

PIER

PENN INSTITUTE *for* ECONOMIC RESEARCH
UNIVERSITY *of* PENNSYLVANIA

The Ronald O. Perelman Center for
Political Science and Economics (PCPSE)
133 South 36th Street
Philadelphia, PA 19104-6297

pier@econ.upenn.edu

<http://economics.sas.upenn.edu/pier>

PIER Working Paper

18-028

Stochastic Games with Hidden States

YUICHI YAMAMOTO
University of Pennsylvania

First Draft: March 29, 2014
This Version: May 19, 2018

<https://ssrn.com/abstract=2563612>

Stochastic Games with Hidden States*

Yuichi Yamamoto[†]

First Draft: March 29, 2014

This Version: May 19, 2018

Abstract

This paper studies infinite-horizon stochastic games in which players observe actions and noisy public information about a hidden state each period. We find a general condition under which the feasible and individually rational payoff set is invariant to the initial prior about the state, when players are patient. 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 uses 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.

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

*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

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. The goal of this paper is to develop some tools which are useful to analyze such a situation.

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 update 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, all players have the same posterior belief μ^t about the current state ω^t . Hence 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. To avoid this complication, various papers in the literature (e.g., Dutta (1995), Fudenberg and Yamamoto (2011b), Hörner, Sugaya, Takahashi, and Vieille (2011)) consider a model which satisfies the *payoff invariance condition*, in the sense that when players are patient, the feasible and individually rational payoff set is invariant to the initial state. In such a model, 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 continuation payoff can be chosen in a flexible way, just as in the standard repeated game. This property helps to discipline players' intertemporal incentives, and the folk theorem can be obtained in general.

We first show that the same result holds even in the infinite-state stochastic game. That is, we prove that the folk theorem holds as long as the payoff invariance condition holds so that the feasible and individually rational payoff set is invariant to the initial prior μ for patient players. The proof is similar to the one in Dutta (1995), but we use the idea of *random blocks* in order to avoid some technical complication coming from infinite states.²

So the remaining question is when this payoff invariance condition holds. For the finite-state case, Dutta (1995) shows that the limit feasible payoff set is indeed invariant if states are *communicating* in that players can move the state from any state to any other state. To see how this condition works, pick an extreme point of the feasible payoff set (say, the welfare-maximizing point). This payoff must be attained by a Markov strategy, so call it the optimal Markov strategy. The communicating states assumption ensures that regardless of the current state, players

¹For 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, 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.

²Interestingly, 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.

can move the state to the one in which they obtain a high payoff; so in the optimal Markov strategy, patient players always attempt to move the state to the one which yields the highest payoff. Using this property, one can show that the state transition induced by the optimal Markov strategy is *ergodic* so 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.

On the other hand, when states are infinite, the communicating states assumption are never satisfied. Indeed, given an initial state, only finitely many states can be reached in finite time, so almost all states are not reachable. So in general players may not be able to move the state to the one which yields a high payoff, and this makes our analysis quite different from the finite-state case. More technically, while there are some sufficient conditions for ergodicity of infinite-state Markov chains (e.g. *Doebelin condition*, see Doob (1953)), these conditions are not satisfied in our setup.³

Despite such complications, we find that under the *full support assumption*, the belief evolution process has a sort of ergodicity, and accordingly the payoff invariance condition holds. The full support assumption requires that regardless of the current state and the current action profile, any signal can be observed and any state can occur tomorrow, with positive probability. Under this assumption, the support of the posterior belief is always the whole state space, i.e., the posterior belief assigns positive probability to every state ω . It turns out that this property is useful to obtain the invariance result.

The proof of invariance of the feasible payoffs is not new, and it directly follows from the theory of partially observable Markov decision process (POMDP). In our model, the feasible payoffs can be computed by solving a Bellman equation in which the state variable is a belief. Such a Bellman equation is known as a POMDP problem, and Platzman (1980) shows that under the full support assumption, a solution to a POMDP problem is invariant to the initial belief. This immediately implies invariance of the feasible payoff set.

On the other hand, we need a new proof technique to obtain invariance of the

³This is essentially because our model is a multi-player version of the partially observable Markov decision process (POMDP). The introduction of Rosenberg, Solan, and Vieille (2002) explains why the POMDP model is intractable.

minimax payoff. The minimax payoff is *not* a solution to a Bellman equation (and hence it is not a POMDP solution), because there is a player who maximizes her own payoff while the others minimize it. The interaction of these two forces complicates the belief evolution, which makes our analysis more difficult than the POMDP problem. To prove invariance of the minimax payoff, we begin with the observation that the minimax payoff (as a function of the initial belief) is the lower envelope of a series of convex curves. Using this convexity, we derive a bound on the variability of the minimax payoffs over beliefs, and then show that this bound is close to zero.

So in sum, under the full support assumption, the payoff invariance condition holds and hence the folk theorem obtains. But the full support assumption is a bit restrictive, and leaves out some economic applications. For example, consider the following natural resource management problem: 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. This example does not satisfy the full support assumption, because the state cannot be the highest one if the fishermen catch too much fish today. Also, games with delayed observations, and even the standard stochastic games (with observable states) do not satisfy the full support assumption.

To address this concern, in Section 5, we show that the payoff invariance condition (and hence the folk theorem) still holds even if the full support assumption is replaced with a weaker condition. 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. This result strengthens the existing results in the POMDP literature; uniform connectedness is more general than various assumptions proposed in the literature.⁴ We also show that the minimax payoff for patient players is invariant to the initial belief under a similar assump-

⁴Such assumptions include 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 E for more details.) The natural resource management problem in this paper is an example which satisfies uniform connectedness but not the assumptions in the literature. Similarly, Examples A1 and A2 in Appendix A satisfies asymptotic uniform connectedness but not the assumptions in the literature.

tion called *robust connectedness*.

Our first assumption, uniform connectedness, is a condition about how the *support* of the belief evolves over time. 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 .) This assumption can be regarded as an analogue of communicating states of Dutta (1995), which requires that players can move the state from any ω to any other $\tilde{\omega}$; but note that uniform connectedness is *not* a condition on the evolution of the belief itself, so it need not imply ergodicity of the belief. Nonetheless we find that this condition implies invariance of the feasible payoff set. A key step in the proof is to find a uniform bound on the variability of feasible payoffs over beliefs with the same support. It turns out that this bound is close to zero, and thus the feasible payoff set is almost determined by the support of the belief. So what matters is how the support changes over time, which suggests that uniform connectedness is useful to obtain the invariance result. Our second assumption, robust connectedness, is also a condition on the support evolution, and has a similar flavor.

Uniform connectedness and robust connectedness are more general than the full support assumption, and satisfied in many economic examples, including the ones discussed earlier. Our folk theorem applies as long as both uniform connectedness and robust connectedness are satisfied.

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

⁵Independently 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.

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 $I = \{1, \dots, N\}$ be the set of players. At the beginning of the game, Nature chooses the state of the world ω^1 from a finite set Ω . The state may change as time passes, and the state in period $t = 1, 2, \dots$ is denoted by $\omega^t \in \Omega$. The state ω^t is not observable to players, and let $\mu \in \Delta\Omega$ 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

⁶An 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.

$g_i^\omega(a) = \sum_{y \in Y} \pi_Y^\omega(y|a) u_i(a, y)$. Here the hidden state ω influences a player's expected payoff through the distribution of y .⁷ Let $g^\omega(a) = (g_i^\omega(a))_{i \in I}$ be the vector of expected payoffs. Let $\bar{g}_i = \max_{\omega, a} |g_i^\omega(a)|$, and let $\bar{g} = \sum_{i \in I} \bar{g}_i$. Also let $\bar{\pi}$ 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 component 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

⁷ Alternatively, we may assume that $g_i^\omega(a)$ is player i 's actual payoff (so the state ω directly influences the payoff) and she does not observe this payoff (so the payoff does not provide extra information about the state). All our results extend to this setup, with no difficulty. Assuming unobservable payoffs is common in the POMDP literature. This assumption is satisfied if we consider a situation in which the game ends with probability $1 - \delta$ after each period, and player i receives all the payoffs after the game ends.

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 $s_i : H \rightarrow \Delta A_i$. 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 $v_i^\omega(\delta, s) = E[(1 - \delta) \sum_{t=1}^{\infty} \delta^{t-1} g_i^\omega(a^t) | \omega, s]$. 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 D 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} , they have the same posterior belief μ^t . Then we can regard this belief μ^t as a common state variable, and so 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. Specifically, given μ^{t-1} , a^{t-1} , and y^{t-1} , the posterior belief μ^t in period t is computed as

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

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

Our solution concept is a sequential equilibrium. Let $\zeta : H \rightarrow \Delta \Omega$ 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 Folk Theorem under Payoff Invariance

3.1 Payoff Invariance Condition

In our model, feasible payoffs are not merely the convex hull of the stage-game payoffs. For example, to maximize the social welfare, an action which yields a low payoff today may be preferred to one which yields a high payoff, if it leads to a better state tomorrow and/or if it gives better signals about the state tomorrow. To capture these effects, we compute the payoff in the infinite-horizon game for each strategy profile s , and define the feasible set as the set of all such payoffs. That is, given the initial prior μ and the discount factor δ , the feasible payoff set is defined as

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

where $\text{co}B$ denote the convex hull of a set B . Here, δ and μ influence the feasible payoff set, as they influence the payoff $v^\mu(\delta, s)$ for a given strategy profile s .

Similarly, given the initial prior μ and the discount factor δ , player i 's *minimax payoff* in the stochastic game is defined to be

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

Note that player i 's sequential equilibrium payoff is at least this minimax payoff, as players do not have private information. The proof is standard and hence omitted. Note also that the minimizer s_{-i} indeed exists; see Appendix D for more details.

In this section, we will prove the folk theorem under the following, *payoff invariance* assumption. Let $d(A, B)$ denote the Hausdorff distance between two sets $A, B \subset \mathbf{R}^N$.

Assumption 1.

- (a) The limit of the feasible payoff set $\lim_{\delta \rightarrow 1} V^\mu(\delta)$ exists and is independent of the initial prior μ ; that is, there is a set $V \subset \mathbf{R}^N$ such that $\lim_{\delta \rightarrow 1} d(V, V^\mu(\delta)) = 0$ for all μ .
- (b) For each i , the limit of the minimax payoff $\lim_{\delta \rightarrow 1} v_i^\mu(\delta)$ exists and is independent of the initial prior μ .

This assumption requires that the feasible payoff set and the minimax payoff be invariant to the initial prior μ , when players are patient. As explained in the introduction, it ensures that even if someone deviates today and manipulates the belief tomorrow, it does not change the feasible payoffs in the continuation game so that we can still discipline players' dynamic incentives effectively. Various papers in the literature on stochastic games (e.g., Dutta (1995), Fudenberg and Yamamoto (2011b), and Hörner, Sugaya, Takahashi, and Vieille (2011)) make a similar assumption.

Take V as in the assumption above. This set V is the limit feasible payoff set; the feasible payoff set $V^\mu(\delta)$ is approximately V for all initial priors μ , when players are patient. Also, let \underline{v}_i denote the limit of the individually rational payoff, that is, let $\underline{v}_i = \lim_{\delta \rightarrow 1} v_i^\mu(\delta)$. Let V^* denote the limit of the feasible and individually rational payoff set, i.e., V^* is the set of all feasible payoffs $v \in V$ such that $v_i \geq \underline{v}_i$ for all i .

Assumption 1 above is not stated in terms of primitives, and in general it is hard to check. In later sections, we will provide sufficient conditions for this assumption.

3.2 Punishment over Random Blocks

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. As they argue, the parameter T must be carefully chosen; specifically, they pick a large T first and then take $\delta \rightarrow 1$, so the minimax phase is not too long relative to the discount factor δ . This ensures that players are indeed willing to minimax a deviator, ex-

pecting a reward after the minimax play. (If we take δ first and then take T large, this punishment mechanism does not work. Indeed, 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, players may not want to play the minimax strategy.)

In stochastic games, the minimax strategy is a strategy for the infinite-horizon game, so we need to carefully think about when players should stop the minimax play and move to the reward phase. When states are finite and observable, Dutta (1995) and Hörner, Sugaya, Takahashi, and Vieille (2011) show that the idea of the T -period punishment mechanism above still works well. A point is that when states are finite, the minimax strategy induces an ergodic state evolution. Thus when $\delta \rightarrow 1$, the average payoff during these T periods approximates the minimax payoff, i.e., even though players play the minimax strategy only for T periods (not infinite periods), the payoff during these punishment periods is as low as the minimax payoff for the infinite-horizon game. Hence a player’s deviation can be deterred using this punishment mechanism.

On the other hand, in our model, it is not clear if such a T -period punishment mechanism works effectively. A problem here is that due to infinite states, the belief evolution induced by the minimax strategy may not be ergodic (although invariance of the minimax payoff suggests a sort of ergodicity). Accordingly, given any large number T , if we take $\delta \rightarrow 1$, the average payoff for the T -period block can be quite different from (in particular, substantially greater than) the minimax payoff in the infinite-horizon game.⁸

To fix this problem, we consider 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

⁸ In the POMDP literature, it is well-known that the payoff in the discounted infinite-horizon problem and the (time-average) payoff in the T -period problem are asymptotically the same if a solution to the discounted problem is invariant to the initial prior in the limit as $\delta \rightarrow 1$, and if the rate of convergence is at most of order $O(1 - \delta)$. (See Hsu, Chuang, and Arapostathis (2006) and the references therein.) Unfortunately, in our setup, the rate of convergence of the feasible payoffs and the minimax payoffs is slower than this bound for some cases, as can be seen in the proof of Proposition A2.

period $t + 1$ is still included in the current random block. Otherwise, the current block terminates. So the random block terminates with probability $1 - p$ each period.

This random block is useful, because it is payoff-equivalent to the infinite-horizon game with the discount factor $p\delta$, due to the random termination probability $1 - p$. So given the current belief μ , if the opponents use the minimax strategy for the initial prior μ and the discount factor $p\delta$ (rather than δ) during the block, then player i 's average payoff during the block never exceeds the minimax payoff $\underline{v}_i^\mu(p\delta)$. 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 effectively punish player i by playing the minimax strategy in the random block.

In the proof of the folk theorem, we pick p close to one, and then take $\delta \rightarrow 1$. This implies that although the random block is long in expectation, players put a higher weight on the continuation payoff after the block than the payoff during the current block. Hence a small variation in continuation payoffs is enough to discipline players' play during the random block. In particular, 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 any history. See the proof in Section 3.4 for more details.

Hörner, Takahashi, and Vieille (2015) also use the idea of random blocks (they call it “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.

3.3 Result

Now we show that if the payoff invariance condition (Assumption 1) holds, the folk theorem obtains. This result encompasses the folk theorem of Dutta (1995) as a special case.

Proposition 1. *Suppose that Assumption 1 holds, and that the limit payoff set V^* is full dimensional (i.e., $\dim V^* = N$). Assume also that public randomization is available. Then for any interior point $v \in V^*$, there is $\bar{\delta} \in (0, 1)$ such that for any $\delta \in (\bar{\delta}, 1)$ and for any initial prior μ , there is a sequential equilibrium with the payoff v .*

In addition to the payoff invariance, the proposition requires the full dimensional assumption, $\dim V^* = N$. This assumption allows us to use player-specific punishments; 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) use this assumption to obtain the folk theorem for repeated games with observable actions.

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. Unfortunately, this result does not 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. 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 very similar to that of Dutta (1995), except that we use random blocks (rather than T -period blocks). 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 mixed minimax strategies will be given in Appendix B.

3.4 Equilibrium with Pure Minimax Strategies

Take an interior point $v \in V^*$. We will construct a sequential equilibrium with the payoff v when δ is close to one. To simplify the notation, we assume that there are only two players. This assumption is not essential, and 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., $w_i(i) \ll v_i$ 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 1 to see how to choose these payoffs $w(i)$. In this figure, the payoffs are normalized so that the limit minimax payoff vector is $\underline{v} = (v_1, v_2) = (0, 0)$.

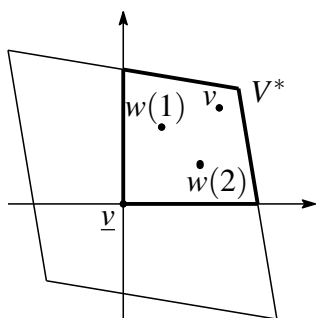


Figure 1: Payoffs $w(1)$ and $w(2)$

Looking ahead, the payoffs $w(1)$ and $w(2)$ can be interpreted as “punishment payoffs.” That is, if player i deviates and players start to punish her, the payoff in the continuation game will be approximately $w(i)$ in our equilibrium. Note that 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) implies that each player i prefers the payoff $w_i(j)$ when she punishes the opponent j to the payoff $w_i(i)$ when she is punished. This ensures that player i is indeed willing to punish the opponent j after j ’s deviation; if she does not, then player i will be punished instead of j , and it lowers player i ’s payoff.

Pick $p \in (0, 1)$ close to one so that the following conditions hold:

- The payoff vectors v , $w(1)$, and $w(2)$ are in the interior of the feasible payoff set $V^\mu(p)$ for each μ .
- $\sup_{\mu \in \Delta\Omega} v_i^\mu(p) < w_i(i)$ for each i .

By the 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 $p\delta$.

Our equilibrium consists of three phases: *regular (cooperative) phase*, *punishment phase for player 1*, and *punishment phase for player 2*. In the regular phase, the infinite horizon is regarded as a series of random blocks. 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, pick some random block, and let μ be the belief and the beginning of the block. 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 use this strategy during the block. If such a pure strategy profile does not exist, 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. After the block, a new block starts and players will behave as above again.

It is important that during the regular phase, after each period t , players' continuation payoffs are always close to the target payoff v . To see why, note first that the average payoff in the current block can be very different from v once the public randomization (which chooses one of the extreme points) realizes. However, when δ is close to one, players do not care much about the payoffs in the current block, and what matters is the payoffs in later blocks, which are exactly v . Hence even after public randomization realizes, the total payoff is still close to v . This property is due to the random block structure, and will play an important role when we check incentive conditions.

As long as no one deviates from the prescribed strategy above, players stay at the regular phase. However, once someone (say, player i) deviates, 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, just as

in the regular phase. In the first K blocks, the opponent (player $j \neq i$) minimaxes player i . 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 $p\delta$. On the other hand, player i maximizes her payoff during these K blocks. After the K blocks, players switch their play in order 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 $w(i)$. The parameter K will be specified later.

If no one deviates from the above play, players stay at this punishment phase forever. Also, 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, then players switch to the punishment phase for player j , in order to punish player j . See Figure 2.

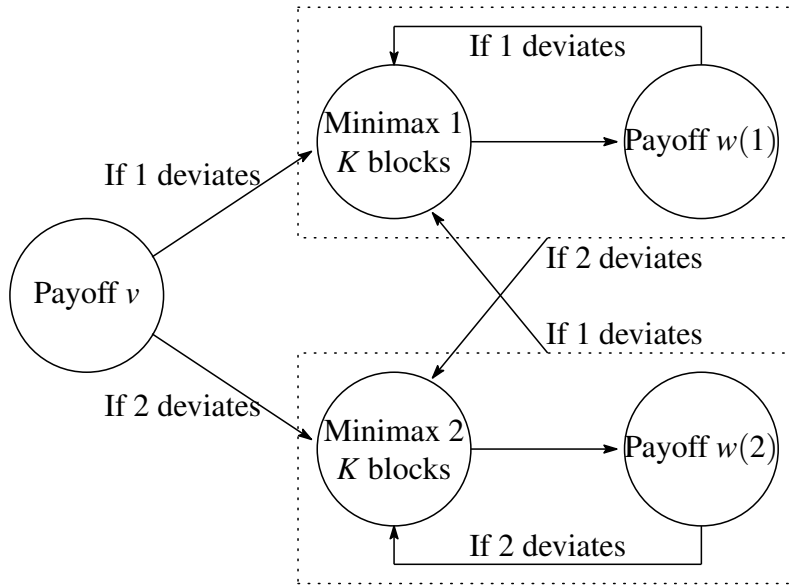


Figure 2: Equilibrium strategy

Now, choose K such that

$$-\bar{g} - \frac{1}{1-p}\bar{g} + \frac{K-1}{1-p}w_i(i) > \bar{g} + \frac{K}{1-p} \sup_{\mu \in \Delta\Omega} v_i^\mu(p) \quad (1)$$

for each i . Note that (1) indeed holds for sufficiently large K , as $\sup_{\mu \in \Delta\Omega} v_i^\mu(p) < w_i(i)$. To interpret (1), suppose that we are now in the punishment phase for player i , in particular a period in which players play the strategy profile with the post-minimax payoff $w(i)$. (1) 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 \bar{g} , and then she will be minmaxed 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 $\frac{K}{1-p} \sup_{\mu \in \Delta\Omega} v_i^\mu(p)$ when $\delta \rightarrow 1$. So the right-hand side of (1) is an upper bound on player i 's unnormalized payoff until the minimax play ends, when she deviates.

On the other hand, if she does not deviate, her payoff today is at least $-\bar{g}$. Also, for the next K periods, she can earn at least $-\frac{1}{1-p}\bar{g} + \frac{K-1}{1-p}w_i(i)$, because we consider the post-minimax play. (Here the payoff during the first block can be lower than $w_i(i)$, as tomorrow may not be the first period of the block. So we use $-\frac{\bar{g}}{1-p}$ as a lower bound on the payoff during this block.) In sum, by not deviating, player i can obtain at least the left-hand side of (1), which is indeed greater than the payoff by deviating.

With this choice of K , by inspection, we can show that the strategy profile above is indeed an equilibrium for sufficiently large δ . The argument is very similar to the one by Dutta (1995) and hence omitted.

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 minmaxes player $j \neq i$. As shown by Fudenberg and Maskin (1986), we can indeed satisfy this indifference condition by perturbing the post-minimax payoff $w_i(j)$ appropriately. See Appendix B for the formal proof.

4 Full Support Assumption

In the previous section, we have shown that the folk theorem holds under the payoff invariance condition. But unfortunately, this assumption is not stated in

terms of primitives, and it is important to better understand when this assumption is satisfied. In this section, we show that the following *full support assumption* is sufficient for the payoff invariance condition:

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 .

In words, the full support assumption requires that any signal y and any state $\tilde{\omega}$ can happen tomorrow with positive probability, regardless of the current state ω and the current action profile a . Under this assumption, the posterior belief is always in the interior of $\Delta\Omega$, that is, after every history, the posterior belief μ^t assigns positive probability to each state ω . It turns out that this property is very useful in order to obtain the payoff invariance.

The full support assumption is easy to check, but unfortunately, it is demanding and leaves out many potential economic applications. For example, this 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 5.3. Also, it rules out even the standard stochastic games (in which the state is observable to players) and the games with delayed observations. To fix this problem, in Section 5, we will explain how to relax the full support assumption.

4.1 Invariance of the Feasible Payoff Set

Let Λ be the set of directions $\lambda \in \mathbf{R}^N$ with $|\lambda| = 1$. For each direction λ , we compute the “score” using the following formula:⁹

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

Roughly speaking, this 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 is simply the highest possible payoff for player i within the feasible payoff set. When $\lambda = (\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}})$, the score is the (normalized) maximal social welfare within the feasible payoff set.

⁹Note that this maximization problem indeed has a solution; see Appendix D for the proof.

For each given discount factor δ , the score can be computed using dynamic programming. Fix a direction λ , and let $f(\mu)$ denote the score given the initial prior μ . Let $\tilde{\mu}(y|\mu, a)$ denote the posterior belief in period two given that the initial prior is μ and players play a and observe y in period one. Then the score function f must solve the following Bellman equation:

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

To interpret this equation, suppose that there are only two players. Let $\lambda = (\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}})$, so that the score $f(\mu)$ represents the maximal social welfare. (2) asserts that the maximal welfare $f(\mu)$ is a sum of the (normalized) welfare today $\lambda \cdot g^\mu(a) = \frac{1}{\sqrt{2}}(g_1^\omega(a) + g_2^\omega(a))$ and the welfare in the continuation game, $f(\tilde{\mu}(y|\mu, a))$. The action a is chosen in such a way that this sum is maximized.

(2) is known as a ‘‘POMDP problem,’’ in the sense that it is a Bellman equation in which the state variable μ is a belief about a hidden state. In the POMDP theory, it is well-known that a solution f is convex with respect to the state variable μ , and that this convexity leads to various useful theorems. For example, Platzman (1980) shows that under the full support assumption, a solution $f(\mu)$ is invariant to the initial belief μ , when the discount factor is close to one. In our context, this implies that when players are patient, the score is invariant to the initial prior μ , and so is the feasible payoff set $V^\mu(\delta)$. Formally, we have the following proposition.

Proposition 2. *Under the full support assumption, for each $\varepsilon > 0$, there is $\bar{\delta} \in (0, 1)$ such that for any $\lambda \in \Lambda$, $\delta \in (\bar{\delta}, 1)$, 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 particular, this implies that for each direction λ , the limit $\lim_{\delta \rightarrow 1} \max_{v \in V^\mu(\delta)} \lambda \cdot v$ of the score is independent of μ ; hence Assumption 1(a) follows.

Note that the limit $\lim_{\delta \rightarrow 1} \max_{v \in V^\mu(\delta)} \lambda \cdot v$ of the score indeed exists, thanks to Theorem 2 of Rosenberg, Solan, and Vieille (2002). Platzman (1980) also shows that the score converges at the rate of $1 - \delta$. So we can replace ε in the above proposition with $O(1 - \delta)$.

4.2 Invariance of the Minimax Payoffs

The following proposition shows that under the full support assumption, even the minimax payoff is invariant to the initial prior:

Proposition 3. *If the full support assumption holds, then Assumption 1(b) holds.*

This result may look similar to Proposition 2, but its proof is substantially different. As noted earlier, Proposition 2 directly follows from the fact that the score function f is a solution to the POMDP problem (2). Unfortunately, the minimax payoff $v_i^\mu(\delta)$ is not a solution to a POMDP problem; this is so because in the definition of the minimax payoff, player i maximizes her payoff while the opponents minimize it. Accordingly, POMDP techniques are not applicable. The proof techniques for the observable-state case does not apply either, as they heavily rely on the assumption that the state space is finite so that one can drive the state to any other state with positive probability in finite time. In the next subsection, we will briefly explain how to prove the result above. The formal proof can be found in Appendix B.

4.3 Outline of the Proof of Proposition 3

Fix a discount factor δ , and let s_{-i}^μ denote the minimax strategy for the initial prior μ . Suppose that the initial prior is $\tilde{\mu}$ but the opponents use the minimax strategy s_{-i}^μ for a different initial prior $\mu \neq \tilde{\mu}$. Let $v_i^{\tilde{\mu}}(s_{-i}^\mu)$ denote player i 's payoff when she takes a best reply in such a situation; that is, let $v_i^{\tilde{\mu}}(s_{-i}^\mu) = \max_{s_i \in S_i} v_i^{\tilde{\mu}}(s_i, s_{-i}^\mu)$. When $\tilde{\mu} = \mu$, this payoff $v_i^{\tilde{\mu}}(s_{-i}^\mu)$ is simply the minimax payoff for the belief μ . That is, $v_i^{\tilde{\mu}}(s_{-i}^\mu) = v_i^\mu(\delta)$. But when $\tilde{\mu} \neq \mu$, the opponents' strategy s_{-i}^μ is different from the minimax strategy $s_{-i}^{\tilde{\mu}}$ for the actual initial prior, and the payoff $v_i^{\tilde{\mu}}(s_{-i}^\mu)$ is greater than the minimax payoff $v_i^{\tilde{\mu}}(\delta)$. Define the *maximal value* \bar{v}_i as the maximum of these payoffs $v_i^{\tilde{\mu}}(s_{-i}^\mu)$ over all $(\mu, \tilde{\mu})$.

It turns out that these payoffs $v_i^{\tilde{\mu}}(s_{-i}^\mu)$ have a nice tractable structure, which allows us to obtain the following result; to make our exposition as simple as possible, here we give only an informal statement. (See Section B.2.3 in the appendix for a more formal version.) Recall that $\bar{\pi}$ is the minimum of $\pi^\omega(y, \tilde{\omega}|a)$.

Key Result: Fix δ . Suppose that $|\bar{v}_i - v_i^\mu(\delta)| \approx 0$ for some interior belief μ such that $\mu(\omega) \geq \bar{\pi}$ for all ω . Then $|\bar{v}_i - v_i^{\tilde{\mu}}(\delta)| \approx 0$ for all beliefs $\tilde{\mu}$.

This result asserts that if the minimax payoff $\underline{v}_i^\mu(\delta)$ approximates the maximal value for *some* interior belief μ , then the minimax payoffs for *for all other* beliefs $\tilde{\mu}$ also approximate the maximal value. So in order to prove Proposition 3, we do not need to evaluate the minimax payoffs for each initial prior μ ; we only need to find *one* interior belief whose minimax payoff approximates the maximal value.

Proof Sketch of Key Result. Pick μ as stated, so that the minimax payoff $\underline{v}_i^\mu(\delta)$ approximates the maximal value. Pick an arbitrary belief $\tilde{\mu} \neq \mu$. Our goal is to show that the minimax payoff for this belief $\tilde{\mu}$ approximates the maximal value.

Consider player i 's payoff $v_i^\mu(s_{-i}^{\tilde{\mu}})$ when the opponents use the minimax strategy for the belief $\tilde{\mu}$ but the actual initial prior is μ . Since the opponents' strategy $s_{-i}^{\tilde{\mu}}$ is different from the minimax strategy for the actual belief μ , this payoff is greater than the minimax payoff $\underline{v}_i^\mu(\delta)$. On the other hand, by the definition, this payoff must be smaller than the maximal value. Hence we have

$$\underline{v}_i^\mu(\delta) \leq v_i^\mu(s_{-i}^{\tilde{\mu}}) \leq \bar{v}_i.$$

Since the minimax payoff $\underline{v}_i^\mu(\delta)$ approximates the maximal value \bar{v}_i , this inequality implies that the payoff $v_i^\mu(s_{-i}^{\tilde{\mu}})$ also approximates the maximal value. This in turn implies that the minimax payoff $\underline{v}_i^{\tilde{\mu}}(\delta) = v_i^{\tilde{\mu}}(s_{-i}^{\tilde{\mu}})$ indeed approximates the maximal value; this last step follows from Lemma B1 in the proof, which asserts that given the opponents' strategy $s_{-i}^{\tilde{\mu}}$, if the payoff $v_i^\mu(s_{-i}^{\tilde{\mu}})$ approximates the maximal value for *some* interior belief μ , then for *all other* beliefs $\hat{\mu}$, the payoff $v_i^{\hat{\mu}}(s_{-i}^{\tilde{\mu}})$ approximates the maximal value. The proof of this lemma relies on the observation that (given the opponent's strategy $s_{-i}^{\tilde{\mu}}$) player i can obtain better payoffs when she has better information about the initial state. *Q.E.D.*

As one can see from the proof sketch above, to obtain the result we want, we relate two minimax payoffs $\underline{v}_i^\mu(\delta)$ and $\underline{v}_i^{\tilde{\mu}}(\delta)$ through the payoff $v_i^\mu(s_{-i}^{\tilde{\mu}})$. This is the value of considering the payoff $v_i^\mu(s_{-i}^{\tilde{\mu}})$.

Given the result above, what remains is to find *one* interior belief whose minimax payoff approximates the maximal value. This can be done by a careful inspection of the maximal value, and the full support assumption is used in this part. See Section B.2.2 in the appendix for more details.

5 Relaxing the Full Support Assumption

We have shown that if the state transition function has full support, Assumption 1 holds so that the folk theorem obtains. However, as noted earlier, the full support assumption is demanding, and rules out many possible applications. To address this concern, in this section, we show that Assumption 1 still holds even if the full support assumption is replaced with a new, weaker condition. Specifically, we show that the feasible payoff set is invariant if the game is *uniformly connected*, and the minimax payoff is invariant if the game is *robustly connected*. Both uniform connectedness and robust connectedness are about how the *support* of the posterior belief evolves over time, and they are satisfied in many economic applications.

5.1 Uniform Connectedness and Feasible Payoffs

5.1.1 Weakly Communicating States

Before we consider the hidden-state model, it is useful to understand when the feasible payoffs are invariant to the initial state in the observable-state case. A key condition is *weakly communicating states*, which requires that there be a path from any state to any other state, except temporary ones. As will be seen, uniform connectedness, which will play a central role in our hidden-state model, is an analogue of this condition.

Let $\Pr(\omega^{T+1} = \omega | \tilde{\omega}, a^1, \dots, a^T)$ denote the probability of the state in period $T + 1$ being ω given the initial state $\tilde{\omega}$ and the action sequence (a^1, \dots, a^T) . A state ω is *globally accessible* if for any initial state $\tilde{\omega}$, there is a natural number T and an action sequence (a^1, \dots, a^T) such that

$$\Pr(\omega^{T+1} = \omega | \tilde{\omega}, a^1, \dots, a^T) > 0. \quad (3)$$

That is, ω is globally accessible if players can move the state to ω from any other state $\tilde{\omega}$.

A state ω is *uniformly transient* if it is not globally accessible and for any pure strategy profile s , there is a natural number T and a globally accessible state $\tilde{\omega}$ so that $\Pr(\omega^{T+1} = \tilde{\omega} | \omega, s) > 0$. Intuitively, uniform transience of ω implies that the state ω is temporary. Indeed, *regardless of players play*, the state cannot

stay there forever and must reach a globally accessible state eventually. As will be explained, this property ensures that the score for a uniformly transient state cannot be too different from the ones for globally accessible states.

States are *weakly communicating* if each state ω is globally accessible or uniformly transient. Figure 3 is an example of weakly communicating states. The state moves along the arrows; for example, there is an action profile which moves the state 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*. It is easy to check that the states ω_1 , ω_2 , and ω_3 are globally accessible, while the states ω_4 and ω_5 are uniformly transient. Note that the uniformly transient states are indeed temporary; once the state reaches a globally accessible state, it never comes back to a uniformly transient state. As can be seen, when states are weakly communicating, the state can go back and forth over all states, except these temporary ones. This condition is a generalization of *communicating states* of Dutta (1995), which requires that all states be globally accessible.

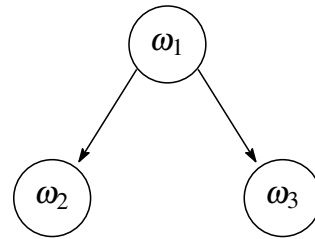
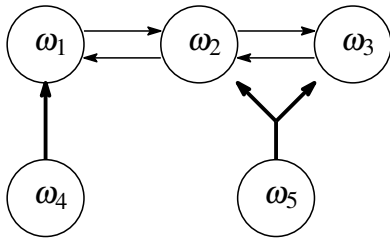


Figure 3: Weakly Communicating States

Figure 4: Two Absorbing States

If states are weakly communicating, the feasible payoff set is invariant to the initial state for patient players. This result is a corollary of Propositions 5 and 7, but a rough idea is as follows. Consider the score toward some direction λ . Let ω be the state which gives the highest score over all initial states, and call this score the *maximal score*. There are two cases to be considered:

Case 1: ω is globally accessible. In this case, given any initial state, players can move the state to ω in finite time with probability one, and can earn the maximal score thereafter. Since payoffs before the state reaches ω are almost negligible for patient players, this implies that regardless of the initial state, the score must be almost as good as the maximal score, and hence the score is indeed invariant to the initial state.

Case 2: ω is not globally accessible. Since states are weakly communicating, ω must be uniformly transient. This means that ω is a temporary state, i.e., if the initial state is ω , the state must eventually reach globally accessible states in finite time, with probability one. Since payoffs before the state reaches globally accessible ones are almost negligible, this implies that there is at least one globally accessible state ω^* whose score is approximately as good as the maximal score. Now, since ω^* is globally accessible, given any initial state, players can move the state to ω^* in finite time with probability one. Hence as in Case 1, we can conclude that regardless of the initial state, the score is almost as good as the maximal score.

On the other hand, if states are not weakly communicating, the feasible payoff set may depend on the initial state, even for patient players. Figure 4 is an example in which states are *not* weakly communicating; it is easy to check that no states are globally accessible, and hence no states are uniformly transient. A key in this example is that we have multiple absorbing states, ω_2 and ω_3 . Obviously, if these two states yield different stage-game payoffs, then the feasible payoff set must depend on the initial state.

5.1.2 Definition of Uniform Connectedness

Since the state variable in our model is a belief μ , a natural extension of weakly communicating states is to assume that there be a path from any belief to any other belief, except temporary ones. But unfortunately, this approach does not work, because such a condition is too demanding and not satisfied in general. A problem is that given an initial prior μ , only finitely many beliefs are reachable in finite time; so almost all beliefs are not reachable from μ , and hence a “globally accessible” belief does not exist in general.

To avoid this problem, we will focus on the evolution of the *support* of the belief, rather than the evolution of the belief itself. Now we do not need to worry about the technical problem above, since there are only finitely many supports. 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, it turns out that this is precisely what we need for invariance of the feasible payoff set.

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 $\Pr(\mu^{T+1} = \tilde{\mu} | \mu, a^1, \dots, a^T)$ denote the probability given that players play the action sequence (a^1, \dots, a^T) in the first T periods. Global accessibility of Ω^* requires that given any current belief μ , players can move the support of the posterior belief to Ω^* (or its subset), by choosing some appropriate action sequence which may depend on μ .¹⁰ This definition can be viewed as an analogue of the global accessibility of the state ω in the observable-state case.

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 , an action sequence (a^1, \dots, a^T) , and a belief $\tilde{\mu}$ whose support is included in Ω^* such that¹¹

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

Global accessibility 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^* . Thanks to this property, the whole state space $\Omega^* = \Omega$ is globally accessible for any game. Also if a set Ω^* is globally accessible, then so is any superset $\tilde{\Omega}^* \supseteq \Omega^*$.

Global accessibility requires that there be a lower bound $\pi^* > 0$ on the probability, while (3) does not. But this difference is not essential; indeed, although it is not explicitly stated in (3), we can always find such a lower bound $\pi^* > 0$ when states are finite. In contrast, we have to explicitly assume the existence of π^* in Definition 2, since there are infinitely many beliefs.¹²

Next, we give the definition of uniform transience of Ω^* . It requires that if the support of the current belief is Ω^* , then *regardless of players' play in the*

¹⁰Here, we define global accessibility and uniform transience using the posterior belief μ^t . In Appendix C, we show that there are equivalent definitions based on primitives. Using these definitions, one can check if a given game is uniformly connected in finitely many steps.

¹¹Replacing the action sequence (a^1, \dots, a^T) in this 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. Also, while the definition above does not provide an upper bound on the number T (so the action sequence can be arbitrarily long), when we check whether a given set Ω^* is globally accessible or not, we can restrict attention to an action sequence with length $T \leq 4^{|\Omega|}$. Indeed, whenever there is an action sequence (a^1, \dots, a^T) which satisfies the property stated here, we can always find an action sequence $(\tilde{a}^1, \dots, \tilde{a}^{\tilde{T}})$ with $\tilde{T} \leq 4^{|\Omega|}$ which satisfies the same property. See Appendix C for more details.

¹²Since there are only finitely many supports, there is a bound π^* which works for all globally accessible sets Ω^* .

continuation game, the support of the posterior belief must reach some globally accessible set with positive probability at some point. Again, this definition can be viewed as an analogue of the uniform transience in the observable-state case.

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 and a belief $\tilde{\mu}$ whose support is globally accessible such that $\Pr(\mu^{T+1} = \tilde{\mu} | \mu, s) > 0$.¹³

As noted earlier, a superset of a globally accessible set is globally accessible. Similarly, as the following proposition shows, a superset of a uniformly transient set is globally accessible or uniformly transient. The proof of the proposition is given in Appendix B.

Proposition 4. *A superset of a globally accessible set is globally accessible. Also, a superset of a uniformly transient set is globally accessible or uniformly transient.*

This result implies that if each singleton set $\{\omega\}$ is globally accessible or uniformly transient, then any subset $\Omega^* \subseteq \Omega$ is globally accessible or uniformly transient. Accordingly, we have two equivalent definitions of uniform connectedness; the second definition is simpler, and hence more useful in applications.

Definition 4. A stochastic game is *uniformly connected* if each subset $\Omega^* \subseteq \Omega$ is globally accessible or uniformly transient. Equivalently, a stochastic game is uniformly connected if each singleton set $\{\omega\}$ is globally accessible or uniformly transient.

Uniform connectedness is more general than the full support assumption. Indeed, if the full support assumption holds, then regardless of the initial prior, the support of the belief in period two is the whole state space Ω ; hence any proper subset $\Omega^* \subset \Omega$ is transient, and the game is uniformly connected.

¹³Again, although the definition here does not provide an upper bound on T , when we check whether a given set Ω^* is uniformly transient or not, we can restrict attention to $T \leq 2^{|\Omega|}$. See Appendix C for more details. The strategy profile s in this definition cannot be replaced with an action sequence (a^1, \dots, a^T) .

5.1.3 Interpretation of Uniform Connectedness: Two-State Case

To better understand the economic meaning of uniform connectedness, we will focus on the two-state case and investigate when the game is uniformly connected and when it is not. It turns out that this question is deeply related to whether or not the state can be revealed by some signals.

So suppose that there are only two states, ω_1 and ω_2 . For simplicity, assume that both states are globally accessible, that is, there is an action profile which moves the state from ω_1 to ω_2 with positive probability, and vice versa. The state ω_1 *can be revealed* if there is a signal sequence which reveals that the state tomorrow is ω_1 for sure. Specifically, we need one of the following conditions: (i) there is ω , a , and y such that $\pi^\omega(y, \omega_1|a) > 0$ and $\pi^{\tilde{\omega}}(y, \omega_2|a) = 0$ for all $\tilde{\omega}$; or (ii) there is ω , a^1 , a^2 , y^1 , and y^2 such that $\pi^\omega(y^1, \omega_2|a^1) > 0$, $\pi^{\tilde{\omega}}(y^1, \omega_1|a^1) = 0$ for all $\tilde{\omega}$, $\pi^{\omega_2}(y^2, \omega_1|a^2) > 0$, and $\pi^{\omega_2}(y^2, \omega_2|a^2) = 0$. The first condition implies that (starting from an interior initial belief μ) if players play a and observe y today, then the state tomorrow will be ω_1 for sure and the posterior puts probability one on it. The second condition allows the possibility that (again, starting from an interior initial belief μ) players cannot directly move the belief to the one which puts probability one on ω_1 , but they can move the belief to the one which puts probability one on ω_2 , and then to the one which puts probability one on ω_1 . The state ω_2 *can be revealed* if a similar condition holds. We consider the following three cases.

Case 1: Both states can be revealed. This case can be view as a generalization of the observable-state case. Here the state need not be observed each period, but it is occasionally observed if players choose right actions. In this case, it is not difficult to show that both $\{\omega_1\}$ and $\{\omega_2\}$ are globally accessible. (Given any initial prior, if players choose right actions, then the state ω is revealed and the support of the posterior indeed reaches $\{\omega\}$.) So uniform connectedness is always satisfied in this case.

Case 2: Only one state can be revealed. Without loss of generality, assume that ω_1 can be revealed. As in the previous case, the set $\{\omega_1\}$ is globally accessible. On the other hand, the set $\{\omega_2\}$ is not globally accessible. This is so because if the initial prior is an interior belief, then regardless of players play, the state ω_2 is never revealed and the support of the posterior cannot reach $\{\omega_2\}$. Hence, for the

game to be uniformly connected, the set $\{\omega_2\}$ must be uniformly transient, which requires us to make an extra assumption. Specifically, in this case, the set $\{\omega_2\}$ is uniformly transient (and hence the game is uniformly connected) if and only if the state ω_2 is not absorbing regardless of players' play, i.e., for each action profile a , we have $\sum_{y \in Y} \pi^{\omega_2}(y, \omega_1 | a) > 0$ so that the state moves from ω_2 to ω_1 with positive probability.¹⁴

Case 3: No states can be revealed. In this case, the sets $\{\omega_1\}$ and $\{\omega_2\}$ are not globally accessible. So for the game to be uniformly connected, both these sets must be uniformly transient, which requires an extra assumption. Specifically, the sets $\{\omega_1\}$ and $\{\omega_2\}$ are uniformly transient (and hence the game is uniformly connected) if and only if the *scrambling condition* holds in the sense that for any initial state ω and for any strategy profile s , there is a signal sequence (y^1, y^2) such that $\Pr(\omega^3 = \tilde{\omega} | \omega^1 = \omega, s, y^1, y^2) > 0$ for each $\tilde{\omega}$.¹⁵ Intuitively, this condition implies that players cannot retain perfect information about the state, in the sense that even if they know the initial state, the posterior belief must become an interior belief in finite time. This scrambling condition can be viewed as a generalization of the full support assumption. Under the full support assumption, the posterior belief becomes an interior belief immediately, after *any* realization of the signal y . Here we need that the posterior becomes an interior belief after *some* realization of the signal sequence.

¹⁴To prove the if part, pick an arbitrary action profile a and let y be such that $\pi^{\omega_2}(y, \omega_1 | a) > 0$. If the initial state is ω_2 and players play a , then with positive probability, this signal y is observed and the posterior puts positive probability on ω_1 , which means that the support of the posterior indeed moves to a globally accessible set (i.e., $\{\omega_1\}$ or Ω). To prove the only if part, suppose not so that there is an action profile a such that $\sum_{y \in Y} \pi^{\omega_2}(y, \omega_1 | a) = 0$. If the initial state is ω_2 and players choose a each period, the posterior belief always puts probability one on ω_2 , so the support stays at $\{\omega_2\}$ forever. Hence the set $\{\omega_2\}$ cannot be uniformly transient.

¹⁵This condition is deeply related to the notion of *scrambling matrices* in ergodic theory. To see this, pick an arbitrary initial state and arbitrary strategy profile, and pick a signal sequence (y^1, y^2) as stated. Let $M = (M_{ij})$ be a two-by-two stochastic matrix which maps the initial prior to the posterior belief in period three, *conditional on this signal sequence* (y^1, y^2) . Specifically, let $M_{ij} = \frac{\Pr(y^1, y^2, \omega^3 = \omega_j | \omega_i, s)}{\Pr(y^1, y^2 | \omega_i, s)}$ for each (i, j) with $\Pr(y^1, y^2 | \omega_i, s) > 0$. For (i, j) with $\Pr(y^1, y^2 | \omega_i, s) = 0$, we let $M_{ij} = M_{\tilde{i}j}$, where $\tilde{i} \neq i$. Then given any initial prior μ , the posterior belief in period three after observing (y^1, y^2) is indeed represented by μM . It is not difficult to see that our scrambling condition holds if and only if this stochastic matrix M is *scrambling*, in the sense that there is j such that $M_{ij} > 0$ for all i . (The proof of the only if part is straightforward. The if part follows from the fact that no states can be revealed and $M_{ij}M_{\tilde{i}j} < 1$ for each j .)

To summarize, uniform connected games are comprised of three different class of games.

- Games in which both states can be revealed. These games can be interpreted as a generalization of the standard stochastic games.
- Games in which only one state can be revealed, and the other state is not absorbing regardless of players' play.
- Games in which both states cannot be revealed, and the scrambling condition holds so that players cannot keep perfect information about the state.

In particular, if no states can be revealed, uniform connectedness reduces to the scrambling condition. Note that the scrambling condition is likely to be satisfied when the state changes stochastically *conditional on the signal realization*. For example, the natural resource management problem in Section 5.3 satisfies the scrambling condition, (and hence uniform connectedness) because the birth rate of fish is random regardless of how much fish was caught today.

On the other hand, when the state transition is deterministic, the scrambling condition is never satisfied. This in particular implies that if no states can be revealed and the state transition is deterministic, uniform connectedness is never satisfied and the payoff invariance condition may not hold. Consider the following example:

Example 1. Suppose that there is only one player, and she has two possible actions, A and B . There are two states, ω_A and ω_B , and the state transition is a deterministic cycle. That is, if the current state is ω_A , the next state is ω_B for sure, and vice versa. The stage-game payoff is 1 if the action matches the state (i.e., $g^{\omega_A}(A) = 1$ and $g^{\omega_B}(B) = 1$), but is -1 otherwise. There is only one signal y^0 , so the signal provides no information about the state.¹⁶ In this game, the scrambling condition is not satisfied, and no states can be revealed. So the singleton set $\{\omega\}$ is neither globally accessible nor uniformly transient, and the game is not uniformly connected. We can also show that the payoff invariance condition does not hold. To see this, note that if the player knows the initial state, then she can earn a

¹⁶Here we implicitly assume that payoffs are not observable until the game ends. See footnote 7 for more discussions about this assumption.

payoff of 1 each period, because she always knows the state. On the other hand, if the player does not know the state (say, the initial prior is $\frac{1}{2}-\frac{1}{2}$), then her expected payoff is 0 each period, because her posterior is always $\frac{1}{2}-\frac{1}{2}$. Accordingly, even if the player is patient, the best payoff in the infinite-horizon game depends on the initial prior.

A point in this example is that even though states are weakly communicating, the initial belief has a non-negligible impact on the posterior belief in a distant future; if the player knows the initial state, she can retain perfect information about the state even after a long time. The scrambling condition rules out such a possibility, and hence ensures the payoff invariance.

The same intuition carries over even when there are more than two states. If each state ω can be revealed, then the game is always uniformly connected. On the other extreme, if signals do not reveal any information (i.e., starting from an interior belief, the posterior is always an interior belief) then uniform connectedness requires the scrambling condition. There are many “intermediate” cases, depending on what information can be revealed. In general, when less states can be revealed, uniform connectedness requires a more restrictive assumption.

5.1.4 Invariance of the Feasible Payoff Set

The following proposition shows that the limit feasible payoff set is invariant, even if the full support assumption in Proposition 2 is replaced with uniform connectedness.

Proposition 5. *Under uniform connectedness, for each $\varepsilon > 0$, there is $\bar{\delta} \in (0, 1)$ such that for any $\lambda \in \Lambda$, $\delta \in (\bar{\delta}, 1)$, μ , and $\tilde{\mu}$,*

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

Hence Assumption 1(a) holds.

The proof of the proposition can be found in Appendix B. To describe a rough idea, pick a direction λ , and consider the score toward this direction λ . When players have better information about the initial state, they obtain higher scores. Hence the score is maximized by an initial prior which puts probability one on

some state. Pick such a state ω , call this score the *maximal score*. There are two cases to be considered:

Case 1: $\{\omega\}$ is globally accessible. In this case, given any initial belief μ , players can move the belief to the one which puts probability one on ω in finite time, and can earn the maximal score thereafter. This implies that the score for any initial prior μ is almost as good as the maximal score.

Case 2: $\{\omega\}$ is not globally accessible. Since the game is uniformly connected, $\{\omega\}$ must be uniformly transient. This means that $\{\omega\}$ is a temporary support. That is, starting from the belief which puts probability one on ω , the belief must eventually reach the ones whose supports are globally accessible, with probability one. Since payoffs before reaching these beliefs are almost negligible, this implies that there is at least *one* belief μ^* whose support (say Ω^*) is globally accessible and whose score is approximately as good as the maximal score. Then Lemma B3 in the proof implies that the same result holds for *all* beliefs with the same support, that is, given any belief $\mu \in \Delta\Omega^*$, the score is approximately as good as the maximal score. Since Ω^* is globally accessible, this immediately implies invariance of the score; indeed, given any initial prior, players can move the belief to some $\tilde{\mu} \in \Delta\Omega^*$ and can approximate the maximal score thereafter.

So the key step in the argument above is Lemma B3, which asserts that if the score for one belief μ^* is approximately as good as the maximal score, the same is true for any belief μ with the same support. This result follows from the fact that players can attain better scores when they have better information about the state. To illustrate the idea, suppose for now that the score for the belief μ^* is exactly (not approximately) equal to the maximal score. Pick an arbitrary belief $\mu \in \Delta\Omega^*$. Then we can find a belief $\tilde{\mu}$ such that the belief μ^* is a convex combination of μ and $\tilde{\mu}$, that is, $\mu^* = \alpha\mu + (1 - \alpha)\tilde{\mu}$ for some $\alpha \in (0, 1)$. Let v and \tilde{v} denote the scores given the beliefs μ and $\tilde{\mu}$, respectively. Then the weighted average of the scores, $\alpha v + (1 - \alpha)\tilde{v}$, must be at least the score for the “muddled” belief μ^* . But since the score for this belief μ^* is exactly equal to the maximal score, the weighted average $\alpha v + (1 - \alpha)\tilde{v}$ must be also equal to the maximal score. Hence both v and \tilde{v} must be equal to the maximal score; i.e., the score for the belief μ is indeed equal to the maximal score.

5.1.5 Uniform Connectedness and Weakly Communicating States

Uniform connectedness is an analogue of weakly communicating states, and these two conditions are deeply related. The following proposition shows that weakly communicating states are necessary for uniform connectedness. The proof can be found in Appendix B.

Proposition 6. *The game is uniformly connected only if states are weakly communicating.*

This proposition implies that if the state transition rule does not satisfy the standard assumption for the standard stochastic game, then the game is never uniformly connected. For example, as noted earlier, if there are multiple absorbing states, then states are not weakly communicating. So the above proposition implies that such a game is never uniformly connected, regardless of the signal structure. Indeed, it is easy to see that if there are multiple absorbing states and if these absorbing states give different stage-game payoffs, then the feasible payoff set must depend on the initial prior, even in the hidden-state model.

For some class of games, the necessary condition above is “tight,” in the sense that it is necessary and sufficient for uniform connectedness. Specifically, we have the following proposition:

Proposition 7. *In stochastic games with observable states, the game is uniformly connected if and only if states are weakly communicating. Similarly, in stochastic games with delayed observations, the game is uniformly connected if and only if states are weakly communicating.*

So in these class of games, if states are weakly communicating, then the feasible payoff set is invariant to the initial prior. This result generalizes the invariance result of Dutta (1995), who assumes communicating states (rather than weakly communicating states).

Note that uniform connectedness is a sufficient condition for invariance of the feasible payoffs, but not necessary. So there are many cases in which uniform connectedness is not satisfied but nonetheless the feasible payoffs are invariant to the initial prior. To cover such cases, in Appendix A, we show that the invariance result holds even if uniform connectedness is replaced with a weaker condition called *asymptotic uniform connectedness*. Asymptotic uniform connectedness is

satisfied in a broad class of games; for example, as shown in Proposition A1, it is satisfied if states are weakly communicating and if states can be statistically distinguished by signals, in that for each fixed action profile a , the signal distributions $\{(\pi_Y^\omega(y|a))_{y \in Y} | \omega \in \Omega\}$ are linearly independent. This result ensures that weakly communicating states “almost always” imply invariance of the feasible payoffs, even in the hidden-state model. Indeed, if states are weakly communicating (here we allow a deterministic state transition) and if the signal space is large enough that $|Y| \geq |\Omega|$, then asymptotic uniform connectedness holds for generic signal distributions.

5.2 Robust Connectedness and Minimax Payoffs

5.2.1 Weak Irreducibility

Again, before studying the hidden-state model, we consider the observable-state case and show that *weak irreducibility* is sufficient for invariance of the minimax payoff. It is useful to understand this weak irreducibility condition, because robust connectedness, which will play a central role in this subsection, is an analogue of this condition for the hidden-state model.

A state ω is *robustly accessible despite i* if for each initial state $\tilde{\omega}$, there is a (possibly mixed) action sequence $(\alpha_{-i}^1, \dots, \alpha_{-i}^{|\Omega|})$ such that for any player i 's strategy s_i , there is a natural number $T \leq |\Omega|$ such that $\Pr(\omega^{T+1} = \omega | \tilde{\omega}, s_i, \alpha_{-i}^1, \dots, \alpha_{-i}^T) > 0$. In words, robust accessibility requires that the opponents can move the state to ω regardless of player i 's play. Clearly, this condition is more demanding than global accessibility introduced in the previous subsection.

A state ω is *avoidable for player i* if it is not robustly accessible despite i and there is player i 's action sequence $(\alpha_i^1, \dots, \alpha_i^{|\Omega|})$ such that for any strategy s_{-i} of the opponent, there is $T \leq |\Omega|$ and a state $\tilde{\omega}$ which is robustly accessible despite i such that $\Pr(\omega^{T+1} = \tilde{\omega} | \omega, \alpha_i^1, \dots, \alpha_i^T, s_{-i}) > 0$. So player i can avoid the state to stay at ω forever; if she chooses a particular action sequence, the state must move to some robustly accessible state with positive probability, regardless of the opponents' play. This condition is somewhat similar to uniform transience of the state, but note that we fix player i 's action sequence in the definition of avoidability. So if player i chooses other actions, the state may stay at ω forever. In contrast, uniform transience requires that the state cannot stay at ω regardless

of players' play. So avoidability of ω does not imply uniform transience of ω .

States are *weakly irreducible for player i* if each state ω is either robustly accessible despite i or avoidable for i . States are *weakly irreducible* if they are weakly irreducible for all players. This condition is somewhat similar to weakly communicating states in the previous subsection, but neither implies the other. Indeed, weakly communicating states need not imply weakly irreducible states, because global accessibility of ω does not imply robust accessibility of ω . Similarly, weakly irreducible states need not imply weakly communicating states, because as mentioned above, avoidability of ω does not imply uniform transience of ω . Note that weak irreducibility here is a generalization of irreducibility of Fudenberg and Yamamoto (2011b), which requires that all states be robustly accessible.

If states are weakly irreducible for player i , then the limit minimax payoff for player i is invariant to the initial state ω . This result follows from Propositions 8 and 9 below, but a rough idea is as follows. Let ω be the initial state which gives the lowest minimax payoff for player i . There are two cases to be considered:

Case 1: ω is robustly accessible. In this case, given any initial state, the opponent can move the state to ω in finite time with probability one, and give the lowest minimax payoff to player i after that. When player i is patient, payoffs before the state reaching ω is almost negligible. Hence for any initial state, player i 's minimax payoff is approximately as low as the lowest one.

Case 2: ω is not robustly accessible. In this case, the state ω is avoidable for player i , so she can "escape" from this worst state. That is, even if the initial state is ω and the opponent plays the minimax strategy, with probability one, player i can move the state to some robustly accessible states in finite time, and after that she earn at least the minimax payoffs for these states. Accordingly, there must be at least one robustly accessible state ω^* whose minimax payoff is approximately as low as the lowest minimax payoff. Then as in the previous case, we can show that for any initial state, the minimax payoff is approximately as low as the lowest one.

5.2.2 Invariance of the Minimax Payoff

Now we consider the hidden-state model, and introduce the notion of *robust connectedness* as an analogue of weak irreducibility. This condition is weaker than the full support assumption but still ensures invariance of the limit minimax payoffs. We first define robust accessibility of the support, which is an analogue of robust accessibility of the state.

Definition 5. 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 a natural number T and an action sequence $(\alpha_{-i}^1, \dots, \alpha_{-i}^T)$ such that for any strategy s_i , there is a natural number $t \leq T$ and a belief $\tilde{\mu}$ with support Ω^* such that¹⁷

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

In the definition above, the support of the resulting belief $\tilde{\mu}$ must be precisely equal to Ω^* . This is more demanding than global accessibility, which allows the support to be a subset of Ω^* . Clearly, robust accessibility of Ω^* implies globally accessibility in the previous subsection.

Next, we define avoidability of the support, which is again an analogue of avoidability of the state ω .

Definition 6. A subset $\Omega^* \subseteq \Omega$ is *avoidable for 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 $(\alpha_i^1, \dots, \alpha_i^T)$ such that for any strategy s_{-i} of the opponents, there is a natural number $t \leq T$ and a belief $\tilde{\mu}$ whose support is robustly accessible despite i such that¹⁸

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

In order to state robust connectedness, we need one more definition:

Definition 7. Supports are *merging* 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

¹⁷As shown in Appendix C, when we check if a given set is robustly accessible, we can restrict attention to $T \leq 4^{|\Omega|}$, without loss of generality. Note also that replacing the action sequence $(\alpha_{-i}^1, \dots, \alpha_{-i}^{4^{|\Omega|}})$ in this definition with a strategy s_{-i} does not relax the condition at all.

¹⁸As shown in Appendix C, when we check if a given set is avoidable, we can restrict attention to $T \leq 4^{|\Omega|}$, without loss of generality.

$\Pr(h^T|\omega, s) > 0$ and such that after the history h^T , the support of the posterior belief induced by the initial state ω is the same as the one induced by the initial prior $\mu = (\frac{1}{|\Omega|}, \dots, \frac{1}{|\Omega|})$.

The merging support condition ensures that regardless of players' play, two different initial priors ω and $\mu = (\frac{1}{|\Omega|}, \dots, \frac{1}{|\Omega|})$ induce posteriors with the same support, after some history. 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 merging support condition holds.

To understand why we need this merging support condition, recall that in the proof of Proposition 3, we consider player i 's payoff $v_i^{\tilde{\mu}}(s_{-i}^{\mu})$ when the opponents play the minimax strategy for some belief μ but the actual initial prior is $\tilde{\mu} \neq \mu$. The full support assumption ensures that after one period, these two beliefs μ and $\tilde{\mu}$ induce posterior beliefs with the same support (the whole state space), and this property plays an important role when we evaluate this payoff. In order to use a similar proof technique, we need a similar property here, and the merging support condition is precisely what we need.

Definition 8. The game is *robustly connected for player i* if supports are merging and each non-empty subset $\Omega^* \subseteq \Omega$ is robustly accessible despite i or avoidable for i . The game is *robustly connected* if it is robustly connected for all players.

Robust connectedness and uniform connectedness may look somewhat similar, but neither implies the other. Indeed, uniform connectedness does not imply robust connectedness because global accessibility of Ω^* does not imply robust accessibility of Ω^* . Also, robust connectedness does not imply uniform connectedness, because avoidability of Ω^* does not imply uniform transience of Ω^* . This is analogous to the fact that in the observable-state case, weakly communicated states does not imply weakly irreducible states, and vice versa.

Robust connectedness is a complicated condition, because it requires the merging support condition, in addition to robust accessibility and avoidability. Accordingly, even when there are only two states, the description of robust connectedness is not as simple as one may wish. However, if we focus on the special case in which no states can be revealed, we can show that the scrambling condition is necessary and sufficient for robust connectedness.¹⁹ As explained in the previ-

¹⁹If there are only two states and none of them can be revealed, the scrambling condition and

ous subsection, the scrambling condition is likely to be satisfied when the state changes stochastically conditional on the signal realization. Note that the scrambling condition is sufficient for uniform connectedness as well, so it is sufficient for both robust connectedness and uniform connectedness.

The following proposition shows that robust connectedness implies invariance of the minimax payoffs. The proof is given in Appendix B.

Proposition 8. *Suppose that the game is robustly connected for player i . Then Assumption 1(b) holds.*

From Propositions 5 and 8, for any games which satisfies both uniform connectedness and robust connectedness, the payoff invariance assumption (Assumption 1) holds, and hence the folk theorem obtains.

5.2.3 Robust Connectedness and Weakly Irreducible States

Robust connectedness is an analogue of weak irreducibility, and these two conditions are closely related. The following proposition shows that weak irreducibility is necessary for robust connectedness. Also, it shows that weak irreducibility is necessary and sufficient for robust connectedness in the standard stochastic games. The proof is very similar to that of Proposition 6 and hence omitted.

Proposition 9. *The game is robustly connected only if the game is weakly irreducible. In particular, for stochastic games with observable states, the game is robustly connected if and only if the game is weakly irreducible.*

Unfortunately, the second result in Proposition 7 does not extend, that is, for stochastic games with delayed observations, weak irreducibility is not sufficient for robust connectedness. 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, specifically, one of the player is randomly selected and her action determines the state tomorrow. For example, if one player

the merging support condition are equivalent. Hence the scrambling condition is necessary for robust connectedness. It is also sufficient for robust connectedness, because under the scrambling condition, each singleton set $\{\omega\}$ is avoidable and the whole state space Ω is robustly accessible.

chooses A and the opponent chooses B , then ω_A and ω_B are equally likely. So regardless of the opponents' play, if a player chooses A , then ω_A will appear with probability at least $\frac{1}{2}$. This implies 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 avoidable. 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 avoidable either.

Note, however, that robust connectedness is just a sufficient condition for invariance of the limit minimax payoff. The following proposition shows that, for stochastic games with delayed observations, weak irreducibility implies invariance of the limit minimax payoff. The proof relies on the fact that there are only finitely many possible posterior beliefs for games with observation delays; see Appendix B.

Proposition 10. *Consider stochastic games with delayed observations, and suppose that the game is weakly irreducible. Then Assumption 1(b) holds.*

5.3 Example: Natural Resource Management

Now we will present an example of natural resource management. This is an example which satisfies uniform connectedness and robust connectedness, but does not satisfy the full support assumption.

Suppose that two fishermen live near a gulf. The state of the world is the number of fish in the gulf, and is denoted by $\omega \in \{0, \dots, K\}$ where K is the maximal capacity. The fishermen cannot directly observe the number of fish, ω , so they have a belief about ω .

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 and hence $y_i = 0$. 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:

$$\omega^{t+1} = \omega^t - (y_1^t + y_2^t) + \varepsilon^t. \quad (4)$$

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 $\Omega = \{0, \dots, K\}$, 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 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. 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.

This example does not satisfy the full support assumption, because the probability of $\omega^{t+1} = K$ is zero if $y_1 + y_2 > 1$. However, as will be explained, uniform

connectedness and robust connectedness are satisfied, so that the payoff invariance condition (and hence the folk theorem) obtains.

To see that this game is indeed uniformly connected, note that this example satisfies the scrambling condition; due to the possibility of natural increase and decrease, given any initial belief μ and given any fishermen's play s , the posterior belief becomes an interior belief (i.e., the support of the posterior becomes the whole state space Ω) if the signal $y = (0, 0)$ is observed for the first K periods. As noted earlier, if the scrambling condition holds, then any singleton set $\{\omega\}$ is uniformly transient, and hence uniform connectedness holds. For the same reason, robust connectedness also holds.

So far we have assumed that $\Pr(\varepsilon|\omega, a, y) > 0$, except the case in which the state does not stay in the space $\{0, \dots, K\}$. Now, modify the model and suppose that $\Pr(\varepsilon = 1|\omega, a, y) = 0$ if $\omega - y_1 - y_2 = 0$ and $a \neq (N, N)$. That is, if the resource is exhausted ($\omega - y_1 - y_2 = 0$) and at least one player tries 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; so the fishermen need to "wait" until the fish grows and the state exceeds this critical level. We still assume that $\Pr(\varepsilon|\omega, a, y) > 0$ for all other cases.

In this new example, players' actions have a significant impact on the state transition, that is, the state *never* increases if the current state is $\omega = 0$ and someone chooses F . This complicates the belief evolution process, and the scrambling condition does not hold anymore. Indeed, if the initial state is $\omega = 0$ and the fishermen choose (F, F) , the belief does not change forever and never become an interior belief.

Nonetheless, the payoff invariance condition (and hence the folk theorem) still holds in this setup. Specifically, we can show that uniform connectedness holds, and thus the feasible payoff set is invariant to the initial prior. Also, while robust connectedness does not hold (indeed, the merging support condition does not hold here), we can compute the minimax payoff for each initial prior and can prove its invariance.

To prove uniform connectedness, note first that each singleton set $\{\omega\}$ is uniformly transient, except $\{0\}$. The reason is exactly the same as in the previous case: Suppose that the initial belief puts probability one on some $\omega \geq 1$. Due to the possibility of natural increase and decrease, if $y = (0, 0)$ is observed for the

first K periods (note that this happens with positive probability regardless of the strategy profile), then the posterior belief becomes an interior belief, and the support reaches the globally accessible set Ω . Hence the set $\{\omega\}$ is indeed uniformly transient.

How about the set $\{0\}$? This set is not uniformly connected, because if the initial prior puts probability one on $\omega = 0$ and someone fishes every period, the posterior belief never changes and the support stays at $\{0\}$ forever. However, we can show that $\{0\}$ is globally accessible. A point is that regardless of the initial prior, the state $\omega = 0$ can be revealed if

- The fishermen do not fish in the first K periods, and then
- Both of them fish and observe $y = (1, 1)$ in the next $K - 1$ periods.

Given any initial prior, after waiting for the first K periods, the posterior belief μ^{K+1} assigns at least probability $\bar{\pi}^K$ on the highest state $\omega = K$ (i.e., $\mu^{K+1}(K) \geq \bar{\pi}^K$). Then if $y = (1, 1)$ is observed in the next period, the posterior belief μ^{K+2} puts probability zero on the highest state $\omega = K$; this is so because the fishermen caught fish more than the natural increase. For the same reason, after observing $y = (1, 1)$ for $K - 1$ periods, the posterior belief puts probability zero on all states but $\omega = 0$, so the state $\omega = 0$ is indeed revealed. Note that the probability of observing $y = (1, 1)$ for $K - 1$ periods is $\mu^{K+1}(K)\bar{\pi}^{K-1} \geq \bar{\pi}^{2K-1}$, so there is a lower bound on the probability of the support reaching $\{0\}$. Hence $\{0\}$ is indeed globally accessible, and thus the game is uniformly connected. This implies that feasible payoffs are invariant to the initial prior.

As noted earlier, we can also show that the minimax payoff is invariant to the initial prior. To see this, note first that a fisherman can obtain at least a payoff of 0 by choosing “Always N .” Hence the limit minimax payoff is at least 0. On the other hand, if the opponent always chooses F , the state eventually reaches $\omega = 0$ with probability one, and thus fisherman i ’s payoff is at most 0 in the limit as $\delta \rightarrow 1$. Thus the limit minimax payoff is 0 regardless of the initial prior.

6 Concluding Remarks

This paper considers a new class of stochastic games in which the state is hidden information. We find that the folk theorem holds when the feasible and individ-

ually rational payoffs are invariant to the initial prior. Then we find sufficient conditions for this payoff invariance condition.

Throughout this paper, we assume that actions are perfectly observable. In an ongoing project, we consider how the equilibrium structure changes when actions are not observable; in this new setup, each player has private information about her actions, and thus different players may have different beliefs. This implies that a player's belief is not public information and cannot be regarded as a common state variable. Accordingly, the analysis of the imperfect-monitoring case is very different from that for the perfect-monitoring case.

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 $\delta \rightarrow 1$ 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.

Appendix A: Extension of Uniform Connectedness

Proposition 5 shows that uniform connectedness ensures invariance of the feasible payoff set. Here we show that the same result holds under 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 states are communicating but nonetheless uniform connectedness does not hold. These examples show that Proposition 7 does not extend to the hidden-state case; the game may not be uniformly connected even if states are communicating.

Example A1. Suppose that there are only two states, $\Omega = \{\omega_1, \omega_2\}$, and that the state evolution is a deterministic cycle, as in Example 1. That is, the state goes to ω_2 for sure if the current state is ω_1 , and vice versa. Assume that there are at least two signals, and 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 $\pi_1 \neq \pi_2$. Assume also that the signal does not reveal the state ω , that is, $\pi_Y^\omega(y|a) > 0$ for all ω , a , and y . As in Example 1, this game does not satisfy the scrambling condition, and no states can be revealed. Hence the game is not uniformly connected.

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

Example A2. 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 $\{\omega_2\}$.

While uniform connectedness does not hold in these examples, the feasible payoffs are still invariant to the initial prior. To describe the idea, consider Example A1. In this example, if the initial state is ω_1 , then the true state must be ω_1 in all odd periods, so the empirical distribution of the signals in odd periods should approximate π_1 with probability close to one. Similarly, if the initial state is ω_2 , the empirical distribution of the public signals in odd periods should approximate π_2 . This suggests 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 must be invariant to the initial prior.

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 A1. 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, \dots, a^{T^*}) such that $\Pr(\mu^{T^*+1} = \tilde{\mu} | \mu, a^1, \dots, a^{T^*}) \geq \pi^*$ 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, \dots, 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 A1. 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 A2 is asymptotically 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 $\{\omega_2\}$.

Next, we give the definition of asymptotic uniform transience, which extends uniform transience.

Definition A2. 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^* .²⁰

Definition A3. 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 A1 and A2 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 the posterior belief in

²⁰Asymptotic uniform transience requires $\tilde{\mu}(\tilde{\omega}) \geq \tilde{\pi}^*$, that is, the posterior belief $\tilde{\mu}$ is not close to the boundary of $\Delta\Omega^*$. We can show that this condition is automatically satisfied in the definition of uniform transience, if $\{\omega\}$ is uniformly transient; so uniform transience implies asymptotic uniform transience.

a distant future. However, the following proposition provides a simple sufficient condition for asymptotic uniform connectedness:

Proposition A1. *The game is asymptotically uniformly connected if states are weakly communicating, and for each action profile a and each proper subset $\Omega^* \subset \Omega$,*

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

In words, the game is asymptotically uniformly connected if states are weakly communicating 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 B 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 states are weakly communicating and the signal space is large enough.

The following proposition shows that the feasible payoff set is indeed invariant to the initial prior if the game is asymptotically uniformly connected.²¹ The proof can be found in Appendix B.

Proposition A2. *If the game is asymptotically uniformly connected, then for each $\varepsilon > 0$, there is $\bar{\delta} \in (0, 1)$ such that for any $\lambda \in \Lambda$, $\delta \in (\bar{\delta}, 1)$, 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 show that the minimax payoff is invariant to the initial prior under a condition weaker than robust connectedness. The idea is quite similar to the one discussed above; we can relax robust accessibility, avoidability, and the merging support condition, just as we did for global accessibility and uniform transience. Details are omitted.

²¹However, unlike Proposition 5, we do not know the rate of convergence, and in particular, we do not know if we can replace ε in the proposition with $O(1 - \delta)$.

Appendix B: Proofs

B.1 Proof of Proposition 1 with Mixed Minimax Strategies

Here we explain how to extend the proof provided in Section 3.4 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 $k \in \{1, \dots, K\}$, let $(\mu^{(k)}, a^{(k)})$ denote the belief and the action profile in the last period of the k th 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 $R_i(\mu^{(1)}, a^{(1)})$, \dots , $R_i(\mu^{(K)}, a^{(K)})$, 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, consider player i 's incentive in the last block of the minimax play. We will ignore the term $R_i(\mu^{(k)}, a^{(k)})$ for $k < K$, as it does not influence player i 's incentive in this block. If we are now in the τ th period of the block, player i 's unnormalized payoff in the continuation game from now on is

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

Here, (μ^t, a^t) denote the belief and the action in the t th period of the continuation game, so the first term of the above display is the expected payoff until the current block ends. The second term is the continuation payoff from the next block; $(1-p)p^{t-1}$ is the probability of period t being the last period of the block, in which

case player i 's continuation payoff is $w_i(j) + \frac{(1-\delta)E[R_i(\mu^t, a^t)]}{\delta(1-p)}$ where the expectation is taken with respect to μ^t and a^t , conditional on that the block does not terminate until period t . We have the term δ^t due to discounting, and we have $\frac{1}{1-\delta}$ 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 minimax blocks. The only difference is that if the current block is the k th block with $k < K$, the corresponding perturbation payoff $R_i(\mu^{(k)}, a^{(k)})$ will not be paid at the end of the current block; it will be paid after the K th block ends. To offset discounting, we have the coefficient $\frac{(1-p\delta)^{K-k}}{\{\delta(1-p)\}^{K-k+1}}$ on $R_i(\mu^{(k)}, a^{(k)})$. To see how it works, suppose that we are now in the second to the last block (i.e., $k = K - 1$). The ‘‘expected discount factor’’ due to the next random block is

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

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 . This discount factor $\frac{\delta(1-p)}{1-p\delta}$ cancels out, thanks to the coefficient $\frac{(1-p\delta)}{\{\delta(1-p)\}^2}$ on $R_i(\mu^{(K-1)}, a^{(K-1)})$. Hence player i is indifferent in all periods during the this block.

So far we have explained that player i is indifferent in all periods during the minimax play. Note also that the perturbed payoff approximates the original payoff $w_i(j)$ for δ close to one, 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.

B.2 Proof of Proposition 3: Invariance of the Minimax Payoffs

We will first prove that the minimax payoffs are invariant to the initial prior for high discount factors. That is, we will show that for any $\varepsilon > 0$, there is $\bar{\delta}$ such that

$$\left| \underline{v}_i^\mu(\delta) - \underline{v}_i^{\tilde{\mu}}(\delta) \right| < \varepsilon \quad (5)$$

for any $\delta \in (\bar{\delta}, 1)$, μ , and $\tilde{\mu}$. After that, we will show that the limit of the minimax payoff exists.

Fix δ , and let s^μ denote the minimax strategy profile given the initial prior μ . As in Section 4.3, let $v_i^{\tilde{\mu}}(s_{-i}^\mu) = \max_{s_i \in S_i} v_i^{\tilde{\mu}}(\delta, s_i, s_{-i}^\mu)$. That is, $v_i^{\tilde{\mu}}(s_{-i}^\mu)$ denotes player i 's maximal payoff when the opponents use the minimax strategy for the belief μ while the actual initial prior is $\tilde{\mu}$. Note that this payoff $v_i^{\tilde{\mu}}(s_{-i}^\mu)$ is convex with respect to the initial prior $\tilde{\mu}$, as it is the upper envelope of the linear functions $v_i^{\tilde{\mu}}(\delta, s_i, s_{-i}^\mu)$ over all s_i .

In Section 4.3, we have defined the *maximal value* as the maximum of these payoffs $v_i^{\tilde{\mu}}(s_{-i}^\mu)$ over all $(\mu, \tilde{\mu})$. But this definition is informal, because the maximum with respect to μ may not exist. To fix this problem, given the opponents' strategy s_{-i}^μ , define

$$\bar{v}_i(s_{-i}^\mu) = \max_{\tilde{\mu} \in \Delta_\Omega} v_i^{\tilde{\mu}}(s_{-i}^\mu),$$

as the maximum of player i 's payoff with respect to the initial prior $\tilde{\mu}$. Then choose μ^* so that

$$\left| \sup_{\mu \in \Delta_\Omega} \bar{v}_i(s_{-i}^\mu) - \bar{v}_i(s_{-i}^{\mu^*}) \right| < 1 - \delta,$$

and call $\bar{v}_i(s_{-i}^{\mu^*})$ the *maximal value*. When δ is close to one, this maximal value indeed approximates the supremum of the payoff $v_i^{\tilde{\mu}}(s_{-i}^\mu)$ over all $(\mu, \tilde{\mu})$. Since $v_i^{\tilde{\mu}}(s_{-i}^{\mu^*})$ is convex with respect to $\tilde{\mu}$, it is maximized when $\tilde{\mu}$ puts probability one on some state. Let ω denote this state, so that $v_i^\omega(s_{-i}^{\mu^*}) \geq v_i^{\tilde{\mu}}(s_{-i}^{\mu^*})$ for all $\tilde{\mu}$.

B.2.1 Step 0: Preliminary Lemma

The following lemma follows from the convexity of the payoffs $v_i^{\tilde{\mu}}(s_{-i}^\mu)$. We will use this lemma repeatedly throughout the proof.

Lemma B1. Take an arbitrary belief μ , and an arbitrary interior belief $\tilde{\mu}$. Let $p = \min_{\tilde{\omega} \in \Omega} \tilde{\mu}(\tilde{\omega})$, which measures the distance from $\tilde{\mu}$ to the boundary of $\Delta\Omega$. Then for each $\hat{\mu} \in \Delta\Omega$,

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

Roughly, this lemma asserts that given the opponents' strategy s_{-i}^{μ} , if player i 's payoff $v_i^{\tilde{\mu}}(s_{-i}^{\mu})$ approximates the maximal value $\bar{v}_i(s_{-i}^{\mu^*})$ for *some* interior initial prior $\tilde{\mu}$, then the same is true for *all other* initial priors $\hat{\mu}$.

More formally, given the opponents' strategy s_{-i}^{μ} , suppose that player i 's payoff $v_i^{\tilde{\mu}}(s_{-i}^{\mu})$ approximates the maximal value $\bar{v}_i(s_{-i}^{\mu^*})$ for *some* interior belief $\tilde{\mu}$ such that $\tilde{\mu}(\tilde{\omega}) \geq \bar{\pi}$ for all $\tilde{\omega}$. The condition $\tilde{\mu}(\tilde{\omega}) \geq \bar{\pi}$ implies that this belief $\tilde{\mu}$ is not too close to the boundary of the belief space $\Delta\Omega$. Then the right-hand side of the inequality in the lemma is approximately zero, as $p \geq \bar{\pi}$. Hence the left-hand side must be approximately zero, which indeed implies that the payoff $v_i^{\hat{\mu}}(s_{-i}^{\mu})$ approximates the maximal value for all $\hat{\mu}$.

In the interpretation above, it is important that the belief $\tilde{\mu}$ is not too close to the boundary of $\Delta\Omega$. If $\tilde{\mu}$ approaches the boundary of $\Delta\Omega$, then p approaches zero so that the right-hand side of the inequality in the lemma becomes arbitrarily large.

Proof. Pick μ , $\tilde{\mu}$, 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 $v_i^{\hat{\omega}}(\delta, s_i, s_{-i}^{\mu}) \leq \bar{v}_i(s_{-i}^{\mu^*}) + (1 - \delta)$ for each $\hat{\omega} \neq \tilde{\omega}$, we obtain

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

Arranging,

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

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

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

Since $\hat{\mu}(\tilde{\omega}) \geq p$, we have

$$\left| \bar{v}_i(s_{-i}^{\mu^*}) + (1 - \delta) - v_i^{\tilde{\omega}}(\delta, s_i, s_{-i}^{\mu}) \right| \leq \frac{\left| \bar{v}_i(s_{-i}^{\mu^*}) + (1 - \delta) - v_i^{\hat{\mu}}(s_{-i}^{\mu}) \right|}{p}. \quad (6)$$

Now, pick an arbitrary $\hat{\mu} \in \Delta\Omega$. Note that (6) holds for each $\tilde{\omega} \in \Omega$. So multiplying both sides of (6) by $\hat{\mu}(\tilde{\omega})$ and summing over all $\tilde{\omega} \in \Omega$,

$$\sum_{\tilde{\omega} \in \Omega} \hat{\mu}(\tilde{\omega}) \left| \bar{v}_i(s_{-i}^{\mu^*}) + (1 - \delta) - v_i^{\tilde{\omega}}(\delta, s_i, s_{-i}^{\mu}) \right| \leq \frac{\left| \bar{v}_i(s_{-i}^{\mu^*}) + (1 - \delta) - v_i^{\hat{\mu}}(s_{-i}^{\mu}) \right|}{p}. \quad (7)$$

Then we have

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

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 (7). *Q.E.D.*

B.2.2 Step 1: Minimax Payoff for Some Belief μ^{**}

In this step, we will show that there is an interior belief μ^{**} whose minimax payoff approximates the maximal value and such that $\mu^{**}(\tilde{\omega}) \geq 0$ for all $\tilde{\omega}$.

To do so, we carefully inspect the maximal value. Suppose that the initial state is ω and the opponents play $s_{-i}^{\mu^*}$. Suppose that player i takes a best reply, which is denoted by s_i , so that she achieves the maximal value $v_i^{\omega}(s_{-i}^{\mu^*})$. As usual, this payoff can be decomposed into the payoff today and the expected continuation payoff:

$$v_i^{\omega}(s_{-i}^{\mu^*}) = (1 - \delta)g_i^{\omega}(\alpha^*) + \delta \sum_{a \in A} \alpha^*(a) \sum_{y \in Y} \pi_Y^{\omega}(y|a) v_i^{\mu}(y|\omega, a)(s_{-i}^{\mu(y|\mu^*, a)}).$$

Here, α^* denotes the action profile in period one induced by $(s_i, s_{-i}^{\mu^*})$. $\mu(y|\omega, a)$ denotes the posterior belief in period two when the initial belief is $\tilde{\mu}^* = \omega$ and players play a and observe y in period one. $\mu(y|\mu^*)$ denotes the posterior belief when the initial belief is μ^* . Given an outcome (a, y) in period one, player i 's continuation payoff is $v_i^{\mu(y|\omega, a)}(s_{-i}^{\mu(y|\mu^*, a)})$, because her posterior is $\mu(y|\omega, a)$ while the opponents' continuation strategy is $s_{-i}^{\mu(y|\mu^*, a)}$. (Note that the minimax strategy is Markov.)

The following lemma shows that there is some outcome (a, y) such that player i 's continuation payoff $v_i^{\mu(y|\omega, a)}(s_{-i}^{\mu(y|\mu^*, a)})$ approximates the maximal value.

Lemma B2. *There is (a, y) such that $\alpha^*(a) > 0$ and such that*

$$\left| v_i^\omega(s_{-i}^{\mu^*}) + (1 - \delta) - v_i^{\mu(y|\omega, a)}(s_{-i}^{\mu(y|\mu^*, a)}) \right| \leq \frac{(1 - \delta)(2\bar{g} + 1)}{\delta}.$$

Proof. Pick (a, y) which maximizes the continuation payoff $v_i^{\mu(y|\omega, a)}(s_{-i}^{\mu(y|\mu^*, a)})$ over all y and a with $\alpha^*(a) > 0$. This highest continuation payoff is at least the expected continuation payoff, so we have

$$v_i^\omega(s_{-i}^{\mu^*}) \leq (1 - \delta)g_i^\omega(\alpha^*) + \delta v_i^{\mu(y|\omega, a)}(s_{-i}^{\mu(y|\mu^*, a)}).$$

Arranging,

$$\left| v_i^\omega(s_{-i}^{\mu^*}) - v_i^{\mu(y|\omega, a)}(s_{-i}^{\mu(y|\mu^*, a)}) \right| \leq \frac{1 - \delta}{\delta} (g_i^\omega(\alpha^*) - v_i^\omega(s_{-i}^{\mu^*})).$$

This implies

$$\left| v_i^\omega(s_{-i}^{\mu^*}) + (1 - \delta) - v_i^{\mu(y|\omega, a)}(s_{-i}^{\mu(y|\mu^*, a)}) \right| \leq \frac{(1 - \delta)(g_i^\omega(\alpha^*) - v_i^\omega(s_{-i}^{\mu^*}) + 1)}{\delta}.$$

Since $g_i^\omega(\alpha^*) - v_i^\omega(s_{-i}^{\mu^*}) \leq 2\bar{g}$, we obtain the desired inequality. *Q.E.D.*

Pick (a, y) as in the lemma above, and let $\mu^{**} = \mu(y|\mu^*, a)$. Then the above lemma implies that

$$\left| v_i^\omega(s_{-i}^{\mu^*}) + (1 - \delta) - v_i^{\mu(y|\omega, a)}(s_{-i}^{\mu^{**}}) \right| \leq \frac{(1 - \delta)(2\bar{g} + 1)}{\delta}.$$

That is, given the opponents' strategy $s_{-i}^{\mu^{**}}$, player i 's payoff $v_i^{\tilde{\mu}}(s_{-i}^{\mu^{**}})$ approximates the maximal value for some belief $\tilde{\mu} = \mu(y|\omega, a)$. Note that under the full support assumption, $\mu(y|\omega, a)[\tilde{\omega}] \geq \bar{\pi}$ for all $\tilde{\omega}$. Hence Lemma B1 ensures that

$$\left| v_i^\omega(s_{-i}^{\mu^*}) + (1 - \delta) - v_i^{\tilde{\mu}}(s_{-i}^{\mu^{**}}) \right| \leq \frac{(1 - \delta)(2\bar{g} + 1)}{\bar{\pi}\delta}$$

for all $\hat{\mu}$. That is, player i 's payoff $v_i^{\hat{\mu}}(s_{-i}^{\mu^{**}})$ approximates the maximal score for *all* initial priors $\hat{\mu}$. In particular, by letting $\hat{\mu} = \mu^{**}$, we can conclude that the minimax payoff for the belief μ^{**} approximates the maximal value. That is,

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

B.2.3 Step 2: Minimax Payoffs for Other Beliefs

Now we will show that the minimax payoff approximates the maximal value for any belief μ , which implies invariance of the minimax payoff.

Pick an arbitrary belief μ . Suppose that the opponents play the minimax strategy s^μ for this belief μ but the actual initial prior is μ^{**} . Then player i 's payoff $v_i^{\mu^{**}}(s_{-i}^\mu)$ is at least the minimax payoff for μ^{**} , by the definition of the minimax payoff. At the same time, her payoff cannot exceed the maximal value $v_i^{\omega}(s_{-i}^{\mu^*}) + (1 - \delta)$. So we have

$$v_i^{\mu^{**}}(s_{-i}^{\mu^{**}}) \leq v_i^{\mu^{**}}(s_{-i}^\mu) \leq v_i^{\omega}(s_{-i}^{\mu^*}) + (1 - \delta).$$

Then from the last inequality in the previous step, we have

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

That is, the payoff $v_i^{\tilde{\mu}}(s_{-i}^\mu)$ approximates the maximal value for some belief $\tilde{\mu} = \mu^{**}$. Then from Lemma B1,

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

for all beliefs $\hat{\mu}$. This implies that the minimax payoff for μ approximates the maximal value, as desired. Hence (5) follows.

B.2.4 Step 3: Existence of the Limit Minimax Payoff

Now we will verify that the limit of the minimax payoff exists. Take i , μ , and $\varepsilon > 0$ arbitrarily. Let $\bar{\delta} \in (0, 1)$ be such that

$$\left| v_i^\mu(\bar{\delta}) - \liminf_{\delta \rightarrow 1} v_i^\mu(\delta) \right| < \frac{\varepsilon}{2} \quad (8)$$

and such that

$$\left| v_i^\mu(\bar{\delta}) - v_i^{\tilde{\mu}}(\bar{\delta}) \right| < \frac{\varepsilon}{2} \quad (9)$$

for each $\tilde{\mu}$. Note that the result in Step 2 guarantees that such $\bar{\delta}$ exists.

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

$$\max_{s_i \in S_i} v_i^\mu(\delta, s_i, s_{-i}^\mu) < \liminf_{\delta \rightarrow 1} v_i^\mu(\delta) + \varepsilon \quad (10)$$

for each $\delta \in (\bar{\delta}, 1)$. 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 $\bar{\delta}$ is worse than the limit inferior of the minimax payoff. Since the minimax strategy for the discount factor $\bar{\delta}$ 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 (10) ensures that the minimax payoff for δ is worse than the limit inferior of the minimax payoff. Since this is true for all $\delta \in (\bar{\delta}, 1)$, the limit inferior is the limit, as desired.

So pick an arbitrary $\delta \in (\bar{\delta}, 1)$, 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 $\bar{\delta}$. To evaluate this payoff, we regard the infinite horizon as a series of random blocks, as in Section 3. The termination probability is $1 - p$, where $p = \frac{\bar{\delta}}{\delta}$. 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 k th random block, play $s_{-i}^{\mu^k}$ where μ^k is the belief in the initial period of the k th block.

Then 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,

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

where $s_{-i}^{\mu^k}$ is the optimal (Markov) strategy in the continuation game from the k th block with belief μ^k . Note that $s_{-i}^{\mu^k}$ may not maximize the payoff during the k th

block, because player i needs to take into account the fact that her action during the k th block influences μ^{k+1} and hence the payoffs after the k th block. But in any case, we have $v_i^{\mu^k}(p\delta, s_i^{\mu^k}, s_{-i}^{\mu^k}) \leq v_i^{\mu^k}(\bar{\delta})$ because $s_{-i}^{\mu^k}$ is the minimax strategy with discount factor $p\delta = \bar{\delta}$. Hence

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

Using (9),

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

Then using (8), we obtain (10).

Note that 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.

B.3 Proof of Proposition 4: Properties of Supersets

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.

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, \dots, y^T) such that if the support of the initial prior is Ω^* and players play s , the signal sequence (y^1, \dots, 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, \dots, 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, \dots, 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.

B.4 Proof of Proposition 5: Score and Uniform Connectedness

We will show that the score is invariant to the initial prior if the game is uniformly connected. Fix δ and the direction λ . For each μ , let s^μ be a pure-strategy profile which solves $\max_{s \in \mathcal{S}} \lambda \cdot v(\delta, s)$. 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 $\lambda \cdot v^\mu(\delta, s^\mu)$. Given δ and λ , the score $\lambda \cdot v^\mu(\delta, s^\mu)$ is convex with respect to μ , as it is the upper envelope of the linear functions $\lambda \cdot v^\mu(\delta, s)$ over all s .

Since the score $\lambda \cdot v^\mu(\delta, s^\mu)$ is convex, it is maximized by some boundary belief. That is, there is ω such that

$$\lambda \cdot v^\omega(\delta, s^\omega) \geq \lambda \cdot v^\mu(\delta, s^\mu) \quad (11)$$

for all μ . Pick such ω . In what follows, the score for this ω is called the *maximal score*.

B.4.1 Step 0: Preliminary Lemmas

We begin with providing two preliminary lemmas. The first lemma is very similar to Lemma B1; it shows that if there is a belief μ whose score approximates the maximal score, then the score for *all other* belief $\tilde{\mu}$ with the same support as μ approximates the maximal score.

Lemma B3. *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 $\Delta\Omega^*$. Then for each $\tilde{\mu} \in \Delta\Omega^*$,*

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

To interpret this lemma, pick some $\Omega^* \subseteq \Omega$, and pick a relative interior belief $\mu \in \Delta\Omega^*$ such that $\mu(\tilde{\omega}) \geq \bar{\pi}$ for all $\tilde{\omega} \in \Omega^*$. Then $p \geq \bar{\pi}$, and thus the lemma above implies

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

for all $\tilde{\mu} \in \Delta\Omega^*$. So if the score $\lambda \cdot v^\mu(\delta, s^\mu)$ for the belief μ approximates the maximal score, then for all beliefs $\tilde{\mu}$ with support Ω^* , the score approximates the maximal score.

The above lemma relies on the convexity of the score, and the proof idea is essentially the same as the one presented in Section 4.3. For completeness, we provide the formal proof below.

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

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

Applying (11) 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 $\lambda \cdot v^\omega(\delta, s^\omega) - \lambda \cdot v^\mu(\delta, s^\mu) > 0$ and $\mu(\tilde{\omega}) \geq p = \min_{\tilde{\omega} \in \Omega^*} \mu(\tilde{\omega})$, we obtain

$$\lambda \cdot v^\omega(\delta, s^\omega) - \lambda \cdot v^{\tilde{\omega}}(\delta, s^\mu) \leq \frac{\lambda \cdot v^\omega(\delta, s^\omega) - \lambda \cdot v^\mu(\delta, s^\mu)}{p}. \quad (12)$$

Pick an arbitrary belief $\tilde{\mu} \in \Delta\Omega^*$. Recall that (12) holds for each $\tilde{\omega} \in \Omega^*$. Multiplying both sides of (12) 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 $\lambda \cdot v^\omega(\delta, s^\omega) \geq \lambda \cdot v^{\tilde{\mu}}(\delta, s^\mu) \geq \lambda \cdot v^\mu(\delta, s^\mu)$,

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

Taking the absolute values of both sides, we obtain the result.

Q.E.D.

The next lemma shows that under global accessibility, players can move the support to a globally accessible set Ω^* by simply mixing all actions each period. Note that π^* in the lemma can be different from the one in the definition of global accessibility.

Lemma B4. *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, \dots, 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, \dots, 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.*

B.4.2 Step 1: Scores for Beliefs with Support Ω^*

As a first step of the proof, we will show that there is a globally accessible set Ω^* such that the score for any belief $\mu \in \Delta\Omega^*$ approximates the maximal score. More precisely, we prove the following lemma:

Lemma B5. *There is a globally accessible set $\Omega^* \subseteq \Omega$ such that for all $\mu \in \Delta\Omega^*$,*

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

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 $\Delta\Omega^*$ 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 a technique similar to the one in the proof of Lemma B2, we can show that the continuation payoff after this history h^T approximates the maximal score. This implies that the score for the belief μ^* approximates the maximal score. Then Lemma B3 ensures that the score for any belief $\mu \in \Delta\Omega^*$ 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 $\Delta\Omega^*$ 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, μ^* assigns at least $\bar{\pi}^T$ to each state $\tilde{\omega} \in \Omega^*$, i.e., $\mu^*(\tilde{\omega}) \geq \bar{\pi}^T$ for each $\tilde{\omega} \in \Omega^*$. This is so because

$$\mu^*(\tilde{\omega}) = \frac{\Pr(\omega^{T+1} = \tilde{\omega} | \omega, h^T)}{\sum_{\hat{\omega} \in \Omega} \Pr(\omega^{T+1} = \hat{\omega} | \omega, h^T)} \geq \Pr(\omega^{T+1} = \tilde{\omega} | \omega, h^T) \geq \bar{\pi}^T$$

where the last inequality follows from the fact that $\bar{\pi}$ is the minimum of the function π .

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 . We decompose the score into the payoffs in the first T periods and the continuation payoff after that:

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

Using (11), $\mu(h^T) = \mu^*$, and $(1 - \delta) \sum_{t=1}^T \delta^{t-1} E[\lambda \cdot g^{\omega^t}(a^t) | \omega^1 = \omega, s^\omega] \leq (1 - \delta^T) \bar{g}$, we obtain

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

Arranging, we have

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

Note that $\Pr(h^T | \omega, s^\omega) \geq \bar{\pi}^T$, because s^ω is a pure strategy. Hence we have

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

Since (11) ensures that the left-hand side is non-negative, taking the absolute values of both sides and using $\lambda \cdot v^\omega(\delta, s^\omega) \geq -\bar{g}$,

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

That is, the score for the belief μ^* approximates the maximal score if δ is close to one. As noted, we have $\mu^*(\tilde{\omega}) \geq \bar{\pi}^T$ for each $\tilde{\omega} \in \Omega^*$. Then applying Lemma B3 to the inequality above, we obtain

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

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

B.4.3 Step 2: Scores for All Beliefs μ

In the previous step, we have shown that the score approximates the maximal score for any belief μ with the support Ω^* . Now we will show that the score approximates the maximal score for all beliefs μ .

Pick Ω^* as in the previous step, so that it is globally accessible. Then pick $\pi^* > 0$ as stated in Lemma B4. So 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}^μ :

- 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 B5 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 B4 ensures that this switch occurs with probability one in finite time and waiting time is almost negligible for patient players. Hence the payoff by this strategy profile \tilde{s}^μ approximates the maximal score. Formally, we have the following lemma.

Lemma B6. *For each μ ,*

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

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

Suppose that the initial prior is μ and players play the strategy profile \tilde{s}^μ . 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}^μ , 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}^μ .

Note that the payoff $v^\mu(\delta, \tilde{s}^\mu)$ by the strategy profile \tilde{s}^μ 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

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

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 B5, 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 $v^* = \lambda \cdot v^\omega(\delta, s^\omega) - \frac{(1-\delta^{2^{|\Omega|}})2\bar{g}}{\delta^{2^{|\Omega|}}\pi^{4^{|\Omega|}}}$. Applying this and $\lambda \cdot g^{\omega^t}(a^t) \geq -2\bar{g}$ to the above equation, we obtain

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

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

$$\lambda \cdot v^\mu(\delta, \tilde{s}^\mu) \geq (1-\delta) \sum_{t=1}^{\infty} \delta^{t-1} \{ \rho^t (-2\bar{g}) + (1-\rho^t) v^* \}. \quad (13)$$

According to Lemma B4, the probability that the support reaches Ω^* within $4^{|\Omega|}$ periods is at least π^* . This implies that the probability that players still randomize all actions in period $4^{|\Omega|} + 1$ is at most $1 - \pi^*$. Similarly, for each natural number n , the probability that players still randomize all actions in period $n4^{|\Omega|} + 1$ is at most $(1 - \pi^*)^n$, that is, $\rho^{n4^{|\Omega|}+1} \leq (1 - \pi^*)^n$. Then since ρ^t is weakly decreasing in t , we obtain

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

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

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

for each $n = 0, 1, \dots$ and $k \in \{1, \dots, 4^{|\Omega|}\}$. Plugging this inequality into (13), we obtain

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

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

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

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

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

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

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

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

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

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}^μ is not the optimal policy) and is at most the maximal score. Then from Lemma B6, we have

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

as desired.

B.5 Proof of Proposition 6: Necessary Condition for Uniform Connectedness

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, \dots, 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, \dots, 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 states are weakly communicating.

B.5.1 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.

B.5.2 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 $\omega_1, \omega_2, \dots, \omega_K$. 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.

B.5.3 Step 3: Uniformly Connected Only If States Are Weakly Communicating

Now we prove that the game is uniformly connected only if states are weakly communicating. 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.

B.6 Proof of Proposition 7

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 singleton sets are uniformly transient. The only if part follows from Proposition 6.

Next, consider stochastic games with delayed observations. Again the only if part follows from Lemma 6, 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, \dots, 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, \dots, y^T) which happens when the state moves from $\tilde{\omega}$ to ω^* . Now, suppose that the initial prior is μ and players play $(a^1, \dots, a^T, s(h^{t-1}))$. Then by the definition, with positive probability, players observe the signal sequence (y^1, \dots, 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 $\mu(\tilde{\omega})\bar{\pi}^{T+1} \geq \frac{\bar{\pi}^{T+1}}{|\Omega|}$ 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 paragraph, we can show that Ω^* is globally accessible, which implies that $\{\omega\}$ is uniformly transient.

B.7 Proof of Proposition 8: Minimax and Robust Connectedness

We will prove only (5). The existence of the limit minimax payoff can be proved just as in Step 3 of the proof of Proposition 3.

Fix δ and i . In what follows, “robustly accessible” means “robustly accessible despite i ,” and “avoidable” means “avoidable for i .”

Let s^μ denote the minimax strategy profile given the initial prior μ . As in the proof of Proposition 3, let $v_i^{\tilde{\mu}}(s_{-i}^\mu) = \max_{s_i \in \mathcal{S}_i} v_i^{\tilde{\mu}}(\delta, s_i, s_{-i}^\mu)$, that is, let $v_i^{\tilde{\mu}}(s_{-i}^\mu)$ denote player i 's payoff when the opponents play the minimax strategy s_{-i}^μ for some belief μ but the actual initial prior is $\tilde{\mu}$. Given the opponents' strategy s_{-i}^μ , let

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

that is, $\bar{v}_i(s_{-i}^\mu)$ is player i 's payoff when the initial prior $\tilde{\mu}$ is the most favorable one, subject to the constraint that $\tilde{\mu}$ and μ have the same support. Then choose μ^* such that

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

We call $\bar{v}_i(s_{-i}^{\mu^*})$ the *maximal value*. The definition of the maximal value here is very similar to that in the proof of Proposition 3, but it is not exactly the same because when we define $\bar{v}_i(s_{-i}^\mu)$, the initial prior $\tilde{\mu}$ is chosen from the set $\Delta(\text{supp}\mu)$.

Since $v_i^{\tilde{\mu}}(s_{-i}^{\mu^*})$ is convex with respect to the initial prior $\tilde{\mu}$, there is a state $\omega \in \text{supp}\mu^*$ such that $v_i^\omega(s_{-i}^{\mu^*}) \geq v_i^{\tilde{\mu}}(s_{-i}^{\mu^*})$ for all $\tilde{\mu} \in \Delta(\text{supp}\mu^*)$. Pick such ω .

B.7.1 Step 0: Preliminary Lemmas

We begin with presenting three preliminary lemmas. The first lemma is a generalization of Lemma B1. The statement is more complicated than Lemma B1, because we focus on a pair of beliefs $(\mu, \tilde{\mu})$ which have the same support. But the implication of the lemma is the same; given the opponents' strategy s_{-i}^μ , if player i 's payoff $v_i^{\tilde{\mu}}(s_{-i}^\mu)$ approximates the maximal value for some relative interior belief $\tilde{\mu} \in \Delta\Omega^*$, then it approximates the maximal value for all beliefs $\hat{\mu} \in \Delta\Omega^*$. The proof of the lemma is very similar to that of Lemma B1, and hence omitted.

Lemma B7. *Pick an arbitrary belief μ , and let Ω^* denote its support. Let $\tilde{\mu} \in \Delta\Omega^*$ be an relative interior belief (i.e., $\tilde{\mu}(\tilde{\omega}) > 0$ for all $\tilde{\omega}$), and let $p = \min_{\tilde{\omega} \in \Omega^*} \tilde{\mu}(\tilde{\omega})$, which measures the distance from $\tilde{\mu}$ to the boundary of $\Delta\Omega^*$. Then for each $\hat{\mu} \in \Delta\Omega^*$,*

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

The next lemma shows that under the merging support condition, given any pure strategy profile s , two posterior beliefs induced by different initial priors ω and μ with $\mu(\omega) > 0$ will have the same support after some history. Also it gives a minimum bound on the probability of such a history.

Lemma B8. *Suppose that the merging support 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 $\Pr(h^T | \omega, s) > (\frac{|\bar{\pi}|}{|A|})^T$ 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 merging support 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 $\tilde{\mu} = (\frac{1}{|\Omega|}, \dots, \frac{1}{|\Omega|})$ and the history h^T . We show that T and h^T here satisfies the desired properties.

Note that $\Pr(h^T | \omega, \tilde{s}) \geq \bar{\pi}^T$, as $\bar{\pi}$ is a pure strategy. This implies that $\Pr(h^T | \omega, s) \geq (\frac{\bar{\pi}}{|\mathcal{A}|})^{4^{|\Omega|}}$, since each period the action profile by s coincides with the one by \tilde{s} with probability at least $\frac{1}{|\mathcal{A}|}$. Also, since $\mu(\omega) > 0$, the support of the belief induced by (ω, h^T) must be 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 last preliminary lemma is a counterpart to Lemma B4. It shows that the opponents can move the support of the belief to a robustly accessible set Ω^* , by simply mixing all actions each period. It also shows that the resulting posterior belief is not too close to the boundary of the belief space $\Delta\Omega^*$.

Lemma B9. *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 $\tilde{\mu} \in \Delta\Omega^*$ such that the posterior belief μ^{T+1} equals $\tilde{\mu}$ with probability at least π^* and such that $\tilde{\mu}(\omega) \geq \frac{1}{|\Omega|} \bar{\pi}^{4^{|\Omega|}}$ for all $\omega \in \Omega^*$.*

Proof. We first show that Ω^* is robustly accessible only if the following condition holds:²² 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 $(a_{-i}^1, \dots, a_{-i}^T)$, and a signal sequence (y^1, \dots, 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 $(a_{-i}^1, \dots, a_{-i}^T)$, then the sequence (y^1, \dots, y^T) realizes with positive probability.
- (ii) If player i plays s_i , the opponents play $(a_{-i}^1, \dots, a_{-i}^T)$, and the signal sequence (y^1, \dots, y^T) realizes, then the state in period $T + 1$ must be in the set Ω^* , regardless of the initial state $\hat{\omega}$ (possibly $\hat{\omega} \neq \omega$).
- (iii) If the initial state is ω , player i plays s_i , the opponents play $(a_{-i}^1, \dots, a_{-i}^T)$, and the signal sequence (y^1, \dots, y^T) realizes, then the support of the belief in period $T + 1$ is the set Ω^* .

²²We 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 $(\alpha_{-i}^1, \dots, \alpha_{-i}^{4^{|\Omega|}})$ which mix all pure actions equally each period.

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 $\mu(\tilde{\omega}) > 0$ 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 $\varepsilon \rightarrow 0$, 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 , $(a_{-i}^1, \dots, a_{-i}^T)$, and (y^1, \dots, 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 $(a_{-i}^1, \dots, a_{-i}^T)$ and the signal sequence $(a_{-i}^1, \dots, a_{-i}^T)$ are observed with probability at least $\mu(\omega) \left(\frac{\bar{\pi}}{|A|^T}\right)^T \geq \frac{1}{|\Omega|} \left(\frac{\bar{\pi}}{|A|^T}\right)^{4|\Omega|}$. Let $\tilde{\mu}$ be the posterior belief in period $T + 1$ in this case. From (iii), $\tilde{\mu}(\omega) \geq \frac{1}{|\Omega|} \bar{\pi}^{4|\Omega|}$ for all $\omega \in \Omega^*$. From (ii), $\tilde{\mu}(\omega) = 0$ for other ω . *Q.E.D.*

B.7.2 Step 1: Minimax Payoff for μ^{**}

As a first step, we will show that there is some belief μ^{**} whose minimax payoff approximates the maximal value. The proof idea is similar to Step 1 in the proof of Proposition 3, but the argument is more complicated because now some signals and states do not occur, due to the lack of the full support assumption. As will be seen, we use the merging support condition in this step.

Recall that the maximal value is achieved when the opponents play the minimax strategy $s_{-i}^{\mu^*}$ for the belief μ^* but the actual initial state is ω . Let s_i^* denote player i 's best reply. Then the maximal value is decomposed into payoffs in the first T periods and the continuation payoff:

$$\begin{aligned} v_i^\omega(s_{-i}^{\mu^*}) &= (1 - \delta) \sum_{t=1}^T \delta^{t-1} E[g_i^{\omega^t}(a^t) | \omega, s_i^*, s_{-i}^{\mu^*}] \\ &\quad + \delta^T \sum_{\tilde{h}^T \in H^T} \Pr(\tilde{h}^T | \omega, s_i^*, s_{-i}^{\mu^*}) v_i^{\mu(\tilde{h}^T | \omega)}(s_{-i}^{\mu(\tilde{h}^T | \mu^*)}). \end{aligned} \quad (14)$$

Here, $\mu(h^T | \omega)$ denotes the posterior in period $T + 1$ when the initial state was ω and the past history was h^T . and $\mu(h^T | \mu^*)$ denotes the posterior when the initial

prior was μ^* rather than ω . The following lemma is a counterpart to Lemma B2: It shows that the continuation payoff $v_i^{\mu(h^T|\omega)}(s_{-i}^{\mu(h^T|\mu^*)})$ approximates the maximal value after some history h^T .

Lemma B10. *There is $T \leq 4^{|\Omega|}$ and h^T such that the two posteriors $\mu(h^T|\omega)$ and $\mu(h^T|\mu^*)$ have the same support and such that*

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

Proof. Since $\mu^*(\omega) > 0$, Lemma B8 ensures that there is a natural number $T \leq 4^{|\Omega|}$ and a history h^T such that $\Pr(h^T|\omega, s_i^*, s_{-i}^{\mu^*}) > (\frac{\bar{\pi}}{|A|})^T$ 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 .

By the definition of \bar{g} , we have $(1 - \delta) \sum_{t=1}^T \delta^{t-1} E[g_i^{\omega^t}(a^t)|\omega, s] \leq (1 - \delta^T)\bar{g}$. Also, since $\mu^*(\omega) > 0$, for each \tilde{h}^T , the support of $\mu(\tilde{h}^T|\omega)$ is a subset of the one of $\mu(\tilde{h}^T|\mu^*)$, which implies $v_i^{\mu(\tilde{h}^T|\omega)}(s_{-i}^{\mu(\tilde{h}^T|\mu^*)}) \leq v_i^\omega(s_{-i}^{\mu^*}) + (1 - \delta)$. Plugging them and $\Pr(h^T|\omega, s_i^*, s_{-i}^{\mu^*}) \geq (\frac{\bar{\pi}}{|A|})^T$ into (14), we have

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

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

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

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

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

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

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

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

Q.E.D.

Let $\mu^{**} = \mu(h^T|\mu^*)$. Then the above lemma implies that

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

That is, given the opponents' strategy $s_{-i}^{\mu^{**}}$, player i 's payoff $v_i^{\tilde{\mu}}(s_{-i}^{\mu^{**}})$ approximates the maximal score for some belief $\tilde{\mu} = \mu(h^T|\omega)$.

From Lemma B10, the support of this belief $\mu(h^T|\omega)$ is the same as the one of μ^{**} . Also, this belief $\mu(h^T|\omega)$ assigns at least probability $\bar{\pi}^{4^{|\Omega|}}$ on each state ω included in its support. Indeed, for such state ω , we have

$$\begin{aligned} \mu(h^T|\omega)[\tilde{\omega}] &= \frac{\Pr(\omega^{T+1} = \tilde{\omega}|\omega, a^1, \dots, a^T)}{\sum_{\hat{\omega} \in \Omega} \Pr(\omega^{T+1} = \hat{\omega}|\omega, a^1, \dots, a^T)} \\ &\geq \Pr(\omega^{T+1} = \tilde{\omega}|\omega, a^1, \dots, a^T) \geq \bar{\pi}^T \geq \bar{\pi}^{4^{|\Omega|}}. \end{aligned}$$

Accordingly, the distance from $\tilde{\mu} = \mu(h^T|\omega)$ to the boundary of $\Delta(\text{supp}\mu^{**})$ is at least $\bar{\pi}^{4^{|\Omega|}}$, and thus Lemma B7 ensures that

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

for all $\hat{\mu} \in \Delta(\text{supp}\mu^{**})$. That is, the payoff $v_i^{\hat{\mu}}(s_{-i}^{\mu^{**}})$ approximates the maximal score for all beliefs $\hat{\mu} \in \Delta(\text{supp}\mu^{**})$. In particular, by letting $\hat{\mu} = \mu^{**}$, we have

$$\left| v_i^\omega(s_{-i}^{\mu^*}) + (1 - \delta) - v_i^{\mu^{**}}(s_{-i}^{\mu^{**}}) \right| \leq \frac{(1 - \delta^{4^{|\Omega|}})2\bar{g}|A|^{4^{|\Omega|}}}{\delta^{4^{|\Omega|}} \bar{\pi}^{(4^{|\Omega|} + 4^{|\Omega|})}} + \frac{(1 - \delta)|A|^{4^{|\Omega|}}}{\bar{\pi}^{(4^{|\Omega|} + 4^{|\Omega|})}}, \quad (15)$$

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

B.7.3 Step 2: Minimax Payoffs when the Support is Robustly Accessible

In this step, we show that the minimax payoff for μ approximates the maximal value for any belief μ whose support is robustly accessible. Again, the proof idea is somewhat similar to Step 2 in the proof of Proposition 3. But the proof here is more involved, because the support of the belief μ^{**} in Step 1 may be different from the one of μ , and thus the payoff $v_i^{\mu^{**}}(s_{-i}^{\mu})$ can be greater than the maximal value.

For a given belief μ , let Δ^μ denote the set of beliefs $\tilde{\mu} \in \Delta(\text{supp}\mu)$ such that $\tilde{\mu}(\tilde{\omega}) \geq \frac{1}{|\Omega|} \bar{\pi}^{4|\Omega|}$ 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 $\Delta(\text{supp}\mu)$.

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}^μ :

- 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_{-i}^μ for the fixed belief μ . From Lemma B9, this switch happens in finite time with probability one regardless of player i 's play. So for δ close to one, payoffs before the switch is almost negligible, that is, 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

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

Hence player i 's payoff against the above strategy \tilde{s}_{-i}^μ cannot exceed K_i^μ by much. Formally, we have the following lemma. Take $\pi^* > 0$ such that it satisfies the condition stated in Lemma B9 for all robustly accessible sets Ω^* . (Such π^* exists, as there are only finitely many sets Ω^* .)

Lemma B11. *For each belief μ whose support is robustly accessible,*

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

Proof. The proof is very similar to that of Lemma B6. Pick a belief μ whose support is robustly accessible. Suppose that the initial prior is μ^{**} , the opponents play \tilde{s}_{-i}^{μ} , 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 B6, we have

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

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

As in the proof of Lemma B6, we have

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

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

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

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

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

Then as in the proof of Lemma B6, the standard algebra shows

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

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

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

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

Q.E.D.

Note that the payoff $v_i^{\mu^{**}}(s_{-i}^{\mu^{**}})$ is at least the minimax payoff $v_i^{\mu^{**}}(s_{-i}^{\mu^{**}})$, as the strategy s_{-i}^{μ} is not the minimax strategy. So we have $v_i^{\mu^{**}}(s_{-i}^{\mu^{**}}) \leq v_i^{\mu^{**}}(s_{-i}^{\mu})$. This inequality and the lemma above imply that

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

At the same time, by the definition of the maximal value, K_i^{μ} cannot exceed $v_i^{\omega}(s_{-i}^{\mu^*}) + (1 - \delta)$. Hence

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

From (15), we know that $v_i^{\mu^{**}}(s_{-i}^{\mu^{**}})$ approximates $v_i^{\omega}(s_{-i}^{\mu^*}) + (1 - \delta)$, so the above inequality implies that K_i^{μ} approximates $v_i^{\omega}(s_{-i}^{\mu^*}) + (1 - \delta)$. Formally, we have

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

Equivalently,

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

where $\tilde{\mu}$ is the belief which achieves K_i^{μ} . This inequality implies that given the opponents' strategy s_{-i}^{μ} , player i 's payoff $v_i^{\tilde{\mu}}(s_{-i}^{\mu})$ approximates the maximal value for some belief $\tilde{\mu}$. Since $\tilde{\mu} \in \Delta^{\mu}$, Lemma B7 ensure that the same result holds for all beliefs with the same support, that is,

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

for all $\hat{\mu} \in \Delta(\text{supp}\mu)$. This in particular implies that the minimax payoff for μ approximates the maximal value.

B.7.4 Step 3: Minimax Payoffs when the Support is Avoidable

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

minimax payoff approximates the maximal value for any belief μ whose support is avoidable.

So pick an arbitrary belief μ whose support is avoidable. 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^μ :

- 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 $v_i^{\mu^t}(s_{-i}^{\mu^t})$. From the previous step, we know 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^μ also approximates the maximal value. Formally, we have the following lemma:

Lemma B12. *For any μ whose support is avoidable,*

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

Proof. The proof is very similar to that of Lemma B11 and hence omitted. *Q.E.D.*

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

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

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

B.8 Proof of Proposition 10

The proof technique is quite similar to that of Proposition 8, 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 8. Let \bar{v}_i be the supremum of the minimax payoffs $v_i^\mu(\delta)$ 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 $v_i^{\mu^*}(\delta)$ 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}^ω :

- 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

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

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 B11 and hence omitted.

Now, since \tilde{s}_{-i}^ω 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 $v_i^{\mu^*}(\delta)$, 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 Lemma B5 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 states are weakly communicating, 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.

B.9 Proof of Proposition A1

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 B13. *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 A1. 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 4, 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 \mathcal{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 $\Omega^* \in \mathcal{O}$; 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 $\Omega \in \mathcal{O}$, then the result immediately holds by setting $\Omega^* = \Omega$. So in what follows, we assume $\Omega \notin \mathcal{O}$. We prove the existence of an asymptotically accessible set $\Omega^* \in \mathcal{O}$ in two steps. In the first step, we show that there is $q > 0$ and $\tilde{\Omega}^* \in \mathcal{O}$ 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 $\tilde{\Omega}^*$), players can move the belief to the one which puts probability at least $1 - \varepsilon$ on some $\Omega^* \in \mathcal{O}$. 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 $\Omega^* \in \mathcal{O}$, 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 B14. *There is $q > 0$ and a set $\tilde{\Omega}^* \in \mathcal{O}$ such that for each initial prior μ , there is a natural number $T \leq |\Omega|$, an action sequence (a^1, \dots, a^T) , and a*

history h^T such that $\Pr(h^T | \mu, a^1, \dots, a^T) \geq \frac{\bar{\pi}^{|\Omega|}}{|\Omega|}$ 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 $\tilde{\Omega}^* \in \mathcal{O}$ which contains at least one globally accessible state $\tilde{\omega}$. Suppose not so that all states in any set $\Omega^* \in \mathcal{O}$ 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 \mathcal{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 $t \rightarrow \infty$.

So there is $\tilde{\Omega}^* \in \mathcal{O}$ 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, \dots, a^T) , and a signal sequence (y^1, \dots, y^T) such that

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

That is, if the initial state is $\hat{\omega}$ and players play (a^1, \dots, 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, \dots, a^T) and (y^1, \dots, y^T) , and let

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

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

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

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

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

as desired.

Q.E.D.

Choose $\tilde{\Omega}^* \in \mathcal{O}$ 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 \mathcal{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 $\Omega^* \in \mathcal{O}$.

Lemma B15. *For each $\varepsilon > 0$ and $q > 0$, there is a natural number T , a set $\Omega^* \in \mathcal{O}$, 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 $\Omega \notin \mathcal{O}$, so any $\Omega^* \in \mathcal{O}$ is a proper subset of Ω . By the assumption, given any $\Omega^* \in \mathcal{O}$ 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.,

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

for each $\bar{\mu} \in \Delta \tilde{\Omega}^*$ and $\underline{\mu} \in \Delta(\Omega \setminus \tilde{\Omega}^*)$. (Here $\|\cdot\|$ denotes the sup norm.) Let $\kappa > 0$ be the minimum of $\kappa(\Omega^*, a)$ over all $\Omega^* \in \mathcal{O}$ 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 $\bar{\mu}$ be the marginal distribution on Ω^1 , that is, $\bar{\mu}(\tilde{\omega}) = \frac{\mu(\tilde{\omega})}{\sum_{\hat{\omega} \in \Omega^1} \mu(\hat{\omega})}$ for each $\tilde{\omega} \in \Omega^1$ and $\bar{\mu}(\tilde{\omega}) = 0$ 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

$$\pi_Y^{\bar{\mu}}(y|a) \geq \pi_Y^{\underline{\mu}}(y|a) + \kappa. \quad (16)$$

Intuitively, (16) 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 $\bar{\mu}$ 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

$$\pi^{\tilde{\omega}}(y, \hat{\omega}|a) = 0 \quad (17)$$

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

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

Here, the second equality comes from (17), and the last inequality from (16). 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 (16) implies

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

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 $\bar{\mu}^2$ 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 $\pi_Y^{\bar{\mu}^2}(y^2|a^2) \geq \pi_Y^{\mu^2}(y^2|a^2) + \kappa$, 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, \dots, 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 \mathcal{O} , by the construction.

Now, choose $\varepsilon > 0$ and $q > 0$ arbitrarily, and then pick T large enough that $\left(\frac{1}{1-\kappa} \right)^T \frac{q}{1-q} \geq \frac{1-\varepsilon}{\varepsilon}$. Then the above posterior belief μ^{T+1} puts probability at least $1 - \varepsilon$ on $\Omega^{T+1} \in \mathcal{O}$. 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 B14, and then choose Ω^* , T , and π^{**} as stated in Lemma B15. 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 $\pi^* = \frac{\bar{\pi}^{|\Omega|} \pi^{**}}{|\Omega|}$, 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 $\Omega^* \in \mathcal{O}$.

B.10 Proof of Proposition A2: Score and Asymptotic Connectedness

Fix δ and λ . Let s^μ and ω be as in the proof of Proposition 5. We begin with two preliminary lemmas. The first lemma shows that the score is Lipschitz continuous with respect to μ .

Lemma B16. For any $\varepsilon \in (0, \frac{1}{|\Omega|})$, μ , and $\tilde{\mu}$ with $|\mu(\tilde{\omega}) - \tilde{\mu}(\tilde{\omega})| \leq \varepsilon$ for each $\tilde{\omega} \in \Omega$,

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

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

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

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

The second preliminary lemma is a counterpart to Lemma B4; 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 B4 and hence omitted.

Lemma B17. 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 B17.

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

B.10.1 Step 1: 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 $\mu \in \Delta\Omega^*(\varepsilon)$ approximates the maximal score.

Lemma B18. *There is an asymptotically accessible set Ω^* such that for any $\mu \in \Delta\Omega^*$,*

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

Then from Lemma B16, there is an asymptotically accessible set Ω^ such that for any $\mu \in \Delta\Omega^*(\varepsilon)$,*

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

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, \dots, y^{T^*}) such that if the initial state is ω and players play s^ω , then the signal sequence (y^1, \dots, 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, \dots, y^{T^*}) . Then as in the proof of Lemma B5, we can prove that

$$\left| \lambda \cdot v^\omega(\delta, s^\omega) - \lambda \cdot v^{\mu^*}(\delta, s^{\mu^*}) \right| \leq \frac{(1 - \delta^T)2\bar{g}}{\delta^T \bar{\pi}^T}. \quad (18)$$

That is, the score with the initial prior μ^* is close to the maximal score. The only difference from Lemma B5 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 B16 implies that these two beliefs μ^* and $\tilde{\mu}^*$ induce similar scores, that is,

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

Plugging this into (18), we obtain

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

That is, the score for the belief $\tilde{\mu}^*$ approximates the maximal score. Then using Lemma B3, we can get the desired inequality. *Q.E.D.*

B.10.2 Step 2: 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}^μ :

- Players randomize all actions equally likely, until the posterior belief becomes an element of $\Delta\Omega^*(\varepsilon)$.
- Once the posterior belief becomes an element of $\Delta\Omega^*(\varepsilon)$ 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 $\Delta\Omega^*(\varepsilon)$; and once it happens, they switch the play to the optimal policy s^{μ^t} in the continuation game. Lemma B18 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 B17 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}^μ approximates the maximal score when the initial prior is μ . Formally, we have the following lemma.

Lemma B19. *For each μ ,*

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

Proof. The proof is essentially the same as that of Lemma B6; we simply replace $4^{|\Omega|}$ in the proof of Lemma B6 with T , and use Lemma B18 instead of Lemma B5. *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 B19,

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

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 C: Uniform Connectedness in Terms of Primitives

In Section 5.1, 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 C1. 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, \dots, a^T) , and a signal sequence (y^1, \dots, y^T) such that the following properties are satisfied.²³

- (i) If the initial state is ω and players play (a^1, \dots, a^T) , then the sequence (y^1, \dots, y^T) realizes with positive probability. That is, there is a state sequence $(\omega^1, \dots, \omega^{T+1})$ 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, \dots, a^T) and observe (y^1, \dots, 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 $(\omega^1, \dots, \omega^{T+1})$ 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 global accessibility here is indeed equivalent to the one stated using beliefs.

Proposition C1. *Definitions 2 and C1 are equivalent.*

²³ 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, \dots, a^T) and (y^1, \dots, y^T) which satisfy (i) and (ii) in Definition C1. 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, \dots, a^t) , and the signal sequence (y^1, \dots, 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 $(a^{t+1}, \dots, a^{\tilde{t}})$ from the original sequence (a^1, \dots, a^T) . Similarly, construct the signal sequence with length $T - (\tilde{t} - t)$. Then these new sequences satisfy (i) and (ii) in Definition C1. 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 C1 implies the one in Definition 2. Take a set Ω^* which is globally accessible in the sense of Definition C1, 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, \dots, a^T) and (y^1, \dots, y^T) as stated in Definition C1. Suppose that the initial prior is μ and players play (a^1, \dots, a^T) . Then clause (i) of Definition C1 guarantees that the signal sequence (y^1, \dots, 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 $\tilde{\omega} \notin \Omega^*$.²⁴ Note that the probability of this signal sequence (y^1, \dots, y^T) is at least

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

where $\Pr(y^1, \dots, y^T | \omega, a^1, \dots, a^T)$ denotes the probability of the signal sequence (y^1, \dots, y^T) given the initial state ω and the action sequence (a^1, \dots, a^T) . This implies that global accessibility in Definition C1 implies the one in Definition 2, by letting $\pi^* \in (0, \frac{1}{|\Omega|} \bar{\pi}^{4|\Omega|})$.

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, \dots, a^T) and a belief $\tilde{\mu}$ whose support is included in Ω^* such that

$$\Pr(\mu^{T+1} = \tilde{\mu} | \mu, a^1, \dots, a^T) \geq \pi^*. \quad (19)$$

Let (y^1, \dots, y^T) be the signal sequence which induces the posterior belief $\tilde{\mu}$ given the initial prior μ and the action sequence (a^1, \dots, a^T) . Such a signal sequence may not be unique, so let \hat{Y}^t be the set of these signal sequences. Then (19) implies that

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

²⁴The reason is as follows. From Bayes' rule, $\mu^{T+1}(\tilde{\omega}) > 0$ only if $\Pr(y^1, \dots, y^T, \omega^{T+1} = \tilde{\omega} | \tilde{\omega}, a^1, \dots, a^T) > 0$ for some $\tilde{\omega}$ with $\mu(\tilde{\omega}) > 0$. But clause (ii) asserts that the inequality does not hold for all $\tilde{\omega} \in \Omega$ and $\tilde{\omega} \notin \Omega^*$.

Arranging,

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

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

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

so that

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

Hence there is some $(y^1, \dots, y^T) \in \hat{Y}^T$ which can happen with positive probability given the initial state ω and the action sequence (a^1, \dots, a^T) . Obviously this sequence (y^1, \dots, y^T) satisfies clause (i) in Definition C1. Also it satisfies clause (ii) in Definition C1, since (y^1, \dots, 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 $s(y^1, \dots, y^{t-1})$ denote the pure action profile induced by s in period t when the past signal sequence is (y^1, \dots, y^{t-1}) .

Definition C2. 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, \dots, y^T) such that for each $\tilde{\omega} \in \Omega^*$, there is a state sequence $(\omega^1, \dots, \omega^{T+1})$ such that $\omega^1 = \omega$, $\omega^{T+1} = \tilde{\omega}$, and $\pi^{\omega^t}(y^t, \omega^{t+1} | s(y^1, \dots, y^{t-1})) > 0$ for all $t \leq T$.²⁵

²⁵Restricting 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 B13, 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.²⁶ 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 C3. 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 4, 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 4 guarantees that they are all uniformly transient.)

Proposition C2. *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 $\Pr(\mu^{T+1} = \tilde{\mu} | \mu, s) > \pi^*$.*

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

²⁶While 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.

the posterior belief will reach a globally accessible set with probability at least $\pi^* = \frac{\bar{\pi}^{2|\Omega|}}{|\Omega|}$.

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 $\bar{\pi}^T$ 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 $\mu(\omega)\bar{\pi}^T \geq \frac{\bar{\pi}^{2|\Omega|}}{|\Omega|} = \pi^*$. 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.²⁷ 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 $X(\Omega^*|\mu, s)$ 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

²⁷Here 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 $\{\omega_1\}$.

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 C3. *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 $\bar{\pi}^{2^{|\Omega|}} > \frac{\varepsilon|\Omega|}{\bar{\pi}^{2^{|\Omega|}}}$, 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 $\Pr(\mu^{T+1} = \tilde{\mu}|\omega, s) \geq \bar{\pi}^{2^{|\Omega|}}$ and $\tilde{\mu}(\tilde{\omega}) \geq \bar{\pi}^{2^{|\Omega|}}$ for all $\tilde{\omega} \in \tilde{\Omega}^*$.

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

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

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 $\tilde{\mu}(\tilde{\omega}) \geq \bar{\pi}^{2^{|\Omega|}} > \frac{\varepsilon|\Omega|}{\bar{\pi}^{2^{|\Omega|}}}$ 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, \dots, a^T) , and a signal sequence (y^1, \dots, y^T) such that the following properties are satisfied:

- (i) If the initial state is $\tilde{\omega}$ and players play (a^1, \dots, a^T) , then the sequence (y^1, \dots, y^T) realizes with positive probability.
- (ii) If players play (a^1, \dots, a^T) and observe (y^1, \dots, 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 $\hat{\mu} \in \Delta \tilde{\Omega}^*$ 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 C1 and hence omitted. This and global accessibility of $\tilde{\Omega}^*$ imply that Ω^* is globally accessible, which is a contradiction. *Q.E.D.*

Appendix D: Existence of Maximizers

Lemma D1. *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) $f : H \rightarrow \mathbf{R}$. For each function $f \in l^\infty$, let Tf be a function such that

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

where $\tilde{\mu}(h^t)$ is the posterior belief of 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 D2. $\max_{v \in V^\mu(\delta)} \lambda \cdot v$ has a solution.

Proof. Identical with that of the previous lemma.

Q.E.D.

Lemma D3. *There is s_{-i} which solves $\min_{s_{-i} \in \mathcal{S}_{-i}} \max_{s_i \in \mathcal{S}_i} v_i^\mu(\delta, s)$.*

Proof. The formal proof is as follows. Pick μ and δ , and let h^t and l^∞ be as in the proof of Lemma D1. For each function $f \in l^\infty$, let Tf be a function such that

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

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 D1. 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 E: 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.