of Proposition B2

Fix δ and λ. Let s^μ and ω be as in the proof of Proposition 4. We begin with two preliminary lemmas. The first lemma shows that the score is Lipschitz continuous with respect to μ.

Lemma C16. *For any $\varepsilon \in (0, \frac{1}{|\Omega|})$, μ, and $\tilde{\mu}$ with $|\mu(\tilde{\omega}) - \tilde{\mu}(\tilde{\omega})| \leq \varepsilon$ for each $\tilde{\omega} \in \Omega$,*

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

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

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

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

The second preliminary lemma is a counterpart to Lemma C4; it shows that the action sequence in the definition of asymptotic accessibility can be replaced with fully mixed actions. The proof is similar to that of Lemma C4 and hence omitted.

Lemma C17. *Suppose that players randomize all actions equally each period. Then for any $\varepsilon > 0$, there is a natural number T and $\pi^* > 0$ such that given any initial prior μ and any asymptotically accessible set Ω^*, there is a natural number $T^* \leq T$ and $\tilde{\mu}$ such that the probability of $\mu^{T^*+1} = \tilde{\mu}$ is at least π^*, and such that $\sum_{\omega \in \Omega^*} \tilde{\mu}(\omega) \geq 1 - \varepsilon$.*

Since there are only finitely many subsets $\Omega^* \subset \Omega$, there is $\tilde{\pi}^* > 0$ such that for each asymptotically uniformly transient Ω^*, $\tilde{\pi}^*$ satisfies the condition stated in the definition of asymptotic uniform transience. Pick such $\tilde{\pi}^* > 0$. Pick $\varepsilon \in (0, \frac{1}{|\Omega|})$ arbitrarily. Then choose a natural number T and $\pi^* > 0$ as in Lemma C17.

For each set Ω^*, let $\triangle\Omega^*(\varepsilon)$ denote the set of beliefs μ such that $\sum_{\tilde{\omega} \in \Omega^*} \mu(\tilde{\omega}) \geq 1 - \varepsilon$.

As in the proof of Proposition 4, the rest of the proof consists of three steps. The first step is to show that for each set Ω^*, if there is a relative interior belief $\mu \in \triangle\Omega^*$ whose score approximates the maximal score, then the score for *every* belief with support Ω^* approximates the maximal score. This result is already proved as Lemma C2 in the proof of Proposition 4 (note that Lemma C2 does not rely on uniform connectedness), and we do not state it here in order to avoid redundancy.

In the second step, we show that there is an asymptotically accessible set Ω^* such that the score for any belief $\mu \in \triangle\Omega^*(\varepsilon)$ approximates the maximal score. The proof idea is similar to the second step in the proof of Proposition 4.

Then in the last step, we show that the score approximates the maximal score for any belief μ. Once again the proof idea is similar to the one in the proof of Proposition 4.

C.12.1 Step 2: Bound on the Scores for All Beliefs in $\Omega^*(\varepsilon)$

In this step, we prove the following lemma, which shows that there is an asymptotically accessible set Ω^* such that the score for any belief $\mu \in \triangle\Omega^*(\varepsilon)$ approximates the maximal score.

Lemma C18. *There is an asymptotically accessible set Ω^* such that for any $\mu \in \triangle\Omega^*$,*

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

Then from Lemma C16, there is an asymptotically accessible set Ω^* such that for any $\mu \in \triangle\Omega^*(\varepsilon)$,

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

Proof. Since the game is asymptotically uniformly connected, $\{\omega\}$ is either asymptotically accessible or asymptotically uniformly transient. We first consider the case in which it is asymptotically accessible. Let $\Omega^* = \{\omega\}$. Then this Ω^* satisfies the desired property, as it contains only the belief $\mu = \omega$, and the score for this belief is exactly equal to the maximal score.

Next, consider the case in which $\{\omega\}$ is asymptotically uniformly transient. In this case, there is an asymptotically accessible set Ω^*, a natural number $T^* \leq T$, 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 resulting posterior belief μ^* satisfies $\sum_{\tilde{\omega} \in \Omega^*} \mu^*[\tilde{\omega}] \geq 1 - \varepsilon$ and $\mu^*[\tilde{\omega}] \geq \tilde{\pi}^*$ for all $\tilde{\omega} \in \Omega^*$. Take such Ω^*, T^*, and $(y^1, \cdots, y^{T^*})$. Then as in the proof of Lemma C3, we can prove that

$$\left|\lambda \cdot v^{\omega}(\delta, s^{\omega}) - \lambda \cdot v^{\mu^*}(\delta, s^{\mu^*})\right| \leq \frac{(1-\delta^T)2\overline{g}}{\delta^T \overline{\pi}^T}. \tag{28}$$

That is, the score with the initial prior μ^* is close to the maximal score. The only difference from Lemma C3 is to replace $2^{|\Omega|}$ with T.

Since $\sum_{\tilde{\omega} \in \Omega^*} \mu^*[\tilde{\omega}] \geq 1 - \varepsilon$ and $\mu^*[\tilde{\omega}] \geq \tilde{\pi}^*$ for all $\tilde{\omega} \in \Omega^*$, there is a belief $\tilde{\mu}^*$ whose support is Ω^* such that $\tilde{\mu}^*[\tilde{\omega}] \geq \tilde{\pi}^*$ for all $\tilde{\omega} \in \Omega^*$, and such that $\tilde{\mu}^*$ is ε-close to μ^* in that $\max_{\tilde{\omega} \in \Omega} |\mu^*(\tilde{\omega}) - \tilde{\mu}^*(\tilde{\omega})| \leq \varepsilon$. Lemma C16 implies that these two beliefs μ^* and $\tilde{\mu}^*$ induce similar scores, that is,

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

Plugging this into (28), we obtain

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

That is, the score for the belief $\tilde{\mu}^*$ approximates the maximal score. Then using Lemma C2, we can get the desired inequality. *Q.E.D.*

C.12.2 Step 3: Bound on the Score for All Beliefs

Here we show that for any belief μ, the score approximates the maximal score. To do so, for each initial belief μ, consider the following strategy profile $\tilde{s}^\mu$:

- Players randomize all actions equally likely, until the posterior belief becomes an element of $\triangle\Omega^*(\varepsilon)$.
- Once the posterior belief becomes an element of $\triangle\Omega^*(\varepsilon)$ in some period t, then players play s^{μ^t} in the rest of the game. They do not change the play after that.

Intuitively, players randomize all actions and wait until the belief reaches $\triangle\Omega^*(\varepsilon)$; and once it happens, they switch the play to the optimal policy s^{μ^t} in the continuation game. Lemma C18 guarantees that the continuation play after the switch to s^{μ^t} approximates the maximal score $\lambda \cdot v^\omega(\delta, s^\omega)$. Also, Lemma C17 ensures that the waiting time until this switch occurs is finite with probability one. Hence for δ close to one, the strategy profile $\tilde{s}^\mu$ approximates the maximal score when the initial prior is μ. Formally, we have the following lemma.

Lemma C19. *For each μ,*

$$|\lambda \cdot v^\omega(\delta, s^\omega) - \lambda \cdot v^\mu(\delta, \tilde{s}^\mu)| \leq \frac{(1-\delta^T)2\overline{g}}{\delta^T \underline{\pi}^T \tilde{\pi}^*} + \frac{(1-\delta^T)3\overline{g}}{\pi^*} + \frac{2\varepsilon\overline{g}|\Omega|}{\tilde{\pi}^*}.$$

Proof. The proof is essentially the same as that of Lemma C5; we simply replace $4^{|\Omega|}$ in the proof of Lemma C5 with T, and use Lemma C18 instead of Lemma C3. *Q.E.D.*

Note that

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

that is, the score for μ is at least $\lambda \cdot v^\mu(\delta, \tilde{s}^\mu)$ and is at most the maximal score. Then from Lemma C19,

$$\begin{aligned} |\lambda \cdot v^\omega(\delta, s^\omega) - \lambda \cdot v^\mu(\delta, s^\mu)| &\leq |\lambda \cdot v^\omega(\delta, s^\omega) - \lambda \cdot v^\mu(\delta, \tilde{s}^\mu)| \\ &\leq \frac{(1-\delta^T)2\overline{g}}{\delta^T \underline{\pi}^T \tilde{\pi}^*} + \frac{(1-\delta^T)3\overline{g}}{\pi^*} + \frac{2\varepsilon\overline{g}|\Omega|}{\tilde{\pi}^*}. \end{aligned}$$

Recall that T and π^* depend on ε but not on δ or λ. Note also that $\tilde{\pi}^*$ does not depend on ε, δ, or 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 the lemma.

Appendix D: Uniform Connectedness in Terms of Primitives

In Section 3.2, we have provided the definition of uniform connectedness. We give an alternative definition of uniform connectedness, and some technical results. We begin with global accessibility.

Definition D1. 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:[20]

(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 $t \leq T$.

As the following proposition shows, the definition of globally accessibility here is indeed equivalent to the one stated using beliefs.

Proposition D1. *Definitions 2 and D1 are equivalent.*

[20] As argued, 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 D1. 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 D1. We can repeat this procedure to show the existence of sequences with length $T \leq 4^{|\Omega|}$ which satisfy (i) and (ii).

Proof. We first show that global accessibility in Definition D1 implies the one in Definition 2. Take a set Ω^* which is globally accessible in the sense of Definition D1, 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 Definition D1. Suppose that the initial prior is μ and players play $(a^1, \cdots, a^T)$. Then clause (i) of Definition D1 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 $\tilde{\omega} \notin \Omega^*$.[21] Note that the probability of this signal sequence $(y^1, \cdots, y^T)$ is at least

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

where $\Pr(y^1, \cdots, y^T | \omega, a^1, \cdots, a^T)$ denotes the probability of the signal sequence $(y^1, \cdots, y^T)$ given the initial state ω and the action sequence $(a^1, \cdots, a^T)$. This implies that global accessibility in Definition D1 implies the one in Definition 2, by letting $\pi^* \in (0, \frac{1}{|\Omega|} \overline{\pi}^{4^{|\Omega|}})$.

Next, we show that the converse is true. Let Ω^* be a globally accessible set in the sense of Definition 2. Pick $\pi^* > 0$ as stated in Definition 2, 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

$$\Pr(\mu^{T+1} = \tilde{\mu} | \mu, a^1, \cdots, a^T) \geq \pi^*. \tag{29}$$

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 (29) implies that

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

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

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 D1. Also it satisfies clause (ii) in Definition D1, 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. *Q.E.D.*

Next, we give the definition of uniform transience in terms of primitives. With an abuse of notation, for each pure strategy profile s, 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 D2. A singleton set $\{\omega\}$ is *uniformly transient* if it is not globally accessible and for any pure strategy profile s, there is a globally accessible set Ω^*, 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 $t\leq T$.[22]

[22]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 as in the proof of Lemma C13, we can construct a strategy profile s^* such that if the initial prior is μ and players play s^*, the support of the posterior belief never reaches a globally accessible set.

In words, $\{\omega\}$ is uniformly 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.[23] It is obvious that the definition of uniform transience above is equivalent to Definition 3, except that here we consider only singleton sets $\{\omega\}$.

Now we are ready to give the definition of uniform connectedness:

Definition D3. A stochastic game is *uniformly connected* if each singleton set $\{\omega\}$ is globally accessible or uniformly transient.

In this definition, we consider only singleton sets $\{\omega\}$. However, as shown by Proposition 1, if each singleton set $\{\omega\}$ is globally accessible or uniformly transient, then any subset $\Omega^* \subseteq \Omega$ is globally accessible or uniformly transient. Hence the above definition is equivalent to the one stated using beliefs.

Before we conclude this appendix, we present two propositions, which hopefully help our understanding of uniformly transient sets. The first proposition shows that if the game is uniformly connected, then the probability of the support moving from a uniformly transient set to a globally accessible set is bounded away from zero uniformly in the current belief. (The proposition considers a special class of uniformly transient sets; it considers a uniformly transient set Ω^* such that any non-empty subset of Ω^* is also uniformly transient. However, this is a mild restriction, and when the game is uniformly connected, any uniformly transient set Ω^* satisfies this condition. Indeed, uniform connectedness ensures that any subset of a uniformly transient set Ω^* is globally accessible or uniformly transient, and Proposition 1 guarantees that they are all uniformly transient.)

Proposition D2. *Let Ω^* be a uniformly transient set such that any non-empty subset of Ω^* is also uniformly transient. Then there is $\pi^* > 0$ such that for any initial prior μ with support Ω^* and for any pure strategy profile s, there is a natural number $T \leq 2^{|\Omega|}$ and a belief $\tilde{\mu}$ whose support is globally accessible such that* $\Pr(\mu^{T+1} = \tilde{\mu} | \mu, s) > \pi^*$.

[23] While we consider an arbitrary strategy profile s in the definition of uniform transience, in order to check whether a set $\{\omega\}$ is uniformly 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 uniform transience of each set $\{\omega\}$ can be done in finite steps.

Proof. Pick Ω^* and μ as stated. Pick an arbitrary pure strategy profile s. It is sufficient to show that given the initial prior μ and the profile s, the support of the posterior belief will reach a globally accessible set with probability at least $\pi^* = \frac{\overline{\pi}^{2^{|\Omega|}}}{|\Omega|}$.

Take a state ω such that $\mu(\omega) \geq \frac{1}{|\Omega|}$. By the definition of Ω^*, the singleton set $\{\omega\}$ is uniformly transient.

Consider the case in which the initial prior puts probability one on ω, and players play s. Since $\{\omega\}$ is uniformly transient, there is a natural number $T \leq 2^{|\Omega|}$ and a history h^T such that the history h^T appears with positive probability and the support of the posterior belief after this history h^T is globally accessible. Take such a history h^T, and let $\tilde{\Omega}^*$ be the support of the posterior belief. Note that this history appears with probability at least $\overline{\pi}^T$ given the initial state ω and the profile s.

Now, consider the case in which the initial prior is μ (rather than the known state ω) and players play s. Still the history h^T occurs with positive probability, because μ puts positive probability on ω. Note that its probability is at least $\mu(\omega)\overline{\pi}^T \geq \frac{\overline{\pi}^{2^{|\Omega|}}}{|\Omega|} = \pi^*$. Note also that the support after the history h^T is globally accessible, because it is a superset of the globally accessible set $\tilde{\Omega}^*$. Hence if the initial prior is μ and players play s, the support of the posterior belief will reach a globally accessible set with probability at least π^*, as desired. *Q.E.D.*

The next proposition shows that if the support of the current belief is uniformly transient, then the support cannot return to the current one forever with positive probability.[24] This in turn implies that the probability of the support being uniformly transient in period T is approximately zero when T is large enough. So when we think about the long-run evolution of the support, the time during which the support stays at uniformly transient sets is almost negligible. Let $X(\Omega^*|\mu, s)$

[24]Here is an example in which the support moves from a globally accessible set to a uniformly transient set. Suppose that there are two states, ω_1 and ω_2, and that the state ω_2 is absorbing. Specifically, the next state is ω_2 with probability $\frac{1}{2}$ if the current state is ω_1, while the state tomorrow is ω_2 for sure if the current state is ω_1. There are three signals, y_1, y_2, and y_3, and the signal is correlated with the state tomorrow. If the state tomorrow is ω_1, the signals y_1 and y_3 realize with probability $\frac{1}{2}$ each. Likewise, If the state tomorrow is ω_2, the signals y_2 and y_3 realize with probability $\frac{1}{2}$ each. So y_1 and y_2 reveal the state tomorrow. It is easy to check that $\{\omega_2\}$ and Ω are globally accessible, and $\{\omega_1\}$ is uniformly transient. If the current belief is $\mu = (\frac{1}{2}, \frac{1}{2})$, then with positive probability, the current signal reveals that the state tomorrow is ω_1, so the support of the posterior belief moves to the uniformly transient set $\{\omega_1\}$.

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.

Proposition D3. *Let Ω^* be a uniformly transient set such that any non-empty subset of Ω^* is also uniformly transient. Then there is $\pi^* > 0$ such that for any initial prior μ whose support is Ω^*, and any pure strategy profile s,*

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

Proof. Suppose not so that for any $\varepsilon > 0$, there is a pure strategy profile s and a belief μ whose support is Ω^* such that $\Pr(X(\Omega^*|\mu,s) < \infty) \geq 1 - \varepsilon$.

Pick $\varepsilon > 0$ small so that $\overline{\pi}^{2^{|\Omega|}} > \frac{\varepsilon|\Omega|}{\overline{\pi}^{2^{|\Omega|}}}$, and choose s and μ as stated above. Choose $\omega \in \Omega^*$ such that $\mu(\omega) \geq \frac{1}{|\Omega|}$. Suppose that the initial state is ω and players play s. Let $X^*(\Omega^*|\omega,s)$ be the random variable which represents the first time in which the support of the posterior belief is Ω^* or its subset. Since $\Pr(X(\Omega^*|\mu,s) < \infty) \geq 1 - \varepsilon$, we must have

$$\Pr(X^*(\Omega^*|\omega,s) < \infty) \geq 1 - \frac{\varepsilon}{\mu(\omega)} \geq 1 - \varepsilon|\Omega|.$$

That is, given the initial state ω and the strategy profile s, the support must reach Ω^* or its subset in finite time with probability close to one.

By the definition of Ω^*, the singleton set $\{\omega\}$ is uniformly transient. So there is $T \leq 2^{|\Omega|}$ and $\tilde{\mu}$ whose support is globally accessible such that $\Pr(\mu^{T+1} = \tilde{\mu}|\omega,s) > 0$. Pick such a posterior belief $\tilde{\mu}$ and let $\tilde{s}$ be the continuation strategy after that history. Let $\tilde{\Omega}^*$ denote the support of $\tilde{\mu}$. Since $\tilde{\mu}$ is the posterior induced from the initial state ω, we have $\Pr(\mu^{T+1} = \tilde{\mu}|\omega,s) \geq \overline{\pi}^{2^{|\Omega|}}$ and $\tilde{\mu}(\tilde{\omega}) \geq \overline{\pi}^{2^{|\Omega|}}$ for all $\tilde{\omega} \in \tilde{\Omega}^*$.

Since $\Pr(\mu^{T+1} = \tilde{\mu}|\omega,s) \geq \overline{\pi}^{2^{|\Omega|}}$ and $\Pr(X^*(\Omega^*|\omega,s) < \infty) \geq 1 - \varepsilon|\Omega|$, we must have

$$\Pr(X^*(\Omega^*|\tilde{\mu},\tilde{s}) < \infty) \geq 1 - \frac{\varepsilon|\Omega|}{\overline{\pi}^{2^{|\Omega|}}}.$$

That is, given the initial belief $\tilde{\mu}$ and the strategy profile $\tilde{s}$, the support must reach Ω^* or its subset in finite time with probability close to one. Then since $\tilde{\mu}(\tilde{\omega}) \geq \overline{\pi}^{2^{|\Omega|}} > \frac{\varepsilon|\Omega|}{\overline{\pi}^{2^{|\Omega|}}}$ for each $\tilde{\omega} \in \tilde{\Omega}^*$, we can show that for each state $\tilde{\omega} \in \tilde{\Omega}^*$, 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 $\tilde{\omega}$ and players play $(a^1, \cdots, a^T)$, then the sequence $(y^1, \cdots, y^T)$ realizes with positive probability.

(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^*, for any initial state $\hat{\omega} \in \tilde{\Omega}^*$ (possibly $\hat{\omega} \neq \tilde{\omega}$).

This result implies that for any initial belief $\hat{\mu} \in \triangle\tilde{\Omega}^*$ players can move the support to Ω^* or its subset with positive probability, and this probability is bounded away from zero uniformly in $\hat{\mu}$; the proof is very similar to that of Proposition D1 and hence omitted. This and global accessibility of $\tilde{\Omega}^*$ imply that Ω^* is globally accessible, which is a contradiction. *Q.E.D.*

Appendix E: Existence of Maximizers

Lemma E1. *For each initial prior μ, discount factor δ, and s_{-i}, player i's best reply s_i exists.*

Proof. The formal proof is as follows. Pick μ, δ, and s_{-i}. 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 $(Tf)(\mu) \leq (T\tilde{f})(\mu)$ for all μ if $f(\mu) \leq \tilde{f}(\mu)$ for all μ. Moreover T is discounting, because letting $(f+c)(\mu) = f(\mu) + c$, the standard argument shows that $T(f+c)(\mu) \leq (Tf)(\mu) + \delta c$ 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}. *Q.E.D.*

Lemma E2. $\max_{v \in V^\mu(\delta)} \lambda \cdot v$ *has a solution.*

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

Lemma E3. *There is* s_{-i} *which solves* $\min_{s_{-i} \in S_{-i}} \max_{s_i \in S_i} v_i^\mu(\delta, s)$.

Proof. The formal proof is as follows. Pick μ and δ, and let h^t and l^∞ be as in the proof of Lemma E1. For each function $f \in l^\infty$, let Tf be a function such that

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

where $\tilde{\mu}(h^t)$ is the posterior belief of 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, 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 the proof of Lemma E1. Then from Blackwell's theorem, the operator T is a contraction mapping and thus has a unique fixed point f^*. The corresponding action sequence is the minimizer s_{-i}. *Q.E.D.*

Appendix F: Hsu, Chuang, and Arapostathis (2006)

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

Suppose that there is one player, two states (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 uniform connectedness in this paper.

References

Arellano, C. (2008): "Default Risk and Income Fluctuations in Emerging Economies," *American Economic Review* 98, 690-712.

Athey, S., and K. Bagwell (2008): "Collusion with Persistent Cost Shocks," *Econometrica* 76, 493-540.

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

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

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

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

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

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

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

Escobar, J., and J. Toikka (2013) "Efficiency in Games with Markovian Private Information," *Econometrica* 81, 1887-1934.

Fudenberg, D., D.K. Levine, and J. Tirole (1985): "Infinite-Horizon Models of Bargaining with One-Sided Incomplete Information," in A.E. Roth (Ed.), *Game-Theoretic Models of Bargaining*. Cambridge University Press, Cambridge, UK.

Fudenberg, D., and E. Maskin (1986): "The Folk Theorem for Repeated Games With Discounting or With Incomplete Information," *Econometrica* 54, 533-554.

Fudenberg, D., and Y. Yamamoto (2010): "Repeated Games where the Payoffs and Monitoring Structure are Unknown," *Econometrica* 78, 1673-1710.

Fudenberg, D., and Y. Yamamoto (2011a): "Learning from Private Information in Noisy Repeated Games," *Journal of Economic Theory* 146, 1733-1769.

Fudenberg, D., and Y. Yamamoto (2011b): "The Folk Theorem for Irreducible Stochastic Games with Imperfect Public Monitoring," *Journal of Economic Theory* 146, 1664-1683.

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

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

Hörner, J., T. Sugaya, S. Takahashi, and N. Vieille (2011): "Recursive Methods in Discounted Stochastic Games: an Algorithm for $\delta \to 1$ and a Folk Theorem," *Econometrica* 79, 1277-1318.

Hörner, J., S. Takahashi, and N. Vieille (2011): "Recursive Methods in Discounted Stochastic Games II: Infinite State Space," *mimeo*.

Hörner, J., S. Takahashi, and N. Vieille (2015): "Truthful Equilibria in Dynamic Bayesian Games," *Econometrica* 83, 1795-1848.

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.

Mailath, G.J., and L. Samuelson (2006): *Repeated Games and Reputations: Long-Run Relationships*. Oxford University Press, New York, NY.

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.