A Foundation for Markov Equilibria in Infinite Horizon Perfect Information Games

V. Bhaskar,* George J. Mailath† and Stephen Morris‡

March 5, 2009

Abstract

We study perfect information games with an infinite horizon played by an arbitrary number of players. This class of games includes infinitely repeated perfect information games, repeated games with asynchronous moves, games with long and short run players, games with overlapping generations of players, and canonical non-cooperative models of bargaining.

We consider two restrictions on equilibria. An equilibrium is *purifiable* if close by behavior is consistent with equilibrium when agents' payoffs at each node are perturbed additively and independently. An equilibrium has *bounded memory* if there exists K such that at most one player's strategy depends on what happened more than K periods earlier. We show that only Markovian equilibria have bounded memory and are purifiable. Thus if a game has at most one long run player, all purifiable equilibria are Markovian.

1 Introduction

Markov equilibria are widely used in the applied analysis of dynamic games, in fields ranging from industrial organization[1] to political economy.[2] Their appeal lies primarily in their simplicity and the sharp predictions obtained.[3]

*University College London

†University of Pennsylvania

‡Princeton University

[1]See Maskin and Tirole (1987, 1988a,b), and Ericson and Pakes (1995).

[2]See Acemoglu and Robinson (2001).

[3]See Duffie, Geanakoplos, Mas-Colell, and McClennan (1994) and Maskin and Tirole (2001) for reviews of these arguments.

However principled reasons for restricting attention to Markov equilibria are limited in the literature.[4]

This paper provides a foundation for Markov strategies for dynamic games with perfect information that rests on two assumptions. First, we make the restriction that all players (except possibly one) must use bounded recall strategies, i.e., strategies that do not depend on the infinite past. Second, we require equilibrium strategies to be "purifiable," i.e., to also constitute an equilibrium of a perturbed game with independent private payoff perturbations in the sense of Harsanyi (1973). Our main result is that Markov equilibria are the only equilibria which are bounded and purifiable.

The purifiability requirement reflects the view that our models are only an approximation of reality, and there is always some private payoff information. We make the modest requirement that there must be some continuous perturbation such that the equilibrium survives. The boundedness requirement is of interest for two distinct reasons. First, in many contexts, it is natural to assume that there do not exist two players who can observe the infinite past: consider, for example, games between a long run player and a sequence of short run players or in games with overlapping generations players. Second, strategies that depend on what happens in the arbitrarily distant past do not seem very robust to memory problems and/or noisy information. While we do not formally model this justification for focussing on bounded memory strategy profiles, we believe it may make them interesting objects of study.[5]

Our argument exploits special features of the games we study: only one player moves at a time and there is perfect information. Perfect information and the purifying payoff perturbations 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. Thus 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 point 5 on page 17.

[4]One exception is Bhaskar and Vega-Redondo (2002) who provides a rationale for Markov equilibria in asynchronous choice games based on complexity costs.

[5]In a different context (repeated games with imperfect public monitoring), Mailath and Morris (2002) and Mailath and Morris (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.

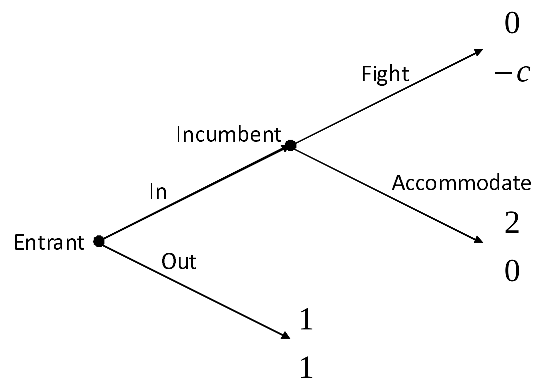

Figure 1: The stage game for the chain store. The top payoff is the payoff to the Entrant.

While perfect information games are special, many economic models fall within the class of this paper, as noted briefly in the abstract and discussed at length on page 7. In any case, the modeling choice between treating dynamic interactions as repeated simultaneous move games or repeated asynchronous move games is often made for tractability and transparency. As argued by Rubinstein (1991), some of our modeling choices should be understood as capturing players' understanding of their situation rather than a literal description of the environment. Our results highlight that bootstrapping payoff irrelevant information into non-Markovian folk theorems is sensitive to how the game is played.

2 A Long-Run Short-Run Player Example

Consider the following example of a repeated perfect information game, 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) must decide 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) must decide whether to accommodate or fight. The stage game is depicted in Figure 1. The short run player maximizes his stage game payoff while the long run player maximizes the discounted sum of payoffs with discount factor δ which is less than but close to 1. Let us assume that the short-run player only observes and thus can only condition on what happened in the last period. The long run player has observes the entire history. We will require equilibria to satisfy sequential rationality – each player must be choosing optimally at every possible history.

Ahn (1997, Chapter 3) has studied this type of game, and shows that for generic values of the discount factor, there is no pure strategy equilibrium where entry is deterred. 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 element of $H = \{OUT, A, F\}$. Consider a trigger strategy type 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 had not played A; 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. For 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 high enough δ.

There are however mixed strategy equilibria where entry is deterred in each period. One such equilibrium has the incumbent playing F with probability $\frac{1}{2}$, independent of history. The entrant is indifferent between IN and OUT at any information set, given incumbent's strategy. He plays OUT at $t = 1$. At $t > 1$ he plays OUT after $a_{t-1} \in \{OUT, F\}$; if $a_{t-1} = A$, he plays IN with probability $q = (1-\delta+(1-\delta)c)/(1-\delta^2+\delta(1-\delta)c)$, where q was chosen to make the incumbent indifferent between accommodating and fighting. In this equilibrium, the entrant's beliefs about the incumbent's response is *identical* after the two one-period histories A and $\{OUT, F\}$. Nevertheless, the entrant plays differently.

We now establish that this mixed strategy equilibrium cannot be purified if we add small shocks to the game's payoffs. So 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 z_i^2 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 entrant is

$$\rho_t : \{OUT, A, F\} \times [0,1] \to \Delta(A_1),$$

while a strategy for the incumbent is

$$\sigma_t : H^t \times [0,1] \to \Delta(A_2)$$

(in principle, it could depend condition on history of past payoff shocks, but this turns out to not matter). Note that ρ_{t+1} does not condition on what happened at $t-1$. Fix a history $h^t = (h_1, h_2, ..., h_t) \in H^t$ with $h_t = IN$ (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},

$$\sigma_t\left(\left(h^{t-1}, IN\right), z_2^t\right) = \sigma_t\left(\left(\tilde{h}^{t-1}, IN\right), z_2^t\right)$$

for almost all z_2^t. So the incumbent does not condition on h^{t-1}. Now the entrant at t also has a payoff shock. Since the incumbent does not condition on h^{t-1},

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

for almost all z_1^t.

We conclude that for any $\varepsilon > 0$, only equilibria in Markov strategies exist. In this context, this implies that the backwards induction outcome of stage game must be played in every period.

3 The Model

3.1 The Perfect Information Game

We consider an infinite dynamic game of perfect information, Γ. The game has a recursive structure and may also have public moves by nature. The set of players is denoted 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. Let $q(s'|s, a)$ denote the probability of state s' following state s when action a is played; thus $q : S \times A \to \Delta(S)$. The initial distribution over states is given by $q(\varnothing)$. Player i has bounded flow payoff $u_i : S \times A \to \mathbb{R}$ and a discount factor $\delta_i \in [0, 1)$. Total payoffs in the game are the discounted sum of flow payoffs. The dynamic game is given by $\Gamma = \left\{S, \mathcal{N}, \iota, q, (u_i)_{i \in \mathcal{N}}\right\}$.

The game starts in a state s_0 at period 0 determined by $q(\varnothing)$ and the history at period $t \geq 1$ is a sequence of states and actions, $H^t = (S \times A)^t$. Some histories may not be feasible: if after a history $h = (s_\tau, a_\tau)_{\tau=0}^t$, the state s has zero probability under $q(\cdot \mid s_t, a_t)$, then that state cannot arise after the history h. Since infeasible histories arise with zero probability and the set of all histories is countable, without loss of generality our notation often ignores the possibility of infeasible histories. Let $H^0 = \{\varnothing\}$ and $H = \cup_{t=0}^{\infty} H^t$; we write h for a typical element of H, $\tau(h)$ for the length of the history (i.e., $\tau(h)$ is the t for which $h \in H^t$), and $H^\infty = (S \times A)^\infty$ for the set of outcomes (infinite histories) with typical element h^∞. We sometimes write (h, s) for $(h, s_{\tau(h)}) = (s_0, a_0; s_1, a_1; \ldots, s_{\tau(h)-1}, a_{\tau(h)-1}; s_{\tau(h)})$, with the understanding that $s = s_{\tau(h)}$. Player i's payoff as a function of outcome, $U_i : H^\infty \to \mathbb{R}$, is

$$U_i(h^\infty) = U_i\left((s_t, a_t)_{t=0}^{\infty}\right) = (1-\delta_i) \sum_{t=0}^{\infty} \delta_i^t u_i(s_t, a_t).$$

A (*behavioral*) *strategy* for player i is a mapping $b_i : H \times S(i) \to \Delta(A)$. Write B_i for the set of strategies of player i. A strategy profile $b = (b_i)_{i \in \mathcal{N}}$ can be understood as a mapping $b : H \times S \to \Delta(A)$, specifying a mixed action at every history. Write $V_i(b \mid h, s)$ for player i's expected continuation utility from the strategy profile b at the history (h, s). This value is given recursively by

$$V_i(b \mid h, s) = \sum_{a \in A} b_{\iota(s)}(a \mid h, s) \Bigg\{ (1-\delta_i) u_i(s, a) + \delta_i \sum_{s' \in S} q(s' \mid s, a) V_i\left(b \mid (h, s, a), s'\right) \Bigg\}.$$

We write $V_i(b) \equiv \sum q(s \mid \varnothing) V_i(b | (\varnothing, s))$ for player i's ex ante utility under strategy profile b.

Definition 1 *A strategy b_i is* Markovian *if for each $s \in S(i)$ and histories $h, h' \in H$ of the same length (i.e., $\tau(h) = \tau(h')$),*

$$b_i(h, s) = b_i(h', s).$$

A Markovian strategy is stationary *if the two histories can be of different lengths.*

Remark 1 (Markov strategies) In this definition, we have taken the notion of *Markovian* strategy as a primitive. The restriction to Markov strategies is often motivated by the desire to restrict behavior to only depend on the payoff relevant aspects of history. Maskin and Tirole (2001) define a *payoff relevant* partition over (same length) histories using payoff relevance,[6] and identify elements of the partition with Markov states. While in general the payoff relevant partition is coarser than the partition induced by the states S, a sufficient condition for the two partitions to agree is that for every pair of states $s, s' \in S(i)$, $u_i(s, a)$ is not an affine transformation of $u_i(s', a)$ (Mailath and Samuelson, 2006, Proposition 5.6.2).[7] ♦

Definition 2 *Strategy profile b is a subgame perfect Nash equilibrium (SPNE) if, for all $s \in S$, $h \in H$, and each $i \in \mathcal{N}$ and $b'_i \in B_i$,*

$$V_i((b_i, b_{-i}) \mid h, s) \geq V_i((b'_i, b_{-i}) \mid h, s). \quad (1)$$

If b is both Markovian and a SPNE, it is said to be a Markov equilibrium.

Many games fit into our general setting. Repeated perfect information game (e.g. Rubinstein and Wolinsky (1995), Takahashi (2005)) fit: the state w tracks where you are in the perfect information stage game; $u_i(w_t, a_t)$ is zero whenever (w_t, a_t) results in a non-terminal node of the stage game, and is the payoff at the terminal node otherwise. Stochastic games where players move sequentially also fit: now the state w would stand either for a node in the perfect information game, or the initial node of one of the perfect information games. Doraszelski and Judd (2007) argue that this class of games is computationally tractable and important. A finite game of perfect information also fits: we simply add a terminal state. Perfect information games played between overlapping generations of players are an example (Kandori (1992), Bhaskar (1998) and Muthoo and Shepsle (2006)). Extensive form games between long run and short run players, as studied in the reputation literature, fit naturally (e.g. Fudenberg and Levine (1989); Ahn (1997, Chapter 3)). A literature examines infinitely repeated games with asynchronous moves, either with a deterministic order of moves (as in Maskin and Tirole (1987, 1988a,b), Lagunoff and Matsui (1997) and Bhaskar and Vega-Redondo (2002)) or with a random order of moves (as in Matsui and Matsuyama (1995)).[8] In both cases, the state is the profile of actions

[6] Loosely, the desired partition is the coarsest partition with the property that to every profile measurable with respect to that partition, each player has a best response measurable with respect to that partition.

[7] Maskin and Tirole (2001) use cardinal preferences to determine payoff equivalence, and hence the presence of affine transformations.

[8] To incorporate the Poisson process of opportunities to change actions, as in Matsui and Matsuyama (1995), we would have to incorporate a richer timing structure into our model. But the extension would be inessential.

	c_1	c_2	d
c_1	$11, 11$	$6, 9$	$-20, 20$
c_2	$9, 6$	$10, 10$	$-20, 20$
d	$20, -20$	$20, -20$	$0, 0$

Figure 2: Payoffs for an augmented prisoner's dilemma.

of players whose actions are fixed, and $u_i(w_t, a_t)$ is the stage game payoff. Canonical non-cooperative model of bargaining where, in each period, one proposer makes an offer and other players decide sequentially whether to accept or reject the offer also fit, with both deterministic order of moves (Rubinstein (1982)) and random order (Chatterjee, Dutta, Ray, and Sengupta (1993)).

Example 1 (An asynchronous move example) Consider the augmented prisoners' dilemma illustrated in Figure 2. With asynchronous moves, player 1 moves in odd periods and player 2 in even periods (since time begins at $t = 0$, player 2 makes the first move). State and action sets are $S = A = \{c_1, c_2, d\}$.

Suppose the initial state is given by c_1. The game admits multiple stationary Markov perfect equilibria, as well as nonstationary Markov perfect equilibria.

There are two stationary pure strategy Markov perfect equilibria: Let $b^* : S \to A$ be the Markov strategy given by $b^*(s) = s$. It is straightforward to verify that b^* is a perfect equilibrium for $\delta \in \left[\frac{1}{2}, \frac{20}{31}\right]$. Let $b^\dagger : S \to A$ be the Markov strategy given by $b^\dagger(c_1) = b^\dagger(c_2) = c_2$ and $b^\dagger(d) = d$. It is straightforward to verify that $b^\dagger$ is a perfect equilibrium for $\delta \in \left[\frac{1}{2}, \frac{2}{3}\right]$.

Finally, denote by $b^\alpha : S \to \Delta(A)$ the Markov strategy given by $b^\alpha(c_1) = \alpha \circ c_1 + (1-\alpha) \circ c_2$, $b^\alpha(c_2) = c_2$, and $b^\alpha(d) = d$. Suppose it is player i's turn. At (h, c_2), the payoff from following b^α is

$$V_i(b^\alpha \mid h, c_2) = 10. \tag{2}$$

At (h, c_1), the payoff from choosing c_1, and then following b^α, is

$$(1-\delta)11 + \delta\alpha\{(1-\delta)11 + \delta V_i(b^\alpha \mid (h, c_1, c_1), c_1)\} \\ + \delta(1-\alpha)\{(1-\delta)6 + \delta V_i(b^\alpha \mid (h, c_1, c_1), c_2)\}, \tag{3}$$

while the payoff from choosing c_2, and then following b^α, is

$$(1-\delta)9 + \delta V_i(b^\alpha \mid (h, c_1), c_2)\}. \tag{4}$$

In order for player i to be willing to randomize, (3) must equal (4), with this common values being $V_i(b^\alpha \mid h, c_1)$. Since $V_i(b^\alpha \mid (h, c_1), c_2) = 10$, (4) implies $V_i(b^\alpha \mid h, c_1) = 9 + \delta$, and solving (3) for α yields

$$\alpha = \frac{(4\delta - 2)}{(5 - \delta)\delta}. \tag{5}$$

This is a well defined probability for $\delta \geq \frac{1}{2}$. Moreover, b^α, for α satisfying (5), is a Markov perfect equilibrium for $\delta \in \left[\frac{1}{2}, \frac{2}{3}\right]$.

For any time t, the nonstationary Markov strategy specifying for periods before or at t, play according to b^*, and for periods after t, play according to b^α, for α satisfying (5), is a Markov perfect equilibrium for $\delta \in (\frac{1}{2}, \frac{2}{3})$. ★

3.2 The Perturbed Game

We now allow for the payoffs in the underlying game to be perturbed, as in Harsanyi (1973). We require that the payoff perturbations 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.[9] At each history (h, s), a payoff perturbation $z \in Z$ is independently drawn according to $\mu^s \in \Delta^*(Z)$. The payoff perturbation is observed by only $\iota(s)$, i.e., the player moving at s. If this player chooses action a, his payoff is augmented by εz^a, where $\varepsilon > 0$. Thus players' stage payoffs in the perturbed game depend only on the current state, action and payoff perturbation (s, a, z) and are given by

$$\tilde{u}_i(s, a, z) = \begin{cases} u_i(s, a) + \varepsilon z^a, & \text{if } i = \iota(s), \\ u_i(s, a), & \text{otherwise.} \end{cases}$$

We denote the perturbed game by $\Gamma(\varepsilon, \mu)$.

To describe strategies, we first describe players' information more precisely. Write $T_i(h, s)$ for the collection of periods at which player i moved (and thus observed a payoff perturbation), i.e.,

$$T_i(h, s) \equiv \{k \in \{0, 1, \ldots, \tau(h)\} \mid \iota(s_k) = i\}.$$

[9]Our analysis only requires that the support be in Z, but notation is considerably simplified by assuming Z is the support.

At the public history (h, s), in the perturbed game player $i = \iota(s)$ has private information

$$z^{T_i(h,s)} \equiv (z_k)_{k \in T_i(h,s)} \in Z^{T_i(h,s)}.$$

A *behavior strategy* for player i in the perturbed game, $\tilde{b}_i$, specifies player i's mixed action $\tilde{b}_i(h, s, z^{T_i(h,s)})$, at every history (h, s) with $s \in S(i)$ and for every realization of i's payoff perturbations $z^{T_i(h,s)}$. The set of all behavior strategies for player i is denoted $\tilde{B}_i$.

A vector of realized payoff perturbations is written $\mathbf{z}$, with the dimension implied by context. Thus, $\mathbf{z} = (z^{T_i(h,s)})_{i=1}^n$ is the vector of *all* realized payoff perturbations at the history (h, s) .

The definition of sequential rationality requires us to have notation to cover unreached information sets. Write

$$T_{-i}\,(h, s) = \bigcup_{j \neq i} T_j\,(h, s)$$

for the set of dates at which player i does *not* observe the payoff perturbations. A belief assessment for player i specifies, for every feasible history $h \in H$ and $s \in S(i)$, a belief

$$\pi_i^{h,s} \in \Delta\left(Z^{T_{-i}(h,s)}\right) \tag{6}$$

over the payoff perturbations $z^{T_{-i}(h,s)}$ that have been observed by other players at history (h, s). Note that, as suggested by the structure of the perturbed game, we require that these beliefs are independent of player i's private payoff perturbations, $z^{T_i(h,s)}$; beyond this requirement, we impose no further restrictions (such as that the payoff shocks are independent across the other players or periods)—see Remark 2.

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

$$\tilde{V}_i(\tilde{b} \mid h, s, \mathbf{z}) = \sum_{a \in A} b_{\iota(s)}(a \mid h, s, z^{T_{\iota(s)}(h,s)}) \Bigg[(1 - \delta_i)\tilde{u}_i(s, a, z_{\tau(h)}) \\ + \delta_i \sum_{s' \in S} q(s' \mid s, a) \int \tilde{V}_i(b \mid (h, s, a), s', (\mathbf{z}, z))\mu^{s'}(dz) \Bigg].$$

Since player i does not know all the coordinates of $\mathbf{z}$, player i's expected payoff from the profile $\tilde{b}$ is given by

$$\int \tilde{V}_i(\tilde{b} \mid h, s, (z^{T_i(h,s)}, z^{T_{-i}(h,s)}))\ \pi_i^{h,s}(dz^{T_{-i}(h,s)}). \tag{7}$$

Definition 3 *Strategy $\tilde{b}_i$ is a sequential best response to $(\tilde{b}_{-i}, \pi_i)$, if for each $h \in H$, $s \in S(i)$, $z^{T_i(h,s)} \in Z^{T_i(h,s)}$, and $\tilde{b}'_i \in \tilde{B}_i$,*

$$\int \tilde{V}_i((\tilde{b}_i, \tilde{b}_{-i}) \mid h, s, (z^{T_i(h,s)}, z^{T_{-i}(h,s)}))\ \pi_i^{h,s}(dz^{T_{-i}(h,s)}) \geq \int \tilde{V}_i((\tilde{b}'_i, \tilde{b}_{-i}) \mid h, s, (z^{T_i(h,s)}, z^{T_{-i}(h,s)}))\ \pi_i^{h,s}(dz^{T_{-i}(h,s)}).$$

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.

Definition 4 *A strategy $\tilde{b}_i$ is* shock history independent *if for all $h \in H$, $s \in S(i)$, and shock histories $z^{T_i(h,s)}, \hat{z}^{T_i(h,s)} \in Z^{T_i(h,s)}$,*

$$\tilde{b}_i(h, s, z^{T_i(h,s)}) = \tilde{b}_i(h, s, \hat{z}^{T_i(h',s)}), \qquad a.a.\ z \in Z,$$

whenever $z_{\tau(h)} = \hat{z}_{\tau(h)}$.

Lemma 1 *If $\tilde{b}_i$ is a sequential best response to any $\tilde{b}_{-i}$, then $\tilde{b}_i$ is a shock history independent strategy.*

Proof. Fix a player i, $h \in H$, $w \in W_i$ and payoff perturbation history $z^{T_i(h,s)}$ with $z_{\tau(h)} = z$. Player i's *next period* expected continuation payoff under $\tilde{b}$ from choosing action a this period, $\mathbf{V}_i(a, \tilde{b}_{-i}, \pi_i \mid h, s)$, is given by

$$\sum_{s'} q(s' \mid s, a) \iint \max_{\tilde{b}_i} V_i(\tilde{b}_i, \tilde{b}_{-i} \mid (h, s, a), s', \mathbf{z}, z')\ \mu^{s'}(dz')\pi_i^{h,s'}(dz^{T_{-i}(h,s)}).$$

Since $\tilde{b}_{-i}$ and $\pi_i^{h,s'}$ do not depend on player i's shocks before period $\tau(h)$, $z^{T_i(h,s)}$, the maximization implies that $\mathbf{V}_i(a, \tilde{b}_{-i}, \pi_i \mid h, s)$ also does not depend on those shocks. Thus, his total utility is

$$(1 - \delta_i)[u_i(s, a) + \varepsilon z^a] + \delta_i \mathbf{V}_i(a, \tilde{b}_{-i}, \pi_i \mid h, s).$$

Since Z has full dimension and μ^s is absolutely continuous, player i can only be indifferent between two actions a and a' for a zero measure set of $z \in Z$. For other z, there is a unique best response, and so it is shock history independent. ■

A shock history independent strategy (ignoring realization of z of measure 0) can be written as

$$\tilde{b}_i : H \times S(i) \times Z \to \Delta(A).$$

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

$$V_i^*(\tilde{b} \mid h, s) = \int \sum_{a \in A} \tilde{b}_{\iota(s)}(a \mid h, s, z) \Bigg[(1-\delta_i)\tilde{u}_i(s, a, z) + \delta_i \sum_{s' \in S} q(s' \mid s, a) V_i^*(\tilde{b} \mid (h, s, a), s') \Bigg] \; \mu^s(dz). \quad (8)$$

It is now immediate from Lemma 1 that beliefs over unreached information sets are essentially irrelevant in the notion of sequential best responses, because, while behavior can in principle depend upon prior payoff shocks, optimal behavior does not.

Lemma 2 *A profile $\tilde{b}$ is a profile of mutual sequential best responses if, and only if, for all i, $\tilde{b}_i$ is shock history independent, and for each $h \in H$, $s \in S(i)$ and $\tilde{b}'_i \in \tilde{B}_i$,*

$$V_i^*((\tilde{b}_i, \tilde{b}_{-i}) \mid h, s) \geq V_i^*((\tilde{b}'_i, \tilde{b}_{-i}) \mid h, s). \quad (9)$$

Remark 2 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 that we may appeal to. Our notion of sequential best response is very weak (not even requiring that the beliefs respect Bayes' rule *on the path of play*). The only requirement is that each player's beliefs over other players' payoff shocks be independent of that player's 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. This requirement 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: this would allow players to have different beliefs about past payoff perturbations depending on their realized current payoff realization.

However, Lemma 2 implies that once we impose mutuality of sequential best responses, any additional restrictions have no restrictive power. It is worth noting why no belief assessment $\pi_i^{h,s}$ appears either in the description of V_i^*, (8), or in Lemma 2: Player i's expected payoff from the profile $\tilde{b}$, given in (7), is the expectation over past payoff shocks of other players, $z^{T_{-i}(h,s)}$, as well as all future payoff shocks. Critically, in this expectation, as implied by the structure of the perturbed game, it is assumed that all future shocks are distributed according to μ, and are independent of all past shocks. ♦

Given Lemma 2 and the discussion in Remark 2, the following definition is natural:

Definition 5 *A* perfect Bayesian equilibrium *is a profile of mutual sequential best responses.*

The definition of *Markovian* shock history independent strategies naturally generalizes that for the unperturbed game: a strategy $\tilde{b}_i$ is *Markovian* if for each $s \in S(i)$, for almost all $z \in Z$, and histories $h, h' \in H$ with $\tau(h) = \tau(h')$,

$$\tilde{b}_i(h, s, z) = \tilde{b}_i(h', s, z).$$

Definition 6 *A shock history independent strategy $\tilde{b}_i$ has K-*recall *if for each $s \in S(i)$, histories $h, h' \in H$ satisfying $\tau(h) = \tau(h') = t$, and almost all $z \in Z$,*

$$\tilde{b}_i(h, s, z) = \tilde{b}_i(h', s, z)$$

whenever $(s_k, a_k)_{k=t-K}^{t-1} = (s_k, a_k)_{k=t-K}^{t-1}$. A strategy b_i has infinite recall *if it does not have K-recall for any K. A Markovian strategy is a 0-recall strategy (there being no restriction on h and h').*

A K-recall strategy is stationary *if the two histories can be of different lengths.*

The following is the key result of the paper.

Lemma 3 *If $\tilde{b}_i$ is a sequential best response to $\tilde{b}_{-i}$ and does not have K-recall, then for some $j \neq i$, $\tilde{b}_j$ does not have $(K+1)$-recall.*

Proof. If $\tilde{b}_i$ does not have K-recall, then there exist h and h' with $\tau(h) = \tau(h') = t \geq K$ and $s \in S(i)$

$$(s_k, a_k)_{k=t-K}^{t-1} = (s'_k, a'_k)_{k=t-K}^{t-1}$$

and

$$\tilde{b}_i(h, s, z) \neq \tilde{b}_i(h', s, z) \tag{10}$$

for a positive measure of z.

Suppose that $\tilde{b}_j$ has $(K+1)$-recall for each $j \neq i$. Since histories (h, s, a) and (h', s, a) agree in the last $K+1$ periods, player i's continuation value from playing action a at (h, s) and at (h', s) is the same, for all s':

$$V_i^*((\tilde{b}_i, \tilde{b}_{-i}) \mid (h, s, a), s'). = V_i^*((\tilde{b}_i, \tilde{b}_{-i}) \mid (h', s, a), s').$$

Hence, player i's total expected utility from choosing action a at *either* (h, s) or (h', s) is

$$(1 - \delta_i)\tilde{u}_i(s, a, z) + \delta_i \sum_{s' \in S} q(s' \mid s, a) V_i^*((\tilde{b}_i, \tilde{b}_{-i}) \mid (h, s, a), s').$$

For almost all z, there will be a unique a maximizing this expression, contradicting our premise (10). ■

Corollary 1 *If $\tilde{b}$ is a perfect Bayesian equilibrium of the perturbed game, then either $\tilde{b}$ is Markovian or at least two players have infinite recall.*

3.3 Purification in the Games of Perfect Information

We now consider the purifiability of rationalizable strategies in the unperturbed game. 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 $h \in H$, $s \in S_i$ and $a \in A$,

$$\int \tilde{b}_i^k(a \mid h, s, z)\mu^s(dz) \to b_i(a \mid h, s)$$

Definition 7 *The strategy profile b is* purifiable *if there exists $\mu : S \to \Delta^*(Z)$ 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 $\Gamma(\mu, \varepsilon^k)$ for each k.*

We immediately have the following:

Proposition 1 *If b is a purifiable SPNE in which no more than one player has infinite recall, then b is Markovian.*

4 Discussion

1. PURIFIABILITY OF MARKOV EQUILIBRIA. All Markovian equilibria are purifiable: we can simply pick the noise distribution to support exactly the Markov behavioral strategy we are purifying.

   We expect suitably regular Markov equilibria to satisfy stronger purifiability properties. It is worth noting that—if we are allowed enough freedom in picking the noise—we can purify anything without any regularity arguments.

   To make this precise, we proceed as follows: We restrict attention to finite (N) players and finite states; and generalize the payoff perturbation so that each player gets a payoff perturbation as each decision node, so $\mu : S \to \Delta^*(Z^N)$. We also weaken the notion of purification, allowing the distribution μ to depend on k.

   Definition 8 *The strategy profile b is* weakly purifiable *if there exists $\varepsilon^k \to 0$ and, for each k, $\mu^k : S \to \Delta^*(Z^N)$, 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 $\Gamma(\mu, \varepsilon^k)$ for each k.*

   Claim 1 *Suppose there is a finite number of players, N, and S is finite. Every Markov equilibrium in the unperturbed game is weakly purifiable.*

   SKETCH OF PROOF.

   Let $b : S \to \Delta(A)$ be any Markov equilibrium of the unperturbed game. Write $A^*(s)$ for the set of actions that are best responses (perhaps weak best responses) at state s. Fix a sequence of Markov strategy profiles b^k such that $b^k \to b$ (i.e., $b^k(a \mid s) \to b(a \mid s)$ for each a and s) with the support of $b^k(\cdot \mid s)$ equal to $A^*(s)$.

   Write $V_i(b' \mid s)$ for the expected payoff to player i from b' in the unperturbed game starting in state s. Recall that Z, a full dimensional compact subset of $\mathbb{R}^A$, is the support of the payoff shocks. Let's make the normalization that the $\mathbf{0}$ vector is in the interior of Z. For each action a, write $Z^*(a, s)$ for the collection of payoff shock profiles favoring action a among those played with positive probability, i.e.,

$$Z^*(a, s) = \left\{ z \in Z^N | z^a_{\iota(s)} > z^{a'}_{\iota(s)} \text{ for all } a' \in A^*(s), \, a' \neq a \right\}.$$

Note that the union of the closures of the sets $(Z^*(a,s))_{a\in A}$ is Z^N, i.e.,

$$\cup_{a\in A}\mathrm{cl}\,(Z^*(a,s)) = Z^N.$$

Write $a^*(z,s)$ for the action satisfying $z \in Z^*(a,s)$. (Such an action is unique for almost all z, choose arbitrarily when it is not unique.) For each $s \in S$ and k, set

$$\Delta_i^k(s) = (1-\delta_i)\sum_{a\in A}\left(b(a \mid s) - b^k(a \mid s)\right)u_i(s,a)$$

Now choose $\varepsilon^k \in \mathbb{R}_+$ and $\mu^k(s) \in \Delta\left(Z^N\right)$ with a strictly positive density such that

$$E_{\mu^k(s)}\left(z_i^{a^*(z,s)}\right) = \Delta_i^k(s)$$

and, for each $a \in A^*(s)$

$$\Pr_{\mu^k(s)}(Z^*(a,s)) = b^k(a|s).$$

(I assert that this is feasible; is this obvious?). Since $b^k \to b$ we can choose these so that $\varepsilon^k \to 0$. Now consider the perturbed game $\Gamma\left(\mu^k,\varepsilon^k\right)$. Consider the strategy profile $\widetilde{b}^k$ given by

$$\widetilde{b}^k_{\iota(s)}(a|s,z) = \begin{cases} 1, \text{ if } z \in Z^*(a,s) \\ 0, \text{ if } z \in Z^*(a',s), \text{ for any } a' \in A^*(s),\ a' \neq a \end{cases}$$

It does not matter what $\widetilde{b}$ does on the (zero measure) boundaries of the sets $Z^*(a,s)$. I claim this is a PBE of the perturbed game (i.e., each $\widetilde{b}_i^k$ is a sequential best response to $\widetilde{b}_{-i}^k$) and that $\widetilde{b}^k$ converges to b.

2. EXISTENCE OF MARKOV EQUILIBRIA. A large literature addresses the question of existence of Markov equilibria: see Duffie, Geanakoplos, Mas-Colell, and McClennan (1994), Escobar (2008) and Doraszelski and Escobar (2008).

3. UNIQUENESS OF MARKOV EQUILIBRIA. There will often be multiple Markov equilibria in our class of games: see, for example, Maskin and Tirole (1988a). But this multiplicity of Markov equilibria does not allow us to sustain any additional outcomes.

4. HARSANYI PURIFICATION. "Purification" has had two multiple meanings in the literature (see Morris (2008)). One question asked in the literature is when can we guarantee that there exists an essentially pure equilibrium in a game by adding noise to payoffs (e.g., Radner and Rosenthal (1982)). It is trivially true that our perturbation ensures that there is an essentially pure equilibrium (we build in enough independence to guarantee that this is the case). We follow Harsanyi (1973) in being interested in the relation between equilibria of the unperturbed game and equilibria of the perturbed game. But our definition of "purifiability" is very weak: we require only that there exists a sequence of equilibria of a sequence perturbed games that converge to the desired behavior. Harsanyi (1973) showed (for static games) the much stronger that (under some regularity conditions) every equilibrium was the limit of a sequence of equilibria in *every* sequence of perturbed games. This suggests a stronger definition of purifiability in our context. Strategy profile b is *Harsanyi purifiable* if, for *every* $\mu : S \to M$ and $\varepsilon^k \to 0$, the $\left(\mu, \varepsilon^k\right)_{k=1}^{\infty}$ perturbed games have a sequence of strategy profiles $\tilde{b}^k$ converging to b, with $\tilde{b}_i^k$ a sequential best response to $\tilde{b}_{-i}^k$ in $\Gamma\left(\mu, \varepsilon^k\right)$ for each i. We conjecture that with additional regularity assumptions, Markovian equilibria will be Harsanyi purifiable: Doraszelski and Escobar (2008) provided conditions for the Harsanyi purifiability of Markovian equilibria; while their class of games they study do not encompass those in the present paper, it seems possible to extend their results.

5. SIMULTANEOUS MOVES. Our results do not extend to games where more than one player moves at a time, e.g. repeated synchronous move games. Mailath and Morris (2002) and Mailath and Olszewski (2008) give examples of finite recall strategy profiles which are strict and therefor purifiable. In this context, one might conjecture a weaker result that purifiability would rule out the "belief-free" strategies developed is the recent literature (Piccione (2002), Ely and Välimäki (2002) and Ely, Hörner, and Olszewski (2005)). Bhaskar, Mailath, and Morris (2008) show that the one period recall strategies of Ely and Välimäki (2002) are purifiable via one period recall strategies in the perturbed game; however, they are purifiable via infinite recall strategies. The purifiability of such belief free strategies via finite recall strategies remains an open question.

6. ENDOGENOUS IDENTIFICATION OF MARKOVIAN STRUCTURE.

We constructed a game where the publicly observed state w was a sufficient statistic for everything payoff relevant in the game. Maskin and Tirole (2001) describe how one can identify the coarsest possible description of a state that is sufficient statistic for payoff relevance in the continuation. If the game we were studying turned out to have a courser payoff relevant, our arguments in favor of Markov perfection would apply to that coarser state space also.

7. INESSENTIAL ELEMENTS IN THE MODELLING. A number of simplifying assumptions were made in our formulation to lighten the notation, and could easily be relaxed. There is no need to have the same support for the noise in different states. The length of time between moves could be random (for example, reflecting a Poisson process of arrival of revision opportunities). Our model could easily be enriched to allow for this.

8. TECHNOLOGICAL RESTRICTION ON MEMORY. We maintained the assumption of perfect information in our formal analysis. In many applications, such as those discussed on page 3.1, it is natural to assume that some players observe only a finite history of play. This will (technologically) rule out equilibria with infinite memory. In principle, allowing imperfect observation of play could create finite memory equilibria that would not have been equilibria with infinite memory. However, the arguments that we presented apply also if players have restricted memory.

9. RATIONALIZABILITY. Our argument was stated for equilibrium, but could be extended to apply to versions of rationalizability for this class of games that maintained sequential rationality and the key feature that players' beliefs about other players' past payoff shocks were not correlated with their current payoff shocks.

References

Acemoglu, D., and P. Robinson (2001): "A Theory of Political Transistions," *American Economic Review*, 91(4), 938–963.

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.

Bhaskar, V., and F. Vega-Redondo (2002): "Asynchronous Choice and Markov Equilibria," *Journal of Economic Theory*, (2), 334–350.

Chatterjee, K., B. Dutta, D. Ray, and K. Sengupta (1993): "A Non-Cooperative Theory of Coalitional Bargaining," *Review of Economic Studies*, 60(2), 463–477.

Doraszelski, U., and J. Escobar (2008): "A Theory of Regular Markov Perfect Equilibria in Dynamic Stochastic Games: Genericity, Stability, and Purification," Discussion paper, Harvard University and Stanford University.

Doraszelski, U., and K. Judd (2007): "Dynamic Stochastic Games with Sequential State-to-State Transitions," Discussion paper, Harvard University.

Duffie, D., J. Geanakoplos, A. Mas-Colell, and A. McClennan (1994): "Stationary Markov Equilibria," *Econometrica*, 62(4), 745–781.

Ely, J. C., J. Hörner, and W. Olszewski (2005): "Belief-Free Equilibria in Repeated Games," *Econometrica*, 73(2), 377–416.

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

Ericson, R., and A. Pakes (1995): "Markov-Perfect Industry Dynamics: A Framework for Empirical Work," *Review of Economic Studies*, 62(1), 53–53.

Escobar, J. (2008): "Existence of Pure and Behavior Strategy Stationary Markov Equilibrium in Dynamic Stochastic Games," Discussion paper, Stanford University.

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

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.

Kandori, M. (1992): "Social Norms and Community Enforcement," *The Review of Economic Studies*, 59(1), 63–80.

Lagunoff, R., and A. Matsui (1997): "Asynchronous Choice in Repeated Coordination Games," *Econometrica*, 6(5), 1467–1477.

Mailath, G., and W. Olszewski (2008): "Folk Theorems with Bounded Recall under (Almost) Perfect Monitoring," PIER Working Paper 08-027, University of Pennsylvania.

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

Mas-Colell, A., M. D. Whinston, and J. Green (1995): *Microeconomic Theory*. Oxford University Press, New York, NY.

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.

Matsui, A., and K. Matsuyama (1995): "An Approach to Equilibrium Selection," *Journal of Economic Theory*, 65(2), 415–434.

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.

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.

Rubinstein, A. (1982): "Perfect Equilibrium in a Bargaining Model," *Econometrica*, 50(1), 97–109.

——— (1991): "Comments on the interpretation of game theory," *Econometrica*, 59, 909–924.

Rubinstein, A., and A. Wolinsky (1995): "Remarks on Inifinitely Repeated Extensive-Form Games," *Games and Economic Behavior*, 9(1), 110–115.

Takahashi, S. (2005): "Infinite Horizon Common Interest Games with Perfect Information," *Games and Economic Behavior*, 53(2), 231–247.