PIER Working Paper 12-003

"A Foundation for Markov Equilibria with Finite Social Memory"

by

V. Bhaskar, George J. Mailath, and Stephen Morris

http://ssrn.com/abstract=1998810

A Foundation for Markov Equilibria with Finite Social Memory[*]

V. Bhaskar[†], George J. Mailath[‡] and Stephen Morris[§]

January 31, 2012

Abstract

We study stochastic games with an infinite horizon and sequential moves played by an arbitrary number of players. We assume that social memory is finite—every player, except possibly one, is finitely lived and cannot observe events that are sufficiently far back in the past. This class of games includes games between a long-run player and a sequence of short-run players and games with overlapping generations of players. Indeed, *any* stochastic game with infinitely lived players can be reinterpreted as one with finitely lived players: Each finitely-lived player is replaced by a successor, and receives the value of the successor's payoff. This value may arise from altruism, but the player also receives such a value if he can "sell" his position in a competitive market. In both cases, his objective will be to maximize infinite horizon payoffs, though his information on past events will be limited.

An equilibrium is *purifiable* if close-by behavior is consistent with equilibrium when agents' payoffs in each period are perturbed additively and independently. We show that only Markov equilibria are purifiable when social memory is finite. Thus if a game has at most one long-run player, all purifiable equilibria are Markov.

1 Introduction

Repeated game theory has shown that *punishment strategies*, strategies contingent on payoff irrelevant histories, greatly expand the set of equilibrium outcomes. Yet in much applied analysis of dynamic games, researchers restrict attention to Markov equilibria, equilibria in which behavior does not depend on payoff irrelevant histories. Arguments for focussing on Markov equilibria include (i) their simplicity; (ii) their sharp predictions; (iii) their role in highlighting the key payoff relevant dynamic incentives; and (iv) their descriptive accuracy in settings where the coordination implicit in payoff irrelevant history dependence does not seem to occur. However, principled reasons for restricting attention to Markov equilibria, based on strategic considerations, are limited.[1]

This paper provides a foundation for Markov strategies for dynamic games that rests on three assumptions. First, we assume that social memory is bounded – every player, except possibly one, cannot observe events that are sufficiently far back in the past. Second, we assume that moves are sequential – the game is such that only one player moves at any point of time. Finally, we require equilibrium strategies to be "purifiable," i.e., nearby strategies must constitute an equilibrium of a perturbed game with independent private payoff shocks in the sense of Harsanyi (1973). Our main result is that Markov equilibria are the only purifiable equilibria in games with sequential moves when social memory is bounded.

The purifiability requirement reflects the view that models are only an approximation of reality, and so there is always some private payoff information. We make the modest requirement that there must be some continuous shock under which the equilibrium survives.

The boundedness of social memory is natural in many contexts, such as games between a long-run player and a sequence of short-run players or in games with overlapping generations of players. Indeed, *any* stochastic game with infinitely lived players can be reinterpreted as one with finitely lived players: Each finitely-lived player is replaced by a successor, and receives the value of the successor's payoff. This value may arise from altruism, but the player also receives such a value if he can "sell" his position in a competitive

---

[1]For asynchronous choice games, Jehiel (1995) and Bhaskar and Vega-Redondo (2002) provide a rationale for Markov equilibria based on complexity costs. Maskin and Tirole (2001) discuss the notion of payoff relevance and the continuity properties of Markov equilibria; we discuss Maskin and Tirole (2001) in Section 5.5. Harsanyi and Selten (1988) provide a justification for Markov equilibrium that has a more axiomatic flavor, based on the notion of subgame consistency.

1

market. In both cases, his objective will be to maximize infinite horizon payoffs, though his information on past events will be limited. Our results also apply if we have long run players who are perfectly informed on past events but are restricted to using bounded memory strategies.[2]

Our argument exploits the special feature of the games we study, whereby only one player moves at a time. Purifying payoff shocks imply that if a player conditions upon a past (payoff irrelevant) event at date t, then some future player must also condition upon this event. Such conditioning is possible in equilibrium only if the strategy profile exhibits infinite history dependence. We thus give the most general version of an argument first laid out by Bhaskar (1998) in the context of a particular (social security) overlapping generations game. This argument does not apply with simultaneous moves since two players may mutually reinforce such conditioning at the same instant, as we discuss in Section 5.3.

2 A Long-Run Player/Short-Run Player Example

Consider the chain store game, played between a long-run player and an infinite sequence of short-run players. In each period, an entrant (the short-run player) decides whether to enter or stay out. If the entrant stays out, the stage game ends; if he enters, then the incumbent (the long-run player) decides whether to accommodate (A) or fight (F). The stage game is depicted in Figure 1.

Each entrant maximizes his stage game payoff, only observing and thus only conditioning on what happened in the previous period. The incumbent maximizes the discounted sum of payoffs, observing the entire history. The incumbent's discount factor δ is between $c/(1 + c)$ and 1. We require equilibria to satisfy sequential rationality—each player is choosing optimally at every possible history.

Ahn (1997, Chapter 3) shows that there is no pure strategy equilibrium where entry is deterred (for generic values of the discount factor). To provide some intuition, restrict attention to stationary strategies. Since the entrant only observes the outcome of the previous period, the entrant's history is an

---

[2]The previous version of this paper (Bhaskar, Mailath, and Morris, 2009) assumed perfectly informed players who were constrained to use finite recall strategies. Strategies that depend on what happens in the arbitrarily distant past do not seem robust to noisy information. In a different context (repeated games with imperfect public monitoring), Mailath and Morris (2002, 2006) show that strategies based on infinite recall are not "robust to private monitoring," i.e, they cease to constitute equilibrium with even an arbitrarily small amount of private noise added to public signals.
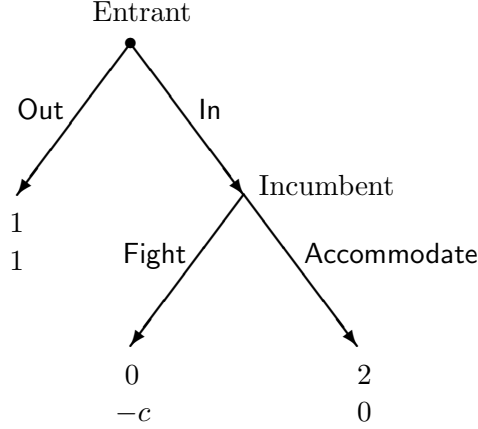
Entrant

Out / In

1
1

Incumbent

Fight / Accommodate

0
$-c$

2
0

Figure 1: The stage game for the chain store. The top payoff is the payoff to the Entrant.

element of $\mathcal{O} = \{\mathsf{Out}, \mathsf{A}, \mathsf{F}\}$. Consider a trigger strategy equilibrium where the entrant enters after accommodation in the previous period, and stays out otherwise. For this to be optimal, the incumbent must play a strategy of the form: F as long as he has not played A in the previous period; A otherwise. Such a strategy is not sequentially rational, because it is not optimal to play A when A had been played in the previous period. In this case, playing A secures a payoff of zero, while a one step deviation to F earns $-(1-\delta)c + \delta$, which is strictly positive for $\delta > c/(1+c)$.

There is however a class of mixed strategy equilibria in which entry is deterred with positive probability in each period. In any equilibrium in this class, the incumbent plays F with probability $\frac{1}{2}$, independent of history. The entrant is indifferent between In and Out at any information set, given the incumbent's strategy. He plays In with probability p at $t = 1$. At $t > 1$ he plays In with probability p after $o_{t-1} \in \{\mathsf{Out}, \mathsf{F}\}$; after $o_{t-1} = \mathsf{A}$, he plays In with probability q, where $q = p + c/[\delta(1 + c)]$. That is, the difference in entry probabilities across histories, $q - p$, is chosen to make the incumbent indifferent between accommodating and fighting. If we choose $p = 0$, then no entry takes place on the equilibrium path. Note that we have a one-dimensional manifold of equilibria in this class. In any such equilibrium, the entrant's beliefs about the incumbent's response is *identical* after the two one-period histories $o_{t-1} = \mathsf{A}$ and $o_{t-1} \in \{\mathsf{Out}, \mathsf{F}\}$. Nevertheless, the entrant plays differently.

3

We now establish that none of these mixed strategy equilibria can be purified if we add small shocks to the game's payoffs. Suppose that the entrant gets a payoff shock $\varepsilon \tilde{z}_1^t$ from choosing Out while the incumbent gets a payoff shock $\varepsilon \tilde{z}_2^t$ from choosing F. We suppose each $\tilde{z}_i^t$ is drawn independently across players and across time according to some known density with support $[0,1]$. The shocks are observed only by the player making the choice at the time he is about to make it. A strategy for the period t entrant is

$$\rho_t : \mathcal{O} \times [0,1] \to \Delta(A_1),$$

while a strategy for the incumbent is

$$\sigma_t : \mathcal{O}^t \times \{\mathsf{In}\} \times [0,1] \to \Delta(A_2)$$

(in principle, it could condition on the history of past payoff shocks, but this turns out not to matter). Note that ρ_{t+1} does not condition on what happened at $t-1$. Fix a history $h^t = (o_1, o_2, \ldots, o_t) \in \mathcal{O}^t$ with $o_t = \mathsf{In}$ (entry at date t) and z_2^t (payoff realization for incumbent). For almost all z_2^t, the incumbent has a unique pure best response. Since ρ_{t+1} does not condition on h^{t-1},

$$\sigma_t((h^{t-1}, \mathsf{In}), z_2^t) = \sigma_t((\hat{h}^{t-1}, \mathsf{In}), z_2^t)$$

for almost all z_2^t and any $\hat{h}^{t-1} \in A^{t-1}$. So the incumbent does not condition on h^{t-1}. Since the entrant at t also has a payoff shock, it has a unique pure best response for almost all payoff shock realizations, and so

$$\rho_t(h^{t-1}, z_1^t) = \rho_t(\tilde{h}^{t-1}, z_1^t)$$

for almost all z_1^t. In particular, the entrant's behavior after F or A in the previous period must be the same.

We conclude that for any $\varepsilon > 0$, only equilibria in Markov strategies exist in the perturbed game. If ε is sufficiently small, the incumbent accommodates for all realizations of his payoff shock, and therefore, with probability one. So the entrant enters with probability one. Thus, in any purifiable equilibrium of the unperturbed game, the backwards induction outcome of the stage game must be played in every period, and an equilibrium is purifiable if and only if it is Markov.

3   The Benchmark Game

We consider a potentially infinite dynamic game, Γ. The game has a recursive structure and may also have moves by nature. The set of players is denoted

4

by $\mathcal{N}$ and the set of states by S, both of which are countable. Only one player can move at any state, and we denote the assignment of players to states by $\iota : S \to \mathcal{N}$. This assignment induces a partition $\{S(i) \mid i \in \mathcal{N}\}$ of S, where $S(i) = \{s \in S \mid \iota(s) = i\}$ is the set of states at which i moves. Let A denote the countable set of actions available at any state; since payoffs are state dependent, it is without loss of generality to assume that the set of actions is state independent.

While states are observed by players, actions need not be. There is monitoring of the actions via signals y drawn from a countable set, Y. The transition function $q : S \times A \to \Delta(Y \times S)$ specifies the probability $q(y, s' \mid s, a)$ of the signal y and next period's state s' as a function of this period's state s and action a.[3] The initial distribution over states is given by $q_0 \in \Delta(S)$. Our notational convention is that period t begins in state s_t, an action a_t is taken (by player $\iota(s_t)$), resulting in the realized signal-state pair (y_t, s_{t+1}).

The assumption that the sets of actions, states, and signals are countable is made for expositional convenience. The model and results naturally extend to continuum spaces; see Section 5.1.

All but perhaps one of the players are finitely-lived. Denote by $\widetilde{T}(s)$ the set of dates at which state s arises with positive probability under q and some specification of actions, $\widetilde{T}(s) := \{t : \Pr\{s_t = s \mid a_0, a_1, \ldots, a_{t-1}\} > 0$ for some $a_0, a_1, \ldots, a_{t-1}\}$. The long-lived player, if present, is player i^*. Player $i \neq i^*$ has a *finite life* described by a first and last date, $0 \leq t_i \leq T_i < \infty$. Consistency with the process determining state transitions requires that

$$\iota(s) = i \Rightarrow \widetilde{T}(s) \subseteq \{t_i, ..., T_i\} .$$

(We take $t_{i^*} = 0$ and $T_{i^*} = \infty$, so that the long-lived player is also covered.)

For $i \neq i^*$, let $\tilde{t}_i = \min_{s \in S(i)} \widetilde{T}(s)$ denote the earliest possible date that player i moves and let $\tilde{T}_i = \max_{s \in S(i)} \widetilde{T}(s)$ denote the latest possible date. Since we would like to allow for the possibility that a player has information regarding events that take place before he moves, $\tilde{t}_i$ can be greater than t_i, the player's birth date.

We assume players only know the period t states and signals while they are alive. Since younger players know only a strict subset of current older players' histories, we say the histories of states and signals are *semi-public*. Denote by $h_{i\tau}$ a player i semi-public history at date τ, $t_i \leq \tau \leq T_i$, i.e.,

$$h_{i\tau} := (s_{t_i}, y_{t_i}, s_{t_i+1}, y_{t_i+1}, \ldots, s_\tau) \in (S \times Y)^{\tau - t_i} \times S.$$

---

[3]If the game has perfect monitoring, then we take $Y = A$, with $y = a$ receiving probability 1 when a is taken.

Note that the period-τ signal is not in the period-τ history; in particular, $h_{it_i} = (s_{t_i})$ so that if player i chooses an action in his or her first period, the only information that player has is the current state. In addition, each player knows his or her own past actions; these histories constitute player i's *private histories.*

If player i moves at date τ, $t_i \leq \tau \leq T_i$, after a semi-public history $h_{i\tau}$, then $s_\tau \in S(i)$. The set of all feasible player i period-τ semi-public histories at which i moves is denoted $H_{i\tau}$. Thus, player i moves in period τ if, and only if, $H_{i\tau} \neq \varnothing$. This set is a strict subset of $(S \times Y)^{\tau - t_i} \times S$. In addition to the requirement $s_\tau \in S(i)$, some signals may have zero probability at some states, and some state transitions have zero probability (such as to a state of a player who is not alive in that period). Given $h_{i\tau}$, $P_{i\tau}(h_{i\tau})$ is the set of player i *private histories* at date τ, i.e., $P_{i\tau}(h_{i\tau}) := \{(a_t)_{\{t:i=\iota(s_t),t<\tau\}} : a_t \in A\}$, with typical element $p_{i\tau}$. At player i's initial move, $P_{i\tau}$ is (as usual) the singleton set consisting of the null history. A period-τ history $h_{i\tau}$ or $(h_{i\tau}, p_{i\tau})$ is *relevant* if player i moves after the history.

We require that there is a uniform bound on the life span of the finitely-lived players:

Assumption 1 *There exists K such that $T_i - t_i \leq K$ for all $i \neq i^*$.*

Player i's flow payoffs are described by a bounded function

$$u_i : S \times Y \times A \to \mathbb{R}.$$

There is only an apparent tension between our definition of payoffs and our assumption that the signal y_τ and the transition from s_τ to $s_{\tau+1}$ are the only information that players $i \neq \iota(s_\tau)$ have about the action taken by player $\iota(s_\tau)$ in period τ. If the game has imperfect monitoring and the ex post flow payoff of a player $i \neq \iota(s_\tau)$ depends nontrivially on the action chosen, then our modeling is consistent with two interpretations: (1) the signal y conveys the same information (for example, part of the signal may be player i's payoff at s_τ); and (2) player i only observes the payoff at T_i. (For the long-lived player, the second interpretation is less natural than the first.)

This formulation allows for both deterministic and stochastic finite horizons: one (or more) of the states may be absorbing, and gives all players a zero payoff. Player i's discount factor is given by δ_i.

One important special case is where a short run player, $i \neq i^*$, receives payoffs only between dates t_i and T_i, so that u_i is identically zero in periods before t_i or after T_i. A second important special case is where players are

6

short-lived but maximize infinite horizon payoffs. This occurs in a dynastic model, where each short-run player is replaced at the end of her life by her descendent, towards whom she has altruistic preferences. If altruism is perfect, then this corresponds to a model with constant discounting. This also arises if the player is the owner of a firm, who is able to sell it on in a competitive capital market, thus capitalizing the present value of his expected profits. In all these cases, a short run player maximizes infinite horizon payoffs, but his information is limited, since he only observes public signals that are realized during his lifetime. Thus any infinite horizon stochastic game with finitely many long lived players can be re-interpreted as a model of short-lived players, and our results will also apply to these.

The game starts in a state s_0 at period 0 determined by q_0. Denote by $\mathcal{H}^\infty \subset (S \times Y \times A)^\infty$ the set of feasible outcomes with typical element $\mathfrak{h}^\infty$; an initial t-period history is denoted $\mathfrak{h}^t$. Player i's payoff as a function of outcome, $U_i : \mathcal{H}^\infty \to \mathbb{R}$, is

$$U_i\left(\mathfrak{h}^\infty\right) = U_i\left((s_t, y_t, a_t)_{t=0}^\infty\right) = \sum_{t=0}^\infty \delta_i^t u_i(s_t, y_t, a_t).$$

A period-τ *behavior strategy* for player i is a mapping

$$b_{i\tau} : \bigcup\nolimits_{h_{i\tau} \in H_{i\tau}} \{\{h_{i\tau}\} \times P_{i\tau}(h_{i\tau})\} \to \Delta(A),$$

and we write $b_i = (b_{i\tau})_{\tau=t_i, H_{i\tau} \neq \varnothing}^{T_i}$ and B_i for the set of strategies of player i.

Many games fit into our general setting:

1. Perfect information games played between overlapping generations of short-lived players. These include the classical consumption-loan model of Samuelson (1958), models of organizations with finitely lived managers (Cremer, 1986) or of legislatures with overlapping terms (Muthoo and Shepsle, 2006). Kandori (1992) and Smith (1992) prove folk theorems for these games, under the assumption that the short-lived players are fully informed about all past events. Bhaskar (1998) and Muthoo and Shepsle (2006) consider the implications of informational restrictions.

2. Extensive form games between a long-lived and a sequence of short-lived players. Such games arise naturally in the reputation literature (e.g., Fudenberg and Levine (1989)). Ahn (1997, Chapter 3) examines the implications of the short lived players having a bounded observation of past histories.

3. Any stochastic game or repeated game with long lived players can be interpreted as one with short-lived players who either have altruistic preferences towards their descendants or who sell their "position" in a competitive market. Our analysis applies to any stochastic game where only one player moves at a time—the asynchronous choice models of oligopoly due to Maskin and Tirole (1987, 1988a,b) are a leading example.

Our next step is to define equilibrium. Player i's expected continuation value from strategy profile b at $(h_{i\tau}, p_{i\tau})$ is defined recursively as follows. If $\iota(s_\tau) = i$, then player i's period-τ value function satisfies

$$
\begin{aligned}
V_i(b \mid h_{i\tau}, p_{i\tau}) &= \sum_{a \in A} b_i(a \mid h_{i\tau}, p_{i\tau}) \sum_{y \in Y, s' \in S} \Big\{ u_i(s_\tau, y, a) \\
&\quad + \delta_i V_i(b \mid (h_{i\tau}, y, s'), (p_{i\tau}, a) \Big\} q(y, s' \mid s_\tau, a) \\
&= \sum_{a \in A} b_i(a \mid h_{i\tau}, p_{i\tau}) \Big\{ \sum_y u_i(s_\tau, y, a) q_Y(y \mid s_\tau, a) \\
&\quad + \delta_i \sum_{y, s'} V_i(b \mid (h_{i\tau}, y, s'), (p_{i\tau}, a)) q(y, s' \mid s_\tau, a) \Big\},
\end{aligned}
$$

where q_Y is the marginal distribution on Y.

An almost identical equation holds when $\iota(s_\tau) \neq i$, with two changes. First, in the specification of the period-$(\tau + 1)$ value function, the private history $p_{i\tau}$ is not augmented by the period-τ action of player $\iota(s_\tau) \neq i$. Second, the distribution over the period τ action, $b_i(a \mid h_{i\tau}, p_{i\tau})$ is replaced by player i's belief over the behavior of player $\iota(s_\tau)$:

$$
b_{\iota(s_\tau)}(a \mid h_{i\tau}, p_{i\tau}) = E[b_{\iota(s_\tau)}(a \mid h_{\iota(s_\tau)\tau}, p_{\iota(s_\tau)\tau}) \mid h_{i\tau}, p_{i,\tau}].
$$

This conditional expectation is well defined for histories $(h_{i\tau}, p_{i\tau})$ on the path of play.

For other histories, we assume the player has some beliefs over the histories observed by the other players. While it is natural to require player i's beliefs over player $\iota(s_\tau)$'s history to respect Bayes' rule when possible, we do not impose this requirement. Instead, we simply require that players have well-defined beliefs at every feasible history, and so the value function is well defined at all feasible histories.

Definition 1 *A strategy profile b is a* perfect Bayes equilibrium (PBE) *if, for all $i \in \mathcal{N}$, $h_{i\tau} \in H_{i\tau}$, $p_{i\tau} \in P_{i\tau}(h_{i\tau})$, and $b'_i \in B_i$,*

$$
V_i((b_i, b_{-i}) \mid h_{i\tau}, p_{i\tau}) \geq V_i((b'_i, b_{-i}) \mid h_{i\tau}, p_{i\tau}). \tag{1}
$$

A strategy profile is a sequentially strict PBE *if for all for all $i \in \mathcal{N}$, $h_{i\tau} \in H_{i\tau}$, $p_{i\tau} \in P_{i\tau}(h_{i\tau})$, $b_i(h_{i\tau}, p_{i\tau})$ is a strict best reply in period τ: that is, for all $b_i' \in B_i$ satisfying $b_i'(h_{i\tau}, p_{i\tau}) \neq b_i(h_{i\tau}, p_{i\tau})$,*

$$V_i((b_i, b_{-i}) \mid h_{i\tau}, p_{i\tau}) > V_i((b_i', b_{-i}) \mid h_{i\tau}, p_{i\tau}).$$

Definition 2 *A strategy b_i is* Markov *if for any two relevant histories $(h_{i\tau}, p_{i\tau})$ and $(h_{i\tau}', p_{i\tau}')$ ending in the same state (i.e., $s_\tau = s_\tau'$),*

$$b_i(h_{i\tau}, p_{i\tau}) = b_i(h_{i\tau}', p_{i\tau}').$$

If b is both Markov and a PBE, it is a Markov perfect equilibrium.

Note that $(h_{i\tau}, p_{i\tau})$ and $(h_{i\tau}', p_{i\tau}')$ are both of length $\tau - t_i$.

Lemma 1 *Every sequentially strict PBE is Markov perfect.*

Proof. Fix a t period history $\mathfrak{h}^t$. By Assumption 1, it is common knowledge that from period $t + K + 1$ onwards, the behavior of players $i \neq i^*$ does not depend on $\mathfrak{h}^t$.[4] This implies that long-lived player's value function from $t + K + 1$ onwards does not depend on $\mathfrak{h}^t$. Thus, if the long-lived player's strategy satisfies sequential strictness, it does not depend on $\mathfrak{h}^t$ after date $t + K + 1$.

We consider first the last player whose behavior could potentially depend on elements of $\mathfrak{h}^t$, namely the player choosing an action in period $t + K$.

For any K-period continuation of $\mathfrak{h}^t$, $\mathfrak{h}^{t+K}$, this player is $j = \iota(s_{t+K})$, with associated semi-public and private histories $(h_{j,t+K}, p_{j,t+K})$. We now argue that

$$b_{j,t+K}\left((h_{jt}, h_{j,(t+1,t+K)}), p_{j,t+K} \right) = b_{j,t+K}\left((\widetilde{h}_{jt}, h_{j,(t+1,t+K)}), \widetilde{p}_{j,t+K} \right) \quad (2)$$

for all $\widetilde{h}_{jt}$ and $\widetilde{p}_{j,t+K}$, where $h_{j,t+K} =: (h_{jt}, h_{j,(t+1,T+K)})$.

We have two cases:

1. Player j is born after period t (i.e., $t_j > t$). In this case, (2) for all $\widetilde{h}_{jt}$ is immediately implied by feasibility;

---

[4] As Example 1 below illustrates, our assumption that the signal y_t is a signal of a_t only, and not of earlier actions, is important in this deduction. The Markovian nature of the process determining states is also important.

2. Player j is born at or before period t (i.e., $t_j \leq t$). Nonetheless, the decision problem facing player j is independent of $\widetilde{h}_{jt}$. Moreover, the decision problem is also independent of player j's private history $p_{j,t+K}$, and so the set of maximizing actions is independent of $\widetilde{h}_{jt}$ and $p_{j,t+K}$. Finally, sequential strictness implies the set of maximizing choices is a singleton, implying (2).

This argument now applies to show that the choice of the player making a choice in period $t + K - 1$ is also independent of the semi-public as well as the complete private history. Proceeding recursively yields the result. ∎

We are not the first to observe this implication of sequential strictness. In the context of repeated games where players move asynchronously, Jehiel (1995) and Bhaskar and Vega-Redondo (2002) have used this logic to conclude that if players have a motivation to reduce the memory requirements of their strategies, even if lexicographically, then they must play Markov strategies. However, sequential strictness is a demanding requirement, and any equilibrium in mixed strategies fails it. For example in the Maskin and Tirole (1988b) model of dynamic price competition where firms move asynchronously, collusive pricing can be sustained via Markov strategies, but this requires randomization off the equilibrium path. Similarly, in the repeated prisoner's dilemma with asynchronous moves, there exists a cooperative Markov perfect equilibrium where any breakdown of cooperation requires a randomized reversion to cooperation. In the following section, we will argue that payoff perturbations and purification yield the appropriate notion of sequential strictness, by eliminating non-Markov equilibria, while retaining mixed Markov equilibria.

4   The Game with Payoff Shocks

We now allow for the payoffs in the underlying game to be perturbed, as in Harsanyi (1973). Unfortunately, the description of the perturbed game is notationally cumbersome. Moreover, the definition of sequential rationality in the perturbed game requires specifying beliefs over histories of both past private actions as well as past private payoff shocks. Fortunately, the structure of the model allows us to finesse many of the details. Section 4.1 gives the formalism of the perturbed model, including beliefs, and ends with Lemma 2, which shows that optimal behavior is independent of the private history of actions and payoff shocks. Section 4.2 shows that all perfect Bayesian equilibria of the perturbed game are Markov.

4.1 The General Structure of the Perturbed Game

We require that the payoff shocks respect the recursive payoff structure of the infinite horizon game, i.e., to not depend upon history except via the state: Let Z be a full dimensional compact subset of $\mathbb{R}^{|A|}$ and write $\Delta^*(Z)$ for the set of measures with support Z generated by strictly positive densities. At each history $\mathfrak{h}^t \in (S \times Y \times A)^t \times S$, a payoff shock $z_i \in Z$ is drawn according to $\mu_i^{s_t} \in \Delta^*(Z)$ for each i.[5] The payoff shocks are independently distributed across players and histories. We write $\mu^{s_t} := \times_i \mu_i^{s_t}$ for the product measure on $Z^{\mathcal{N}}$. The complete history, including payoff shocks is denoted $\widetilde{\mathfrak{h}}^t \in (S \times Y \times A \times Z^{\mathcal{N}})^t \times S \times Z^{\mathcal{N}}$. We emphasize that the period-t state and payoff shock profile are in $\widetilde{\mathfrak{h}}^t$. If player $\iota(s)$ chooses action a, i's payoff is augmented by εz_i^a, where $\varepsilon > 0$ is a positive constant and z_i^a is player i's private payoff shock under action a. Thus, players' stage payoffs in the perturbed game depend only on the current state, signal, action, and payoff shock (s, y, a, z), and are given by

$$\tilde{u}_i(s, y, a, z_i) = u_i(s, y, a) + \varepsilon z_i^a.$$

Each player i only observes his/her private payoff shock in the periods t when i is alive, i.e., $t_i \leq t \leq T_i$.[6] We denote the perturbed game by $\Gamma(\varepsilon, \mu)$.

To describe strategies, we first describe players' information more precisely. Write $\mathbf{z}_i(\widetilde{\mathfrak{h}}^t) =: \mathbf{z}_i^t$ for the sequence of payoff shocks realized for player i along $\widetilde{\mathfrak{h}}^t$, $z_{it}(\widetilde{\mathfrak{h}}^t)$ for player i's current shock (thus $z_{it}(\widetilde{\mathfrak{h}}^t)$ is the last element of the sequence $\mathbf{z}_i(\widetilde{\mathfrak{h}}^t)$), and $\mathbf{z}(\widetilde{\mathfrak{h}}^t)$ for the sequence of payoff shock profiles realized for all players in $\widetilde{\mathfrak{h}}^t$.

At the semi-public history h_{it}, a player i period-t private history is $\tilde{p}_{it} := (p_{it}, (z_{i\tau})_{\tau=t_i}^t) \in P_{it}(h_{it}) \times Z^{t-t_i+1}$. A *behavior strategy* for player i, $\tilde{b}_i$, in the perturbed game specifies player i's mixed action $\tilde{b}_i(h_{it}, \tilde{p}_{it})$, at every relevant history $(h_{it}, \tilde{p}_{it})$, i.e., with $s_t \in S(i)$ and for every specification of player i's actions and realization of i's payoff shocks. The set of all behavior strategies for player i is denoted $\widetilde{B}_i$.

Each player i will maximize expected payoffs given beliefs over the unknown aspects of history. A *belief assessment* for player i specifies, for every relevant history $(h_{it}, \tilde{p}_{it})$, a belief $\pi_i^{(h_{it}, \tilde{p}_{it})}$ over histories $\widetilde{\mathfrak{h}}^t$, that is,

$$\pi_i^{(h_{it}, \tilde{p}_{it})} \in \Delta\left((S \times Y \times A \times Z^{\mathcal{N}})^t \times S \times Z^{\mathcal{N}}\right).$$

---

[5] Our analysis only requires that the shock distributions have densities with full dimensional compact supports. The assumption of common support is made to simplify notation.

[6] Our assumption, made to simplify notation, that all players receive payoff shocks in all periods (and not just in the periods they are alive) is without loss of generality.

Since the distribution of a player's private payoff shock after the history $\mathfrak{h}^t$ is completely determined by the state s_t and players' private payoff shocks are independent, player i's beliefs over the unknown aspects of history are independent of the realization of these private payoff shocks. In addition, a player's past actions should not affect player i's own beliefs (Fudenberg and Tirole (1991) refer to this type of condition as "no-signaling-what-you-don't-know").

We are thus led to the following maintained assumption on belief assessments:

Assumption 2 *Every player i's belief assessment satisfies*

1. *the implied beliefs over other players' semi-public and private histories are independent of player i's private payoff shocks and past actions p_{it}; and*

2. *player i's beliefs assign probability zero to the event that the history $\tilde{\mathfrak{h}}^t$ is inconsistent with $(h_{it}, \tilde{p}_{it})$.*

Beyond Assumption 2, we impose no further restrictions (such as consistency with Bayes' rule on the equilibrium path and independence of payoff shocks across other players or periods); we return to this issue after we introduce the notion of a sequential best response.

Player i's ex post value is recursively given by, for a given strategy profile $\tilde{b}$,

$$\widetilde{V}_i(\tilde{b} \mid \tilde{\mathfrak{h}}^t) = \sum_{a \in A} \tilde{b}_{\iota(s_t)}(a \mid h_{\iota(s_t)t}, \tilde{p}_{\iota(s_t)t}) \sum_{y \in Y, s' \in S} \left\{ \tilde{u}_i(s_t, y, a, z_{it}) \right.$$
$$\left. + \delta_i \int \widetilde{V}_i(\tilde{b} \mid \tilde{\mathfrak{h}}^t, y, s', a, z') \, d\mu^{s'}(z') \right\} q(y, s' \mid s, a).$$

Player i's expected payoff from the profile $\tilde{b}$ is given by

$$\int \widetilde{V}_i(\tilde{b} \mid \tilde{\mathfrak{h}}^t) \, d\pi_i^{(h_{it}, \tilde{p}_{it})}(\tilde{\mathfrak{h}}^t). \tag{3}$$

Definition 3 *Strategy $\tilde{b}_i$ is a* sequential best response *to $(\tilde{b}_{-i}, \pi_i)$, if for each $h_{it} \in H_{it}$, $\tilde{p}_{it} = (p_{it}, (z_{i\tau})_{\tau=t_i}^t) \in P_{it}(h_{it}) \times Z^{t-t_i+1}$, and $\tilde{b}_i' \in \tilde{B}_i$,*

$$\int \widetilde{V}_i((\tilde{b}_i, \tilde{b}_{-i}) \mid \tilde{\mathfrak{h}}^t) \, d\pi_i^{(h_{it}, \tilde{p}_{it})}(\tilde{\mathfrak{h}}^t) \geq \int \widetilde{V}_i((\tilde{b}_i', \tilde{b}_{-i}) \mid \tilde{\mathfrak{h}}^t) \, d\pi_i^{(h_{it}, \tilde{p}_{it})}(\tilde{\mathfrak{h}}^t).$$

Strategy $\tilde{b}_i$ is a sequential best response *to $\tilde{b}_{-i}$ if strategy $\tilde{b}_i$ is a sequential best response to $(\tilde{b}_{-i}, \pi_i)$ for some π_i.*

Because the perturbed game has a continuum of possible payoff shocks in each period, and players may have sequences of unreached information sets, there is no standard solution concept to which we may appeal. Our notion of sequential best response is very weak (not even requiring that beliefs respect Bayes' rule *on the path of play*). Assumption 2 does require each player's beliefs over other players' payoff shocks be independent of his own shocks. For information sets on the path of play, this requirement is implied by Bayes' rule. Tremble-based refinements imply such a requirement at all information sets, though they may imply additional restrictions across information sets. Assumption 2 is not implied by the notion of "weak perfect Bayesian equilibrium" from Mas-Colell, Whinston, and Green (1995), where no restrictions are placed on beliefs off the equilibrium path (which would allow players to have different beliefs about past payoff shocks depending on their realized current payoff realization).

In principle, a strategy for a player i depends on the fine details of the private histories that the player observes, i.e. his past payoff shocks and his past actions. Lemma 2 shows that any sequential best response must ignore such fine details, although it may depend upon the player's current payoff shock.

Definition 4 *A strategy $\tilde{b}_i$ is a* current shock strategy *if for all $h_{it} \in H_{it}$, and private histories $(p_{it}, (z_{i\tau})_{\tau=t_i}^{t})$, $(p'_{it}, (z'_{i\tau})_{\tau=t_i}^{t}) \in P_{it}(h_{it}) \times Z^{t-t_i+1}$, if $z_{it} = z'_{it} = z$, then for almost all $z \in Z$,*

$$\tilde{b}_i(h_{it}, (p_{it}, (z_{i\tau})_{\tau=t_i}^{t})) = \tilde{b}_i(h_{it}, (p'_{it}, (z'_{i\tau})_{\tau=t_i}^{t})).$$

Lemma 2 *If $\tilde{b}_i$ is a sequential best response to $\tilde{b}_{-i}$, then $\tilde{b}_i$ is a current shock strategy.*

Proof. Fix a player i, $h_{it} \in H_{it}$, and private history , $\tilde{p}_{it} = (p_{it}, (z_{i\tau})_{\tau=t_i}^{t})$. Denote i's beliefs by π_i. Player i's *next period* expected continuation payoff under $\tilde{b}$ from choosing action a this period, $\mathbf{V}_i(a, \tilde{b}_{-i}, \pi_i \mid (h_{it}, \tilde{p}_{it}))$, is given by

$$\sum_{y,s'} q(y, s' \mid s, a) \iint \max_{\hat{b}_i} \widetilde{V}_i(\hat{b}_i, \tilde{b}_{-i} \mid \tilde{\mathfrak{h}}^t, y, s', a, z') \, d\mu^{s'}(z') \, d\pi_i^{(h_{it}, \tilde{p}_{it})}(\tilde{\mathfrak{h}}^t).$$

Since $\tilde{b}_{-i}$ and the beliefs implied by $\pi_i^{(h_{it}, \tilde{p}_{it})}$ over other players' semi-public and private histories are independent of $\tilde{p}_{it}$, the maximization implies that $\mathbf{V}_i(a, \tilde{b}_{-i}, \pi_i \mid (h_{it}, \tilde{p}_{it}))$ also does not depend on player i's private history.

Thus, player i's total utility from the action a,

$$u_i(s, y, a) + \varepsilon \tilde{z}_i^a + \delta_i \mathbf{V}_i(a, \tilde{b}_{-i}, \pi_i \mid (h_{it}, \tilde{p}_{it})),$$

is independent of player i's private history. Since μ^s is absolutely continuous, player i can only be indifferent between two actions a and a' on a zero measure set of $z \in Z$. For other z, there is a unique best response, and so it is independent of the private history before the current shock. ∎

4.2 Mutual Sequential Best Responses are Markov

A current shock strategy (ignoring realizations of z of measure 0) can be written as

$$\tilde{b}_i : \cup_{t_i \leq t \leq T_i} H_{it} \times Z \to \Delta(A).$$

If *all* players are following current shock strategies, we can recursively define value functions for a given strategy profile $\tilde{b}$ that do not depend on *any* payoff shock realizations:

$$V_i^*(\tilde{b}|h_{it}) = \int \sum_{a \in A} \tilde{b}_{\iota(s_t)}(a|h_{it}, z_{\iota(s_t)}) \sum_{y \in Y, s' \in S} \left[\tilde{u}_i(s_t, y, a, z_i) \right.$$
$$\left. + \delta_i V_i^*(\tilde{b}|h_{it}, y, s') \right] q(y, s'|s_t, a) \, d\mu^{s_t}(z), \quad (4)$$

where

$$\tilde{b}_{\iota(s_t)}(a|h_{it}, z_{\iota(s_t)}) = E[\tilde{b}_{\iota(s_t)}(a|h_{\iota(s_t)t}, z_{\iota(s_t)})|h_{it}, z] \quad (5)$$

is i's prediction of player $\iota(s_t)$'s behavior given i's semi-public history and the period t payoff shock. As suggested by (5), V_i^* does depend, in general, on non-payoff-shock aspects of i's belief assessments. As will be clear, this dependence does not arise in equilibrium, and so we economize on notation by suppressing the potential dependence of V_i^* on beliefs. For future reference, player i's ex post value from the action a given the payoff shock z_i (the value of the second summation) is denoted by

$$\widetilde{V}_i^*(a, z_i; \tilde{b}|h_{it}) := \sum_{y \in Y, s' \in S} \left[\tilde{u}_i(s_t, y, a, z_i) + \delta_i V_i^*(\tilde{b}|h_{it}, y, s') \right] q(y, s'|s_t, a).$$

It is straightforward to verify that $\widetilde{V}_i^*$ is the expectation of $\widetilde{V}_i$, conditional on h_{it}, when all players are following current shock strategies.

Lemma 2 implies that beliefs over private histories are essentially irrelevant in the notion of sequential best responses, because, while behavior can in principle depend upon private histories, optimal behavior does not. We restate Lemma 2 in a more convenient form using the V_i^* notation:

14

Lemma 3 *A profile $\tilde{b}$ is a profile of mutual sequential best responses if, and only if, for all i, $\tilde{b}_i$ is a current shock strategy, and for each $h_{it} \in H_{it}$, and $\tilde{b}'_i \in \tilde{B}_i$,*

$$V_i^*((\tilde{b}_i, \tilde{b}_{-i}) \mid h_{it}) \geq V_i^*((\tilde{b}'_i, \tilde{b}_{-i}) \mid h_{it}). \tag{6}$$

Given Lemma 3 and the discussion in Section 4.1, the following definitions are natural:

Definition 5 *A* perfect Bayesian equilibrium *(PBE) is a profile of mutual sequential best responses (which is necessarily a profile of current shock strategies).*

A profile $\tilde{b}$ of current shock strategies is an essentially sequentially strict *equilibrium if, for all $i \in \mathcal{N}$, $h_{it} \in H_{it}$, for almost all payoff shocks $z_i \in Z$, the action $\tilde{b}_i(h_{it}, z_i)$ is pure and is the unique maximizer of player i's ex post value from the action a given the payoff shock z_i, $\widetilde{V}_i^*(a, z_i; \tilde{b}|h_{it})$.*

A current shock strategy $\tilde{b}_i$ is Markov *if for for almost all $z_i \in Z$, feasible histories $h_{it}, h'_{it} \in H_{it}$ satisfying $s_t = s'_t$,*

$$\tilde{b}_i(h_{it}, z_i) = \tilde{b}_i(h'_{it}, z_i).$$

If $\tilde{b}$ is both Markov and a PBE, it is a Markov perfect equilibrium.

After this considerable notational journey, we are led to a key result of the paper (with a gratifyingly short proof).

Proposition 1 *Every PBE of the perturbed game is essentially sequentially strict, and so is Markov perfect.*

Proof. Since flow payoffs are given by

$$\tilde{u}_i(s, y, a, z_i) = u_i(s, y, a) + \varepsilon z_i^a,$$

the equality $\widetilde{V}_i^*(\hat{a}, z_i; \tilde{b}|h_{it}) = \widetilde{V}_i^*(\tilde{a}, z_i; \tilde{b}|h_{it})$ implies

$$\varepsilon(z_i^{\hat{a}} - z_i^{\tilde{a}}) = \sum_{y \in Y, s' \in S} [u_i(s_t, y, \tilde{a}) - u_i(s_t, y, \hat{a})]$$

$$+ \delta_i V_i^*(\tilde{b}|h_{it}, y, s') [q(y, s'|s_t, \tilde{a}) - q(y, s'|s_t, \hat{a})].$$

Since the set of actions is countable, the set of values of z_i for which player i can be indifferent between any two actions is of Lebesgue measure zero. Thus, for almost all z_i, the set of maximizers must be a singleton, and the profile is essentially sequentially strict.

The proof that every essentially strict equilibrium is Markov perfect is almost identical to that of Lemma 1, and so is omitted. ∎

4.3 Purification

We now consider the purifiability of equilibria in the unperturbed game. *Purification* has several meanings in the literature (see Morris (2008)). One question asked in the literature is when can we guarantee that every equilibrium is essentially pure by adding noise to payoffs (e.g., Radner and Rosenthal (1982))? As we have seen in Proposition 1, our shocks ensure that in the perturbed game, any equilibrium must be essentially pure.

We follow Harsanyi (1973) in being interested in the relation between equilibria of the unperturbed game and equilibria of the perturbed game. But the definition of *purifiability* that we require for our main result is very weak: we require only that there exists a sequence of equilibria of a sequence of perturbed games that converge to the desired behavior.

Fix a strategy profile b of the unperturbed game. We say that a sequence of current shock strategies $\tilde{b}_i^k$ in the perturbed game converges to a strategy b_i in the unperturbed game if expected behavior (taking expectations over shocks) converges, i.e., for each $h_{it} \in H_{it}$ and $a \in A$,

$$\int \tilde{b}_i^k(a \mid h_{it}, z_i) \, d\mu^s(z) \to b_i(a \mid h_{it}). \tag{7}$$

Definition 6 *The strategy profile b is* weakly purifiable *if there exists a sequence $\{(\mu^k, \varepsilon^k)\}_{k=1}^{\infty}$, with $\mu^k : S \to \Delta^*(Z)$ and $\varepsilon^k \to 0$, such that there is a sequence of profiles $\{\tilde{b}^k\}_{k=1}^{\infty}$ converging to b, with $\tilde{b}^k$ a perfect Bayesian equilibrium of the perturbed game $\Gamma(\mu^k, \varepsilon^k)$ for each k.*

Since the supporting sequence of private payoff shocks is allowed to depend on the strategy profile b, and the distribution μ^k is itself indexed by k, this notion of purifiability is almost the weakest possible.[7] Our notion crucially maintains the recursive payoff structure of the infinite horizon game (in particular, we require that the payoff shocks are intertemporally independent). Allowing for intertemporally dependent payoff shocks violates the spirit of our analysis.

A stronger notion of purification, closer to the spirit of Harsanyi (1973), is the following:

Definition 7 *The strategy profile b is* Harsanyi purifiable *if for every sequence $\{(\mu^k, \varepsilon^k)\}_{k=1}^{\infty}$, where $\mu^k : S \to \Delta^*(Z)$ and $\varepsilon^k \to 0$, there is a sequence*

---

[7]It is also worth noting that we only require pointwise convergence in (7). For infinite horizon games, we may ask for uniform (in h_{it}) convergence, as is done in the positive result (Theorem 3) in Bhaskar, Mailath, and Morris (2008). Negative results are of course stronger with pointwise convergence.

of profiles $\{\tilde{b}^k\}_{k=1}^{\infty}$ converging to b, with $\tilde{b}^k$ a perfect Bayesian equilibrium of the perturbed game $\Gamma(\mu^k, \varepsilon^k)$ for each k.

Clearly, if a profile is Harsanyi purifiable, then it is weakly purifiable. The following is immediate from Section 4.2:

Proposition 2 *Every weakly purifiable PBE is Markov.*

The logic behind this proposition is straight-forward. Proposition 1 implies that in any perturbed game, any PBE is Markov. Thus, if b is not Markov, given an arbitrary sequence $< \mu^k, \varepsilon^k >_{k=1}^{\infty}$ with $\varepsilon^k \to 0$, we cannot find a sequence of PBE of the perturbed games converging to b.

5 Extensions and Discussion

5.1 Continuum Action and State Spaces

The assumption that the sets of actions, states, and signals are countable was made for expositional convenience. The model and results naturally extend to continuum spaces, as we now describe.

Suppose the set A is a compact subset and Y and S are measurable subsets of finite dimensional Euclidean spaces. Transitions and monitoring are described by a mapping (probability kernel) $q : S \times A \times \mathcal{F} \to [0,1]$, where $\mathcal{F}$ is the collection of events (measurable subsets) of $Y \times S$, $q(s,a,F)$ is a measurable function of (s,a) for each $F \in \mathcal{F}$, as well as a probability measure over $\mathcal{F}$ for each (s,a); we write $q^{s,a}$ for this measure. The flow payoffs are bounded and continuous functions of (s,y,a). The value functions are as before (with integrals replacing summations in the calculation of expectations), as is the definition of PBE. A profile b in the unperturbed game is an *essentially sequentially strict PBE* if (as in Definition 5) $b_i(h_{i\tau}, p_{i\tau})$ is a strict best reply for almost all signals. The proof that every essentially sequentially strict equilibrium is Markov perfect is along the same lines as the proof of Lemma 1.

To keep things simple in our discussion of the perturbed game, we now suppose actions, states and signals are all one dimensional (though the following extends in an obvious manner to more dimensions). Each player i's payoff perturbations is indexed by the one dimensional variable z_i. Flow payoffs are given by

$$u_i(s,y,a) + \varepsilon v_i(s,y,a,z_i),$$

where $\varepsilon > 0$ and $v_i(s, ., ., .)$ is a parameterization of the payoff perturbation. In particular, we assume

$$\bar{v}_i(s, a, z_i) := \int v_i(s, y, a, z_i) \, dq^{s,a}(y, s') \tag{8}$$

is either strictly supermodular or strictly submodular in (a, z_i). Note that we do *not* make any similar assumption on the payoff function u_i.

Each player's payoff shock z_i is, for each history $\mathfrak{h}^t$, distributed according to $\mu_i^{s_t}$ with common interval support $Z \subset \mathbb{R}$. As for the countable case, we assume the payoff shocks are continuously and independently distributed across players and histories, and every player's belief assessment is assumed to satisfy Assumption 2. The notions of sequential best response (Definition 3) and current shock strategy (Definition 4) are unchanged, and Lemma 2 continues to hold with essentially the same proof (again, after summations are replaced by the appropriate integrals). Similarly, Lemma 3 still holds, and the notions of PBE and Markov perfect equilibrium continue to be given by Definition 5.

We now argue that Proposition 1 holds, that is, every PBE is essentially sequentially strict and so is Markov perfect. When players follow a profile $\tilde{b}$, player i's ex post payoff from choosing action $\hat{a}$ at (h_{it}, z_i) is given by

$$\widetilde{V}_i^*(a, z_i; \tilde{b}|h_{it}) = \int\limits_{(y,s') \in Y \times S} u_i(s_t, y, a) + \varepsilon v_i(s_t, y, a, z_i)$$

$$+ \delta_i V_i^*(\tilde{b}|h_{it}, y, s') \, dq^{s_t, a}(y, s')$$

$$= W_i(a; \tilde{b}|h_{it}) + \varepsilon \bar{v}_i(s_t, a, z_i),$$

where

$$W_i(a; \tilde{b}|h_{it}) := \int\limits_{(y,s') \in Y \times S} u_i(s_t, y, a) + \delta_i V_i^*(\tilde{b}|h_{it}, y, s') \, dq^{s_t, a}(y, s')$$

and $\bar{v}_i$ is defined in (8). Note that the ex post payoff function $\widetilde{V}_i^*$ inherits the strict super- or submodularity in (a, z_i) of $\bar{v}_i$.

Define $\phi_i(z_i|h_{it}, \tilde{b}) := \arg\max_{a' \in A} \widetilde{V}_i^*(a', z_i; \tilde{b}|h_{it})$, i.e., $\phi_i(z_i|h_{it}, \tilde{b})$ is the correspondence describing the maximizers of $\widetilde{V}_i^*$. A PBE $\tilde{b}$ will be essentially sequentially strict if, for all i and h_{it}, the correspondence $\phi_i(z_i|h_{it}, \tilde{b})$ is a singleton for almost all values of z_i.

Suppose $\widetilde{V}_i^*$ is strictly supermodular (a similar argument applies if it is strictly submodular). It is standard that the correspondence $\phi_i(\cdot|h_{it}, \tilde{b})$ is strictly increasing in the following sense: for all $z_i < z_i'$, and for all $a \in \phi_i(z_i|h_{it}, \tilde{b})$ and $a' \in \phi_i(z_i'|h_{it}, \tilde{b})$, we have $a < a'$. This implies that the convex hulls of $\phi_i(z_i|h_{it}, \tilde{b})$ and $\phi_i(z_i'|h_{it}, \tilde{b})$ are disjoint for all $z_i \neq z_i'$, and so $\phi_i(z_i|h_{it}, \tilde{b})$ must be singleton-valued for all but a countable number of z_i. Hence, every PBE is essentially sequentially strict.

Finally, Proposition 2 holds in the current setting (for any sensible notion of convergence in the definition of weak purification).

We now consider some applications of this result:

1. In the asynchronous choice oligopoly models of Maskin and Tirole (1988a, 1987), players choose quantities, so that action sets and states are subsets of the real line. Suppose now that the payoff perturbations, z_i, arise due to cost shocks, that are independently and identically distributed in every period. Thus $v_i(a, y, z_i, s) = -az_i$, so the submodularity condition is satisfied. Thus our argument justifies the restriction to Markov perfect equilibria in the context of this model.

2. In Maskin and Tirole (1988b), firms choose prices and the products are homogeneous. Payoff shocks in this setting do not generate the required variability in current period output, since when a firm prices above its competitor (as occurs in some parts of the Edgeworth cycle), in that period, the firm makes no sales. On the other hand, in the differentiated products version of Eaton and Engers (1990), a firm's quantities are decreasing in the firm's own price, and thus the payoff perturbation is strictly supermodular. Our argument therefore applies in this case.

3. In Samuelson's (1958) consumption loan model, agents live for two periods, and a young agent must choose a transfer to the old agent, a real number. With bounded social memory, there exists a pure strategy equilibrium that supports efficient transfers. If we perturb payoffs, so that there are shocks to the marginal rate of substitution of the young agent, between consumption when young and consumption when old, then the supermodularity condition is satisfied. Thus only the Markov perfect equilibrium survive these perturbations.[8]

---

[8]For reasons of space we omit the details of this argument. They are available from the authors on request.

More generally, in many economic applications, supermodularity or submodularity of the payoff function in the payoff shock parameter and the actions is a natural assumption. Thus the purification justification for Markov equilibrium can be readily applied, to many economic examples of dynamic games, across a variety of fields in economics.

5.2   Alternative Informational Assumptions on Signals

It is critical for our result that the signal in any period, y_t, depends on a_t only, and not on earlier actions.

Example 1 There are overlapping generations with all players being finitely-lived and living for two periods. Agents are indexed by $\tau \in \mathbb{N}$ with $t_\tau = \tau - 1$ and $T_\tau = \tau + 1$. Each player chooses $a \in \{0, 1\}$. The period-τ player receives a zero payoff in all periods except period $\tau + 1$, where the payoff $u(a_\tau, a_{\tau+1})$ satisfies

$$u(0, 1) > u(1, 1) > u(0, 0) > u(1, 0).$$

At the end of period $\tau - 1$, agent τ observes a signal $y_{\tau-1}$ given by

$$y_{\tau-1} = a_{\tau-2} + a_{\tau-1},$$

where we initialize the recursion by specifying $a_{-1} = 1$.

A pure strategy for players $\tau \geq 1$ is a function $\sigma_t : \{0, 1, 2\} \rightarrow \{0, 1\}$, while a strategy for player 0 is an action choice from $\{0, 1\}$ (since player 0 has a null history). A Markov strategy is an action choice (since the state space is a singleton). The game has a unique Markov perfect equilibrium, and in this equilibrium all players choose 0.

Consider the following (non-Markov) strategy profile: every player $\tau \geq 1$ plays the strategy σ given by $\sigma(0) = \sigma(1) = 0$ and $\sigma(2) = 1$, and player 0 chooses 1. We argue that this profile is a sequentially strict PBE, even though there is a uniform bound on players' life spans and the profile is not Markov.

The choice of 1 at $y_\tau = 2$ is optimal for all players, since $u(1, 1) > u(0, 0)$. Moreover, choosing 0 at $y_\tau = 0$ is optimal since it yields $u(0, 0)$ while choosing 1 yields $u(1, 0)$.

The critical value of y is $y = 1$. The value of $y_\tau = 1$ arises from two different private histories: $(a_{\tau-2}, a_{\tau-1}) = (0, 1)$ and $(a_{\tau-2}, a_{\tau-1}) = (1, 0)$. If it is the former, the action choice a_τ is pivotal in determining $a_{\tau+1}$ under σ, while under the latter history, it is not (since $a_{\tau+1} = 0$ independent of a_τ). If player τ assigns sufficient probability to the action history

20

	A	F
IN	$2, 0$	$0, -c$
OUT	$1, 1+z$	$1, 1$

Figure 2: A simultaneous move version of the Chain Store game.

$(a_{\tau-2}, a_{\tau-1}) = (1, 0)$, then $a_\tau = 0$ is strictly optimal. Finally, beliefs such that the probability that $(a_{\tau-2}, a_{\tau-1}) = (1, 0)$ equals one can be derived as the limit of a sequence of trembles requiring only a single agent to deviate from the equilibrium path of play. Since this equilibrium is strict, it can admits a Harsanyi purification – if ε is sufficiently small, in the perturbed game there exists an equilibrium where each player plays as in the above equilibrium of unperturbed game, for all realizations of his payoff shock.
★

This example illustrates that cooperation and non-Markovian behavior can be sustained as an equilibrium when there is less information on past events, as compared to the case when there is more information.

5.3 Simultaneous Move Games

Propositions 1 and 2 do not extend to games where more than one player moves at a time, e.g. repeated synchronous move games. It is easy to construct counterexamples, and indeed Mailath and Olszewski (2011) proves a folk theorem using strict, and hence purifiable, finite recall strategy profiles. For an illustrative example, consider the modification of the chain store game, where players move simultaneously, with payoffs given in Figure 2. We assume that $z > 0$, so that (OUT,F) is not a Nash equilibrium of the stage game.

The game is played between an incumbent (the long run player) who chooses from {A,F} and a sequence of entrants (short run players). The entrant born at date $t-1$ observes the action profile played at $t-1$, and plays the game at date t. The incumbent observes the entire history. Consider the strategy profile where actions at date t only depend upon the action profile played at $t-1$. The entrant plays OUT at $t = 1$, and at any date t if (OUT,F) is played at $t - 1$; otherwise, he plays IN. The incumbent plays F at $t = 1$, and and at any date t if (OUT,F) is played at $t - 1$; otherwise, he plays IN. If $(1 - \delta)z < \delta$, the incumbent has no incentive to deviate to A along the

equilibrium path, and this profile is a sequentially strict PBE; more precisely, for any beliefs over histories that each entrant may have, the strategy profile satisfies sequential rationality. Since all players have strict incentives at every information set, it is easy to show that this equilibrium admits a Harsanyi purification. However, the unique Markov Perfect equilibrium of this game has players playing the unique Nash equilibrium of the stage game, (IN,A), at every history.[9]

The strategy profile just constructed has the property that the actions at date t depend upon the action profile at $t-1$, i.e. upon the actions of *both* the entrant and the incumbent. This feature prevents the incumbent's unilateral manipulation of history that precludes non-Markovian equilibrium play in the sequential chain store of Section 2. In the simultaneous-move chain store, the incumbent cannot restore the (OUT,F) continuation path after a play of (IN,A), because the profile specifies continued play of (IN,A) after (IN,F). Note that under such a specification in the sequential chain store of Section 2, sequential rationality would force the incumbent to play A after the first play of IN, eliminating any incentive for the entrants to play OUT. In contrast, under simultaneous moves, on the equilibrium path, the current entrant believes the incumbent will choose F (which is sequentially rational for the incumbent, since simultaneous moves prevent the incumbent's current action from depending on the current action of the entrant).

One can, of course, increase the set of Markov equilibria in the chain store game by expanding the set of states, by allowing them to depend upon the previous period's outcome. However, with sequential moves, such a "spurious" Markov strategy profile will not be sequentially strict. As we show in Section 5.5, it may not admit a Harsanyi purification.

Although purifiability clearly has less bite in simultaneous move games, it is possible that it may allow us to restrict the set of equilibria. For example, one conjecture is that it might rule out the "belief-free" strategies recently introduced by Piccione (2002) and Ely and Välimäki (2002). Bhaskar, Mailath, and Morris (2008) show that the one period recall strategies of Ely and Välimäki (2002) are not purifiable via one period recall strategies in the perturbed game; however, they are purifiable via infinite recall strategies.[10]

---

[9]Liu (2011) and Liu and Skrzypacz (2011) consider simultaneous move games played between a long run player and short run players with finite social memory. They show that such games can give rise to interesting dynamics in the presence of a reputational type for the long run player.

[10]A similar argument shows that in the chain store example of Section 2, if entrants have unbounded memory, the one-period recall mixed strategy equilibrium is purifiable (Bhaskar, Mailath, and Morris, 2009, Example 3).

The purifiability of belief free strategies via finite recall strategies remains an open question.

Recent work of Peski (2009, 2012) gives conditions under which all repeated game equilibria are sequences of stage game equilibria. Peski assumes a continuum of signals that are sufficiently rich, even though the stage game has finitely many actions. The result uses a purification argument, and a requirement that strategies are measurable with respect to a finite partition of histories. Sequences of stage game equilibria are the only Markov equilibria in repeated games, and thus these papers give alternative conditions under which only Markovian equilibria survive.

5.4 Purification of Stationary Markov Equilibria

Our primary interest in this paper is to explore the extent to which purification justifies restricting attention to Markov equilibria. In this section, we provide a partial converse: for a class of models, every stationary Markov equilibrium can be purified. Our central Assumption 1 implies that the game necessarily has a countable infinity of players, and our notion of state accommodates this infinity. The unperturbed game is a dynamic stochastic game with a countable set of players, states, signals, and actions, while the perturbed game adds the further complication of a continuum of private payoff shocks. We are not aware of any standard theorems on existence and purification for our perturbed game setting.

Our strategy therefore is to show the existence and purifiability of Markov perfect equilibria in a class of stationary games, with a countable set of players and states. We do this by showing an equivalence between the equilibria of this underlying game, Γ, and of a related game $\widehat{\Gamma}$ that has finitely many states and players. Thus the existence and purification results for the latter can be extended to the underlying game, Γ.

As we have already noted, any finite stochastic game can be reinterpreted as a game with infinitely many short run players who have the same payoff function as the long run player in the original game. The payoff assumption will be valid either if the short run player can "sell" her position to a successor, or if the she is altruistic towards her successor. We formalize this as follows.

An infinite horizon finite (sequential) stochastic game is described by the collection $\widehat{\Gamma} := \{W, \widehat{N}, \hat{\imath}, A, \hat{q}, (\delta_i, u_i)_{i \in \widehat{N}}\}$, where W is the finite set of states, $\widehat{N}$ is the finite set of long-lived players, $\hat{\imath} : W \to \widehat{N}$ identifies the mover at each state, A is the finite set of actions, $\hat{q} : W \times A \to \Delta(W)$ is a stationary transition function, and finally, player $i \in \widehat{N}$ maximizes the

discounted (by δ_i) expected value of the infinite sequence of flow payoffs given by $u_i : W \times A \to \mathbb{R}$.

Note that, for simplicity, we assume the game has perfect monitoring, and so there are no public signals y.

To map the game $\widehat{\Gamma}$ into a game Γ covered by Section 3, we begin by setting $S := W \times \mathbb{N}_0$, where $\mathbb{N}_0 := \{0, 1, 2, \dots\}$ denotes the set of periods. Transitions are described by $q : S \times A \to \Delta(S)$, where $q(w, t, a) := \hat{q}(w, a)f_{t+1}$ and f_{t+1} is a degenerate distribution on $\mathbb{N}_0$ assigning probability one to $t + 1$.

Fix $K \geq 1$. For $i \in \widehat{N}$, associate a countably infinite number of short-lived players, with player $j \in \mathbb{N}_0$ having flow payoff function u_i, birth date $t_j = jK$, and death date $T_j = (j + 1)K - 1$.[11] Denote the jth short-lived version of player i by $j(i)$, and the set of such players by $\in \mathcal{N}(i)$. The countable collection of short-lived players is then given by $\mathcal{N} := \cup_{i \in \widehat{N}} \mathcal{N}(i)$. The assignment of players to states $s = (w, t)$ is $\iota(w, t) := j_t(\iota(w))$, where j_t satisfies $j_t K \leq t < (j_t + 1)K$. Finally, all short-lived players associated with player i in $\widehat{\Gamma}$ share player i's discount factor δ_i (in particular, player $j(i)$ is altruistic with respect to future generations $j'(i), j' > j$). We call the game Γ *an infinite player version of a finite stochastic game.*

In both Γ and $\widehat{\Gamma}$, a stationary Markov strategy profile is a mapping $\tilde{b} : W \to \Delta(A)$. It is immediate that the stationary Markov equilibria of Γ and $\widehat{\Gamma}$ are identical (the non-Markov equilibria may differ since the players in $\widehat{\Gamma}$ have additional information on payoff-irrelevant histories). The existence of stationary Markov perfect equilibria in the game $\widehat{\Gamma}$ is well established in the literature. Thus stationary Markov equilibria exist in the game Γ. As in Section 4.1, player i's payoff shocks are described by the continuous measure $\mu_i^w \in \Delta^*(Z)$, and we denote the perturbed games by $\widehat{\Gamma}(\varepsilon, \mu)$ and $\Gamma(\varepsilon, \mu)$. As for $\Gamma(\varepsilon, \mu)$, sequential best replies in $\widehat{\Gamma}(\varepsilon, \mu)$ are current shock strategies (finite memory plays no role in the proof of Lemma 2), and so the stationary Markov equilibria of $\widehat{\Gamma}(\varepsilon, \mu)$ and $\Gamma(\varepsilon, \mu)$ are identical.

The game $\widehat{\Gamma}$ is parameterized by the collection of utility functions, $u_i : W \times A \to \mathbb{R}$, $i \in \widehat{N}$, i.e., by a point in $\mathbb{R}^{|W \times A \times \widehat{N}|}$. By Doraszelski and Escobar (2010, Theorem 2), for almost all payoffs in $\mathbb{R}^{|W \times A \times \widehat{N}|}$, any stationary Markov equilibrium of $\tilde{\Gamma}$ is Harsanyi-purifiable. In view of the equivalence between the Markov equilibria of Γ and $\tilde{\Gamma}$, and between the Markov equilibria of the perturbed version of these two games, this implies that for almost all payoffs in $\mathbb{R}^{|W \times A \times \bar{\mathcal{N}}|}$, any stationary Markov equilibrium of Γ can be

---

[11]More general patterns of birth and death dates, consistent with Assumption 1, can easily be accommodated, at the cost of more complicated notation.

purified.

5.5   The Notion of Markov State

One criticism made of the notion of Markov equilibria is that it can be made arbitrarily permissive by expanding the set of states. In particular, consider a non-Markov strategy profile that plays differently across different histories or information sets. Such a profile can be made into a Markov profile by expanding the set of states, so that distinct information sets induce distinct states. However, since our Harsanyi-purification result holds only for games with generic payoffs, there is no guarantee that the resulting equilibria can be purified. Indeed, it seems likely that genericity will be violated since if the two "spuriously" distinct states are labelled s and s', $u_i(a,s) = u_i(a,s')$. The following two examples, that are based on the chain store game, are illustrative.

Assume, as in Section 2, that the short run player only observes the outcome in the previous period, which belongs to the set $\mathcal{O} = \{\mathsf{Out}, \mathsf{A}, \mathsf{F}\}$. This outcome was assumed to be not payoff relevant in the current period, while the set of states W has two elements: the initial node, w_0 where the entrant moves and w_1, where the incumbent moves. Now suppose that the state space is $W \times \mathcal{O}$, i.e. we augment the state space by also including the previous period's outcome. The mixed strategy equilibrium where entry is deterred is now a Markov equilibrium. However, it cannot be Harsanyi-purified: It suffices to consider any payoff perturbation μ that does not depend upon $\mathcal{O}$, i.e. where $\mu^{w_i o} = \mu^{w_i \tilde{o}}$ for any $o, \tilde{o} \in \mathcal{O}$. Since neither state transitions nor payoffs depend on the current element of $\mathcal{O}$, an equilibrium that conditions upon $\mathcal{O}$ cannot be purified under these perturbations. This also implies that the mixed strategy equilibrium cannot be weakly purified when we require the payoff perturbations to not depend on $\mathcal{O}$.

On the other hand, if the state allows a way of encoding the infinite history, then such Markov equilibria can be purified even if the states do not directly affect payoffs. To see this, suppose the state space is given by $\{w_0, w_1\} \times \{\mathsf{a}, \mathsf{f}\}$, where w_0 and w_1 are as described in the previous paragraph. The states a and f encode history, with the current encoding being f if any entry had been met with F, and a if a single entry had been followed by A. Consider the Markov strategy profile where the incumbent plays f at F and A at a, while the entrant enters at a and stays out at f. This is a Markov equilibrium where each player has strict incentives at each information set, and it is easy to show that it can be Harsanyi purified. Our justification for restricting attention to Markov equilibria rests on informational restrictions

– in particular, bounded memory. Thus if states allow players to encode infinite histories, then mutual conditioning upon these states can be sustained even if these states do not directly affect payoffs.

These two examples illustrate the difference between our approach, based on informational considerations and payoff perturbations and that of Maskin and Tirole (2001), that is based entirely on payoff considerations. Loosely, Maskin and Tirole (2001) consider stochastic games with perfect information on histories and use payoff equivalence to induce a partition over histories of the same length. The set of Markov states is defined to be the coarsest partition over histories such that for every profile measurable with respect to that partition, each player has a best response measurable with respect to that partition. Under the Maskin-Tirole definition, in both the above examples, in the unique Markov equilibrium the backwards induction outcome is played in each period. However, if the state allows the encoding of infinite histories, then a trigger strategy type equilibrium can be purified.

References

AHN, I. (1997): "Three Essays on Repeated Games Without Perfect Information," Ph.D. thesis, University of Pennsylvania.

BHASKAR, V. (1998): "Informational Constraints and the Overlapping Generations Model: Folk and Anti-Folk Theorems," *Review of Economic Studies*, 65(1), 135–149.

BHASKAR, V., G. J. MAILATH, AND S. MORRIS (2008): "Purification in the Infinitely-Repeated Prisoners' Dilemma," *Review of Economic Dynamics*, 11(3), 515–528.

——— (2009): "A Foundation for Markov Equilibria in Infinite Horizon Perfect Information Games," PIER Working Paper No. 09-029, University of Pennsylvania.

BHASKAR, V., AND F. VEGA-REDONDO (2002): "Asynchronous Choice and Markov Equilibria," *Journal of Economic Theory*, (2), 334–350.

CREMER, J. (1986): "Cooperation in Ongoing Organizations," *The Quarterly Journal of Economics*, 101(1), 33.

DORASZELSKI, U., AND J. ESCOBAR (2010): "A Theory of Regular Markov Perfect Equilibria in Dynamic Stochastic Games: Genericity, Stability, and Purification," *Theoretical Economics*, 5(3), 369–402.

EATON, J., AND M. ENGERS (1990): "Intertemporal Price Competition," *Econometrica*, 58(3), 637–659.

ELY, J. C., AND J. VÄLIMÄKI (2002): "A Robust Folk Theorem for the Prisoner's Dilemma," *Journal of Economic Theory*, 102(1), 84–105.

FUDENBERG, D., AND D. K. LEVINE (1989): "Reputation and Equilibrium Selection in Games with a Patient Player," *Econometrica*, 57(4), 759–778.

FUDENBERG, D., AND J. TIROLE (1991): "Perfect Bayesian Equilibrium and Sequential Equilibrium," *Journal of Economic Theory*, 53(2), 236–260.

HARSANYI, J. C. (1973): "Games with Randomly Disturbed Payoffs: A New Rationale for Mixed-Strategy Equilibrium Points," *International Journal of Game Theory*, 2(1), 1–23.

HARSANYI, J. C., AND R. SELTEN (1988): *A General Theory of Equilibrium in Games*. MIT Press, Cambridge.

JEHIEL, P. (1995): "Limited Horizon Forecast in Repeated Alternate Games," *Journal of Economic Theory*, 67, 497–519.

KANDORI, M. (1992): "Repeated Games Played by Overlapping Generations of Players," *The Review of Economic Studies*, 59(1), 81.

LIU, Q. (2011): "Information Acquisition and Reputation Dynamics," *Review of Economic Studies*, 78(4), 1400–1425.

LIU, Q., AND A. SKRZYPACZ (2011): "Limited Records and Reputation," Columbia University and Graduate School of Business, Stanford.

MAILATH, G. J., AND S. MORRIS (2002): "Repeated Games with Almost-Public Monitoring," *Journal of Economic Theory*, 102(1), 189–228.

——— (2006): "Coordination Failure in Repeated Games with Almost-Public Monitoring," *Theoretical Economics*, 1, 311–340.

MAILATH, G. J., AND W. OLSZEWSKI (2011): "Folk Theorems with Bounded Recall under (Almost) Perfect Monitoring," *Games and Economic Behavior*, 71(1), 174–192.

MAS-COLELL, A., M. D. WHINSTON, AND J. GREEN (1995): *Microeconomic Theory*. Oxford University Press, New York, NY.

27

MASKIN, E., AND J. TIROLE (1987): "A Theory of Dynamic Oligopoly, III: Cournot Competition," *European Economic Review*, 31(4), 947–968.

——— (1988a): "A Theory of Dynamic Oligopoly I: Overview with Quantity Competition and Large Fixed Costs," *Econometrica*, 56(3), 549–569.

——— (1988b): "A Theory of Dynamic Oligopoly, II: Price Competition, Kinked Demand Curves, and Edgeworth Cycles," *Econometrica*, 56(3), 571–599.

——— (2001): "Markov Perfect Equilibrium I. Observable Actions," *Journal of Economic Theory*, 100(2), 191–219.

MORRIS, S. (2008): "Purification," in *The New Palgrave Dictionary of Economics Second Edition*, ed. by S. Durlauf, and L. Blume, pp. 779–782. Macmillan Palgrave.

MUTHOO, A., AND K. SHEPSLE (2006): "Information, Institutions and Constitutional Arrangements," Discussion paper, Harvard University.

PESKI, M. (2009): "Asynchronous Repeated Games with Rich Private Monitoring and Finite Past," Department of Economics, University of Toronto.

——— (2012): "An Anti-Folk Theorem for Finite Past Equilibria in Repeated Games with Private Monitoring," *Theoretical Economics*, 7(1), 25–55.

PICCIONE, M. (2002): "The Repeated Prisoner's Dilemma with Imperfect Private Monitoring," *Journal of Economic Theory*, 102(1), 70–83.

RADNER, R., AND R. ROSENTHAL (1982): "Private Information and Pure-Strategy Equilibria," *Mathematics of Operations Research*, 7(3), 401–409.

SAMUELSON, P. A. (1958): "An Exact Consumption-Loan Model of Interest With or Without the Social Contrivance of Money," *Journal of Political Economy*, 66(6), 467–482.

SMITH, L. (1992): "Folk Theorems in Overlapping Generations Games," *Games and Economic Behavior*, 4(3), 426–449.