# Finite State Equilibria in Dynamic Games


Michihiro Kandori *
*Faculty of Economics*
*University of Tokyo*

Ichiro Obara †
*Department of Economics*
UCLA


June 23, 2007


**Abstract**

An equilibrium in an infinite horizon stochastic game is called a *finite state equilibrium*, if each player's action *on the equilibrium path* is given by an automaton with a finite state space. We provide a complete characterization of this class of equilibria and provide a recursive computational method to check the equilibrium conditions. This encompasses the majority of existing works on repeated games with private monitoring.


## 1 Introduction

An equilibrium in an infinite horizon game is called a *finite state equilibrium*, if each player's action *on the equilibrium path* is given by an automaton with a finite state space. We provide a complete characterization of this class of equilibria and provide a recursive computational method to determine if a given profile of finite automata (one for each player to determine his action on the path of play) can constitute a sequential (thus a finite) equilibrium.

Although we can allow private types for each player, driven by some stochastic process, the major area of application of our result is repeated games with private monitoring. The present paper provides a unifying general theory to encompass the majority of the existing work as most of them are based on some form of finite state equilibria.

The belief-based approach by Sekiguchi [9], Bhaskar and Obara [1] considers an equilibrium which coincides with a trigger strategy on the path


*e-mail: kandori@e.u-tokyo.ac.jp

†e-mail: iobara@econ.ucla.edu

of play (hence a finite state equilibrium), but is different from the trigger strategy off the equilibrium path. The present paper can be regarded as a generalization of those papers.

The belief-free approach by Ely and Välimäki [3] considers an equilibrium which can be implemented by finite state automatons on and off the path of play. Proposition 4 in Ely, Horner and Olszewski [2](bang-bang property) shows that most of belief-free equilibrium payoffs (indeed all of them if the discount factor is close to 1) can be obtained by a finite state equilibrium with only two states. Matsushima [6] and Hörner and Olszewski [4] considers a class of equilibria which generalizes belief-free equilibrium, but retains similar properties. They are also finite equilibria. An exception which does not employ a finite equilibrium is Piccione [8], whose equilibrium path requires infinite (countably many) states. However, the result by Ely, Horner and Olszewski shows that Piccione's equilibrium payoff can be obtained by a finite state equilibrium.

A more recent paper by Phelan and Skrzypacz [7] also proposes an algorithm to compute a class of stationary finite state equilibria. Later we discuss differences between their approach and our approach after presenting our theorem.

## 2 Model

This paper considers infinite horizon stochastic games with private monitoring. The stage game is defined by players, actions, types, private signals, and payoffs. Let $N = \{1, ..., n\}$ be the set of players. For each player, there is a finite set of possible types $\Theta_i$ and a finite set of actions $A_i$. Actions are not observable. Player $i$ observes a private signal $s_i \in S_i$. Let $\pi(s|a)$ be the probability of private signal profile $s$ given action profile $a$. We assume that, for each player $i$, the marginal distribution of $\pi$ on $s_i$ is a full support distribution for every $a \in A$. The standard imperfect public monitoring games fit into our framework, so we can apply our method to the private strategy equilibria (Kandori and Obara [5]). Player $i$'s payoff is determined by his action, his private type, and his private signal. Denote player $i$'s payoff by $u_i(\theta_i, a_i, s_i)$. Let $g_i(\theta_i, a) = \sum_{s_i} u_i(\theta_i, a_i, s_i)\, \pi(s_i|a)$ be player $i$'s expected payoff given $a$ when his type is $\theta_i$.

The game continues for infinite number of periods. Type profile $\theta = (\theta_1, ..., \theta_n) \in \Theta$ follows a Markov process characterized by the transition functions $p(\theta'|\theta, a) \in [0, 1]$ with some initial distribution $p' \in \triangle\Theta$. Player $i$'s period $t-$history is $h_i^t = (\theta_i^1, a_i^1, s_i^1, ..., \theta_i^{t-1}, a_i^{t-1}, s_i^{t-1}) \in H_i^t$. Player $i$'s pure strategy $\sigma_i \in \Sigma_i$ is a mapping from $\cup_{t=1}^{\infty} H_i^t$ to $A_i$. Private strategy profile $\sigma$, combined with $p'$, generates a probability measure on the set of infinite sequences of private types and actions. Player $i$'s expected discounted average payoff $V_i(\sigma, p') = (1-\delta) E[\sum_{t=1}^{\infty} \delta^{t-1} g_i(\theta_i^t, a^t) | \sigma, p']$ is computed with respect to this measure. We analyze a sequential equilibrium of this infinite horizon stochastic games with private monitoring.

This model includes many type of economic models. As already mentioned, this model becomes a repeated game with private monitoring when there is no type. If types are drawn only once in the beginning and fixed throughout the game $(p(\theta|\theta, a) = 1$ for all $(\theta, a))$, then it becomes the standard model of reputation with one-sided or two-sided incomplete information. If types are drawn independenty across time and players, then it becomes is the standard model of repeated adverse selection. Note that our model also allows a middle ground in these two extreme cases: the case where types are neither perfectly persistent nor perfectly independent over time.

We focus on a class of equilibria where *equilibrium behavior* can be described by finite state automatons. An automaton consists of three components: states, a transition rule, a mapping from states to actions for each type. Let $Q_i(\theta_i)$ be a finite set of states for player $i$ whose type is $\theta_i$. Let $Z_i$ be all feasible type-state pairs: $z_i = (\theta_i, q_i) \in Z_i$ if and only if $q_i \in Q_i(\theta_i)$. A mapping $f_i : Z_i \to A_i$ describes player $i$'s play for each type-state pair of player $i$. Note that we can assume that players choose a pure action in each state without loss of generality. For example, when player $i$ plays action $a_i'$ and $a_i''$ with an equal probability at state $\theta_i$, we can "split" $\theta_i$ into two states $\theta_i'$ and $\theta_i''$ and assume that $a_i'$ $(a_i'')$ is played in $\theta_i'$ $(\theta_i'')$ (and the mixing probabilities can be encoded in the state transition probabilities).

Conditional on $\theta_i$ being player $i$'s current type, the probability that player $i$ is in type-state pair $z_i = (\theta_i, q_i) \in Z_i$ is given by his private signal $s_i$ and type-state pair $(\theta_i', q_i')$ in the previous period. This probability is denoted $m_i^{\theta_i}(q_i|\theta_i', q_i', s_i) \in \triangle Q_i(\theta_i)$.

An automaton for player $i$, defined as $M_i = \left( \left\{ Q_i(\theta_i), m_i^{\theta_i} \right\}_{\theta_i \in \Theta_i,}, \ f_i \right)$, still does not describe player $i$'s behavior completely for two reasons. First, it does not specify which state to start. We allow players to use a correlation device, which is independent of $\theta$, to generate the initial joint distribution over $Z$. The idea is that a mediator sends a private message to each player, which players use to determine their initial state combined with their type realizations. This generates an initial distribution $\mu$ over $Z$, whose marginal distribution on $\Theta$ is $p'$. Secondly, note that each automaton does not specify which action to choose when player $i$ does not follow the course of action suggested by the automaton. Note that a specification of the actions on the path of play corresponds to a *reduced normal form strategy.* A reduced normal form strategy for player $i$ specifies the sequence of actions of player $i$ for any strategy profile of the opponents. This comes from our assumption that private signal $s_i$ has full support. An automaton only defines a reduced normal form strategy, but does not specify what to play off the equilibrium path. Thus automatons must be complemented by a strategy off the equilibrium path to make a full-fledged strategy. This is the topic for the next subsection.

### 2.1 Belief-Based Strategy and Finite State Equilibrium

Fix a profile of automatons $M$. Given $M_{-i}$, player $i$ can compute his belief about the other players' type-state pair profile at *every* private history combined with his knowledge of the initial distribution $\mu$. Denote player $i$'s belief by $\mu_i \in \triangle Z_{-i}$. Consider a function $\rho_i : \triangle Z_i \rightarrow A_i$. Now this function defines a full-fledged strategy given any initial belief. We call such a strategy *belief-based strategy.* If we combine player $i$'s automaton with any belief-based strategy $\rho_i$ by assuming that player $i$'s behavior follows $\rho_i$ once he deviates from $M_i$, then we have a complete strategy for player $i$. Denote this strategy for player $i$ by $\sigma_i (M_i, \rho_i)$. For a given profile of finite state automaton $(M_1, ..., M_n)$, suppose that $(\sigma_1 (M_1, \rho_1), ..., \sigma_n (M_n, \rho_n))$ is a (correlated) sequential equilibrium with some initial distribution $\mu$ on $Z$. We call such an equilibrium a *finite state equilibrium*, and say that $(M_1, ..., M_n)$ generates a finite equilibrium with initial distribution $\mu$. This equilibrium notion generalizes mixed trigger strategy equilibria used in Sekiguchi [9] and Bhaskar and Obara [1].

## 3 Belief-Based Approach to Finite State Equilibrium

We ask the following question: when does a particular profile of finite state automatons generate a finite state equilibrium? When players' behavior is based on finite state automatons, their continuation strategies are completely summarized by their types and states. Therefore what is relevant to player $i$'s decision at each point of time is his belief over $Z_{-i}$. Since $Z_{-i}$ is a finite set, player $i$'s belief is always confined in the finite dimensional simplex. This makes our analysis and computation easier.

Our approach to attack this question is as follows. Since player $i$ cares only about the other players' types and states, player $i$'s decision problem becomes a dynamic programming problem where the state variable is his belief over $Z_{-i}$. Solving this dynamic programming problem, we can derive the optimal action correspondence from player $i$'s belief to player $i$'s optimal actions. From this correspondence, we generate a set of sequentially optimal belief-based strategies. Then we check consistency, namely, whether an action assigned by an automaton is indeed optimal at every history on the equilibrium path. To do so, we need to make sure that, when player $i$'s state is $q_i$, player $i$'s belief on $Z_{-i}$ is contained in the set of beliefs where $f_i(q_i)$ is optimal, whichever private history has brought player $i$ to state $q_i$ .

We start with a relatively simple case with no type. That is, the underlying game is a repeated game with private monitoring. In this case, an automaton for player $i$ is simply denoted by $M_i = (Q_i, m_i, \; f_i)$. Fix a set of automaton $(M_1, ...., M_n)$. Suppose that player $i$'s belief is $\mu_i(h_i^t) \in \triangle Q_{-i}$ in period $t$. This determines the distribution of $a_{-i}$ in the current period. Player $i$'s belief in the next period $\mu_i(h_i^t)$ is affected by his action and private signal in the current period. Let $V_i^t(\mu_i(h_i^{t+1}))$ be player $i$'s optimal payoff given $\mu_i(h_i^t)$. Since stage games are bounded above and there is discounting, $V_i^t$ satisfies the following dynamic programming equation.

$$V_i^t\left(\mu_i\left(h_i^t\right)\right) = \max_{a_i \in A_i} (1-\delta) E^{\mu_i\left(h_i^t\right)}[g_i(a_i, f_{-i}(q_{-i}))] + \delta E^{\mu_i\left(h_i^t\right)}[V_i^{t+1}\left(\mu_i\left(h_i^{t+1}\right)\right)|a_i]$$

where $\mu_i\left(h_i^{t+1}\right) \in \triangle Q_{-i}$ is computed by Bayes' rule as follows:

$$\mu_i\left(h_i^{t+1}\right)(q_{-i}) = \frac{\sum_{q'_{-i}} \sum_{s^t_{-i}} \mu_i\left(h_i^t\right)(q'_{-i}) \pi\left(s^t | a_i^t, f_{-i}\left(q'_{-i}\right)\right) \prod_{j \neq i} m_j\left(q_j | q'_j, s_j^t\right)}{\sum_{q'_{-i}} \sum_{s^t_{-i}} \mu_i\left(h_i^t\right)(q'_{-i}) \pi\left(s^t | a_i^t, f_{-i}\left(q'_{-i}\right)\right)}.$$

In the stationary form, this becomes

$$V_i(\mu_i) = \max_{a_i \in A_i} (1-\delta) E^{\mu_i}[g_i(a_i, f_{-i}(q_{-i}))] + \delta E^{\mu_i}[V_i(\chi_i[\mu_i, a_i, s_i])|a_i]$$

where $\chi_i : \triangle Q_i \times A_i \times S_i \rightarrow \triangle Q_{-i}$ is given by

$$\chi_i[\mu_i, a_i, s_i](q_{-i}) = \frac{\sum_{q'_{-i}} \sum_{s_{-i}} \mu_i(q'_{-i}) \pi\left(s | a_i, f_{-i}\left(q'_{-i}\right)\right) \prod_{j \neq i} m_j\left(q_j | q'_j, s_j\right)}{\sum_{q'_{-i}} \sum_{s_{-i}} \mu_i(q'_{-i}) \pi\left(s | a_i, f_{-i}\left(q'_{-i}\right)\right)}$$

We call this function $V_i(\mu_i)$ a *belief-based value function.*

By solving this dynamic programming problem, we can obtain a correspondence from player $i$'s belief to player $i$'s optimal actions. Let $\tau_i : A_i \rightrightarrows \triangle Q_{-i}$ be the correspondence defined by $\mu_i \in \tau_i(a_i)$ if and only if $a_i$ is one of the optimal actions given $\mu_i$. Thus $\tau_i$ is the (lower) inverse of the optimal policy correspondence, defined for each $a_i \in A_i$.

We like to verify whether player $i$'s automaton $M_i$ assigns optimal actions on the equilibrium path. For $M_i$ to be optimal, then player $i$'s belief $\mu_i$ at each $q_i$ needs to satisfy $\mu_i \in \tau_i(f_i(q_i))$. Note that player $i$'s belief at each state is not unique because different histories can lead player $i$ to the same state. Let $B_i(q_i, \mu)$ be the set of beliefs which may arise at state $q_i$ when players are playing *according to* $M$ and the initial joint distribution is $\mu \in \triangle Q$.

Then we have the following theorem.

**Theorem 1** *A finite-state automaton profile $M$ generates a finite state equilibrium with initial joint distribution $\mu \in \triangle Q$ if and only if $B_i(q_i, \mu) \subset \tau_i(f_i(q_i))$ for all $q_i \in Q_i$ and $i$.*

**Proof.** Suppose that $M$ generates a finite state equilibrium. If $B_i(q_i, \mu) \notin \tau_i(f_i(q_i))$ for some $q_i$ and $i$, then there exists some private history $h_i^t$ after which player $i$'s state is $q_i$ and player $i$'s belief $\mu_i\left(h_i^t\right)$ is not in $\tau_i(f_i(q_i))$. Then $f_i(q_i) = a_i$ cannot be an action to satisfy sequential rationality at this history, this is a contradiction.

Conversely, suppose that $B_i(q_i, \mu_1) \subset \tau_i(f_i(q_i))$ for all $q_i \in Q_i$ and $i$. Let $\rho_i : \triangle Q_{-i} \rightarrow A_i$ be any selection from the optimal policy correspondence of the above dynamic programming problem for player $i$. Consider a complete strategy profile $\sigma = (\sigma_1(M_1, \rho_1), ..., \sigma_n(M_n, \rho_n))$. This strategy satisfies one-shot deviation constraints off the equilibrium path by definition. One-shot deviation constraints on the equilibrium path are also satisfied because $B_i(q_i, \mu) \subset \tau_i(f_i(q_i))$ for all $q_i \in Q_i$ and $i$. Since all one-shot deviation constraints are satisfied, this strategy profile constitutes a sequential equilibrium. ■

**Remark.**

1. Belief-free equilibrium is included in this class of equilibria. In particular, $\tau_i(f_i(q_i)) = \triangle Q_{-i}$ for all $q_i$ holds for belief-free equilibria, i.e. the action assigned to player $i$ (and player $i$'s continuation strategy) on the equilibrium path is sequentially optimal independent of his belief about the other players' states.

2. Phelan and Skrzypacz [7] analyzes a similar problem, but the following points distinguish our approach from theirs. They consider the class of sequential equilibria that can be represented by a profile of finite state automatons *on and off the equilibrium path.* On the other hand, we only require that some finite state automaton profile describes on the equilibrium path behavior. Continuation strategies off the equilibrium path are given by belief-based strategies which are not necessary finite-state automatons.

There is a simpler method to obtain the same result. To verify the conditions in Theorem 1, one needs to know (the inverse of) the optimal policy correspondence ($\tau_i$) for the belief-based dynamic programming problem and also needs to compute all possible beliefs (the set $B_i(q_i, z)$) explicitly. Note that the optimal policy correspondence is obtained by the optimal belief-based value function $V_i(\mu_i)$. The next theorem shows that we can actually use that information only (i.e., without computing belief dynamics) to check if the given automata constitutes a finite state equilibrium. To state the theorem, we need to introduce a couple of concepts. Let $v_i(q)$ be player $i$'s payoff generated by automata $(M_1, ...., M_n)$ when the initial state profile is $q = (q_1, ..., q_n)$. This function can be easily obtained by solving a $|Q|$ simultaneous (dynamic programming) equations, one for each $v_i(q), q \in Q$. Let $W_i(q_i, \mu_i) = \sum_{q_{-i}} v_i(q)\mu_i(q_{-i})$. Note that this is a linear function of player $i$'s belief $\mu_i(q_{-i})$. $W_i(q_i, \mu_i)$ is the average payoff to player $i$ when $i$'s current state and belief are $q_i$ and $\mu_i$ respectively (assuming that all players follow the given finite state automata $(M_1, ..., M_n)$). By definition, $W_i(q_i, \mu_i) \leq V_i(\mu_i)$. Define $\Xi_i(q_i) \equiv \{\mu_i | W_i(q_i, \mu_i) = V_i(\mu_i)\}$. This $\Xi_i(q_i)$ is the set of player $i$'s beliefs where $M_i$, with initial state $q_i$, indeed solves the dynamic programming problem.

Now we are ready to state our next theorem.

**Theorem 2** *Fix a finite state automata $(M_1, ..., M_n)$. If there is a joint distribution of the states $\mu \in \Delta Q$ such that $\mu_i(q_i) \in \Xi_i(q_i)$ (for all $q_i$ with $\sum_{q_{-i}} \mu(q) > 0$) for every $i$, then $(M_1, ..., M_n)$ generates a finite state equilibrium with initial joint distribution $\mu$.*

**Proof.** The conditions show that $W_i(q_i, \mu_i(q_i)) = V_i(\mu_i(q_i))$ for all $q_i$ with $\sum_{q_{-i}} \mu(q) > 0$. This means that, for each realization of $q$ from initial distribution $\mu$, the players are mutually taking best replies. Hence the given automata $(M_1, ..., M_n)$ specify the equilibrium path of play. ■

A crucial condition in this theorem is that for some range of beliefs, the automaton payoff $W_i(q_i, \mu_i)$ coincides with the optimal value $V_i(\mu_i)$. This theorem shows that there is a (conceptually) easy way to verify if any given finite state automata $(M_1, ..., M_n)$ can constitutes an equilibrium. Note that here we only need to calculate players' *initial beliefs* (associated with the initial joint distribution $\mu$), while Theorem 1 requires to compute all possible beliefs that can arise over time (the set $B_i(q_i, \mu)$). Also notice that one can easily obtain $W_i(q_i, \mu_i)$ for any automata. This is just to solve a system of finitely many linear equations, so a closed form solution can always be obtained (by Cramer's rule). The optimal belief-based value function $V_i(\mu_i)$ can be numerically computed by the standard successive approximation method. Once they are obtained, we only need to check if they coincide on appropriate regions.

The following lemma may be useful in checking the condition $W_i(q_i, \mu_i) = V_i(\mu_i)$.

**Lemma 3** *The optimal belief-based value function $V_i(\mu_i)$ is convex. Moreover, if $W_i(q_i, \mu_i^k) = V_i(\mu_i^k)$ for $k = 1, 2$, then $W_i(q_i, \lambda\mu_i^1 + (1-\lambda)\mu_i^2) = V_i(\lambda\mu_i^1 + (1-\lambda)\mu_i^2)$ for all $\lambda \in [0, 1]$.*

**Proof.** Convexity of $V_i$ is standard. Let $u_i(\sigma_i, \mu_i)$ be the average expected payoff to player $i$ when (i) $i$ use a repeated game strategy $\sigma_i$ and (ii) other players follow the given finite state automaton and their initial state distribution is given by $\mu_i$. Note that, as an expected value, $u_i(\sigma_i, \mu_i)$ is linear in $\mu_i$. Now take any two beliefs $\mu_i^k$, $k = 1, 2$ and any $\lambda \in [0, 1]$. By definition, $V_i(\lambda\mu_i^1 + (1-\lambda)\mu_i^2) = u_i(\sigma_i^*, \lambda\mu_i^1 + (1-\lambda)\mu_i^2)$ for some $\sigma_i^*$. By the linearity of $u_i$, we have

$$\begin{aligned} V_i(\lambda\mu_i^1 + (1-\lambda)\mu_i^2) &= \lambda u_i(\sigma_i^*, \mu_i^1) + (1-\lambda) u_i(\sigma_i^*, \mu_i^2) \\ &\leq \lambda V_i(\mu_i^1) + (1-\lambda) V_i(\mu_i^2) \end{aligned}$$

where the inequality comes from the fact that $V_i$ is the optimal value. Hence $V_i$ is convex. Now we proceed to show the second part of the lemma. By the convexity, we know that $V_i(\lambda\mu_i^1 + (1-\lambda)\mu_i^2) \leq \lambda V_i(\mu_i^1) + (1-\lambda)V_i(\mu_i^2) = W_i(q_i, \lambda\mu_i^1 + (1-\lambda)\mu_i^2)$, where the equality comes form the linearity of $W_i(q_i, \mu_i)$ in $\mu_i$ and our premise $W_i(q_i, \mu_i^k) = V_i(\mu_i^k)$, $k = 1, 2$. However, the optimality of $V_i(\lambda\mu_i^1 + (1-\lambda)\mu_i^2)$ implies $V_i(\lambda\mu_i^1 + (1-\lambda)\mu_i^2) \geq W_i(q_i, \lambda\mu_i^1 + (1-\lambda)\mu_i^2)$. Hence we have $W_i(q_i, \lambda\mu_i^1 + (1-\lambda)\mu_i^2) = V_i(\lambda\mu_i^1 + (1-\lambda)\mu_i^2)$. ■

Our procedure to compare $V_i(\mu_i)$ with $W_i(q_i, \mu_i)$, although conceptually simple, has some potential drawback. The problem is that the numerical method to compute $V_i(\mu_i)$ may not converge in a finite time, so what we can obtain in a finite time, say $V_i^T(\mu_i)$, may not be equal to the true value function $V_i(\mu_i)$. Then we may not know if $W_i(q_i, \mu_i)$ is *exactly* equal to $V_i(\mu_i)$. In the next section, we introduce a stronger notion of finite state equilibria to overcome this problem which is inherent in the numerical approach.

### 3.1 General Case with Private Types

This generalizes to the case where players have private types. Player $i$'s current belief $\mu_i$ over $Z_{-i}$ determines the distribution of $a_{-i}$ in the current period and affect his belief in the next period. Player $i$'s belief in the next period is also affected by his action, his private signal, and his type in the next period. Denote this belief by $\chi_i\left[\mu_i, \theta_i', a_i, s_i\right]$. Then player $i$'s dynamic programming problem for $\theta_i$ becomes

$$V^{\theta_i}(\mu_i) = \max_{a_i \in A_i} (1-\delta) E^{\mu_i}[g_i(\theta_i, a_i, f_{-i}(z_{-i}))] + \delta E^{\mu_i}[V^{\theta_i'} \chi_i[\mu_i, \theta_i', a_i, s_i])|a_i, \theta_i]$$

Note that now the value function also depend on player $i$'s type because types are payoff relevant and his type contains information about the other players' types. Solving this programming problem, we can obtain correspondences $\tau_i^{\theta_i} : A_i \rightrightarrows \triangle Z_{-i}, \theta_i \in \Theta_i$. Let $B_i(z_i, \mu)$ be the set of beliefs which may arise at type-state pair $z_i$ when players are playing according to $M$ and the initial joint distribution is $\mu \in \triangle Z$

**Theorem 4** *A finite-state automaton profile $M$ generates a finite state equilibrium with initial joint distribution $\mu \in \triangle Z$ if and only if $B_i(z_i, \mu) \subset \tau_i^{\theta_i}(f_i(q_i))$ for all $z_i = (\theta_i, q_i) \in Z_i$ and $i$.*

## 4 More Analytical Approach to Finite State Equilibrium

Our Theorems 1, 2 and 4 provide *numerical* recursive methods to find a finite state equilibrium. Given a profile of automata describing the equilibrium path, these theorems require that the belief-based value function $V_i(\mu_i)$ to be numerically computed by the standard recursive method with a finite dimensional state space. There are two potential problems. First, the state variable $\mu_i \in \triangle Q_{-i}$ is a continuous variable. There are many numerical methods to cope with this problem, one of which is to compute the value function on a finitely many grid points (see, for example, Ch. 12 of K. Judd (1998) *Numerical Methods in Economics*, MIT press.) (Incidentally, the grid method may be easy if $\mu_i$ is one-dimensional (two-player games with two states for each player), but it becomes computationally infeasible when $|Q_{-i}|$ is large.). Second, as we have already discussed, $V_i(\mu_i)$ is not exactly obtained in a finite step.

In this section, we introduce a special class of finite state equilibrium, called strong finite state equilibrium, to overcome these problems.

**Definition 5** *A finite state equilibrium associated with (finite state) automata* $(M_1, ...., M_n)$ *is said to be a* ***strong finite state equilibrium*** *if* $(M_1, ...., M_n)$ *specifies equilibrium strategies both on* and off *the path of play.*

Phelan and Skrzypacz [7] provide necessary and sufficient conditions for a strong finite state equilibrium when the automata satisfy what they call information depreciation condition. Their method requires first to look at a game without the starting date ($t = 0, \pm 1, \pm 2, ...$) and also to keep track of the dynamics of beliefs. Here we present much simpler necessary and sufficient conditions without requiring the information depreciation condition. The example of the next section has the property that the mixture of the grim trigger strategy and permanent defection are played on the path of play, but the information depreciation condition is not satisfied.

For a given automata $(M_1, ...., M_n)$, let $v_i(q)$ be player $i$'s payoff generated by the automata when the initial state profile is $q = (q_1, ..., q_n)$. We will find an equilibrium where continuation path of play is always given by automata $(M_1, ...., M_n)$ with some initial state $q$. Let $W_i(q_i, \mu_i) = \sum_{q_{-i}} v_i(q)\mu_i(q_{-i})$. Define the upper envelope of those functions

$$W_i^*(\mu_i) = \max_{q_i \in Q_i} W_i(q_i, \mu_i).$$

To interpret this, let $M_i(q_i)$ be player $i$'s (contingent) action path generated by automaton $M_i$ with initial state $q$. Function $W_i^*(\mu_i)$ is the (constrained) optimal payoff when player $i$'s path of play is restricted to $M_i(q_i)$, $q_i \in Q_i$. For each $\mu_i$, let $\mathcal{U}_i(\mu_i)$ be the set of states where $M_i(q_i)$ is constrained optimal:

$$\mathcal{U}_i(q_i) = \{\mu_i | W_i^*(\mu_i) = W_i(q_i, \mu_i)\}.$$

Note that $W_i(q_i, \mu_i)$ is linear in $\mu_i$, so that (i) the upper envelope $W_i^*(\mu_i)$ is a piecewise linear function (which can be calculated very easily) and (2) $\mathcal{U}_i(q_i)$ is a convex polyhedron.

In what follows we derive the conditions under which the constrained optimization coincides with the true optimal (so that $W_i^*(\mu_i)$ is equal to the belief-based (truly optimal) value function $V_i(\mu_i)$). Let $V_i(a_i, \mu_i)$ be player $i$'s discounted average payoff when his belief is $\mu_i$ and he plays $a_i$ in the current period, followed by the constrained optimal path from $M_i(q_i)$, $q_i \in Q_i$. That is,

$$V_i(a_i, \mu_i) = \sum_{q_{-i}} \mu_i(q_{-i}) \left[ (1-\delta) \sum_{a_{-i}} g_i(a_i, f_{-i}(q_{-i})) + \delta \sum_{s_i} W_i^*(\chi_i[\mu_i, a_i, s_i]) \pi_i(s_i | a_i, f_{-i}(q_{-i})) \right]$$

Define $T_i W_i^* (\mu_i)$ by $T_i W_i^* (\mu_i) = \max_{a_i \in A_i} V_i(a_i, \mu_i)$.

If we can show that $T_i W_i^* (\mu_i) = W_i^*(\mu_i)$, then $W_i^*(\mu_i) = V_i(\mu_i)$ follows from the uniqueness of the value function. Then, by Theorem 2, what we need is just an appropriate initial distribution of the states to make sure that each player's belief is in $\mathcal{U}_i(q_i)$ when he is recommended to start at $q_i$. To show that $T_i W_i^* (\mu_i) = W_i^*(\mu_i)$, we just need to check that, for each $\mu_i$, player $i$'s optimal action to solve $\max_{a_i \in A_i} V_i(a_i, \mu_i)$ and his continuation behavior is consistent with the behavior specified by his finite state automaton for some initial state. If this is true, then $\max_{a_i \in A_i} V_i(a_i, \mu_i)$ is achieved by one of $M_i(q_i)$. Since $\max_{a_i \in A_i} V_i(a_i, \mu_i)$ is larger than $W_i^* (\mu_i)$ by definition, we can conclude that $W_i^* (\mu_i) = T_i W_i^* (\mu_i)$.

Note that $T_i W_i^* (\mu_i) = W_i^*(\mu_i)$ does not need to be verified for every belief on $Q_{-i}$. Instead, we can work with a subset of beliefs from which one cannot escape on and off the equilibrium path once your belief is in the set. Formally, we need to introduce the notion of *belief-closed* set.

**Definition 6** *A subset of player i's beliefs $\Xi_i \subset \Delta Q_{-i}$ is* belief-closed *if beliefs, once in this set, never exit from this set on and off the equilibrium path:*

$$\forall a_i, s_i \ \ \mu_i \in \Xi_i \implies \mu_i' = \chi_i[\mu_i, a_i, s_i] \in \Xi_i.$$

Now we are ready to state a set of necessary and sufficient conditions for a given automata $(M_1, ...., M_n)$ to generate a finite state equilibrium.

**Theorem 7** *A finite state automata $(M_1, ...., M_n)$ generates a strong finite state equilibrium if and only if*

1. *there is a belief-closed set $\Xi_i$ for every $i \in N$ such that, for every $\mu_i \in \Xi_i$, there exists $q_i$ that satisfies the following two properties:*

   *(a) $f_i(q_i)$ maximizes $V_i(a_i, \mu_i)$*

(b) $\chi_i[\mu_i, f_i(q_i), s_i] \in \mathcal{U}_i(q_i')$ *for all* $s_i$ *and all* $q_i' \in \cup suppm_i(\cdot|q_i, f_i(q_i), s_i)$[1]

2. *there is an initial joint distribution* $\mu$ *on* $Q$ *such that* $\mu_i(q_i) \in \mathcal{U}_i(q_i)$ *for every* $q_i \in Q_i$ *with* $\mu_i(q_i) > 0$.

**Proof.** Suppose that 1-(a) and 1-(b) are satisfied. Then we have $T_i W_i^*(\mu_i) = W_i(q_i, \mu_i)$ for some $q_i$. Notice that $T_i W_i^*(\mu_i) \geq W_i^*(\mu_i)$ by definition. Since $W_i^*(\mu_i)$ is the upper envelope of $W_i(q_i, \mu_i)$, this implies that $W_i(q_i, \mu_i) = W_i^*(\mu_i)$ for this $q_i$. Hence $W_i^*$ is a fixed point of $T_i$, thus it must coincide with the true belief-based value function $V_i$ on $\Xi_i$ (because the value function is unique by Blackwell's theorem). Clearly this finite state automata $(M_1, ...., M_n)$ generates a finite state equilibrium given the joint distribution on $Q$ specified in (2).

On the other hand, suppose that $(M_1, ...., M_n)$ generates a finite state equilibrium with some initial joint distribution $\mu$. Clearly (2) is a necessary condition for this finite state automata to be optimal. Define $\Xi_i$ to be the set of all possible beliefs of player $i$ on and off the equilibrium path given this automata. For any $\mu_i \in \Xi_i$, $M_i$ must specify the optimal strategy for some state $q_i$ by definition. Therefore 1-(a) and 1-(b) follow. ■

# 5 Examples

## 5.1 Bhaskar and Obara (2002)

We apply the above technique to provide a complete characterization of finite state equilibria from Bhaskar and Obara [1]. Consider the standard prisoners' dilemma stage game with the following payoff table

| | $C$ | $D$ |
|---|---|---|
| $C$ | $1, 1$ | $-1, 2$ |
| $D$ | $2, -1$ | $0, 0$ |

Actions are not observable. Player $i$ observes a private signal $s_i \in \{c, d\}$ about player $j \neq i$'s action. Player $i'$s signal is said to be correct when $s_i = c$ $(d)$ for $a_i = C$ $(D)$. Otherwise it is not correct. Assume that both players observe correct signals with probability $1 - 2\varepsilon - \eta$, only one player does with probability $\varepsilon > 0$, and both observe incorrect signals with probability $\eta > 0$.

Consider a set of symmetric automatons $M_i = (Q_i, m_i, f_i), i = 1, 2$ which is defined by $Q_i = \{x, y\}$, $m_i(x|x, c) = 1$, $m_i(x|x, d) = m_i(x|y, s_i) = 0$ for

[1] supp$m_i(\cdot|q_i, a_i, s_i)$ is the subset of $Q_i$ which arise with positive probability given $(q_i, a_i, s_i)$ (the collection of $q_i'$ such that $m_i(q_i'|q_i, a_i, s_i) > 0$) .

any $s_i$, $f(x) = C$, and $f(y) = D$. Note that $M_i(x)$ is realization equivalent to the standard trigger strategy and $M_i(y)$ is to the strategy which plays $D$ independent of history. Clearly, $M = (M_1, M_2)$ generates a finite state equilibrium if $(y, y)$ is the initial state. The question is when it generates a finite state equilibrium with a joint distribution that puts positive probability on $(x, x)$. Bhaskar and Obara [1] (originally Sekiguchi [9]) shows that $(M_1, M_2)$ generates a finite state equilibrium with nondegenerate independent distribution when $\delta$ is in some middle range and the monitoring is almost perfect, i.e. $\varepsilon$ and $\eta$ is small enough. Our goal is to *characterize* a range of parameters $(\varepsilon, \delta, \eta)$ where $(M_1, M_2)$ generates a finite state equilibrium with some joint distribution (with positive weight on $(x, x)$).

Let's first calculate $W^*(\mu_i)$. We focus on player 1 as the game is symmetric. Let $V_{q_1 q_2}$ be player 1's discounted average payoff when player 1's state is $q_1$ and player 2's state is $q_2$. Then

$$\begin{aligned}
V_{xx} &= (1-\delta) + \delta\left\{(1-2\varepsilon-\eta)V_{xx} + \varepsilon\left(V_{xy}+V_{yx}\right) + \eta V_{yy}\right\} \\
V_{xy} &= (1-\delta)(-1) + \delta\left\{(\varepsilon+\eta)V_{xy} + (1-\varepsilon-\eta)V_{yy}\right\} \\
V_{yx} &= (1-\delta)2 + \delta\left\{(\varepsilon+\eta)V_{yx} + (1-\varepsilon-\eta)V_{yy}\right\} \\
V_{yy} &= (1-\delta)0 + \delta V_{yy}
\end{aligned}$$

Therefore

$$\begin{aligned}
V_{xx} &= \frac{(1-\delta)(1-\delta\eta)}{\{1-\delta(1-2\varepsilon-\eta)\}\{1-\delta(\varepsilon+\eta)\}} \\
V_{xy} &= \frac{-(1-\delta)}{1-\delta(\varepsilon+\eta)} \\
V_{yx} &= \frac{2(1-\delta)}{1-\delta(\varepsilon+\eta)} \\
V_{yy} &= 0
\end{aligned}$$

Let $\mu$ be player 1's belief that player 2 is in state $x$. Then $W(x,\mu) = \mu V_{xx} + (1-\mu)V_{xy}$ and $W(y,\mu) = \mu V_{yx} + (1-\mu)V_{yy}$. It is easy to show that $W(x,\mu) \geq W(y,\mu)$ if and only if $\mu \geq \mu^*$, where $\mu^*$ is defined as:

$$\begin{aligned}
\mu^* &= \frac{V_{xy}}{V_{xy}+V_{yx}-V_{xx}} \\
&= \frac{1-\delta(1-2\varepsilon-\eta)}{\delta(1-2\varepsilon-2\eta)}
\end{aligned}$$

Since we are looking for a finite state equilibrium where $x$ is used with positive probability, $\mu^* \leq 1$ is necessary for our purpose. This leads to the following constraint on $\varepsilon$ and $\eta$,

$$2-\frac{1}{\delta}\geq 4\varepsilon+3\eta \tag{1}$$

Given that this condition is satisfied, $W^*(\mu)=W(x,\mu)$ when $\mu\in[\mu^*,1]$ and $W^*(\mu)=W(y,\mu)$ when $\mu\in[0,\mu^*]$, i.e. it is better to be in cooperative state $x$ if and only if he is enough confident that the other player is in state $x$.

Let's check the condition for Theorem 7. Player 1's belief is tracked by $\chi[\mu,a,s]$, which takes the following forms for this problem:

$$\begin{aligned}
\chi[\mu,C,c] &= \frac{\mu(1-2\varepsilon-\eta)}{\mu(1-\varepsilon-\eta)+(1-\mu)(\varepsilon+\eta)}\\
\chi[\mu,C,d] &= \frac{\mu\varepsilon}{\mu(\varepsilon+\eta)+(1-\mu)(1-\varepsilon-\eta)}\\
\chi[\mu,D,c] &= \frac{\mu\varepsilon}{\mu(1-\varepsilon-\eta)+(1-\mu)(\varepsilon+\eta)}\\
\chi[\mu,D,d] &= \frac{\mu\eta}{\mu(\varepsilon+\eta)+(1-\mu)(1-\varepsilon-\eta)}
\end{aligned}$$

We need to find a belief closed set $\Xi\subset[0,1]$ where 1-(a),(b) is satisfied for each $\mu\in\Xi$. Note first that both $\chi[\mu,D,c]$ and $\chi[\mu,D,d]$ is smaller than $\mu$ (because of 1). As in Bhaskar and Obara [1], we can show that $D$ solves $\max_a V(a,\mu)$ for $\mu\in[0,\mu^*]$ (the detail is omitted). Thus 1-(a) and 1-(b) is satisfied for $\mu\in[0,\mu^*]$ (with $q=y$).

For $\mu>\mu^*$, we like to show that $M(x)$ is optimal. Since $\chi[\mu,C,c]$ is crosses the 45° line once from the above, $\mu$ converges monotonically to $\mu'$ that satisfies $\chi[\mu',C,c]=\mu'$ after $(C,c)$,...,$(C,c)$. Hence $\Xi$ must include beliefs arbitrarily close to $\mu'$. Then it is necessary that (A) $\mu'\geq\mu^*$ and (B) $\chi[\mu',C,d]\leq\mu^*$ for 1-(a) and 1-(b) to be satisfied. If one of these conditions are violated, then $M(x)$ cannot be optimal in an small neighborhood of $\mu'$.

Conversely, (A) and (B) are sufficient for 1-(a) and 1-(b) to hold. First, they immediately imply 1-(b) if we take $\Xi=[0,\mu']$.[2] Furthermore, we can show that $C$ solves $\max_a V(a,\mu')$ (1-(a)) for any $\mu\in[0,\mu']$. It is clear that we can find an initial distribution to satisfy (2) of Theorem 2 (for example, independent randomization according to $\mu^*$).

To sum up, the only conditions we need to verify are (A) and (B). Since

$$\mu'=\frac{1-3\varepsilon-2\eta}{1-2\varepsilon-2\eta}$$

[2] Indeed we can use a larger set $\mu\in[\mu^*,\mu'']$ where $\mu''$ is the critical $\mu$ such that $\chi[\mu'',C,d]=\mu^*$.

these conditions boil down to

$$\frac{1-3\varepsilon-2\eta}{1-2\varepsilon-2\eta} \geq \frac{1-\delta\left(1-2\varepsilon-\eta\right)}{\delta\left(1-2\varepsilon-2\eta\right)}$$
$$\frac{\frac{1-3\varepsilon-2\eta}{1-2\varepsilon-2\eta}\varepsilon}{\frac{1-3\varepsilon-2\eta}{1-2\varepsilon-2\eta}\left(\varepsilon+\eta\right)+\left(1-\frac{1-3\varepsilon-2\eta}{1-2\varepsilon-2\eta}\right)\left(1-\varepsilon-\eta\right)} \leq \frac{1-\delta\left(1-2\varepsilon-\eta\right)}{\delta\left(1-2\varepsilon-2\eta\right)}$$

The first condition is

$$\begin{aligned} \frac{1-3\varepsilon-2\eta}{1-2\varepsilon-2\eta} &\geq \frac{1-\delta\left(1-2\varepsilon-\eta\right)}{\delta\left(1-2\varepsilon-2\eta\right)} \\ \delta\left(2-5\varepsilon-3\eta\right) &\geq 1 \end{aligned}$$

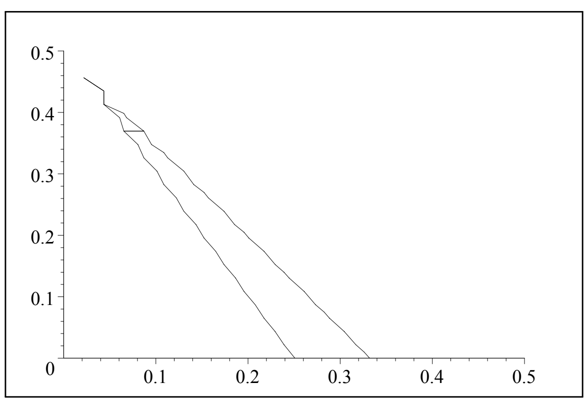
The second condition is

$$\begin{aligned} & 2\varepsilon+\eta-6\varepsilon\eta-4\varepsilon^2-2\eta^2 \\ \geq\ & \delta(3\varepsilon+\eta-12\varepsilon\eta-11\varepsilon^2-3\eta^2 \\ & +16\varepsilon^2\eta+10\varepsilon\eta^2+8\varepsilon^3+2\eta^3) \end{aligned}$$

There exists $\delta$ to satisfy both if and only if

$$\varepsilon+\eta-11\varepsilon\eta-7\varepsilon^2-4\eta^2+26\varepsilon^2\eta+18\varepsilon\eta^2+12\varepsilon^3+4\eta^3 \geq 0.$$

The region of $(\varepsilon, \eta)$ to satisfy this condition is as follows.

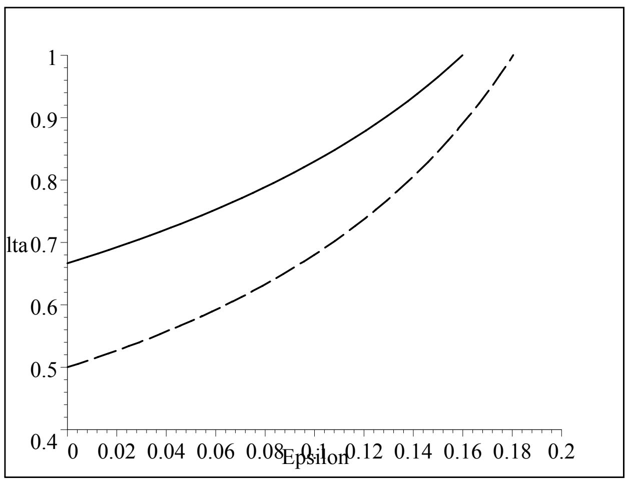



If $\varepsilon$ is close enough to 0, then this condition is satisfied for any $\eta$. This is the case where the monitoring is close to almost public monitoring. On the other hand, it is violated for a middle range of $\varepsilon$ however small $\eta$ is. Intuitively, if $\eta$ is close to 0, then an observation $c$ is a strong signal that the other player is observing $d$. Thus it becomes difficult to sustain a trigger strategy.

Consider the special case in which private signals are independent $(\eta = \varepsilon^2)$. Then the above inequalities become

$$\begin{aligned} \delta\left(2-5\varepsilon-3\varepsilon^2\right) &\geq 1 \\ 2-3\varepsilon-6\varepsilon^2-2\varepsilon^3 &\leq \delta(3-10\varepsilon-4\varepsilon^2+13\varepsilon^3+10\varepsilon^4+2\varepsilon^5) \end{aligned}$$

Below we plot these two inequalities. If $(\varepsilon, \delta)$ falls in the area between the solid line and the dashed line, then $M$ generates a finite state equilibrium.




For a large $\varepsilon$, $M$ can generate a finite state equilibrium with an independent mixture of $x$ and $y$ even when players are very patient (however this is a very inefficient equilibrium). As $\varepsilon$ gets smaller, a range of $\delta$ where such equilibrium exists go down. The efficiency improves when $\varepsilon$ is small so that "punishment" is not accidentally triggered and/or $\delta$ is small so that eventual breakdown of cooperation does not affect the current discounted average payoff much. The efficient payoff can be approached as $(\varepsilon, \delta) \to \left(0, \frac{1}{2}\right)$ within this area. This is the efficient equilibrium obtained in Bhaskar and Obara [1]. On the other hand, the above picture captures the whole range of $(\varepsilon, \delta)$ for which a finite state equilibrium (including very inefficient ones) can be generated.

## References


[1] V. Bhaskar and I. Obara. Belief-based equilibria in the prisoners' dilemma with private monitoring. *Journal of Economic Theory*, 102:16–39, 2002.

[2] J. C. Ely, J. Hörner, and W. Olszewski. Belief-free equilibria in repeated games. *Econometrica*, 73(2):377–415, 2005.

[3] J. C. Ely and J. Välimäki. A robust folk theorem for the prisoner's dilemma. *Journal of Economic Theory*, 102(1):84–105, 2002.

[4] J. Hörner and W. Olszewski. The folk theorem for games with private almost-perfect monitoring. mimeo, 2004.

[5] M. Kandori and I. Obara. Efficiency in repeated games revisited: The role of private strategies. *Econometrica*, 74(2):499–519, 2006.

[6] H. Matsushima. Repeated games with private monitoring: Two players. *Econometrica*, 72:823–852, 2004.

[7] C. Phelan and A. Skrzypacz. Private monitoring with infinite histories. mimeo, 2006.

[8] M. Piccione. The repeated prisoner's dilemma with imperfect private monitoring. *Journal of Economic Theory*, 102(1):70–83, 2002.

[9] T. Sekiguchi. Efficiency in repeated prisoner's dilemma with private monitoring. *Journal of Economic Theory*, 76:345–361, 1997.