Weakly Belief-Free Equilibria in Repeated Games with Private Monitoring

KANDORI, Michihiro*
Faculty of Economics, University of Tokyo

Sept 23, 2008

Abstract

Repeated games with imperfect private monitoring have a wide range of applications, but a complete characterization of all equilibria in this class of games has yet to be obtained. The existing literature has identified a relatively tractable *subset* of equilibria. The present paper introduces the notion of *weakly belief-free equilibria* for repeated games with imperfect private monitoring. This is a tractable class which subsumes, as a special case, a major part of the existing work (the belief-free equilibria), and it is shown that this class can outperform (in terms of efficiency) the equilibria identified by previous work.

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

1 Introduction

Repeated games with imperfect private monitoring have a wide range of applications, but a complete characterization of all equilibria in this class of games has yet to be obtained. The existitng literature has identified a relatively tractable *subset* of equilibria. Since an equilibrium in a repeated game represents a self-enforcing agreement in a long term relationship, the current state of our knowledge implies that we are still trying to understand how much cooperation can be sustained when agents interact over time under private monitoring.

The present paper demonstrates a new way to construct an equilibrium in repeated games with imperfect private monitoring, which can outperform the equilibria identified by the previous literature. Specifically, I generalize the notion of belief-free equilibria (Ely and Valimaki (2002) and Ely, Horner, and Olszewski (2005), EHO hereafter), which has played a major role in the existing literature, and show that the resulting *weakly belief-free equilibria* continue to possess a nice recursive structure. I then apply this concept to a repeated prisoners' dilemma game with private monitoring and construct a simple equilibrium, which outperforms the equilibria identified by previous work. The superior performance is due to the fact that the equilibrium partially embodies the essential mechanism to achieve efficiency in repeated games with imperfect monitoring (the transfer of continuation payoffs across players, as in Fudenberg, Levine, and Maskin (1994)). In addition, the equilibrium is in very simple pure strategies, and it is robust in the sense that players' actions are always *strict* best replies.[1] This is in contrast to belief-free equilibria, which rely on judiciously chosen mixed strategies and provide only *weak* incentive to follow the equilibrium actions.

A repeated game is a dynamic game where the same set of agents play the same game (the stage game) over an infinite time horizon. Economists and game theorists have successfully employed this class of models to examine how self-interested agents manage to cooperate in long-term relationships. A repeated game is said to have (imperfect) *private monitoring* if agents' actions are not directly observable and each agent receives imperfect *private* information (a *private signal*) about the opponents' actions. This class of games has a number of important potential applications. A leading example is a price competition game where firms may offer secret price cuts to

[1] In contrast, it is essential that players are indifferent over a set of actions in belief-free equilibria. Bhaskar (2000) argues that this is a problematic feature, because such an equilibrium may not be purified (in the sense of Harsanyi) by a plausible payoff purturbation. See Section 5 for more discussion.

their customers. In such a situation, each firm's sales level serves as the private signal, which imperfectly reveals the rivals' pricing behavior. Despite the wealth of potential applications, however, this class of games is not fully understood[2]. This stands in sharp contrast to the case where players share the same information (repeated games with *perfect* or imperfect *public* monitoring), where the set of equilibria are fully characterized and efficient outcomes can be sustained under a mild set of conditions (the Folk theorems) [3]. The main difficulty in the private monitoring case comes from the fact that players have *diverse* information about each other's behavior.

In the perfect or public monitoring case, players always share a mutual understanding about what they are going to do in the future. In the private monitoring case, however, each player has to draw statistical inferences about the opponents' future action plans, because they depend on an unobservable history of the opponents' private signals. The inferences quickly become complicated over time, even if players adopt relatively simple strategies. Hence, checking the equilibrium condition is in general a demanding task in repeated games with private monitoring (see Kandori (2002)). As a result, the complete characterization of all equilibria in this class of games is still unknown. The existing literature has only identified a rather tractable subset of equilibria.

To deal with this difficulty, the existing literature has adopted two alternative approaches. One is the *belief-based approach,* which looks at the case where the inference problem is relatively tractable. The literature either confines attention to special examples (see, for example, Sekiguchi (1997), Bhaskar and van Damme (2002), Bhaskar and Obara (2002)), or employs computational methods (see the important recent contribution by Phelan and Skrzypacz (2008)). More successful so far has been the other approach, the *belief-free approach*, which bypasses the complexity of inference altogether by constructing an equilibrium where players do not have to draw any statistical inferences at all. Let us denote player i's action and private signal in period t by $a_i(t)$ and $\omega_i(t)$. Note that, in gen-

[2] If communication is allowed, it is known that the Folk theorm holds in private monitoring repeated games (Compte (1998) and Kandori and Matsushima (1998)). Recent literature, including the present article, mainly explore the possibility of cooperation under no communication. This is important because in a major applied area (collusion) communication is explicitly prohibited by anti-trust laws.

[3] See the *self-generation* condition of Abreu, Pearce and Stacchetti (1990), and the *Folk theorems* of Fudenberg and Maskin (1986) for the public monitoring case and Fudenberg, Levine, and Maskin (1994) for the imperfect public monitoring case. A survey of repeated game literature can be found in Kandori (2008) (for an accessible overview), and in Mailath and Samuelson (2006) (for a comprehensive exposition).

eral, each player i's continuation strategy at time $t+1$ is determined by his *private history* $h_i^t = (a_i(1), \omega_i(1), ..., a_i(t), \omega_i(t))$. The belief-free approach constructs an equilibrium where player i's continuation strategy is a best reply to the opponents' continuation strategies *for any realization* of $h_{-i}^t = (a_{-i}(1), \omega_{-i}(1), ..., a_{-i}(t), \omega_{-i}(t))$, thereby making player i's belief over h_{-i}^t irrelevant. Such an equilibrium is called a *belief-free equilibrium* (defined formally by EHO). The core of this approach was provided by the influential works of Piccione (2002), Obara (1999), and Ely and Valimaki (2002). This idea was later substantially generalized by Matsushima (2004), EHO (2005), Horner and Olszewski (2006), and Yamamoto (2007). EHO show that the set of belief-free equilibria can be characterized by a simple recursive method similar to that of Abreu, Pearce and Stacchetti (1990).

In the present paper, I propose a weakening of the belief-free conditions, leading to a set of equilibria which are still tractable and are capable of sustaining a larger payoff set. Note that the belief-free conditions imply that, at the beginning of period $t+1$, player i does not have to form beliefs over $h_{-i}^t = (a_{-i}(1), \omega_{-i}(1), ..., a_{-i}(t), \omega_{-i}(t))$. In contrast, I require that player i does not need to form beliefs over $(a_{-i}(1), \omega_{-i}(1), ..., a_{-i}(t))$, *omitting the last piece of information* $\omega_{-i}(t)$ *from the belief-free requirement.* This says that player i does not have to know the opponents' histories *up to the previous actions.* However, player i does need to understand correctly that, for each possible action profile $a(t)$, the private signals in the previous period are distributed according to $p(\omega(t)|a(t))$. I call equilibria with this property *weakly belief-free equilibria.*

To show that weakly belief-free equilibria have a recursive structure, I depart from the tradition of looking at the continuation *payoff sets.* In the perfect or imperfect monitoring cases (Abreu, Pearce, and Stacchetti (1990)) as well as in the belief-free approach in the private monitoring case (see EHO (2005)), it has been a common practice to keep track of the set of continuation payoffs to exploit the recursive structure. In contrast, I introduce the notion of *reduced games* and examine their recursive structure. A reduced game at time t is a game with the same set of players and actions as in the stage game. The reduced game payoff to player i under action profile a is defined to be i's continuation payoff when (i) players adopt the current continuation strategies and (ii) the current action profile is a. When players use one-period memory strategies, current actions fully specify the continuation strategies, so that the reduced game payoff to player i is represented as a simple function $u_i^t(a)$. In this case, the weakly belief-free equilibria can be characterized by the property that *players always play a correlated equilibrium of the reduced game* after any history. In general, players' continuation strategies depend on the past history as well as the current action. Let θ_i be a state variable to summarize player i's private history. In the general case, a reduced game payoff to player i is represented as $v_i^t(a|\theta_1, ..., \theta_N)$, and the weakly belief-free equilibria are characterized by the property that players always play a *Bayesian correlated equilibrium* of the reduced game.

The paper is organized as follows. Section 2 presents the basic model and defines weakly belief-free equilibria. Then, weakly belief-free equilibria are characterized by a recursive method, for the one-period memory case (Section 3) and for the general case (Section 4). Section 5 presents an example of a one-period memory weakly belief-free equilibrium which outperforms the belief-free equilibria. This example 'embeds' the chicken game as the reduced game in the repeated prisoners' dilemma. Section 6 provides a brief discussion about the advantage of weakly-belief equilibria over belief-free equilibria. Appendices A - D contain technical details of the example in Section 5.

2 The Model

Let us first define the stage game. Let A_i be the (finite) set of actions for player $i = 1, ..., N$ and define $A = A_1 \times \cdots \times A_N$. I primarily consider the case with imperfect private monitoring, where each player i observes her own action a_i and private signal $\omega_i \in \Omega_i$. (My formulation, however, accommodates the imperfect public monitoring case: see footnote 4.) We denote $\omega = (\omega_1, ..., \omega_N) \in \Omega = \Omega_1 \times \cdots \times \Omega_N$ and let $p(\omega|a)$ be the probability of private signal profile ω given action profile a (we assume that Ω is a finite set). It is also assumed that no player can infer which actions were taken (or not taken) for sure; that is, I suppose that given any $a \in A$, each $\omega_i \in \Omega_i$ occurs with positive probability.[4] We denote the marginal distribution of ω_i as $p_i(\omega_i|a)$. Player i's realized payoff is determined by her own action and signal, and denoted $\pi_i(a_i, \omega_i)$. Hence her *expected* payoff is given by

$$g_i(a) = \sum_{\omega \in \Omega} \pi_i(a_i, \omega_i) p(\omega|a).$$

This formulation ensures that the realized payoff π_i conveys no more information than a_i and ω_i do. The stage game is to be played repeatedly over

[4] I do not require that the joint distribution of the private signals has full support. Our assumption accomodates the case of *imperfect public monitoring*, where all players receive the same signal with probability one (hence the event where players receive different signals has zero probability).

an infinite time horizon $t = 1, 2, ...,$ and each player i's discounted payoff is given by $\sum_{t=1}^{\infty} g_i(a(t))\delta^{t-1}$, where $\delta \in (0, 1)$ is the discount factor and $a(t) \in A$ is the action profile at time t. A mixed action for player i is denoted by $\alpha_i \in \Delta(A_i)$, where $\Delta(A_i)$ is the set of probability distributions over A_i. With an abuse of notation, we denote the expected payoff and signal distribution under a mixed action profile $\alpha = (\alpha_1, ..., \alpha_N)$ by $g_i(\alpha)$ and $p(\omega|\alpha)$ respectively.

A *private history* for player i up to time t is the record of player i's past actions and signals, $h_i^t = (a_i(1), \omega_i(1), ..., a_i(t), \omega_i(t)) \in H_i^t \equiv (A_i \times \Omega_i)^t$. To determine the initial action of each player, we introduce a dummy initial history (or *null history*) h_i^0 and let H_i^0 be a singleton set $\{h_i^0\}$. A pure strategy s_i for player i is a function specifying an action after any history: formally, $s_i : H_i \to A_i$, where $H_i = \cup_{t \geq 0} H_i^t$. Similarly, a (behaviorally) mixed strategy for player i is denoted by $\sigma_i : H_i \to \Delta(A_i)$.

A *continuation strategy* for player i after private history h_i^t is denoted by $\sigma_i[h_i^t]$, defined as (i) $\sigma_i[h_i^t](h_i^0) = \sigma_i(h_i^t)$ and (ii) for any other history $h_i \neq h_i^0$, $\sigma_i[h_i^t](h_i) = \sigma_i(h_i^t h_i)$, where $h_i^t h_i$ represents a history obtained by attaching h_i after h_i^t. For any given strategy profile $\sigma = (\sigma_1, ..., \sigma_N)$ and any private history profile $h^t = (h_1^t, ..., h_N^t)$, let $BR(\sigma_{-i}[h_{-i}^t])$ be the set of best reply strategies for player i against $\sigma_{-i}[h_{-i}^t]$. EHO (2005) defined a *belief-free* strategy profile as follows.

Definition 1 *A strategy profile σ is* belief-free *if for any h^t and i, $\sigma_i[h_i^t] \in BR(\sigma_{-i}[h_{-i}^t])$.*

Note that the above requirement implies that the current continuation strategy for a player is a best reply for *any realization of* private histories of other players. In this sense, in a belief-free equilibrium players never need to compute beliefs over opponents' private histories. EHO (2005) showed that belief-free equilibria are tractable in the sense that a recursive method similar to that of Abreu, Pearce and Stacchetti (1990) can be employed to obtain a complete characterization of belief-free equilibrium payoffs.

In the present paper, I propose a weakening of the belief-free conditions, leading to a set of equilibria which are still tractable and manage to sustain a larger payoff set. Note that the belief-free conditions imply that, at the beginning of period $t + 1$, player i does not have to form beliefs over $h_{-i}^t = (a_{-i}(1), \omega_{-i}(1), ..., a_{-i}(t), \omega_{-i}(t))$. In contrast, I require that player i does not need to form beliefs over $(a_{-i}(1), \omega_{-i}(1), ..., a_{-i}(t))$, *omitting the last piece of information $\omega_{-i}(t)$ from the belief-free requirement.*

Now let us formalize the above idea. Fix any strategy profile σ and history profile $h^t = (a(1), \omega(1), ..., a(t), \omega(t))$. At the end of period t, player

i has observed his private history $h_i^t = (a_i(1), \omega_i(1), ..., a_i(t), \omega_i(t))$. Given this information, what would player i's belief over the opponents' continuation strategies be, *if he knew* the opponents' private histories *up to the actions in the previous period* $(a_{-i}(1), \omega_{-i}(1), ..., a_{-i}(t))$? This is given by the probability mixture of continuation strategy profiles of the opponents,

$$\sigma_{-i}[a_{-i}(1), \omega_{-i}(1), ..., a_{-i}(t), \omega'_{-i}(t)] \text{ for } \omega'_{-i}(t) \in \Omega_{-i},$$

each of which is chosen with conditional probability $p_{-i}(\omega'_{-i}(t)|a(t), \omega_i(t))$. Let us denote the probability distribution thus defined over the opponents' continuation strategies by $\overline{\sigma}_{-i}[a_{-i}(1), \omega_{-i}(1), ..., a_{-i}(t)|h_i^t]$.[5]

Definition 2 *A strategy profile* σ *is* weakly belief-free *if for any* $h^t = (a(1), \omega(1), ..., a(t), \omega(t))$ *and* i, $\sigma_i[h_i^t] \in BR(\overline{\sigma}_{-i}[a_{-i}(1), \omega_{-i}(1), ..., a_{-i}(t)|h_i^t])$.

This definition says that, under a weakly belief-free strategy profile, player i in period $t+1$ does not have to know the opponents' histories *up to the previous actions* $(a_{-i}(1), \omega_{-i}(1), ..., a_{-i}(t))$ to calculate his optimal continuation strategy. He may, however, need to form some beliefs over the previous signals $\omega_{-i}(t)$. More precisely, he may need to understand correctly that, for each possible action profile $a(t)$, the private signals in the previous period are distributed according to $p(\omega(t)|a(t))$. In the subsequent sections, I characterize the set of weakly belief-free equilibria.

3 One-Period Memory

In this section, we consider weakly belief-free equilibria with one-period memory[6]. This is a particularly tractable class which subsumes a major segment of the belief-free equilibria identified by Ely and Valimaki (2002) and EHO (2005) as a special case. We say that player i's strategy has *one-period memory* if it specifies the current (mixed) action $\alpha_i(t)$ to be taken depending only on $a_i(t-1)$ and $\omega_i(t-1)$.

[5] Formally, $\overline{\sigma}_{-i}[a_{-i}(1), \omega_{-i}(1), ..., a_{-i}(t)|h_i^t] = \sum_{\omega'_{-i}(t)} \sigma_{-i}[a_{-i}(1), \omega_{-i}(1), ..., a_{-i}(t), \omega'_{-i}(t)] p_{-i}(\omega'_{-i}(t)|a(t), \omega_i(t))$. Note that this depends on h_i^t only through $(a_i(t), \omega_i(t))$. Note also that this is a *correlated strategy profile* of the opponents, when $N \geq 3$. Below we assume that the best reply correspondence BR_i is defined over the domain of correlated strategy profiles of the opponents.

[6] Deviations to general strategies (not neccessarily with one-period memory) are allowed, so that we are *not* weakening the usual equilibrium conditions.

Definition 3 *A* one-period memory strategy *for player i is defined by an initial (mixed) action $\alpha_i(1)$ and transition rules $m_i^t : A_i \times \Omega_i \to \Delta(A_i)$, $t = 1, 2, \ldots$ The probability of $a_i(t+1)$ given $a_i(t)$ and $\omega_i(t)$ under m_i^t is denoted by*

$$m_i^t(a_i(t+1)|a_i(t), \omega_i(t)).$$

The set of all one-period memory transition rules for player i is denoted by M_i.

Under a one-period memory strategy profile, at each moment t, the current action profile $a(t)$ determines the continuation play (independent of previous history)[7]. Hence, we can define $u_i^t(a(t))$ as the (average) expected continuation payoff to player i. The function $u_i^t(a(t))$ can be regarded as a payoff in a game which has the same action sets as the stage game. Let us call the game defined by $(u_i^t,\ A_i)_{i=1,\ldots,N}$ a **reduced game**. This enables us to view a repeated game as *a sequence of reduced games*, and I will analyze its recursive structure. This is an important departure from the previous literature (Abreu, Pearce and Stacchetti (1990) and EHO (2005)) which views a repeated game as a sequence of continuation *payoff sets* and exploits its recursive structure.

Before stating my characterization, it is necessary to define a couple of concepts. In what follows, we will analyze a reduced game played with a (partial) correlation device $\omega \in \Omega$. First, a realization of the correlation device, $\omega = (\omega_1, \ldots, \omega_N)$, is generated according to a certain probability distribution, after which each player i observes ω_i. Depending on ω_i, player i chooses an action a_i (possibly by using a mixing device). This process generates a joint distribution over (a, ω), which will be denoted by $q(a, \omega)$. I would like to define the situation where players take mutual best replies, or the situation where $q(a, \omega)$ is a *correlated equilibrium* of reduced game u. In contrast to the standard definition of correlated equilibrium, which only considers a joint distribution over *actions* (interpreted as "recommendations"), the definition here considers the situation where each player receives recommended actions *and some additional information*, ω_i.

Formally, I say that a probability distribution q on $A \times \Omega$ is a *correlated equilibrium* of game $u : A \to \mathbf{R}^N$, when

$$\forall i \forall a_i \forall \omega_i \forall a_i' \sum_{a_{-i}, \omega_{-i}} u_i(a) q(a, \omega) \geq \sum_{a_{-i}, \omega_{-i}} u_i(a_i', a_{-i}) q(a, \omega). \qquad (1)$$

[7] Continuation strategies after $t+1$ are determined, under a one-period memory strategy profile, by $a(t)$ and $\omega(t)$. As the latter is generated by $p(\omega(t)|a(t))$, $a(t)$ alone determines the contingent action plans of players after $t+1$.

This admits the same interpretation as the standard condition for correlated equilibrium with respect to actions. Consider the marginal distribution of (a_i, ω_i), given by $q_i(a_i, \omega_i) \equiv \sum_{a_{-i},\omega_{-i}} q(a, \omega)$. If (a_i, ω_i) does not arise with a positive probability (i.e., $q_i(a_i, \omega_i) = 0$), then the above inequality is automatically satisfied (as the both sides are equal to 0). Otherwise dividing both sides of inequality (1) by $q_i(a_i, \omega_i)$ reduces it to

$$\mathrm{E}\left[u_i(a_i, \widetilde{a}_{-i})|a_i, \omega_i\right] \geq \mathrm{E}\left[u_i(a_i', \widetilde{a}_{-i})|a_i, \omega_i\right],$$

where the expectation is taken with respect to the conditional probability $\Pr(a_{-i}|a_i, \omega_i) = \sum_{\omega_{-i}} q(a, \omega)/q_i(a_i, \omega_i)$. In our situation where each player receives a recommended action *and some additional information* ω_i, condition (1) thus ensures that player i has an incentive to follow the recommended action a_i. The set of correlated equilibria of game u is denoted by $C(u)$:

$$C(u) \equiv \{q \in \Delta(A \times \Omega)|\text{ Condition (1) holds.}\}, \tag{2}$$

where $\Delta(A\times\Omega)$ is the set of probability distribution over $A\times\Omega$. A standard result for the set of correlated equilibria carries over to our formulation: From (1), we can see that $C(u)$ is convex. As it plays a vital role in what follows, I state it here.

Lemma 4 *For any $u : A \to \mathbf{R}^N$, $C(u)$ is convex.*

Now consider one-period memory transition rules $m = (m_1, ..., m_N) \in M = M_1 \times \cdots \times M_N$. Given such a profile and our monitoring structure $p(\omega|a)$, the probability of $(a(t+1), \omega(t))$ given $a(t)$ is determined as follows. This will play the role of the correlation device for the reduced game at time t.

Definition 5 *The action-signal distribution given $a(t)$ under one-period memory strategy profile m is defined by*

$$q^m(a(t+1), \omega(t)|a(t)) \equiv \prod_{i=1}^{N} m_i(a_i(t+1)|a_i(t), \omega_i(t))p(\omega(t)|a(t)). \tag{3}$$

Its marginal distribution of $a(t+1)$ is the law of motion *under m and defined by*

$$p^m(a(t+1)|a(t)) \equiv \sum_{\omega(t)\in\Omega} q^m(a(t+1), \omega(t)|a(t)). \tag{4}$$

Now I am ready to introduce my equilibrium concept.

Definition 6 *A set of reduced games $U \subset \{u | u : A \to \mathbf{R}^N\}$ is* ***self-generating*** *if, for any $u \in U$, there exist $v \in U$ and a one-period memory transition rule profile $m \in M$ such that*

$$\forall a \;\; u(a) = (1-\delta)g(a) + \delta \sum_{a' \in A} v(a')p^m(a'|a) \tag{5}$$

and

$$\forall a \;\; q^m(\cdot, \cdot|a) \in C(v), \tag{6}$$

where $C(u)$, q^m and p^m are defined by (2), (3), and (4).

Note that condition (6) is the key requirement that players are always playing a correlated equilibrium of the reduced game, *on and off the path of play*. We claim that the equilibrium payoffs associated with a self-generating set can be achieved by equilibria of the repeated game. Let $N(u)$ be the Nash equilibrium payoff set associated with game u. Then, one obtains the following complete characterization of one-period memory belief-free equilibria, which is similar to Abreu, Pearce and Stacchetti (1990) (note, however, that the present recursive characterization is given in terms of *reduced games*, in contrast to continuation *payoff sets* in APS).

Theorem 7 *Let $U \subset \{u | u : A \to \mathbf{R}^N\}$ be self-generating and bounded in the sense that there exists $K > 0$ such that $|u_i(a)| < K$ for all i, $u \in U$, and a. Then, any point in*

$$N(U) \equiv \bigcup_{u \in U} N(u)$$

can be achieved as the average payoff of a one-period memory weakly belief-free *sequential equilibrium. The set of all* one-period memory weakly belief-free *sequential equilibrium payoff profiles is given by $N(U^*)$, where U^* is the largest (in the sense of set inclusion) bounded self-generating set.*

Note that a one-period memory weakly belief-free sequential equilibrium has the following features; (i)for each player i, the distribution of $a_i(t)$ depends only on $\omega_i(t-1)$ and $a_i(t-1)$ in each stage $t > 1$ and (ii) players always play a correlated equilibrium of the repeated game after any history (on and off the path of play). Also note that, if a (partial) correlation device is available at the beginning of the repeated game, the set of one-period memory belief-free sequential equilibrium is given by $C(U) \equiv \bigcup_{u \in U} C(u)$ (i.e., the correlated equilibria associated with reduced games $u \in U$).

Proof. For any $u \in U$, repeated application of (5) induces a sequence of reduced games $\{u^t\}$ and one-period memory strategies $\{m^t\}$ that satisfy

$$\forall a \;\; u^t(a) = (1-\delta)g(a) + \delta \sum_{a' \in A} u^{t+1}(a')p^{m^t}(a'|a),$$

and

$$\forall a \;\; q^{m^t}(\cdot,\cdot|a) \in C(u^{t+1}), \tag{7}$$

for $t = 1, 2, ...$ with $u^1 = u$. Hence, for any $T(>2)$, we have

$$u(a) = (1-\delta)\left\{ g(a) + E\left[\sum_{t=2}^{T-1} g(a(t))\delta^{t-1} + u^T(a(t+1))\delta^{T-1} \middle| a \right] \right\}.$$

The expectation $E[\cdot|a]$ presumes that the distribution of $a(t+1)$ given $a(t)$ is $p^{m^t}(a(t+1)|a(t))$ with $a(1) = a$. As u^T is bounded, we can take the limit $T \to \infty$ to get

$$u(a) = (1-\delta)\left\{ g(a) + E\left[\sum_{t=2}^{\infty} g(a(t))\delta^{t-1} \middle| a \right] \right\}. \tag{8}$$

Hence $u(a)$ can be interpreted as the average payoff profile when the players choose a today and follow one-period memory strategy profile m^t, $t = 1, 2, ...$ Let α be a (possibly mixed) Nash equilibrium of game u, and let σ be the strategy where α is played in the first period and the players follow m^t, $t = 1, 2, ...$ By construction σ achieves an average payoff of $u(\alpha)$ (the expected payoff associated with α), and we show below that it is a sequential equilibrium because after any history no player can gain from a one-shot unilateral deviation[8]. In the first period, no one can gain by a one-shot unilateral deviation from α because it is a Nash equilibrium of game u. For stage $t > 0$, take any player i and any private history for her $(a_i^0(1), , , , a_i^0(t-1), \omega_i^0(1), ..., \omega_i^0(t-1))$. Let $\mu(a(t-1))$ be her belief about last period's action profile given her private history. Then, her belief about the current signal distribution is

$$q(a(t), \omega(t)) = \sum_{a(t-1) \in A} q^{m^{t-1}}(a(t), \omega(t)|a(t-1))\mu(a(t-1)).$$

(Note that under σ other players' continuation strategies do not depend on their private histories *except for* their current signals.) Let $v \in U$ be the

[8] The standard dynamic programming result shows that this implies that no (possibly infinite) sequence of unilateral deviations is profitable.

continuation payoff in stage t (including stage t's payoff). Then, condition (7) for self-generation, $q^m(\cdot,\cdot|a) \in C(v)$ for all a, and the convexity of the correlated equilibrium set $C(v)$ implies $q \in C(v)$. This means that player i cannot gain by one-shot unilateral deviation at this stage.

Conversely, given any one-period memory weakly belief-free sequential equilibrium, one can calculate a sequence of reduced games u^t, $t = 1, 2, ...$ It is straightforward to check that $U' \equiv \{u^t | t = 1, 2, ...\}$ is a self-generating set bounded by $K \equiv \max_{i.a} |g_i(a)|$. Since a union of self-generating sets bounded by K is also self-generating and bounded by K, we conclude that the set of all one-period memory weakly belief-free sequential equilibrium payoff profiles is given by $N(U^K)$, where U^K is the largest (in the sense of set inclusion) self-generating set bounded by K. Now consider any self-generating set U which is bounded (not necessarily by K). The first part of this proof shows that U is actually bounded by K (as any $u \in U$ is an average payoff profile of the repeated game). This implies $U^K = U^*$, which completes the proof. ■

4 General Strategies

In this section, I consider weakly belief-free equilibria in fully general strategies. For the purpose of this section, it is convenient to represent a strategy in the following way. For each player i, we specify

- a set of states Θ_i
- an initial state $\theta_i(1) \in \Theta_i$
- (mixed) action choice for each state, $\rho_i : \Theta_i \to \Delta(A_i)$
- state transition $\tau_i : \Theta_i \times A_i \times \Omega_i \to \Delta(\Theta_i)$. This determines the probability distribution of the next state $\theta_i(t+1)$ based on the current state $\theta_i(t)$, current action $a_i(t)$, and current private signal $\omega_i(t)$.

I call $ms_i \equiv (\Theta_i, \theta_i(1), \rho_i, \tau_i)$ a *machine strategy*. All strategies can trivially be represented as a machine strategy, when we set Θ_i equal to the set of all histories for player i: $\Theta_i = H_i$.[9] The action choice and transition rule are assumed to be time-independent, but this is without loss

[9]In this case, the transition rule τ_i is deterministic: given $\theta_i(t) = h_i(t)$, $a_i(t)$, and $\omega_i(t)$, τ_i assigns probability one to $\theta_i(t+1) = (\theta_i(t), a_i(t), \omega_i(t))$. The initial state should be the null history $\theta_i(t) = h_i^0$.

of generality. We can always include the current time in the state variable θ_i (as $\theta_i = (\widehat{\theta}_i, t)$).

Under a machine strategy profile $ms = (ms_i, ..., ms_N)$, if we fix a current state profile $\theta(t) \in \Theta \equiv \Theta_1 \times \cdots \times \Theta_N$, continuation strategies are fully specified. For each ms, we can compute the continuation payoff to player i, when (i) all players' continuation strategies are specified by ms given $\theta(t)$ and (ii) the current action profile is $a(t)$. Denote this by $v_i(a(t)|\theta(t))$. Note well that the function v_i is *defined over all* $a(t)$, some of which may be outside of the support of the current mixed action specified by ms under $\theta(t)$. This makes v_i a useful tool to check the profitability of one-shot deviations from the given strategy profile ms. We call the function $v : A \times \Theta \to \Re^N$ an **ex-post reduced game**.

If a machine strategy profile ms is a weakly belief-free equilibrium, in each period t, players are taking mutual best replies *for each* $(\theta(t-1), a(t-1))$.[10] Given $(\theta(t-1), a(t-1))$, the machine strategy profile under consideration provides some joint distribution of $\omega(t-1)$, $\theta(t)$, and $a(t)$, denoted by $r(\omega(t-1), \theta(t), a(t))$. Given a realization of $\omega_i(t-1)$, $\theta_i(t)$, and $a_i(t)$ (interpreted as a recommended action) of this distribution r, player i must be happy to choose $a_i(t)$. This can be regarded as a *correlated equilibrium of a Bayesian game*, where types θ, recommended actions a, and some additional information ω are generated by a *joint* distribution $r(\omega, \theta, a)$, and the ex-post payoff function is given by the ex-post reduced game $v_i(a|\theta)$.

Definition 8 *Probability distribution r over $\Omega \times \Theta \times A$ is a* ***Bayesian correlated equilibrium*** *of ex-post reduced game v when*

$$\forall i \forall a_i \forall \omega_i \forall \theta_i \forall a'_i \quad \sum_{a_{-i}, \omega_{-i}, \theta_{-i}} v_i(a|\theta) r(\omega, \theta, a) \geq \sum_{a_{-i}, \omega_{-i}, \theta_{-i}} v_i(a'_i, a_{-i}|\theta) r(\omega, \theta, a). \tag{9}$$

The set of Bayesian correlated equilibria of the ex-post reduced game v is denoted by

$$BC(v) = \{r \in \Delta(\Omega \times \Theta \times A) | \textit{Condition (9) holds.}\} \tag{10}$$

The defining condition (9) shows that BC is a convex set, which plays an important role in what follows. The following notation clarifies how $r(\omega(t-1), \theta(t), a(t))$ is determined in the repeated game.

[10] This is because, under a machine strategy profile, previous history $(a(1), \omega(1), ..., a(t-2), \omega(t-2))$ affects the continuation strateies at time t only when it affects $\theta(t-1)$.

Definition 9 *The* state-action-signal distribution *given* $\theta(t-1)$*,* $a(t-1)$ *under machine strategy profile ms is defined by*

$$q^{ms}(\omega(t-1), \theta(t), a(t)|\theta(t-1), a(t-1)) =$$

$$\sum_{\omega(t-1)} \prod_{i=1}^{N} \rho_i(a_i(t)|\theta_i(t)) \tau_i(\theta_i(t)|\theta_i(t-1), a_i(t-1), \omega_i(t-1)) p(\omega(t-1)|a(t-1)), \tag{11}$$

where ρ_i *and* τ_i *are the action choice and state transition rule of* ms_i*. Its marginal distribution of* $(\theta(t), a(t))$ *is the* law of motion *under ms and is defined by*

$$p^{ms}(\theta(t), a(t)|\theta(t-1), a(t-1)) \equiv \sum_{\omega(t-1)\in\Omega} q^{ms}(\omega(t-1), \theta(t), a(t)|\theta(t-1), a(t-1)). \tag{12}$$

Now I am ready to state my main characterization conditions.

Definition 10 *An ex-post reduced game* $v_i(a|\theta)$*,* $i = 1, ..., N$ *is* ***self-generating*** *if there exists a machine strategy profile ms (defined over states* $\theta \in \Theta$*) such that*

$$\forall i \forall a \forall \theta \quad v_i(a|\theta) = (1-\delta) g_i(a) + \delta \sum_{a'\in A} v_i(a'|\theta') p^{ms}(\theta', a'|\theta, a) \tag{13}$$

and

$$\forall a \forall \theta \quad q^{ms}(\cdot, \cdot, \cdot|\theta, a) \in BC(v), \tag{14}$$

where q^{ms}*,* p^{ms}*, and* $BC(v)$ *are defined by (11), (12), and (10).*

In contrast to the formulation in Section 3, where we considered a set U of reduced games, here we consider a single function profile v. In Section 3, we needed to consider a set of reduced games to allow the possibility that the one-period memory transition rule is time-dependent (hence a set of reduced games $\{u^t|t = 1, 2, ...\}$ is associated with an equilibrium). Here, we can confine our attention to a single function profile v, because state θ can encode time (as $\theta = (\widehat{\theta}, t)$) and a single function profile $v(\cdot|\theta)$ can represent potentially time-dependent ex-post reduced games.

Given an ex-post reduced game $v = v(a|\theta)$, let $N(v)$ be the set of Nash equilibrium payoff profiles of game $g(a) = v(a|\theta)$ for some θ. Suppose that v is self-generating and $w \in N(v)$ is obtained as a Nash equilibrium of game $g(a) = v(a|\theta)$. Then, w is obtained as a machine strategy equilibrium where the initial state is θ. Formally, we obtain the following characterization result.

Theorem 11 *Let v be a self-generating ex-post reduced game, which is bounded in the sense that there exists $K > 0$ such that $|v_i(a|\theta)| < K$ for all i, a, and θ. Then any $w \in N(v)$ is a weakly belief-free equilibrium payoff profile. Conversely, any weakly belief-free equilibrium payoff profile is an element of $N(v)$, for some bounded self-generating ex-post reduced game v.*

The proof is basically the same as in Section 3. Condition (14) says that players are always playing a Bayesian correlated equilibrium of the ex-post reduced game *after any history (on and off the path of play)*, and it implies that one-shot deviations from the machine strategy profile do not pay. Hence, the standard dynamic programming argument shows that players are always choosing mutual best replies.

Remark 12 *Given a weakly belief-free equilibrium machine strategy profile, we can calculate the associated ex-post reduced game $v(a|\theta)$. The original (pure or mixed) equilibrium payoff in the repeated game is given by a Nash equilibrium of $v(a|\theta(1))$, where $\theta(1)$ is the initial state profile of the given machine strategies. However, the weakly belief-free requirement implies that any Nash equilibrium payoff profile of $v(a|\theta)$ for* any *θ (not necessarily the initial one) is also an equilibrium payoff profile of the repeated game. This comes from the following fact. Consider the strategy profile defined by (i) the initial action profile is an equilibrium of game $g(a) = v(a|\theta)$ and (ii) the continuation play is given by the machine strategy profile. As the machine strategy profile is weakly belief-free, the strategy profile thus constructed satisfies the property that one-shot deviations are never profitable (hence it is an equilibrium).*

Remark 13 *Theorem 11 can be extended to the case where there is a correlation device at the beginning of the repeated game. It is also straightforward to incorporate a public randomization device at each moment in time.*

Example: The belief-free (hence by definition weakly belief-free) equilibrium of Kandori and Obara (2006) (a private strategy equilibrium in a imperfect *public* monitoring game) is an example of Theorem 11. Unlike the belief-free equilibrium in Ely and Valimaki (2002), this equilibrium does not have the one-period memory property. Both Kandori and Obara (2006) and Ely and Valimaki (2002) can be represented by a machine strategy profile with two states $\Theta_i = \{P, R\}$. In Ely and Valimaki (2002), distinct pure actions are played in different states (coupled with a stochastic state transition function, which effectively implements a mixed action in each period).

Hence, the previous action can be identified with the previous state[11], and it completely determines (along with the private signal in the previous period) the current action. This shows that Ely and Valimaki has a one-period memory strategy. In contrast, Kandori and Obara, probability mixtures over the same set of actions are played in all states. Hence, it is not possible to identify the current state with the current action. The same action-signal pair $(a_i(t), \omega_i(t))$ would lead to a different mixed action in $t+1$, depending on the current state $\theta_i(t)$. Hence Kandori and Obara's model does not employ one-period memory strategies, but it has a belief-free equilibrium so that players are always playing an *ex-post equilibrium* (a special case of Bayesian correlated equilibrium) of $v(a|\theta)$.

5 An Example: The Chicken Game in the Repeated Prisoners' Dilemma

In this section, I present a simple example of a one-period memory belief-free equilibrium, where the set U in our characterization (Definition 6) is a singleton. This example shows that a weakly belief-free equilibrium can have the following desirable properties: (i) it can be in very simple pure strategies, (ii) players always have a strict incentive to follow the equilibrium action, and (iii) it can outperform the equilibria identified by previous work. The equilibrium in this example also has an interesting property that it "embeds" the chicken game (as the reduced game) in a repeated prisoner's dilemma game. The stage game has the following prisoner's dilemma structure:

	C	D
C	$1,1$	$-1/6, 3/2$
D	$3/2, -1/6$	$0,0$

.

For computational purposes, I have normalized the payoffs in such a way that the maximum and minimum payoffs are 1 and 0 respectively, but it

[11] Ely and Valimaki consider a prisoners' dilemma game, and actions C (resp. D) is played in state R (resp. P). When a player deviates, say in state R, to play D, we can specify that the continuation strategy is given by the one where current state is P. (Hence, the player behaves as if the current state were P, after deviating to D) By the belief-free conditions, players are taking mutual best replies after such a deviation (guaranteeing that the specified strategy profile constitutes a sequential equilibrium). Hence in Ely and Valimaki we can assume that players are following one-period memory strategies not only on the path of play, but also after deviations.

may be easier to consider

	C	D
C	$6,6$	$-1,9$
D	$9,-1$	$0,0$

,

which is proportional to the first payoff table. Each player's private signal has binary outcomes, $\omega_i = G, B, i = 1, 2$. The signal profile distribution depends on the current action profile and it is denoted by $p(\omega_1, \omega_2 | a_1, a_2)$. The relationship between current action and signal profiles (the monitoring structure) is as follows:

$(C, C) \Longrightarrow$

$\omega_1 \diagdown \omega_2$	G	B
G	$1/3$	$1/3$
B	$1/3$	0

$(D, C) \Longrightarrow$

$\omega_1 \diagdown \omega_2$	G	B
G	$1/8$	$1/2$
B	$1/4$	$1/8$

$(C, D) \Longrightarrow$

$\omega_1 \diagdown \omega_2$	G	B
G	$1/8$	$1/4$
B	$1/2$	$1/8$

$(D, D) \Longrightarrow$

$\omega_1 \diagdown \omega_2$	G	B
G	0	$2/5$
B	$2/5$	$1/5$

This set of distributions admits the following natural interpretation. When both players cooperate, they can avoid a mutually bad outcome (B, B). If one player defects, with a high probability (1/2), the defecting player enjoys a good outcome (G) while the other player receives a bad one (B). Finally, when both player defect, they cannot achieve a mutually good outcome (G, G).

I have made some entries in the above tables equal to 0 (so that the example has "moving supports") to simplify the analysis, but, as I will formally show in what follows, this is inessential to the main results (i.e., similar results are obtained even though I make those entries non-zero, small numbers).

Let us consider the following simple (and intuitive) one-period memory transition rule:

$$a_i(t) = \begin{cases} C & \text{if } \omega_i(t-1) = G \\ D & \text{if } \omega_i(t-1) = B \end{cases} . \tag{15}$$

The *reduced game* payoff for profile a, denoted $u_i(a)$, is defined to be the average payoff when a is played in the initial period and then players follow the above strategy. Since the transition rule is time-independent, the reduced game is also time-independent. Let us denote the reduced game payoffs by

	C	D
C	x, x	α, β
D	β, α	y, y

.

For example, $(u_1(C, D), u_2(C, D)) = (\alpha, \beta)$. Since the same reduced game u is played in each period, the dynamic programming value equation in the self-generation condition (5) reduces to a simple system of equations

$$\forall i \forall a \quad u_i(a) = (1-\delta)g(a) + \delta \sum_{a' \in A} u_i(a')p^m(a'|a),$$

where $p^m(a'|a)$ denotes the transition probability of current and subsequent actions under our strategy (15). By symmetry, this further reduces to a system of four equations with four unknowns:

$$\begin{cases} x = (1-\delta) + \delta\frac{1}{3}(x + \alpha + \beta) \\ y = \delta(\frac{1}{5}y + \frac{2}{5}(\alpha + \beta)) \\ \alpha = (1-\delta)(-\frac{1}{6}) + \delta(\frac{1}{8}x + \frac{1}{4}\alpha + \frac{1}{2}\beta + \frac{1}{8}y) \\ \beta = (1-\delta)\frac{3}{2} + \delta(\frac{1}{8}x + \frac{1}{2}\alpha + \frac{1}{4}\beta + \frac{1}{8}y) \end{cases}.$$

For example, $x = u_1(C, C)$ is associated with current payoff $1 = g_1(C, C)$, and given the current action profile (C, C) and the transition rule (15), the continuation payoff is $x = u_1(C, C)$, $\alpha = u_1(C, D)$, or $\beta = u_1(D, C)$ with probability $1/3$. Hence we have the first equality $x = (1-\delta) \times 1 + \delta\frac{1}{3}(x+\alpha+\beta)$. The rest admit similar interpretations. When $\delta = 0.99$, for example, we have the following solutions:

$$\begin{cases} x = 0.641\,26 \\ y = 0.627\,89 \\ \alpha = 0.629\,14 \\ \beta = 0.642\,5. \end{cases} \tag{16}$$

Note first that we have $u_1(D, C) = \beta > x = u_1(C, C)$ and $u_1(C, D) = \alpha > y = u_1(D, D)$, which means that the reduced game is a "Chicken Game", where (D, C) and (C, D) are Nash equlibria. The high discount factor is responsible for the fact that these four payoffs are close to each other. This is because the transition rule (15) defines an irreducible and

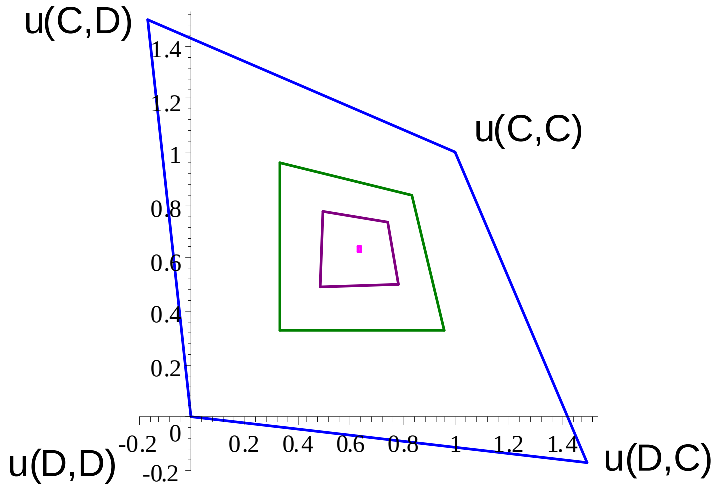


Figure 1: Reduced Games: From outer to inner, $\delta = 0$, $\delta = 4/7$, $\delta = 4/5$, $\delta = 0.99$.

aperiodic Markov chain over the stage game action profiles (C, C), (C, D), (D, C), and (D, D), and the average payoff to player i given any initial action profile (i.e., x, y, α, or β) tends to, as $\delta \to 1$,

$$g^* = \sum_a g_i(a)\mu^*(a), \tag{17}$$

where μ^* is the unique (ergodic) stationary distribution of the Markov chain.

Note that the reduced game coincides with the Prisoner's Dilemma game (i.e., the stage game) when $\delta = 0$. As we have seen, all four payoff profiles of the reduced game tend to the single point (g^*, g^*) given by (17), as $\delta \to 1$. Numerical computation shows that, when $\delta > 4/7$, the reduced game becomes a chicken game ($\alpha > y$, and $\beta > x$) (See Figure 1).

I will show that, when the discount factor is sufficiently large (specifically, $\delta \geq 0.989\,54$), the action profile distribution after any history becomes a correlated equilibrium of the reduced game (hence the one-period memory transition rule (15), coupled with a suitable initial action profile, is a weakly belief-free equilibrium).

Now let us check the incentive constraints. Namely, we will examine when the second requirement for self-generation (6) is satisfied. This is the requirement that players always play a correlated equilibrium of the reduced game after any history. Under our one-period memory strategy (15), the joint distribution of current actions depends on the previous action profile in the following way:

$a(t-1)=(C,C) \Longrightarrow$

$a_1(t)\diagdown a_2(t)$	C	D
C	1/3	1/3
D	1/3	0

$a(t-1)=(D,C) \Longrightarrow$

$a_1(t)\diagdown a_2(t)$	C	D
C	1/8	1/2
D	1/4	1/8

$a(t-1)=(C,D) \Longrightarrow$

$a_1(t)\diagdown a_2(t)$	C	D
C	1/8	1/4
D	1/2	1/8

$a(t-1)=(D,D) \Longrightarrow$

$a_1(t)\diagdown a_2(t)$	C	D
C	0	2/5
D	2/5	1/5

I will show that, when the discount factor is sufficiently high, all these joint distributions are correlated equilibria of the reduced (Chicken) game u.

Before going into the details, let me provide some intuition about how the equilibrium in this example works. The above tables show that, once (C,D) is played, (D,C) follows with a large probability, thereby punishing player 2 who initially played D. Note that player 1, who was cheated initially, benefits from the transition. Hence the equilibrium strategy here provides incentives by the *transfer of continuation payoffs* (taking away some continuation payoff from the deviator and giving it to the victim). As I will elaborate on later, this is an essential mechanism to achieve efficiency in repeated games with imperfect monitoring (Fudenberg, Levine and Maskin (1994)). After (D,C) is played, action profile largely goes back and forth between (C,D) and (D,C). As the discount factor increases, this provides a large impact on the average payoffs, and the reduced game payoff set, which is a prisoners' dilemma game payoff set when $\delta = 0$, is "compressed" in the northwest-southeast directions and the reduced game becomes a Chicken game (see Figure 1) for large δ. Since (C, D) and (D, C) are Nash equilibria of the Chicken game, a joint distribution of actions which places relatively large probabilities to these profiles can be a correlated equilibrium. Our strategy, summarized in the tables, above indeed has this property. Hence it can be a weakly-belief free equilibrium.

Let us now examine the incentive constraint (6) in detail. Note that the reduced game is strategically equivalent to

$1\backslash 2$	C	D
C	$0, 0$	$\alpha - y, \beta - x$
D	$\beta - x, \alpha - y$	$0, 0$

(18)

(In general, a game $g_i(a)$, $i = 1, ..., N$ is strategically equivalent to game $g_i(a) + K_i(a_{-i})$, which means that both games have the same best response correspondences and hence the same (Nash or correlated) equilibria.) For the reduced game to be a chicken game, we need to have

$$x < \beta \text{ and } y < \alpha. \tag{19}$$

The game (18) is in turn strategically equivalent to (just multiply the payoffs by $\frac{1}{\beta - x}$)

$1\backslash 2$	C	D
C	$0, 0$	$z, 1$
D	$1, z$	$0, 0$

, (20)

where

$$z \equiv \frac{\alpha - y}{\beta - x}.$$

Hence the correlated equilibria are completely characterized by this single quantity z. Recall that, under transition rule (15), the strategy profile distribution is given by

$1\backslash 2$	C	D
C	$p(G, G\|a)$	$p(G, B\|a)$
D	$p(B, G\|a)$	$p(B, B\|a)$

, (21)

where a is the action profile *in the previous period*. I identify the condition under which this is a correlated equilibrium of (20) (and therefore of the original reduced game u) for *all* a.

In the general model in Section 3, I defined correlated equilibrium with respect to joint distributions over (a, ω) (see (1)). In the equilibrium considered here, there is a one-to-one correspondence between $a(= a(t))$ and

$\omega(=\omega(t-1))$ (see (15)), and as a result I can apply the standard definition of correlated equilibrium with respect to distributions of a alone. Generally speaking, a distribution over action profile $q(a)$ is a correlated equilibrium of the game v, if[12]

$$\forall i \forall a_i \forall a_i' \sum_{a_{-i}} v_i(a) q(a) \geq \sum_{a_{-i}} v_i(a_i', a_{-i}) q(a).$$

Now I apply this condition for player $i=1$, recommended action $a_i = C$, and a possible deviation $a_i' = D$, where v is equal to the transformed reduced game (20). This condition is expressed as

$$v_i(C,C)q(C,C) + v_i(C,D)q(C,D) \geq v_i(D,C)q(C,C) + v_i(D,D)q(C,D),$$

or

$$0 \cdot q(C,C) + z \cdot q(C,D) \geq 1 \cdot q(C,C) + 0 \cdot q(C,D).$$

When action distribution $q(a)$ is given by table (21), this reduces to

$$p(G,B|a)z \geq p(G,G|a). \tag{22}$$

Similarly, we have the following incentive constraints for the proposed strategy profile distribution to be a correlated equilibrium.

$$\text{Condition for player 1 to choose } D\text{:} \quad p(B,G|a) \geq p(B,B|a)z \tag{23}$$

$$\text{Condition for player 2 to choose } C\text{:} \quad p(B,G|a)z \geq p(G,G|a) \tag{24}$$

$$\text{Condition for player 2 to choose } D\text{:} \quad p(G,B|a) \geq p(B,B|a)z \tag{25}$$

As we have $p(G,B|a) \neq 0$ and $p(B,G|a) \neq 0$, the correlated equilibrium conditions (22) - (25) reduce to

$$\min_a \min \left\{ \frac{p(G,B|a)}{p(B,B|a)}, \frac{p(B,G|a)}{p(B,B|a)} \right\} \geq z \geq \max_a \max \left\{ \frac{p(G,G|a)}{p(G,B|a)}, \frac{p(G,G|a)}{p(B,G|a)} \right\} \tag{26}$$

with the understanding that $\frac{p(G,B|a)}{p(B,B|a)}, \frac{p(B,G|a)}{p(B,B|a)} = \infty > z$ when $p(B,B|a) = 0$. Note that the minimum on the left hand side is attained by $a = (D,C)$, (C,D), and (D,D), and it is equal to $\frac{1/4}{1/8} = \frac{2/5}{1/5} = 2$. The maximum on the

[12] A standard argument shows that this is equivalent to

$$\mathrm{E}\left[u_i(a_i, \widetilde{a}_{-i})|a_i\right] \geq \mathrm{E}\left[u_i(a_i', \widetilde{a}_{-i})|a_i\right].$$

See the explanation after condition (1).

right hand side is attained by $a = (C, C)$, and it is equal to $\frac{1/3}{1/3} = 1$. Hence, the current action profile distribution is always a correlated equilibrium (given any action profile in the previous period) of the reduced game iff

$$2 \geq z \geq 1, \tag{27}$$

and the incentive constraints (22) - (25) are satisfied with strict inequality when $2 > z > 1$. If $\delta = 0.99$, this is indeed satisfied, because we have $z = \frac{\alpha - y}{\beta - x} = 1.0027$. Numerical computation shows that the crucial equilibrium condition (27) is satisfied when $\delta \geq 0.98954$.

Note that the incentive constraint (26) is satisfied with strict inequalities when $\delta = 0.99$. This relationship is unchanged if I slightly modify the signal structure so that $p(\omega_1, \omega_2|a)$ is always strictly positive (because $z = \frac{\alpha - y}{\beta - x}$ is a continuous function of those parameters). In this sense the example here is robust: All major conclusions hold if I make $p(\omega_1, \omega_2|a)$ strictly positive for all (ω_1, ω_2) and all a. More generally, all major results here continue to hold even if we slightly perturb the stage game payoffs, the signal distributions, or the discount factor.

Since the incentive constraints in (27) are satisfied with strict inequalities, each player is always taking a *strict* best reply action (given the future strategy profile). This is in sharp contrast to the equilibria obtained by Ely and Valimaki (2002) or EHO (2005), whose essential feature is that at least one player is indifferent between some actions. I would argue that the weakly belief-free equilibrium identified here has certain advantages over the belief-free equilibria. First, thanks to the strict incentives, the same equilibrium strategy specified here (15) works for all near-by games. In contrast, the mixing probability in a belief-free equilibrium is fine-tuned to the structure of the game. If the payoff, discount factor, or monitoring structure changes, the belief-free equilibrium strategy changes.

Secondly, Bhaskar (2000) argues that belief-free equilibria are unrealistic because they may not be justified by the Harsanyi-type purification argument (with independent perturbations to the stage payoffs). A follow-up paper by Bhaskar, Mailath and Morris (2008) partially confirms this conjecture. They consider one-period memory belief-free strategies *a la* Ely-Valimaki in a *perfect monitoring* repeated *prisoners' dilemma* game (note that the Ely-Valimaki belief-free equilibrium applies to perfect as well as imperfect private monitoring). They show that those strategies cannot be purified by one-period memory strategies, but can be purified by infinite memory strategies. They conjecture that purification fails for any finite memory strategy (so that the purification is possible, but only with substantially more complex strategies). They also conjecture that similar results hold for the imperfect private monitoring case. The equilibrium here is free from the Bhaskar critique.

Finally, let me provide some welfare comparison between the weakly belief-free equilibrium and belief-free equilibria in this example. The best symmetric correlated equilibrium payoff associated with our reduced game is given by (32) in the Appendix A, and, when $\delta = 0.99$, it is equal to

$$\frac{z}{2+z}x + \frac{1}{2+z}\alpha + \frac{1}{2+z}\beta = 0.637\,64. \tag{28}$$

When we confine our attention to Nash equilibria (as opposed to correlated equilibria) of the repeated game, note that there are two asymmetric pure strategy Nash equilibria. Those corresponds to (D, C) and (C, D) in the reduced game, where players receive payoffs (α, β) or (β, α), where $\alpha = 0.629\,14$ or $\beta = 0.642\,5$. As is clear from (16) and the explanation thereafter, those payoffs are close to the best correlated equilibrium payoff (28).

Appendix B shows that our equilibrium lies above the Pareto frontier of all belief-free equilibria in this game, identified by EHO. Let us summarize the results of this section as follows:

Proposition 14 *Let $u_i(a), i = 1, 2$ and $p(a(t+1)|a(t))$ be the reduced game payoffs and law of motion associated with the one-period memory transition rule (15). When $\delta \geq 0.989\,54$, $p(\cdot|a)$ is a correlated equilibrium of the reduced game u for each a, and hence any (Nash or correlated) equilibrium payoff profile of v is a weakly belief-free equilibrium payoff profile of the repeated game. Furthermore, when $\delta \geq 0.989\,54$,*

- *u is a chicken game, and*
- *any weakly belief-free equilibrium payoff profile with transition rule (15) lies above the Pareto frontier of the belief-free equilibrium payoff set.*

In the next section, I will provide a detailed explanation why the weakly belief-free equilibrium outperforms belief-free equilibria.

6 Comparison Between Belief-Free and Weakly Belief-Free Equilibria

In this section I provide some comparisons between belief-free and weakly-belief free equilibria, in terms of the repeated prisoners' dilemma game of the previous section. First, I will show that the notion of a reduced game, introduced in this paper to analyze weakly belief-free equilibria, is useful in understanding the essential properties of belief-free equilibria. The explanation here reveals how weakly belief-free equilibrium generalizes the requirements of belief-free equilibrium. Secondly, we elaborate on why weakly belief-free equilibrium can outperform belief-free equilibria. Thirdly, I show that the weakly belief-free equilibrium of the previous section has a better dynamic stability property than belief-free equilibrium. In particular, it is shown that the weakly belief-free equilibrium strategies can invade a population of belief-free equilibrium strategies (so that the latter is not evolutionarily stable).

A leading example of a belief-free equilibrium is provided by Ely and Valimaki (2002) for the repeated prisoners' dilemma game with private monitoring. They construct judiciously chosen one-period memory mixed strategies, so that the reduced game has the following form:

	C	D
C	R, R	P, R
D	R, P	P, P

, $\quad R > P$

Note that player 1's payoff is completely determined by player 2's action: When 2 plays C, for example, player 1's payoff is always R, irrespective of 1's own action. The same is true for player 2's payoffs. Hence, for any realization of opponent's action, a player is always indifferent between C and D. This implies that *any joint distribution of actions is a correlated equilibrium of this reduced game*, and, in particular, the judiciously chosen mixed strategy equilibrium to induce this reduced game is indeed an equilibrium. Moreover, since the prescribed strategy of a player in this reduced game is always a best reply *irrespective of the beliefs over the opponent's action*, the equilibrium is completely belief-free. The present paper reveals that this is not the only way for players to follow the a correlated equilibria of the reduced game.

Next, I turn to the welfare comparison. Let us examine how the best belief-free equilibrium payoffs (R, R) in the reduced game above are determined. Those payoffs are associated with initial action profile (C, C). In the belief-free equilibrium, player 1 has an incentive to play C because a deviation to D induces player 2 to play D with a large probability. Since monitoring is imperfect, even though (C, C) is played a "bad" signal arises with a positive probability and such a punishment is triggered. Figure 2 (a) shows the directions of punishment in the belief-free equilibrium. The figure shows that when one player is punished, the other player's payoff

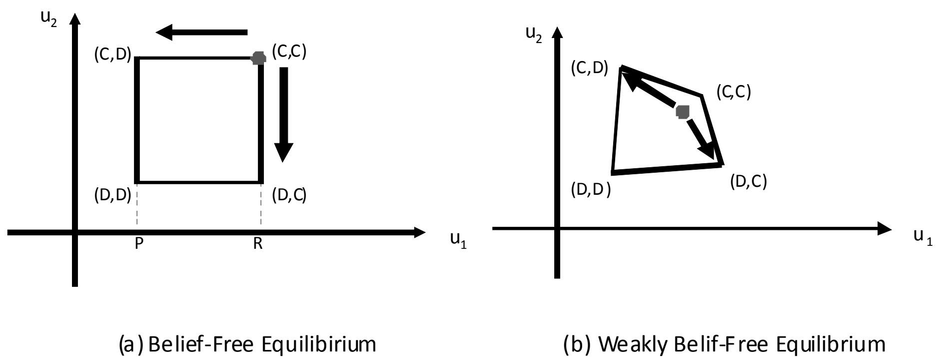


Figure 2: Reduced Game Payoff Set and the Directions of Punishment

cannot be increased. This implies that the total payoff of the players is reduced, and as a result the belief-free equilibrium suffers from a heavy welfare loss. On the other hand, if a player's payoff is increased when the opponent is punished, the loss of total payoff is mitigated (and, if done correctly, the loss can completely vanish, as the Fudenberg-Levine-Maskin (1994) folk theorem shows). The weakly belief-free equilibrium in the previous section embodies such *transfers of continuation payoffs* (although not as perfectly as the Fudenverg-Levine-Maskin equilibria do), and therefore it does better than the belief-free equilibria. The asymmetric punishment mechanism is embodied in the following way. The best symmetric weakly belief-free equilibrium in the example of the previous section is a correlated equilibrium in the reduced "Chicken" game, which mixes (C, C), (C, D) and (D, C). The major directions of punishment are shown in Figure 2 (b): As we saw in the previous section, the equilibrium transition rule (15) together with our information structure $p(\omega|a)$ imply that players alternate between (C, D) and

(D, C) with a large probability. Hence, the weakly belief-free equilibrium partially embodies the transfer of continuation payoffs, an essential tool to achieve efficiency in repeated games with imperfect monitoring.

Lastly, let me compare the belief-free and weakly belief-free equilibrium strategies in terms of their dynamic stability. Namely, I address their evolutionary stability. My argument is roughly summarized as follows. The belief-free equilibrium strategy has a property that any strategy is a best reply against it. This makes the belief-free equilibrium strategy susceptible to invasion by the weakly belief-free equilibrium strategy. In contrast, the weakly belief-free equilibrium strategy specifies a strict best reply in each period, and therefore it is immune to invasion. In what follows, I will formally establish this claim.

Recall the standard definition of evolutionarily stable strategies (ESS)[13]. Consider a symmetric two-person game, where $\pi(s', s)$ denotes a payer's payoff when he and his opponents play strategies s' and s respectively. A strategy s is an evolutionarily stable strategy (ESS) if, for any other strategy $s' \neq s$,

1. $\pi(s, s) \geq \pi(s', s)$, and
2. $\pi(s, s) = \pi(s', s) \Rightarrow \pi(s, s') > \pi(s', s')$.

This is the condition that a population of s cannot be invaded by a small fraction of players employing alternative strategy s'. After the invasion, the population consists of a large fraction of s and a small fraction of s'. When those strategies are randomly paired, the above two conditions ensures that the expected payoff to s is strictly higher than that to s' (and therefore the former dominates the population, by means of imitation of the superior strategy, for example). Now consider the belief-free and weakly belief free equilibrium examined above. Consider first the situation where the belief-free equilibrium strategy is the incumbent strategy s, which is invaded by the weakly belief-free equilibrium strategy s'. Recall the crucial property of the belief-free equilibrium strategy: against this strategy, a player is always indifferent between C and D. In other words, *any strategy is a best reply (yielding the same payoff) against the belief-free equilibrium strategy.* This makes the belief-free equilibrium particularly susceptible to the invasion of another strategy. Thus, we obtain the first part of Condition 2: $\pi(s, s) = \pi(s', s)$. Since s' is an equilibrium strategy, we have $\pi(s, s') \leq \pi(s', s')$.[14]

[13] See Kandori (1997) and Weibull (1995) for accessible exposition.

[14] Actually the inequality is strict, because against the weakly belief-free equilibrium strategy s', player has a strict incentive to follow actions specified by s'.

Hence Condition 2 above is violated and *the belief-free equilibrium is not an ESS*. In particular, it can be invaded by the weakly belief-free equilibrium.

In contrast, *the weakly belief-free equilibrium strategy is an ESS*. This is essentially because a player always has a strict incentive to follow the equilibrium action, and therefore an invasion by a small fraction of alternative strategies is impossible. To formally establish this claim, however, we need to pay attention to the following rather minor technical details. Consider the situation where the weakly belief-free equilibrium strategy is the incumbent strategy s, facing an invasion of another strategy s'. If s' is the strategy which specifies a different action plan than s only after the player's own deviations, however, it yields the same payoff as s does. Hence Condition 2 above is violated ($\pi(s,s) = \pi(s',s)$ and $\pi(s,s') = \pi(s',s')$), and we cannot establish that the weakly belief-free equilibrium strategy s is an ESS. To exclude formally such an inessential technicality, let us cast the question in terms of *reduced normal forms*. Two strategies of a player are *equivalent* if they always yield the same payoff against any given strategy by the opponent. The strategies s and s' considered above, which only differ after the player's own deviations, are equivalent. When we regard a set of strategies which are equivalent to each other (an equivalence class) as one strategy, we obtain a game with a smaller strategy space. The resulting game is called the reduced normal form game. Now let s be the weakly belief-free equilibrium strategy in the reduced normal form. If s' specifies a different action on the path of play, it fares strictly worse than s, and we have $\pi(s',s) < \pi(s,s)$ (because players have a strict incentive to follow the equilibrium action in any period). It remains to show that any reduced normal form strategy $s' \neq s$ specifies a different action than s on the path of play. This is true because the marginal distribution of private signals has full support. This implies that the opponent's deviations are never detected, and the only unreached information sets of a player are the ones that can be reached by his own deviations.[15] Hence, the weakly belief-free equilibrium strategy s is an ESS in the reduced normal form, because $\pi(s',s) < \pi(s,s)$ for any $s' \neq s$.

[15] Suppose there were an information set which is reached by a deviation of the opponent. Let s' be a strategy specifying a different action on this information set than s does. Note that s and s' can be different strategies in the reduced normal form, because they can yield different payoffs against a new strategy of the opponent that specifies the deviation to reach that information set. Against the original strategy of the opponent s, however, both s and s' would yiled the same payoff, and we could not obtain the desirable conclusion $\pi(s,s') < \pi(s',s')$ (we would instead obtain $\pi(s,s') = \pi(s',s')$).

7 Appendix A: The Correlated Equilibria in the Chicken Game

Note that the reduced game is strategically equivalent to

$1\backslash 2$	C	D
C	$0,0$	$\alpha - y, \beta - x$
D	$\beta - x, \alpha - y$	$0,0$

(29)

(In general, a game $g_i(a)$, $i = 1, ..., N$ is strategically equivalent to the game $\gamma g_i(a) + K_i(a_{-i})$ ($\gamma > 0$), which means that both games have the same best response correspondence and hence the same (Nash or correlated) equilibria.) For the reduced game to be a chicken game, we need $x < \beta$ and $y < \alpha$.

The game (29) is in turn strategically equivalent to (just multiply the payoffs by $\frac{1}{\beta - x}$)

$1\backslash 2$	C	D
C	$0,0$	$\frac{\alpha-y}{\beta-x}, 1$
D	$1, \frac{\alpha-y}{\beta-x}$	$0,0$

(30)

In general, when we have a chicken game

$1\backslash 2$	C	D
C	$0,0$	$z,1$
D	$1,z$	$0,0$

(31)

(where $z > 0$), the extremal correlated equilibria are (more explanation to be added)

$1\backslash 2$	C	D
C	$\frac{z}{2+z}$	$\frac{1}{2+z}$
D	$\frac{1}{2+z}$	0

(32)

$1\backslash 2$	C	D
C	0	$\frac{z}{1+2z}$
D	$\frac{z}{1+2z}$	$\frac{1}{1+2z}$

(33)

$1\backslash 2$	C	D
C	0	1
D	0	0

(34)

$1\backslash 2$	C	D
C	0	0
D	1	0

(35)

$1\backslash 2$	C	D
C	$\frac{z^2}{(1+z)^2}$	$\frac{z}{(1+z)^2}$
D	$\frac{z}{(1+z)^2}$	$\frac{1}{(1+z)^2}$

(36)

Note that (34), (35), and (36) correspond to pure and mixed Nash equilibria of the reduced game. Hence the "best" correlated equilibrium (32) is

$1\backslash 2$	C	D
C	$\frac{\frac{\alpha-y}{\beta-x}}{2+\frac{\alpha-y}{\beta-x}}$	$\frac{1}{2+\frac{\alpha-y}{\beta-x}}$
D	$\frac{1}{2+\frac{\alpha-y}{\beta-x}}$	0

(37)

8 Appendix B: An Upper Bound of the Belief-free Equilibrium Payoffs

In this section, I compare the example in Section 5 with the belief-free equilibrium payoffs identified by EHO (2005). To explain their characterization of the belief-free equilibrium payoffs, I first introduce the notion of *regime* $\mathcal{A}$ and an associated value $M_i^{\mathcal{A}}$. Using those concepts, I then find an upper bound for the belief-free equilibrium payoffs.

A regime $\mathcal{A} = \mathcal{A}_1 \times \mathcal{A}_2$ is a product of non-empty subsets of the stage game action sets, $\mathcal{A}_i \subset A_i$, $\mathcal{A}_i \neq \emptyset$, $i = 1, 2$. In each period of a belief-free equilibrium, players typically have multiple best-reply actions and they are played with positive probabilities. A regime corresponds to the set of such actions. For each regime $\mathcal{A}$, define a number $M_i^{\mathcal{A}}$ as follows.

$$M_i^{\mathcal{A}} = \sup v_i$$

such that for some mixed action α_{-i} whose support is $\mathcal{A}_{-i}$

and $x_i : \mathcal{A}_{-i} \times \Omega_{-i} \to \Re_+$

$$v_i \geq g(a_i, \alpha_{-i}) - \sum_{a_{-i}, \omega_{-i}} x_i(a_{-i}, \omega_{-i}) p_{-i}(\omega_{-i}|a_i, a_{-i}) \alpha_{-i}(a_{-i})$$

for all a_i with equality if $a_i \in \mathcal{A}_i$, where $p_{-i}(\omega_{-i}|a_i, a_{-i})$ is the marginal distribution of ω_{-i} given action profile (a_i, a_{-i}). Intuitively, the positive number x_i represents the reduction in player i's future payoffs. Note that a belief-free equilibrium has the property that player i's payoff is solely determined by the opponent's strategy. This is why the reduction in i's future payoffs, x_i, depends the opponent's action and signal (a_{-i}, ω_{-i}). Note also that the opponent's action a_{-i} is restricted to the component $\mathcal{A}_{-i}$ of the current regime $\mathcal{A} = \mathcal{A}_i \times \mathcal{A}_{-i}$. The above set of inequalities ensures that player i's best reply actions in the current period correspond to set $\mathcal{A}_i$, a component of the regime $\mathcal{A} = \mathcal{A}_i \times \mathcal{A}_{-i}$. Hence, the value $M_i^{\mathcal{A}}$ is closely related to the best belief-free payoff when the current regime is $\mathcal{A}$ (a more precise explanation will be given below).

Now let V^* be the limit set of belief-free equilibrium payoffs when $\delta \to 1$. EHO (2005) provides an explicit formula to compute V^*. For our purpose here, I only sketch the relevant part of their characterization to obtain a bound for V^*. In Section 4.1, EHO partitioned all games into three classes, *the positive*, *the negative*, and *the abnormal* cases (for our purpose here, we do not need to know their definitions). Their Proposition 6 shows that the abnormal case obtains *only if* one of the players has a dominant action in the stage game yielding the same payoff against all actions of the other player. Clearly, this is not the case in our example with the prisoner's dilemma stage game, so our example is in either the positive or negative case[16]. If it is in the negative case, EHO's Proposition 5 shows that the only belief-free equilibrium is the repetition of the stage game Nash equilibrium, yielding $(0, 0)$ in our example.

If our example is in the positive case, Proposition 5 in EHO implies that the limit set of belief-free equilibrium payoffs can be calculated as follows:

$$V^* = \bigcup_p \prod_{i=1,2} \left[\sum_{\mathcal{A}} p(\mathcal{A}) m_i^{\mathcal{A}},\ \sum_{\mathcal{A}} p(\mathcal{A}) M_i^{\mathcal{A}} \right], \tag{38}$$

where $m_i^{\mathcal{A}}$ is some number (for our purpose here, we do not need to know its definition) and p is a probability distribution over regimes $\mathcal{A}$. The union is taken with respect to all probability distributions p such that the intervals in the above formula (38) are well defined (i.e., $\sum_{\mathcal{A}} p(\mathcal{A}) m_i^{\mathcal{A}} \leq \sum_{\mathcal{A}} p(\mathcal{A}) M_i^{\mathcal{A}}$, $i = 1, 2$). The point to note is that V^* is a union of product sets (rectangles), and the efficient point (upper-right corner) of each rectangle is a convex combination of $(M_1^{\mathcal{A}}, M_2^{\mathcal{A}})$.

[16] With some calculation, we can determine which case applies to our example, but this is not necessary to derive our upper bound payoff.

The above characterization (38) of V^* implies, in the positive case, the belief-free equilibrium payoffs satisfy the following bound

$$(v_1, v_2) \in V^* \Longrightarrow v_1 + v_2 \le \max_{\mathcal{A}} M_1^{\mathcal{A}} + M_2^{\mathcal{A}}, \tag{39}$$

where maximum is taken over all possible regimes (i.e., for all $\mathcal{A} = \mathcal{A}_1 \times \mathcal{A}_2$ such that $\mathcal{A}_i \subset A_i$, $\mathcal{A}_i \neq \emptyset$, $i = 1, 2$).

In what follows, I estimate $M_1^{\mathcal{A}} + M_2^{\mathcal{A}}$ for each regime $\mathcal{A}$. In our example, $A_i = \{C, D\}$, so that $\mathcal{A}_i = \{C\}, \{D\}$, or $\{C, D\}$. Before examining each regime, I first derive some general results. Consider a regime $\mathcal{A}$ where $C \in \mathcal{A}_i$. In this case, the incentive constraint in the definition of $M_i^{\mathcal{A}}$ reduces to

$$\begin{aligned} v_i &= g(C, \alpha_{-i}) - \sum_{a_{-i}, \omega_{-i}} x_i(a_{-i}, \omega_{-i}) p_{-i}(\omega_{-i}|C, a_{-i}) \alpha_{-i}(a_{-i}) & (40) \\ &\ge g(D, \alpha_{-i}) - \sum_{a_{-i}, \omega_{-i}} x_i(a_{-i}, \omega_{-i}) p_{-i}(\omega_{-i}|D, a_{-i}) \alpha_{-i}(a_{-i}). & (41) \end{aligned}$$

This inequality (41) can be rearranged as

$$\begin{aligned} &\sum_{a_{-i}, \omega_{-i}} x_i(a_{-i}, \omega_{-i}) p_{-i}(\omega_{-i}|C, a_{-i}) \left(\frac{p_{-i}(\omega_{-i}|D, a_{-i})}{p_{-i}(\omega_{-i}|C, a_{-i})} - 1 \right) \alpha_{-i}(a_{-i}) \\ \ge \; & g(D, \alpha_{-i}) - g(C, \alpha_{-i}). \end{aligned} \tag{42}$$

Now let

$$L^* = \max_{\omega_{-i}, a_{-i}} \frac{p_{-i}(\omega_{-i}|D, a_{-i})}{p_{-i}(\omega_{-i}|C, a_{-i})}$$

be the maximum likelihood ratio to detect player i's deviation from C to D. The preceding inequality (42) and $L^* - 1 > 0$ imply[17]

$$\sum_{a_{-i}, \omega_{-i}} x_i(a_{-i}, \omega_{-i}) p_{-i}(\omega_{-i}|C, a_{-i}) \alpha_{-i}(a_{-i}) \ge \frac{g(D, \alpha_{-i}) - g(C, \alpha_{-i})}{L^* - 1}.$$

Plugging this into the definition (40) of v_i, we obtain

$$v_i \le g(C, \alpha_{-i}) - \frac{g(D, \alpha_{-i}) - g(C, \alpha_{-i})}{L^* - 1}.$$

[17] Note that, as long as player i's action affects the distribution of the opponent's signal (which is certainly the case in our example), there must be some ω_{-i} which becomes more likely when player i deviates from C to D. Hence, we have $L^* > 1$.

This is essentially the formula identified by Abreu, Milgrom and Pearce (1991). The reason for welfare loss (the second term on the right hand side), is that players are sometimes punished simultaneously in belief-free equilibria. The welfare loss associated with simultaneous punishment was originally pointed out by Radner, Myerson, and Maskin (1986). Recall that $M_i^{\mathcal{A}}$ is obtained as the supremum of v_i with respect to x_i and α_{-i} whose support is $\mathcal{A}_{-i}$. (Note that the right hand side of the above inequality, in contrast, does not depend on x_i.) Hence, we have

$$M_i^{\mathcal{A}} \leq \sup g(C, \alpha_{-i}) - \frac{g(D, \alpha_{-i}) - g(C, \alpha_{-i})}{L^* - 1}, \tag{43}$$

where the supremum is taken *over all* α_{-i} *whose support is* $\mathcal{A}_{-i}$.

Now we calculate the maximum likelihood ratio L^* and determine the right hand side of the above inequality (43). In our example, when $a_{-i} = C$, $\max_{\omega_{-i}} \frac{p_{-i}(\omega_{-i}|D,a_{-i})}{p_{-i}(\omega_{-i}|C,a_{-i})}$ is equal to (as our example is symmetric, consider $-i = 2$ without loss of generality)

$$\frac{p_2(\omega_2 = B|D, C)}{p_2(\omega_2 = B|C, C)} = \frac{\frac{1}{2} + \frac{1}{8}}{1/3} = \frac{15}{8}.$$

When $a_{-i} = D$, $\max_{\omega_{-i}} \frac{p_{-i}(\omega_{-i}|D,a_{-i})}{p_{-i}(\omega_{-i}|C,a_{-i})}$ is equal to

$$\frac{p_2(\omega_2 = B|D, D)}{p_2(\omega_2 = B|C, D)} = \frac{2/5 + 1/5}{1/4 + 1/8} = \frac{8}{5}.$$

As the former is larger, we conclude $L^* = \frac{15}{8}$. Plugging this into (43), we obtain the following upper bounds of $M_i^{\mathcal{A}}$.

1. When $C \in \mathcal{A}_i$ and $\mathcal{A}_{-i} = \{C\}$,

$$\begin{aligned} M_i^{\mathcal{A}} &\leq g(C, C) - \frac{g(D, C) - g(C, C)}{\frac{15}{8} - 1} \\ &= 1 - \frac{1/2}{\frac{15}{8} - 1} = \frac{3}{7}. \end{aligned}$$

2. When $C \in \mathcal{A}_i$ and $\mathcal{A}_{-i} = \{D\}$,

$$\begin{aligned} M_i^{\mathcal{A}} &\leq g(C, D) - \frac{g(D, D) - g(C, D)}{\frac{15}{8} - 1} \\ &= -\frac{1}{6} - \frac{1/6}{\frac{15}{8} - 1} = -\frac{5}{14}. \end{aligned}$$

3. When $C \in \mathcal{A}_i$ and $\mathcal{A}_{-i} = \{C, D\}$, the larger upper bound in the above two cases applies, so that we have

$$M_i^{\mathcal{A}} \leq \frac{3}{7}.$$

Given those bounds, we are ready to estimate $M_1^{\mathcal{A}} + M_2^{\mathcal{A}}$ for each regime $\mathcal{A}$.

Case (i), where $C \in \mathcal{A}_i$ for $i = 1, 2$: The above analysis (Cases 1 and 3) shows

$$M_1^{\mathcal{A}} + M_2^{\mathcal{A}} \leq \frac{6}{7}.$$

Case (ii), where $C \in \mathcal{A}_i$ and $\mathcal{A}_{-i} = \{D\}$: Our Case 2 shows $M_i^{\mathcal{A}} \leq -\frac{5}{14}$. In contrast, $M_{-i}^{\mathcal{A}}$ is simply achieved by $x_{-i} \equiv 0$ (as D is the dominant strategy in the stage game) so that

$$M_{-i}^{\mathcal{A}} = \sup_{\alpha_i} g(D, \alpha_i) = g(D, C) = \frac{3}{2}.$$

Hence, we have

$$M_1^{\mathcal{A}} + M_2^{\mathcal{A}} \leq \frac{3}{2} - \frac{5}{14} = \frac{8}{7}.$$

Case (iii), $\mathcal{A} = \{D\} \times \{D\}$: Since D is the dominant action in the stage game, $M_i^{\mathcal{A}}$ is achieved by $x_i \equiv 0$. Moreover, the opponent's action is restricted to $\mathcal{A}_{-i} = \{D\}$, so that we have $M_i^{\mathcal{A}} = g(D, D) = 0$. Hence,

$$M_1^{\mathcal{A}} + M_2^{\mathcal{A}} = 0.$$

Let me summarize our discussion above. If our example is in the negative case as defined by EHO, the only belief-free equilibrium payoff is $(0, 0)$. Otherwise, our example is in the positive case, where the sum of belief-free equilibrium payoffs $v_1 + v_2$ (in the limit as $\delta \to 1$) is bounded above by the maximum of the upper bounds found in Cases (i)-(iii), which is equal to $\frac{8}{7}$. Altogether, those results show that any limit belief-free equilibrium payoff profile (as $\delta \to 1$) $(v_1, v_2) \in V^*$ satisfies $v_1 + v_2 \leq \frac{8}{7}$.

To complete our argument, I now examine the belief free equilibrium payoffs *for a fixed discount factor* $\delta < 1$. Let $V(\delta)$ be the set of belief-free equilibrium payoff profiles for discount factor $\delta < 1$. The standard argument[18] shows that this is monotone increasing in δ (i.e. $V(\delta) \subset V(\delta')$ if $\delta < \delta'$). Hence, we have $V(\delta) \subset V^*$, so that for any discount factor δ, all belief-free equilibrium payoffs $(v_1, v_2) \in V(\delta)$ satisfy $v_1 + v_2 \leq \frac{8}{7}$. Now recall that in our example, our one-period memory transition rule (15) is an equilibrium if $\delta \geq 0.989\,54$, with reduced game given by

	C	D
C	x, x	α, β
D	β, α	y, y

. (44)

The numerical analysis in Appendix D shows $x, y, \alpha, \beta > 0.6$ for $\delta \geq 0.989\,54$. Hence, the total payoff *in any entry* in our reduced game payoff table (44) exceeds 1.2, which is larger than the upper bound for the total payoffs associated with the belief-free equilibria, $\frac{8}{7} \doteqdot 1.14$. This implies that *all of our equilibria lie above the Pareto frontier of the belief-free equilibrium payoff set.*

The reason is as follows. Our analysis shows that, by choosing an equilibrium of the reduced game (44) in the first period and then following our one-period memory transition rule (15), (i) we obtain a (strongly recursive) equilibrium of the repeated game and (ii) the average repeated game payoffs are equal to the chosen equilibrium payoff of the reduced game. When $\delta \geq 0.989\,54$, the reduced game (44) is a chicken game. Hence, if we assume that there is no correlation device, (C, D) and (D, C) (the Nash equilibrium of the chicken game (44)) correspond to our equilibria. When public randomization or partial correlation devices are available at the beginning of the game, we can also achieve additional outcomes, i.e., the correlated equilibria of the reduced game (44). In any case, all those equilibria lie above the Pareto frontier of the belief-free equilibrium payoff set.

[18] The proof is as follows. Suppose we terminate the repeated game under $\delta' > \delta$ randomly in each period with probability $1 - \frac{\delta}{\delta'}$ and start a new game (and repeat this procedure). In this way, we can decompose the repeated game under δ' into a series of randomly terminated repeated games, each of which has effective discount factor equal to $\delta' \times \frac{\delta}{\delta'} = \delta$. Hence, any equilibrium (average) payoff under δ can also be achieved under $\delta' > \delta$. This argument presupposes that public randomization is available (to terminate the game). Even without public radomization, however, our conclusion $V(\delta) \subset V^*$ also holds, because (i) the set of belief-free payoff profiles $V(\delta)$ is smaller without public randomization and (ii) the *same* limit payoff set V^* obtains with or without public randomization (see the online appendix to EHO (2004)).

9 Appendix C: The Review Strategy Equilibria in the Example

Matsushima (2004) shows a larger payoff set can be sustained by extending the idea of the belief-free equilibrium by means of *review strategies.* A review strategy equilibrium treats T consecutive stage games as if they were a single stage game, or a *block stage game*, and applies the belief-free equilibrium technique to the sequence of such block stage games. Matsushima showed that, under certain conditions, approximate efficiency can be obtained by the review strategies, even if the observability is quite limited. This is substantially generalized in a recent paper by Fong, Gossner, Horner and Sannikov (2008). In this section, I show that their review strategies do not work in my example.

Matsushima showed that approximate efficiency can be achieved by review strategies in repeated prisoners' dilemma games, provided that the private signals are independently distributed conditional on the action profile and an unobservable common shock. This requirement is expressed as

$$p(\omega_1, \omega_2|a) = \sum_{\theta \in \Theta} q_1(\omega_1|a_1, a_2, \theta) q_2(\omega_2|a_1, a_2, \theta) f(\theta|a_1, a_2), \qquad (45)$$

where $\theta \in \Theta$ is the hidden common shock, *and* for all i and a_i,

$$q_i(\cdot|a_i, a_{-i}, \theta), \text{ for } a_{-i} \in A_{-i} \text{ and } \theta \in \Theta, \text{ are linearly independent.}$$

The latter requirement is satisfied only if $|\Omega_i| \geq |A_{-i}| \times |\Theta|$. Since we have $|\Omega_i| = |A_{-i}| = 2$ in our example, the requirement is satisfied only if $|\Theta| = 1$. In such a case, the first requirement (45) implies that the private signals are conditionally independent $p(\omega_1, \omega_2|a) = q_1(\omega_1|a_1, a_2)q_2(\omega_2|a_1, a_2)$, and this is clearly not the case in our example. Hence, Matsushima's review strategy results do not apply to our example.

Although Matsushima's conditions admit a clear meaning (conditional independence on a hidden common shock), generically it is not satisfied in the space of monitoring structures. Fong, Gossner, Horner and Sannikov (2008) show that efficiency is approximately achieved in repeated prisoners' dilemma games under a different set of assumptions, which are satisfied by a positive measure of monitoring structures. A key assumption is their Assumption 2, which they call "positively correlated scores". In the two-player case, it reduces to the following requirements. Label one signal of player i as 1^i where, for every action of player i, 1^i is always at least as likely when player $j \neq i$ plays C than when player j plays D . The signal

1^i can then be interpreted as a "good" signal about j's cooperation. It is straightforward to check that in my example, the good signal 1^i indeed corresponds to $\omega_i = G$. In this case, they state that their Assumption 2 reduces to the requirement that good signals are positively correlated when both players cooperate, i.e.

$$\Pr(\omega_i = G|C, C, \omega_j = G) > \Pr(\omega_i = G|C, C).$$

In my example, the joint distribution of private signals given action profile (C, C) is

$\omega_1 \diagdown \omega_2$	G	B
G	1/3	1/3
B	1/3	0

,

and obviously good signals are *not* positively correlated. (We see that their condition is violated because $\Pr(\omega_i = G|C, C, \omega_j = G) = 1/2 < \Pr(\omega_i = G|C, C) = 2/3$.) Hence we conclude that their efficiency results in review strategies, as they stand, do not apply to my example.

10 Appendix D: The Reduced Game Payoffs for $\delta \geq 0.98954$

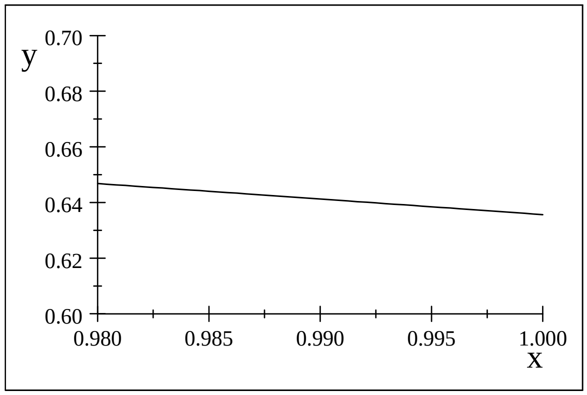
In our example, (15) can be a weakly belief-free equilibrium transition rule when $\delta \geq 0.989\,54$. Here we numerically show that for this range all equilibrium payoffs dominate the belief-free equilibria. The reduced game payoffs in table (44) are obtained as a solution to the system of dynamic programming equations

$$\left\{ \begin{array}{c} x = (1-\delta) + \delta\frac{1}{3}(x + \alpha + \beta) \\ y = \delta(\frac{1}{5}y + \frac{2}{5}(\alpha + \beta)) \\ \alpha = (1-\delta)(-\frac{1}{6}) + \delta(\frac{1}{8}x + \frac{1}{4}\alpha + \frac{1}{2}\beta + \frac{1}{8}y) \\ \beta = (1-\delta)\frac{3}{2} + \delta(\frac{1}{8}x + \frac{1}{2}\alpha + \frac{1}{4}\beta + \frac{1}{8}y) \end{array} \right. .$$

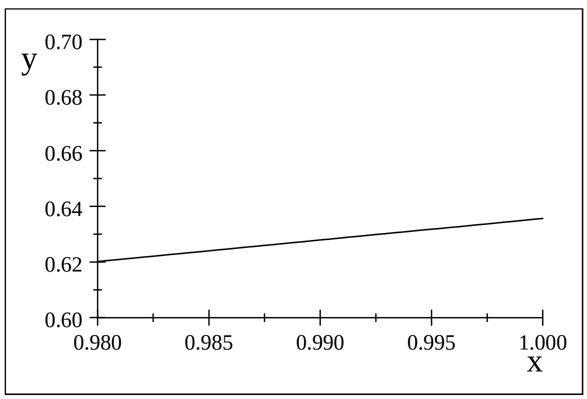
Using Maple, we obtain

$x = \frac{1}{3}\frac{7\delta^2+91\delta-180}{17\delta-60}$, $\beta = -\frac{1}{6}\frac{7\delta^3+285\delta^2-1632\delta+2160}{(17\delta-60)(4+\delta)}$, $y = \frac{2}{3}(7\delta - 48)\frac{\delta}{17\delta-60}$, and $\alpha = -\frac{1}{6}\frac{1448\delta-395\delta^2-240+7\delta^3}{(17\delta-60)(4+\delta)}$.

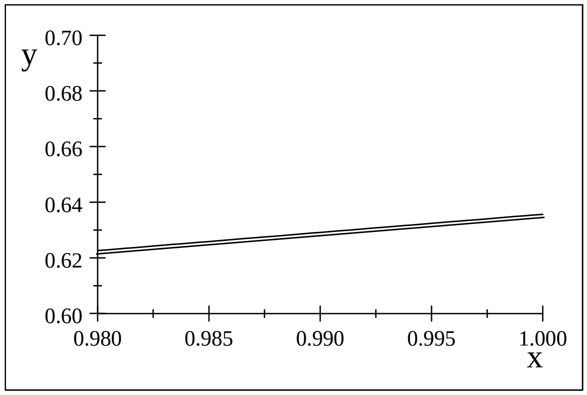
Plotting $x(\delta) = \frac{1}{3} \frac{7\delta^2+91\delta-180}{17\delta-60}$



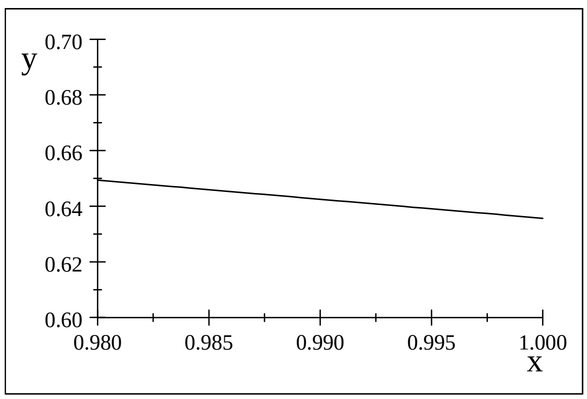
Plotting $y(\delta) = \frac{2}{3} (7\delta - 48) \frac{\delta}{17\delta-60}$



Plotting $\alpha(\delta) = -\frac{1}{6} \frac{1448\delta-395\delta^2-240+7\delta^3}{(17\delta-60)(4+\delta)}$



Plotting $\beta(\delta) = \beta = -\frac{1}{6}\frac{7\delta^3+285\delta^2-1632\delta+2160}{(17\delta-60)(4+\delta)}$



Hence, we have numerically confirmed that $x, y, \alpha, \beta > 0.6$ holds for $\delta \geq 0.989\,54$.

References

Abreu, D., P. Milgrom, and D. Pearce (1991): "Information and Timing in Repeated Partnerships," *Econometrica,* 59, 1713-1733.

Abreu, D., D. Pearce, and E. Stacchetti (1990): "Toward a Theory of Discounted Repeated Games with Imperfect Monitoring," *Econometrica,* 58, 1041-1063.

Bhaskar, V. (2000): "The Robustness of Repeated Game Equilibria to Incomplete Payoff Information", University of Essex.

Bhaskar, V. and E. van Damme. (2002): "Moral Hazard and Private Monitoring," *Journal of Economic Theory*, 102, 16-39.

Bhaskar, V., G. Mailath, and S. Morris (2008): "Purification in the Infinitely-Repeated Prisoner's Dilemma", *Review of Economic Dynamics*, 11, 515-528 .

Bhaskar, V. and I. Obara (2002): "Belief-Based Equilibria in the Repeated Prisoners' Dilemma with Private Monitoring," *Journal of Economic Theory*, 102, 40-69.

Compte, O. (1998): "Communication in Repeated Games with Imperfect Private Monitoring", *Econometrica*, 66, 597-626.

Ely, J.C., J. Hörner, and W. Olszewski (2004): "Dispensing with Public Randomization in Belief-Free Equilibria", http://www.kellogg.northwestern.edu//faculty//horner/ehoappendix.pdf

Ely, J.C., J. Hörner, and W. Olszewski (2005): "Belief-free Equilibria in Repeated Games," *Econometrica*, 73, 377-415.

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

Fong, K, O. Gossner, J. Horner, and Y. Sannikov (2008) "Efficiency in a Repeated Prisoners' Dilemma with Imperfect Private Monitoring,", in preparation.

Fudenberg, D., D. K. Levine, and E. Maskin (1994): "The Folk Theorem with Imperfect Public Information," *Econometrica,* 62, 997-1040.

Fudenberg, D., and E. Maskin (1986): "The Folk Theorem in Repeated Games with Discounting or with Incomplete Information," *Econometrica,* 54, 533-54.

Hörner, J. and W. Olszewski (2006): "The Folk Theorem for Games with Private Almost-Perfect Monitoring," *Econometrica,* 74, 1499-1544.

Kandori, M. (1997): "Evolutionary Game Theory in Economics", in Kreps, D. M. and K. F. Wallis (eds.), *Advances in Economics and Econometrics: Theory and Applications*, Vol. I, Cambridge: Cambridge University Press.

Kandori, M. (2002): "Introduction to Repeated Games with Private Monitoring," *Journal of Economic Theory*, 102, 1-15.

Kandori, M. (2008): "Repeated Games", in *New Palgrave Dictionary of Economics, 2nd edition*, Palgrave Macmillan.

Kandori, M. and I. Obara (2006): "Efficiency in Repeated Games Revisited: The Role of Private Strategies", *Econometrica*, 74, 499-519.

Kandori, M. and H. Matsushima (1998): "Private Observation, Communication and Collusion", *Econometrica*, 66, 627-652.

Mailath, G. and L. Samuelson (2006): *Repeated Games and Reputations*, Oxford University Press.

Matsushima, H. (2004): "Repeated Games with Private Monitoring: Two Players", *Econometrica*, 72, 823-852.

Obara, I. (1999): "Private Strategy and Efficiency: Repeated Partnership Game Revisited," Unpublished Manuscript, University of Pennsylvania.

Phelan, C. and A. Skrzypacz (2008) "Beliefs and Private Monitoring", http://www.stanford.edu/~skrz/phelanskrzypacz.pdf

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

Radner, R., R. Myerson, and E. Maskin (1986): "An Example of a Repeated Partnership Game with Discounting and with Uniformly Inefficient Equilibria," *Review of Economic Studies*, 53, 59-69.

Sekiguchi, T. (1997): "Efficiency in the Prisoner's Dilemma with Private Monitoring", *Journal of Economic Theory*, 76, 345-361.

Yamamoto, Y (2006): "Efficiency Results in N Player Games with Imperfect Private Monitoring", *Journal of Economic Theory*, 135, 382-413.

Weibull, J. (1997): Evolutionary Game Theory, MIT Press.