Theoretical Economics 14 (2019), 1115–1167 1555-7561/20191115

Stochastic games with hidden states

Yuichi Yamamoto
Department of Economics, University of Pennsylvania

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 their 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.

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

JEL classification. C72, C73.

1. Introduction

When agents have a long-run relationship, underlying economic conditions may change over time. A leading example is a repeated Bertrand competition with stochastic demand shocks. Rotemberg and Saloner (1986) explore optimal collusive pricing when random demand shocks are independent and identically distributed (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 that 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

Yuichi Yamamoto: yyam@sas.upenn.edu
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.

© 2019 The Author. Licensed under the Creative Commons Attribution-NonCommercial License 4.0. Available at http://econtheory.org. https://doi.org/10.3982/TE3068

or slumps in the Bertrand model) is given and players have some posterior belief μ^t about the state. Players simultaneously choose actions, and then a public signal y and the next hidden state ω^{t+1} are randomly drawn. After observing the signal y, players updates their posterior belief using Bayes' rule and then go to the next period. The signal y can be informative about both the current and next states, which ensures that our formulation accommodates a wide range of economic applications, including games with delayed observations and a combination of observed and unobserved states.

We assume that actions are perfectly observable, so 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 to past work that assumes finite states (Dutta (1995), Fudenberg and Yamamoto (2011b), and Hörner et al. (2011a)).[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 et al. (2011a)) consider a model that 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 that in Dutta (1995), but we use the idea of random blocks 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

[1]For the infinite-state case, the existence of Markov perfect equilibria is extensively studied. See recent work by Duggan (2012) and Levy (2013), and an excellent survey by Dutta and Sundaram (1998). In contrast to this literature, we consider general non-Markovian equilibria. Hörner et al. (2011b) 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.

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

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 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 that 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.

Unfortunately, when states are infinite, the communicating states assumption is 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 that 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., the Doeblin 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 a 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 the POMDP problem is invariant to the initial belief. This immediately implies invariance of the feasible payoff set.

In contrast, for the minimax payoffs, we need a new proof technique to obtain invariance. 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

[3]This is essentially because our model is a multi-player version of the partially observable Markov decision process (POMDP). The introduction of Rosenberg et al. (2002) explains why the POMDP model is intractable.

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 assumption 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 in 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. In 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 are satisfied in many economic examples, including those 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

[4]Such assumptions include renewability of Ross (1968), reachability–detectability of Platzman (1980), and Assumption 4 of Hsu et al. (2006). (There is a minor error in Hsu et al. (2006); see Appendix E in the working paper version (Yamamoto (2018)) for more details.) The natural resource management problem in this paper is an example that satisfies uniform connectedness but not the assumptions in the literature. Similarly, Example A1 in Appendix A satisfies asymptotic uniform connectedness, but not the assumptions in the literature.

folk theorem for the case of observable actions. Fudenberg and Yamamoto (2011b) and Hörner et al. (2011a) 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 et al. (2015) consider repeated Bayesian games in which the state changes as time goes and players have private information about the current state each period. They look at equilibria in which players report their private information truthfully, which means that the state is perfectly revealed before they choose actions each period.[6]

In contrast, in this paper, players have only limited information about the true state and the state is not perfectly revealed.

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

2. Setup

2.1 *Stochastic games with hidden states*

Let $I = \{1, \dots, N\}$ 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 we let $\mu \in \triangle\Omega$ be the common prior about ω^1.

In each period t, players move simultaneously, with player $i \in I$ choosing an action a_i from a finite set A_i. Let $A \equiv \times_{i\in I} A_i$ be the set of action profiles $a = (a_i)_{i\in I}$. Actions are perfectly observable and, in addition, players observe a public signal y from a finite set Y. Then players go to the next period $t+1$, with a (hidden) state ω^{t+1}. The distribution of y and ω^{t+1} depends on the current state ω^t and the current action profile $a \in A$; let $\pi^\omega(y, \tilde{\omega}|a)$ denote the probability that players observe a signal y and the next state becomes $\omega^{t+1} = \tilde{\omega}$, given $\omega^t = \omega$ and a. In this setup, a public signal y can be informative about the current state ω *and* the next state $\tilde{\omega}$, because the distribution of y may depend on ω and y may be correlated with $\tilde{\omega}$. Let $\pi^\omega_Y(y|a)$ denote the marginal probability of y.

Player i's payoff in period t is a function of the current action profile a and the current public signal y, and is denoted by $u_i(a, y)$. Then her expected stage-game payoff conditional on the current state ω and the current action profile a is $g^\omega_i(a) = \sum_{y\in Y} \pi^\omega_Y(y|a)u_i(a, y)$. Here the hidden state ω influences a player's expected payoff

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

[6]An exception is Sections 4 and 5 of Hörner et al. (2015), who 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.

through the distribution of y.[7] Let $g^\omega(a) = (g_i^\omega(a))_{i\in I}$ be the vector of expected payoffs. Also let $\overline{\pi}$ be the minimum of $\pi^\omega(y, \tilde{\omega}|a)$ over all $(\omega, \tilde{\omega}, a, y)$ such that $\pi^\omega(y, \tilde{\omega}|a) > 0$.

Our formulation encompasses the following examples.

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

- *Stochastic games with delayed observations.* Let $Y = \Omega$ and assume that $\pi_Y^\omega(y|a) = 1$ for $y = \omega$. That is, assume that the current signal y^t reveals the current state ω^t. So players observe the state *after* they move.

- *Observable and unobservable states.* Assume that ω consists of two components, ω_O and ω_U, and that the signal y^t perfectly reveals the first 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 that 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 et al. (2010), and the uncertainty about the market demand can be described as an unobservable state.

In the infinite-horizon stochastic game, players have a common discount factor $\delta \in (0, 1)$. Let $(\omega^\tau, a^\tau, y^\tau)$ be the state, the action profile, and the public signal in period τ. Then the history up to period $t \geq 1$ is denoted by $h^t = (a^\tau, y^\tau)_{\tau=1}^t$. Let H^t denote the set of all h^t for $t \geq 1$ and let $H^0 = \{\varnothing\}$. Let $H = \bigcup_{t=0}^\infty H^t$ be the set of all possible histories. A strategy for player i is a mapping $s_i : H \to \triangle A_i$ Let S_i be the set of all strategies for player i and let $S = \times_{i\in I} S_i$. Given a strategy s_i and history h^t, let $s_i|_{h^t}$ be the continuation strategy induced by s_i after history h^t.

Let $v_i^\omega(\delta, s)$ denote player i's average payoff in the stochastic game when the initial prior puts probability 1 on ω, the discount factor is δ, and players play strategy profile s. That is, let $v_i^\omega(\delta, s) = E[(1-\delta)\sum_{t=1}^\infty \delta^{t-1} g_i^{\omega^t}(a^t)|\omega, s]$. Similarly, let $v_i^\mu(\delta, s)$ denote player i's average payoff when the initial prior is μ. Note that for each initial prior μ, discount factor δ, and s_{-i}, player i's best reply s_i exists; see Appendix D in the working paper version (Yamamoto (2018)) for the proof. Let $v^\omega(\delta, s) = (v_i^\omega(\delta, s))_{i\in I}$ and $v^\mu(\delta, s) = (v_i^\mu(\delta, s))_{i\in I}$.

[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.

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 rewritten as follows. In period 1, 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

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

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 $\zeta : H \to \triangle\Omega$ be a belief system; i.e., $\zeta(h^t)$ is the posterior about ω^{t+1} after history h^t. A belief system ζ is *consistent with the initial prior* μ if there is a completely mixed strategy profile s such that $\zeta(h^t)$ is derived by Bayes' rule in all on-path histories of s. Since actions are observable, given the initial prior μ, a consistent belief is unique at each information set that 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 that yields a low payoff today may be preferred to one that 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

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

where $\operatorname{co} B$ denotes 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, is omitted. Note also that the minimizer s_{-i} indeed exists; see Appendix D in the working paper version (Yamamoto (2018)) for more details.

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

Assumption 1. (i) The limit of the feasible payoff set $\lim_{\delta\to 1} V^\mu(\delta)$ exists and is independent of the initial prior μ; that is, there is a set $V \subset \boldsymbol{R}^N$ such that $\lim_{\delta\to 1} d(V, V^\mu(\delta)) = 0$ for all μ.

(ii) For each i, the limit of the minimax payoff $\lim_{\delta\to 1} \underline{v}_i^\mu(\delta)$ 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 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 et al. (2011a)) 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 $\underline{v}_i = \lim_{\delta\to 1} \underline{v}_i^\mu(\delta)$. 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 is not stated in terms of primitives and, in general, it is hard to check. In later sections, we 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 is minimaxed for T periods and then those who minimaxed are 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 $\delta \to 1$, 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, expecting 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 0, 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 et al. (2011a) 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 $\delta \to 1$, 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.

However, 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 $\delta \to 1$, 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 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 1. (Note that taking p close to 1 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 1 and then take $\delta \to 1$. 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.

[8] In the POMDP literature, it is well known that the payoff in the discounted infinite-horizon problem and the (time-averaged) 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 $\delta \to 1$ *and* if the rate of convergence is at most $O(1 - \delta)$. (See Hsu et al. (2006) and the references therein.) Unfortunately, in out setup, the rate of convergence of the feasible payoffs and the minimax payoffs is slower than this bound for some cases. See the discussion about asymptotic uniform connectedness in Appendix A in the working paper version (Yamamoto (2018)).

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 of Proposition 1 for more details.

Hörner et al. (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 so as to measure payoffs by misreporting. In contrast, in this paper, the random blocks ensure that playing the minimax strategy over the block indeed approximates 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 $\overline{\delta} \in (0,1)$ such that for any $\delta \in (\overline{\delta}, 1)$ and for any initial prior μ, there is a sequential equilibrium with the payoff v.*

The proof of the proposition can be found in Appendix B.1. It is very similar to that of Dutta (1995) except that we use random blocks (rather than T-period blocks).

Proposition 1 imposes the full dimensional assumption, $\dim V^* = N$. This assumption allows us to use player-specific punishments, in the sense 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, and, hence, we need the full dimensional assumption.

4. Full support assumption

In the previous section, we showed 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 $\triangle\Omega$, 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 so as 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 such example 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 explain how to relax the full support assumption.

4.1 *Invariance of the feasible payoff set*

Let Λ be the set of directions $\lambda \in \boldsymbol{R}^N$ with $|\lambda| = 1$. For each direction λ, we compute the "score" using the 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 $\lambda = (1/\sqrt{2}, 1/\sqrt{2})$, the score is the (normalized) maximal social welfare within the feasible payoff set.

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 2 given that the initial prior is μ and players play a, and observe y in period 1. Then the score function f must solve the Bellman

[9]Note that this maximization problem indeed has a solution; see Appendix D in the working paper version (Yamamoto (2018)) for the proof.

equation

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

To interpret this equation, suppose that there are only two players. Let $\lambda = (1/\sqrt{2}, 1/\sqrt{2})$, so that the score $f(\mu)$ represents the maximal social welfare. Equation (1) asserts that the maximal welfare $f(\mu)$ is a sum of the (normalized) welfare today, $\lambda \cdot g^{\mu}(a) = (g_1^{\omega}(a) + g_2^{\omega}(a))/\sqrt{2}$, 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.

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

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

In particular, this implies that for each direction λ, the limit score $\lim_{\delta \to 1} \max_{v \in V^{\mu}(\delta)} \lambda \cdot v$ is independent of μ; hence, Assumption 1(i) follows.

Note that the limit $\lim_{\delta \to 1} \max_{v \in V^{\mu}(\delta)} \lambda \cdot v$ of the score indeed exists, thanks to Theorem 2 of Rosenberg et al. (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. The formal proof can be found in Appendix B.3.

Proposition 3. *If the full support assumption holds, then Assumption 1(ii) 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 (1). Unfortunately, the minimax payoff $\underline{v}_i^{\mu}(\delta)$ 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 her opponents minimize it. Accordingly, POMDP techniques are not applicable. The proof techniques for the observable-state

case do 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 briefly explain how to prove the result above.

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, $v_i^{\tilde{\mu}}(s_{-i}^{\tilde{\mu}}) = \underline{v}_i^{\tilde{\mu}}(\delta)$. 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 $\underline{v}_i^{\tilde{\mu}}(\delta)$. Define the *maximal value* $\overline{v}_i$ 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 Step 2 in Appendix B.3 for a more formal version.) Recall that $\overline{\pi}$ is the minimum of $\pi^{\omega}(y, \tilde{\omega}|a)$.

Key Result Fix δ. Suppose that $|\overline{v}_i - \underline{v}_i^{\mu}(\delta)| \approx 0$ for some interior belief μ such that $\mu(\omega) \geq \overline{\pi}$ for all ω. Then $|\overline{v}_i - \underline{v}_i^{\tilde{\mu}}(\delta)| \approx 0$ 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 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 $\underline{v}_i^{\mu}(\delta)$. At the same time, by the definition, this payoff must be smaller than the maximal value. Hence, we have

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

Since the minimax payoff $\underline{v}_i^{\mu}(\delta)$ approximates the maximal value $\overline{v}_i$, 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, i.e., player i's best payoff is convex with respect to the initial belief μ. □

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 for this part. See Step 1 in Appendix B.3 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 plays 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

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

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

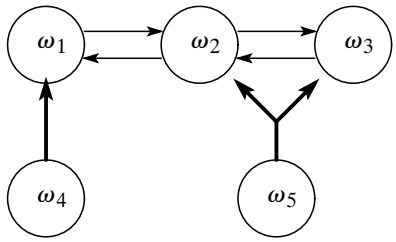

FIGURE 1. Weakly communicating states.

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 those for globally accessible states.

States are *weakly communicating* if each state ω is globally accessible or uniformly transient. Figure 1 is an example of weakly communicating states. The state moves along the arrows; for example, there is an action profile that moves the state from ω_1 to ω_2 with positive probability. Each thick arrow is a move that 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* in Dutta (1995), which requires that all states be globally accessible.

If states are weakly communicating, the feasible payoff set is invariant to the initial state for patient players. This result directly follows from Proposition 5,[10] but a rough idea is as follows. Consider the score toward some direction λ. Let ω be the state that 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 1 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 1. Since payoffs before the state reaches globally accessible states 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 1. Hence,

[10]To apply Proposition 5, note that in stochastic games with observable states, weakly communicating states imply uniform connectedness. See Proposition 7 in the working paper version (Yamamoto (2018)) for details.

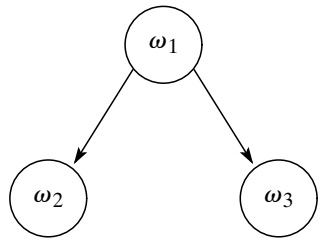

FIGURE 2. Two absorbing states.

as in Case 1, we can conclude that regardless of the initial state, the score is almost as good as the maximal score.

If states are not weakly communicating, the feasible payoff set may depend on the initial state, even for patient players. Figure 2 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 is 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 is 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 focus on the evolution of the *support* of the belief, rather than the evolution of the belief itself. A point is that there are only finitely many supports, so it is natural to expect that there be a globally accessible support. 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 that may depend on μ.[11] This definition can be viewed as an analogue of the global accessibility of the state ω in the observable-state case.

[11]Here, we define global accessibility and uniform transience using the posterior belief μ^t. In Appendix C in the working paper version (Yamamoto (2018)), 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.

Definition 2. A nonempty 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 that[12]

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

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 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 (2) does not. But this difference is not essential; indeed, although it is not explicitly stated in (2), 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.[13]

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 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 $\Pr(\mu^{T+1} = \tilde{\mu} | \mu, s) > 0$.[14]

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 is rather straightforward and can be found in Appendix B.2.

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.*

[12]Replacing 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 that satisfies the condition stated in the definition, we can find an action sequence that 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, we can restrict attention to an action sequence with length $T \leq 4^{|\Omega|}$. Indeed, whenever there is an action sequence $(a^1, \dots, a^T)$ that satisfies the property stated here, we can always find an action sequence $(\tilde{a}^1, \dots, \tilde{a}^{\tilde{T}})$ with $\tilde{T} \leq 4^{|\Omega|}$ that satisfies the same property. See Appendix C in the working paper version (Yamamoto (2018)) for more details.

[13]Since there are only finitely many supports, there is a bound π^* that works for all globally accessible sets Ω^*.

[14]Again, although the definition here does not provide an upper bound on T, when we check whether a given set Ω^* is uniformly transient, we can restrict attention to $T \leq 2^{|\Omega|}$. See Appendix C in the working paper version (Yamamoto (2018)) for more details. The strategy profile s in this definition cannot be replaced with an action sequence $(a^1, \dots, a^T)$.

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 2 is the whole state space Ω; hence, any proper subset $\Omega^* \subset \Omega$ is transient and the game is uniformly connected.

5.1.3 *Interpretation of uniform connectedness: Two-state case* To better understand the economic meaning of uniform connectedness, we 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 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 that 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 that reveals that the state tomorrow is ω_1 for sure. Specifically, we need one of the following conditions: (i) there are ω, a, and y such that $\pi^{\omega}(y, \omega_1|a) > 0$ and $\pi^{\tilde{\omega}}(y, \omega_2|a) = 0$ for all $\tilde{\omega}$; (ii) there are 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 1 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 state that puts probability 1 on ω_1, but they can move the belief to the state that puts probability 1 on ω_2 and then to the state that puts probability 1 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. However, 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.[15]

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 $\tilde{\omega}$.[16] 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.

To summarize, uniform connected games are composed 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.

[15]To 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 1 on ω_2, so the support stays at $\{\omega_2\}$ forever. Hence, the set $\{\omega_2\}$ cannot be uniformly transient.

[16]This 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 that maps the initial prior to the posterior belief in period 3, *conditional on this signal sequence* (y^1, y^2). Specifically, let

$$M_{ij} = \frac{\Pr\big(y^1, y^2, \omega^3 = \omega_j|\omega_i, s\big)}{\Pr\big(y^1, y^2|\omega_i, s\big)}$$

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 3 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 a 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.)

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 were caught today.

Unfortunately, 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.[17] 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 payoff of 1 each period, because she always knows the state. Alternatively, if the player does not know the state (say, the initial prior is 0.5–0.5), then her expected payoff is 0 each period, because her posterior is always 0.5–0.5. 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 nonnegligible 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.

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

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

Hence, Assumption 1(i) holds.

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

The proof of the proposition can be found in Appendix B.4. 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 that puts probability 1 on some state. Pick such a state ω and 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 that puts probability 1 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 that puts probability 1 on ω, the belief must eventually reach the beliefs whose supports are globally accessible, with probability 1. 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 $\mu \in \triangle\Omega^*$, 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 $\tilde{\mu} \in \triangle\Omega^*$ 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. The proof of this lemma relies on the fact that the score is convex with respect to μ so that players can attain better scores when they have better information about the state. See the formal proof for more details.

5.1.5 *Uniform connectedness and weakly communicating states* Uniform connectedness is an analogue of weakly communicating states for the hidden state model. These two conditions are actually identical for some special cases; Proposition 7 in the working paper version (Yamamoto (2018)) shows that in stochastic games with observable states or delayed observations, the game is uniformly connected if and only if states are weakly communicating.

Does this result extend to a more general setup? In Proposition 6 in the working paper version (Yamamoto (2018)), we show that the "only if" part remains true in any games, that is, the game is uniformly connected only if states are weakly communicating. However, the "if" part does not extend, so in general, the uniform connectedness is more demanding than requiring states to be weakly communicating.

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 explain 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, 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 plays 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 $\tilde{\omega}$ that 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 in 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 Proposition 6,[18] but a rough idea is as follows. Let ω be the initial state that 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 1 and give the lowest minimax payoff to player i after that. When player i is patient, payoffs before the state reaching ω are 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 1, player i can move the state to some robustly accessible states in finite time, and after that she earns 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 nonempty 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 that[19]

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

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 global 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

[18]Again, to apply Proposition 6, note that in the standard stochastic games, weak irreducibility implies robust connectedness.

[19]As shown in Appendix C in the working paper version (Yamamoto (2018)), when we check whether 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.

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 that[20]

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

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 $\Pr(h^T | \omega, s) > 0$ 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 support induced by the initial prior $\mu = (1/|\Omega|, \dots, 1/|\Omega|)$.

The merging support condition ensures that regardless of players' play, two different initial priors ω and $\mu = (1/|\Omega|, \dots, 1/|\Omega|)$ 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. To enable the use of 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 nonempty 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 do 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

[20]As shown in Appendix C in the working paper version (Yamamoto (2018)), when we check whether a given set is avoidable, we can restrict attention to $T \leq 4^{|\Omega|}$, without loss of generality.

connectedness.[21] As explained in the previous 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.5.

Proposition 6. *Suppose that the game is robustly connected for player i. Then Assumption 1(ii) holds.*

From Propositions 5 and 6, for any game that satisfies both uniform connectedness and robust connectedness, the payoff invariance assumption (Assumption 1) holds and, hence, the folk theorem obtains.

Of course, robust connectedness is only a sufficient condition for invariance of the minimax payoffs, and there are many games in which robust connectedness is not satisfied but yet the minimax payoff is invariant. For example, in stochastic games with delayed observations and weakly irreducible states, robust connectedness may not hold, but the minimax payoff is always invariant. See Section 5.2.3 in the working paper version (Yamamoto (2018)).

5.3 *Example: Natural resource management*

Now we present an example of natural resource management. This is an example that 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 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

[21] If there are only two states and neither of them can be revealed, the scrambling condition and 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.

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 formula

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

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 are 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 we 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 is 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 grow 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 no longer holds. Indeed, if the initial state is $\omega = 0$ and the fishermen choose (F, F), the belief does not change forever and and never becomes 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 1 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 1 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 $\overline{\pi}^K$ on the highest state $\omega = K$ (i.e., $\mu^{K+1}(K) \geq \overline{\pi}^K$). Then if $y = (1, 1)$ is observed in the next period, the posterior belief μ^{K+2} puts probability 0 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 0 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 $\mu^{K+1}(K)\overline{\pi}^{K-1} \geq \overline{\pi}^{2K-1}$, 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. Also, if the opponent always chooses F, the state eventually reaches $\omega = 0$ with probability 1 and, thus, fisherman i's payoff is at most 0 in the limit as $\delta \to 1$. 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 individually 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.

Appendix A: Extension of Uniform Connectedness

Proposition 5 shows that uniform connectedness ensures invariance of the feasible payoff set. Here we explain that uniform connectedness can be relaxed further. Consider the following example.

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. $\Diamond$

While uniform connectedness does not hold in this example, 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 1. 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 1, 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 1 on ω_1, regardless of the initial prior. As we explain, this property is enough to establish invariance of the feasible payoff set. Formally, asymptotic accessibility is defined as follows.

Definition A1. A nonempty 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 0 on ω, then the posterior belief in the second period puts probability close to 1 on ω. This ensures that there is a uniform bound T on the length T^* of the action sequence.

In the same vein, we can define *asymptotic uniform transience* as an extension of uniform transience. The game is *asymptotically uniformly connected* if each set Ω^* is asymptotically accessible or asymptotically uniformly connected. Asymptotic uniform connectedness is weaker than uniform connectedness, but still ensures invariance of the feasible payoff set. Asymptotic uniform connectedness is satisfied in many examples, and, in particular, we can show that asymptotic uniform connectedness holds if states are weakly communicating and if the signal distributions $\{\pi_Y^\omega(a) | \omega \in \Omega\}$ are linearly independent for each a. See Appendix A in the working paper version (Yamamoto (2018)) for more details.

Appendix B: Proofs

B.1 *Proof of Proposition 1: The folk theorem*

B.1.1 *Equilibrium with pure minimax strategies* We first consider the case in which the minimax strategies are pure strategies. Take an interior point $v \in V^*$. We construct a sequential equilibrium with the payoff v when δ is close to 1. 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) The payoff vector $w(i)$ 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 3 to see how to choose these payoffs $w(i)$. In this figure, the payoffs are normalized so that the limit minimax payoff vector is $\underline{v} = (\underline{v}_1, \underline{v}_2) = (0, 0)$.

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

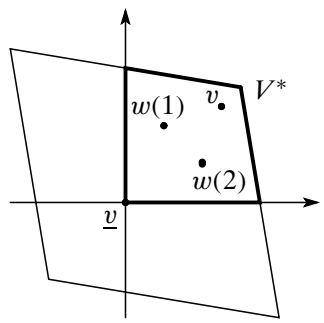

FIGURE 3. Payoffs $w(1)$ and $w(2)$.

is punished. This ensures that player i is indeed willing to punish 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 1 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.
- For each i, $\sup_{\mu \in \triangle\Omega} \underline{v}_i^\mu(p) < w_i(i)$.

By continuity, if the discount factor δ is close to 1, 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 that 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 that 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 ends, a new block starts and the players 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 1, 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 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 $p\delta$. Player i maximizes her payoff during these K blocks. After the K blocks, players switch their play so as to achieve the post-minimax payoff $w(i)$; that is, in each random block, players play a pure strategy profile s that 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 is 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, so as to punish player j. See Figure 4.

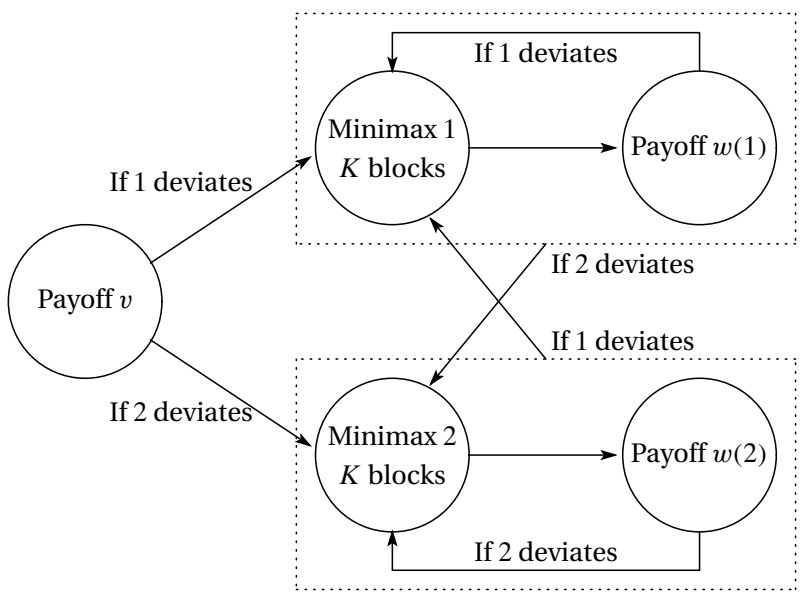

FIGURE 4. Equilibrium strategy.

Now, choose K such that

$$-\overline{g} - \frac{1}{1-p}\overline{g} + \frac{K-1}{1-p} w_i(i) > \overline{g} + \frac{K}{1-p} \sup_{\mu \in \triangle\Omega} \underline{v}_i^{\mu}(p) \tag{3}$$

for each i. Note that (3) indeed holds for sufficiently large K, as $\sup_{\mu\in\triangle\Omega} \underline{v}_i^{\mu}(p) < w_i(i)$. To interpret (3), 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 $w(i)$. Equation (3) ensures that player i's deviation today is not profitable for δ close to 1. To see why, suppose that player i deviates today. Then her stage-game payoff today is at most $\overline{g}$ and then she will be minimaxed for the next K random blocks. Since the expected length of each block is $1/1-p$, the (unnormalized) expected payoff during the minimax phase is at most

$$\frac{K}{1-p}\sup_{\mu\in\triangle\Omega} \underline{v}_i^{\mu}(p)$$

when $\delta \to 1$. So when she deviates, the right-hand side of (3) is an upper bound on player i's unnormalized payoff until the minimax play ends.

Alternatively, if she does not deviate, her payoff today is at least $-\overline{g}$. Also, for the next K periods, she can earn at least

$$-\frac{1}{1-p}\overline{g} + \frac{K-1}{1-p}w_i(i),$$

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 $-(\overline{g}/1-p)$ 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 (3), 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 that in Dutta (1995) and, hence, is omitted.

B.1.2 *Equilibrium with mixed minimax strategies* Now we consider the case in which the minimax strategies are mixed strategies. In this case, we need to modify the above equilibrium construction and make player i indifferent over all actions when she minimaxes player $j \neq i$. This can be done by perturbing the post-minimax payoff $w_i(j)$ appropriately, as is explained below. The idea here is very similar to Fudenberg and Maskin (1986).

We first explain how to perturb the continuation 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 kth block of the minimax play. Then the perturbed continuation payoff is defined as

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

That is, the continuation payoff is now the original value $w_i(j)$ plus the K perturbation terms $R_i(\mu^{(1)}, a^{(1)}), \dots, R_i(\mu^{(K)}, a^{(K)})$ with some coefficient.

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 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 τth period of the block, player i's unnormalized payoff in the continuation game from now on is

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

Here, (μ^t, a^t) denotes the belief and the action in the tth 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

$$w_i(j) + \frac{(1-\delta)E\big[R_i(\mu^t, a^t)\big]}{\delta(1-p)},$$

where the expectation is taken with respect to μ^t and a^t, conditional on the block not terminating until period t. We have the term δ^t due to discounting and we have $1/(1-\delta)$ so as to convert the average payoff to the unnormalized payoff. The above payoff can be rewritten as

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

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 kth 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 Kth block ends. To offset discounting, we have the coefficient $(1-p\delta)^{K-k}/\{\delta(1-p)\}^{K-k+1}$ 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

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

Here the first term on the left-hand side comes from the fact that the length of the next block is 1 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 2 with probability $p(1-p)$, in which case discounting due to the next block is δ^2. This discount factor $\delta(1-p)/(1-p\delta)$ cancels out, thanks to the coefficient $(1-p\delta)/\{\delta(1-p)\}^2$ on $R_i(\mu^{(K-1)}, a^{(K-1)})$. Hence, player i is indifferent in all periods during 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 1, because the perturbation terms are of order $1-\delta$. Hence, for sufficiently large δ, the perturbed payoff vector is in the feasible payoff set and all other incentive constraints are still satisfied.

B.2 *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.3 *Proof of Proposition 3: Invariance of the minimax payoffs*

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

$$\left|\underline{v}_i^{\mu}(\delta)-\underline{v}_i^{\tilde{\mu}}(\delta)\right|<\varepsilon \tag{4}$$

for any $\delta\in(\overline{\delta},1)$, μ, and $\tilde{\mu}$. After that, we 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 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

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

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

$$\Big|\sup_{\mu\in\triangle\Omega} \overline{v}_i\big(s_{-i}^{\mu}\big) - \overline{v}_i\big(s_{-i}^{\mu^*}\big)\Big| < 1-\delta$$

and call $\overline{v}_i(s_{-i}^{\mu^*})$ *the maximal value.* When δ is close to 1, 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 1 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 $\tilde{\mu}$.

Step 0. Preliminary lemma The following lemma follows from the convexity of the payoffs $v_i^{\tilde{\mu}}(s_{-i}^{\mu})$. We 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 $\triangle\Omega$. Then for each $\hat{\mu} \in \triangle\Omega$,*

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

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 $\overline{v}_i(s_{-i}^{\mu^*})$ 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 $\overline{v}_i(s_{-i}^{\mu^*})$ for *some* interior belief $\tilde{\mu}$ such that $\tilde{\mu}(\tilde{\omega}) \geq \overline{\pi}$ for all $\tilde{\omega}$. The condition $\tilde{\mu}(\tilde{\omega}) \geq \overline{\pi}$ implies that this belief $\tilde{\mu}$ is not too close to the boundary of the belief space $\triangle\Omega$. Then the right-hand side of the inequality in the lemma is approximately 0, as $p \geq \overline{\pi}$. Hence, the left-hand side must be approximately 0, 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 $\triangle\Omega$. If $\tilde{\mu}$ approaches the boundary of $\triangle\Omega$, then p approaches 0 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

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

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

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

Arranging yields

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

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

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

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

$$\left|\overline{v}_i\left(s_{-i}^{\mu^*}\right) + (1-\delta) - v_i^{\tilde{\omega}}\left(\delta, s_i, s_{-i}^{\mu}\right)\right| \leq \frac{\left|\overline{v}_i\left(s_{-i}^{\mu^*}\right) + (1-\delta) - v_i^{\tilde{\mu}}\left(s_{-i}^{\mu}\right)\right|}{p}. \tag{5}$$

Now pick an arbitrary $\hat{\mu} \in \triangle\Omega$. Note that (5) holds for each $\tilde{\omega} \in \Omega$. So multiplying both sides of (5) by $\hat{\mu}(\tilde{\omega})$ and summing over all $\tilde{\omega} \in \Omega$ yields

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

Then we have

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

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 (6). □

Step 1. Minimax payoff for some belief μ^{**} In this step, we 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:

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

Here, α^* denotes the action profile in period 1 induced by $(s_i, s_{-i}^{\mu^*})$; $\mu(y|\omega, a)$ denotes the posterior belief in period 2 when the initial belief is $\tilde{\mu}^* = \omega$ and players play a and observe y in period 1; $\mu(y|\mu^*)$ denotes the posterior belief when the initial belief is μ^*. Given an outcome (a, y) in period 1, 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. Let $\overline{g}_i = \max_{\omega,a}|g_i^\omega(a)|$ and let $\overline{g} = \sum_{i\in I}\overline{g}_i$.

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

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

Proof. Pick (a,y) that 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

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

Arranging yields

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

This implies

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

Since $g_i^\omega(\alpha^*) - v_i^\omega(s_{-i}^{\mu^*}) \le 2\overline{g}$, we obtain the desired inequality. □

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

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

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, $\mu(y|\omega,a)[\tilde{\omega}] \ge \overline{\pi}$ for all $\tilde{\omega}$. Hence, Lemma B1 ensures that

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

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,

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

Step 2. Minimax payoffs for other beliefs Now we 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, by the definition of the minimax payoff, player i's payoff $v_i^{\mu^{**}}(s_{-i}^{\mu})$ is at least the minimax payoff for μ^{**}.

At the same time, her payoff cannot exceed the maximal value $v_i^{\omega}(s_{-i}^{\mu^*}) + (1-\delta)$. So we have

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

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

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

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,

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

for all beliefs $\hat{\mu}$. This implies that the minimax payoff for μ approximates the maximal value, as desired. Hence, (4) follows.

Step 3. Existence of the limit minimax payoff Now we verify that the limit of the minimax payoff exists. Take i, μ, and $\varepsilon > 0$ arbitrarily. Let $\overline{\delta} \in (0,1)$ be such that

$$\left|\underline{v}_i^{\mu}(\overline{\delta}) - \liminf_{\delta\to 1} \underline{v}_i^{\mu}(\delta)\right| < \frac{\varepsilon}{2} \tag{7}$$

and such that

$$\left|\underline{v}_i^{\mu}(\overline{\delta}) - \underline{v}_i^{\tilde{\mu}}(\overline{\delta})\right| < \frac{\varepsilon}{2} \tag{8}$$

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

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

$$\max_{s_i\in S_i} v_i^{\mu}\left(\delta, s_i, s_{-i}^{\mu}\right) < \liminf_{\delta\to 1} \underline{v}_i^{\mu}(\delta) + \varepsilon \tag{9}$$

for each $\delta \in (\overline{\delta}, 1)$. That is, we show that when the true discount factor is δ, player i's best payoff against the minimax strategy for the discount factor $\overline{\delta}$ is worse than the limit inferior of the minimax payoff. Since the minimax strategy for the discount factor $\overline{\delta}$ is not necessarily the minimax strategy for δ, the minimax payoff for δ is less than $\max_{s_i\in S_i} v_i^{\mu}(\delta, s_i, s_{-i}^{\mu})$. Hence, (9) ensures that the minimax payoff for δ is worse than the limit inferior of the minimax payoff. Since this is true for all $\delta \in (\overline{\delta}, 1)$, the limit inferior is the limit, as desired.

So pick an arbitrary $\delta \in (\overline{\delta}, 1)$ and compute $\max_{s_i \in S_i} v_i^\mu(\delta, s_i, s_{-i}^\mu)$, player i's best payoff against the minimax strategy for the discount factor $\overline{\delta}$. To evaluate this payoff, we regard the infinite horizon as a series of random blocks, as in Section 3. The termination probability is $1-p$, where $p = \overline{\delta}/\delta$. Then, since s_{-i}^μ is Markov, playing s_{-i}^μ in the infinite-horizon game is the same as playing the following strategy profile:

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

Then 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,

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

where $s_i^{\mu^k}$ is the optimal (Markov) strategy in the continuation game from the kth block with belief μ^k. Note that $s_i^{\mu^k}$ may not maximize the payoff during the kth block, because player i needs to take into account the fact that her action during the kth block influences μ^{k+1} and, hence, the payoffs after the kth block. But in any case, we have $v_i^{\mu^k}(p\delta, s_i^{\mu^k}, s_{-i}^{\mu^k}) \le \underline{v}_i^{\mu^k}(\overline{\delta})$ because $s_{-i}^{\mu^k}$ is the minimax strategy with discount factor $p\delta = \overline{\delta}$. Hence,

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

Using (8),

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

Then using (7), we obtain (9).

Note that this proof does not assume public randomization. Indeed, random blocks are useful to compute the payoff by the strategy s_{-i}^μ, but the strategy s_{-i}^μ itself does not use public randomization.

B.4 *Proof of Proposition 5: Score and uniform connectedness*

We 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 that solves $\max_{s \in S} \lambda \cdot v(\delta, s)$. That is, s^μ is the profile that 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

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

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

Step 0. Preliminary lemmas We begin by 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* beliefs $\tilde{\mu}$ 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 $\triangle\Omega^*$. Then for each $\tilde{\mu} \in \triangle\Omega^*$,*

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

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

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

for all $\tilde{\mu} \in \triangle\Omega^*$. 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.

Lemma B3 relies on the convexity of the score, and the proof idea is essentially the same as that presented in Section 4.3. For completeness, we provide the formal proof.

Proof of Lemma B3. Pick an arbitrary belief μ and let Ω^* be the support of μ. Pick $\tilde{\omega} \in \Omega^*$ arbitrarily. Then we have

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

Applying (9) to the above inequality, we obtain

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

Arranging gives

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

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

$$\lambda \cdot v^{\omega}(\delta, s^{\omega}) - \lambda \cdot v^{\tilde{\omega}}(\delta, s^{\mu}) \leq \frac{\lambda \cdot v^{\omega}(\delta, s^{\omega}) - \lambda \cdot v^{\mu}(\delta, s^{\mu})}{p}. \tag{11}$$

Pick an arbitrary belief $\tilde{\mu} \in \triangle\Omega^*$. Recall that (11) holds for each $\tilde{\omega} \in \Omega^*$. Multiplying both sides of (11) by $\tilde{\mu}(\tilde{\omega})$ and summing over all $\tilde{\omega} \in \Omega^*$ gives

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

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

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

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

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 that 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 $1/|A|^T$ 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 $1/|A|^T \cdot \pi^*$. This probability is bounded from 0o for all μ and, hence, the proof is completed. □

Step 1. Scores for beliefs with support Ω^* As a first step of the proof, we show that there is a globally accessible set Ω^* such that the score for any belief $\mu \in \triangle\Omega^*$ approximates the maximal score. More precisely, we prove the following lemma.

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

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

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

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

Proof of Lemma B5. First, consider the case in which $\{\omega\}$ is globally accessible. Let $\Omega^* = \{\omega\}$. Then this set Ω^* satisfies the desired property, because $\triangle\Omega^*$ contains only the belief $\mu = \omega$, and the score for this belief is exactly equal to the maximal score.

Next consider the case in which $\{\omega\}$ is uniformly transient. Take T, h^T, μ^*, and Ω^* as stated above. By the definition, the support of μ^* is Ω^*. Also, μ^* assigns at least $\overline{\pi}^T$ to each state $\tilde{\omega} \in \Omega^*$, i.e., $\mu^*(\tilde{\omega}) \geq \overline{\pi}^T$ for each $\tilde{\omega} \in \Omega^*$. This is so because

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

where the last inequality follows from the fact that $\overline{\pi}$ 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:

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

Using (10), $\mu(h^T) = \mu^*$, and $(1-\delta)\sum_{t=1}^{T} \delta^{t-1} E[\lambda \cdot g^{\omega^t}(a^t) | \omega^1 = \omega, s^\omega] \leq (1-\delta^T)\overline{g}$, we obtain

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

Arranging, we have

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

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

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

Since (10) ensures that the left-hand side is nonnegative, taking the absolute values of both sides and using $\lambda \cdot v^\omega(\delta, s^\omega) \geq -\overline{g}$ yields

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

That is, the score for the belief μ^* approximates the maximal score if δ is close to 1. As noted, we have $\mu^*(\tilde{\omega}) \geq \overline{\pi}^T$ for each $\tilde{\omega} \in \Omega^*$. Then applying Lemma B3 to the inequality above, we obtain

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

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

Step 2. Scores for all beliefs μ In the previous step, we showed that the score approximates the maximal score for any belief μ with the support Ω^*. Now we 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 moves to Ω^* (or its subset) within $4^{|\Omega|}$ periods with probability at least π^*, regardless of the initial prior.

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

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

That is, players wait until the support of the belief reaches Ω^*, and once it happens, they switch the play to the optimal policy s^{μ^t} in the continuation game. Lemma 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 1 in finite time and that waiting time is almost negligible for patient players. Hence, the payoff of this strategy profile $\tilde{s}^\mu$ approximates the maximal score. Formally, we have the following lemma. The proof is mechanical and can be found in Section S.1 in the Supplemental Material, available in a supplementary file on the journal website, http://econtheory.org/supp/3068/supplement.pdf.

Lemma B6. *For each μ,*

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

Note that

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

that is, the score for μ is at least $\lambda \cdot v^{\mu}(\delta, \tilde{s}^{\mu})$ (this is because $\tilde{s}^{\mu}$ is not the optimal policy) and is at most the maximal score. Then from Lemma B6, we have

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

as desired.

B.5 *Proof of Proposition 6: Minimax and robust connectedness*

We prove only (4). 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 $v_i^{\tilde{\mu}}(s_{-i}^{\mu}) = \max_{s_i \in S_i} v_i^{\tilde{\mu}}(\delta, s_i, s_{-i}^{\mu})$, 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

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

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

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

We call $\overline{v}_i(s_{-i}^{\mu^{*}})$ 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 $\overline{v}_i(s_{-i}^{\mu})$, the initial prior $\tilde{\mu}$ is chosen from the set $\triangle(\operatorname{supp}\mu)$.

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

Step 0. Preliminary lemmas We begin by 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 $(\mu, \tilde{\mu})$ that 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 $\tilde{\mu} \in \triangle\Omega^{*}$, then it approximates the maximal value for all beliefs $\hat{\mu} \in \triangle\Omega^{*}$. The proof of the lemma is very similar to that of Lemma B1 and, hence, is omitted.

Lemma B7. *Pick an arbitrary belief μ and let Ω^* denote its support. Let $\tilde{\mu} \in \triangle\Omega^*$ be a 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 $\triangle\Omega^*$. Then for each $\hat{\mu} \in \triangle\Omega^*$,*

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

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 $\mu(\omega) > 0$ 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 $\Pr(h^T|\omega, s) > (|\overline{\pi}|/|A|)^T$ and such that the support of the posterior belief induced by the initial state ω and the history h^T is identical with that 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 that is chosen with probability at least $1/|A|$ 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 that induced by the initial prior $\tilde{\mu} = (1/|\Omega|, \ldots, 1/|\Omega|)$ and the history h^T. We show that T and h^T here satisfy the desired properties.

Note that $\Pr(h^T|\omega, \tilde{s}) \geq \overline{\pi}^T$, as $\overline{\pi}$ is a pure strategy. This implies that

$$\Pr\left(h^T|\omega, s\right) \geq \left(\frac{\overline{\pi}}{|A|}\right)^{4^{|\Omega|}},$$

since each period the action profile by s coincides with the action profile by $\tilde{s}$ with probability at least $1/|A|$. 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. □

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 $\triangle\Omega^*$.

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 $\tilde{\mu} \in \triangle\Omega^*$ such that the*

posterior belief μ^{T+1} equals $\tilde{\mu}$ with probability at least π^* and such that $\tilde{\mu}(\omega) \geq \overline{\pi}^{4^{|\Omega|}}/|\Omega|$ 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 $T \leq 4^{|\Omega|}$, a pure action sequence $(a_{-i}^1, \ldots, a_{-i}^T)$, and a signal sequence $(y^1, \ldots, y^T)$ such that the following properties are satisfied:

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

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

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

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

Now we prove the lemma. Fix an arbitrary prior μ and pick ω such that $\mu(\omega) \geq 1/|\Omega|$. Then for each s_i, choose T, $(a_{-i}^1, \ldots, a_{-i}^T)$, and $(y^1, \ldots, y^T)$ as stated in the above condition. Property (i) ensures that if the initial prior is μ, player i plays s_i, and the opponents mix all actions equally, the action sequence $(a_{-i}^1, \ldots, a_{-i}^T)$ and the signal sequence $(a_{-i}^1, \ldots, a_{-i}^T)$ are observed with probability at least

$$\mu(\omega)\left(\frac{\overline{\pi}}{|A|^T}\right)^T \geq \frac{1}{|\Omega|}\left(\frac{\overline{\pi}}{|A|^T}\right)^{4^{|\Omega|}}.$$

Let $\tilde{\mu}$ be the posterior belief in period $T+1$ in this case. From (iii), $\tilde{\mu}(\omega) \geq \overline{\pi}^{4^{|\Omega|}}/|\Omega|$ for all $\omega \in \Omega^*$. From (ii), $\tilde{\mu}(\omega) = 0$ for other ω. □

Step 1. Minimax payoff for μ^{**} As a first step, we show that there is some belief μ^{**} whose minimax payoff approximates the maximal value. The proof idea is similar to

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

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 see, 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:

$$\begin{aligned} v_i^\omega\big(s_{-i}^{\mu^*}\big) &= (1-\delta)\sum_{t=1}^{T}\delta^{t-1}E\big[g_i^{\omega^t}\big(a^t\big)|\omega, s_i^*, s_{-i}^{\mu^*}\big] \\ &\quad + \delta^T \sum_{\tilde{h}^T\in H^T} \Pr\big(\tilde{h}^T|\omega, s_i^*, s_{-i}^{\mu^*}\big) v_i^{\mu(\tilde{h}^T|\omega)}\big(s_{-i}^{\mu(\tilde{h}^T|\mu^*)}\big). \end{aligned} \tag{12}$$

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 $T \le 4^{|\Omega|}$ and h^T such that the two posteriors $\mu(h^T|\omega)$ and $\mu(h^T|\mu^*)$ have the same support, and such that*

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

Proof. Since $\mu^*(\omega) > 0$, Lemma B8 ensures that there is a natural number $T \le 4^{|\Omega|}$ and a history h^T such that $\Pr(h^T|\omega, s_i^*, s_{-i}^{\mu}) > (1\overline{\pi}/|A|)^T$, and such that the two posterior beliefs $\mu(h^T|\omega)$ and $\mu(h^T|\mu^*)$ have the same support. Pick such T and h^T.

By the definition of $\overline{g}$, we have $(1-\delta)\sum_{t=1}^{T}\delta^{t-1}E[g_i^{\omega^t}(a^t)|\omega, s] \le (1-\delta^T)\overline{g}$. Also, since $\mu^*(\omega) > 0$, for each $\tilde{h}^T$, the support of $\mu(\tilde{h}^T|\omega)$ is a subset of the support of $\mu(\tilde{h}^T|\mu^*)$, which implies $v_i^{\mu(\tilde{h}^T|\omega)}(s_{-i}^{\mu(\tilde{h}^T|\mu^*)}) \le v_i^\omega(s_{-i}^{\mu^*}) + (1-\delta)$. Plugging them and $\Pr(h^T|\omega, s_i^*, s_{-i}^{\mu^*}) \ge (\overline{\pi}/|A|)^T$ into (12), we have

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

Arranging yields

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

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

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

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

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

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

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

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

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

$$\mu\big(h^T|\omega\big)[\tilde{\omega}] = \frac{\Pr\big(\omega^{T+1} = \tilde{\omega}|\omega, a^1, \ldots, a^T\big)}{\sum_{\hat{\omega}\in\Omega}\Pr\big(\omega^{T+1} = \hat{\omega}|\omega, a^1, \ldots, a^T\big)}$$
$$\geq \Pr\big(\omega^{T+1} = \tilde{\omega}|\omega, a^1, \ldots, a^T\big) \geq \overline{\pi}^T \geq \overline{\pi}^{4^{|\Omega|}}.$$

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

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

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

$$\big|v_i^\omega\big(s_{-i}^{\mu^*}\big) + (1-\delta) - v_i^{\mu^{**}}\big(s_{-i}^{\mu^{**}}\big)\big| \leq \frac{\big(1-\delta^{4^{|\Omega|}}\big)2\overline{g}|A|^{4^{|\Omega|}}}{\delta^{4^{|\Omega|}}\overline{\pi}^{(4^{|\Omega|}+4^{|\Omega|})}} + \frac{(1-\delta)|A|^{4^{|\Omega|}}}{\overline{\pi}^{(4^{|\Omega|}+4^{|\Omega|})}}, \tag{13}$$

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

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 that 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 $\tilde{\mu} \in \triangle(\operatorname{supp}\mu)$ such that $\tilde{\mu}(\tilde{\omega}) \geq \overline{\pi}^{4^{|\Omega|}}/|\Omega|$ for all $\tilde{\omega} \in \operatorname{supp}\mu$. Intuitively, Δ^{μ} is the set of all beliefs $\tilde{\mu}$ with the same support as μ, except those that are too close to the boundary of $\triangle(\operatorname{supp}\mu)$.

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

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

Intuitively, the opponents wait until the belief reaches Δ^{μ}, and once it happens, they switch the play to the minimax strategy s_{-i}^{μ} for the fixed belief μ. From Lemma B9, this switch happens in finite time with probability 1 regardless of player i's play. So for δ close to 1, payoffs before the switch are 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

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

Hence, player i's payoff against the above strategy $\tilde{s}_{-i}^{\mu}$ 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,*

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

The proof of this lemma is mechanical and very similar to that of Lemma B6, and can be found in Section S.2 of the Supplemental Material.

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

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

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,

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

From (13), 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

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

Equivalently,

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

where $\tilde{\mu}$ is the belief that 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 ensures that the same result holds for all beliefs with the same support, that is,

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

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

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 $\tilde{s}_i^\mu$:

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

Intuitively, player i waits until the support of the posterior belief becomes robustly accessible, and once it happens, she plays a best reply to the opponents' continuation

strategy $s_{-i}^{\mu^t}$, where μ^t is the belief when the switch happens. (Here the opponents' continuation strategy is the minimax strategy $s_{-i}^{\mu^t}$, since the strategy s_{-i}^{μ} is Markov and induces the minimax strategy in every continuation game.) Note that player i's continuation payoff after the switch is exactly equal to the minimax payoff $v_i^{\mu^t}(s_{-i}^{\mu^t})$. 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 1, player i's payoff by playing the above strategy $\tilde{s}_i^{\mu}$ also approximates the maximal value. Formally, we have the following lemma. The proof is very similar to that of Lemma B11 and, hence, is omitted.

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

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

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

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

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.

References

Arellano, Cristina (2008), "Default risk and income fluctuations in emerging economies." *American Economic Review*, 98, 690–712. [1116]

Athey, Susan and Kyle Bagwell (2008), "Collusion with persistent cost shocks." *Econometrica*, 76, 493–540. [1119]

Bagwell, Kyle and Robert W. Staiger (1997), "Collusion over the business cycle." *RAND Journal of Economics*, 28, 82–106. [1115]

Besanko, David, Ulrich Doraszelski, Yaroslav Kryukov, and Mark Satterthwaite (2010), "Learning by doing, organizational forgetting, and industry dynamics." *Econometrica*, 78, 453–508. [1120]

Doob, Joseph L. (1953), *Stochastic Processes*. Wiley Publications in Statistics. John Wiley & Sons. [1117]

Duggan, John (2012), "Noisy stochastic games." *Econometrica*, 80, 2017–2045. [1116]

Dutta, Prajit K. (1995), "A folk theorem for stochastic games." *Journal of Economic Theory*, 66, 1–32. [1116, 1118, 1122, 1123, 1124, 1129, 1146]

Dutta, Prajit K. and Rangarajan K. Sundaram (1998), "The equilibrium existence problem in general Markovian games." In *Organizations With Incomplete Information: Essays in Economic Analysis: A Tribute to Roy Radner* (Mukul Majumdar, ed.), 159–207, Cambridge University Press, New York, New York. [1116]

Escobar, Juan F. and Juuso Toikka (2013), "Efficiency in games with Markovian private information." *Econometrica*, 81, 1887–1934. [1119]

Fudenberg, Drew and Eric Maskin (1986), "The folk theorem for repeated games with discounting or with incomplete information." *Econometrica*, 54, 533–554. [1122, 1124]

Fudenberg, Drew and Yuichi Yamamoto (2010), "Repeated games where the payoffs and monitoring structure are unknown." *Econometrica*, 78, 1673–1710. [1119]

Fudenberg, Drew and Yuichi Yamamoto (2011a), "Learning from private information in noisy repeated games." *Journal of Economic Theory*, 146, 1733–1769. [1119]

Fudenberg, Drew and Yuichi Yamamoto (2011b), "The folk theorem for irreducible stochastic games with imperfect public monitoring." *Journal of Economic Theory*, 146, 1664–1683. [1116, 1119, 1122, 1136]

Haltiwanger, John and Joseph E. Harrington Jr. (1991), "The impact of cyclical demand movements on collusive behavior." *RAND Journal of Economics*, 22, 89–106. [1115]

Hörner, Johannes, Takuo Sugaya, Satoru Takahashi, and Nicolas Vieille (2011a), "Recursive methods in discounted stochastic games: An algorithm for $\delta \to 1$ and a folk theorem." *Econometrica*, 79, 1277–1318. [1116, 1119, 1122, 1123]

Hörner, Johannes, Satoru Takahashi, and Nicholas Vieille (2015), "Truthful equilibria in dynamic Bayesian games." *Econometrica*, 83, 1795–1848. [1119, 1124]

Hörner, Johannes, Satoru Takahashi, and Nicolas Vieille (2011b), "Recursive methods in discounted stochastic games II: Infinite state space." Unpublished paper, Yale University, Department of Economics. [1116]

Hsu, Shun-Pin, Dong-Ming Chuang, and Ari 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. [1118, 1123]

Kandori, Michihiro (1991), "Correlated demand shocks and price wars during booms." *Review of Economic Studies*, 58, 171–180. [1115]

Levy, Yehuda (2013), "Discounted stochastic games with no stationary Nash equilibrium: Two examples." *Econometrica*, 81, 1973–2007. [1116]

Platzman, Loren K. (1980), "Optimal infinite-horizon undiscounted control of finite probabilistic systems." *SIAM Journal on Control and Optimization*, 18, 362–380. [1117, 1118, 1126]

Renault, Jérôme and Bruno Ziliotto (2014), "Hidden stochastic games and limit equilibrium payoffs." Unpublished paper, arXiv:1407.3028v2. [1119]

Rosenberg, Dinah, Eilon Solan, and Nicolas Vieille (2002), "Blackwell optimality in Markov decision processes with partial observation." *Annals of Statistics*, 30, 1178–1193. [1117, 1126]

Ross, Sheldon M. (1968), "Arbitrary state Markovian decision processes." *Annals of Mathematical Statistics*, 39, 2118–2122. [1118]

Rotemberg, Julio and Garth Saloner (1986), "A supergame-theoretic model of price wars during booms." *American Economic Review*, 76, 390–407. [1115]

Shapley, Lloyd (1953), "Stochastic games." *Proceedings of the National Academy of Sciences*, 39, 1095–1100. [1118]

Wiseman, Thomas (2005), "A partial folk theorem for games with unknown payoff distributions." *Econometrica*, 73, 629–645. [1119]

Wiseman, Thomas (2012), "A partial folk theorem for games with private learning." *Theoretical Economics*, 7, 217–239. [1119]

Yamamoto, Yuichi (2018), "Stochastig Games with Hidden States." Available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2563612. [1118, 1120, 1122, 1123, 1125, 1129, 1130, 1131, 1135, 1137, 1138, 1139, 1143]

Co-editor Simon Board handled this manuscript.

Manuscript received 13 November, 2017; final version accepted 2 December, 2018; available online 17 December, 2018.