Sequential Decision Making for Intelligent Agents
Papers from the AAAI 2015 Fall Symposium

How Is Cooperation/Collusion Sustained in Repeated Multimarket Contact with Observation Errors?

Atsushi Iwasaki[1], **Tadashi Sekiguchi**[2], **Shun Yamamoto**[3], **Makoto Yokoo**[3]
1: University of Electro-Communications, iwasaki@is.uec.ac.jp
2: Kyoto University, sekiguchi@kier.kyoto-u.ac.jp
3: Kyushu University, {syamamoto@agent., yokoo@}inf.kyushu-u.ac.jp

Abstract

This paper analyzes repeated multimarket contact with observation errors where two players operate in multiple markets simultaneously. Multimarket contact has received much attention from the literature of economics, management, and information systems. Despite vast empirical studies that examine whether multimarket contact fosters cooperation/collusion, little is theoretically known as to how players behave in an equilibrium when each player receives a noisy observation of other firms' actions. This paper tackles an essentially realistic situation where the players do not share common information; each player may observe a different signal (*private* monitoring). Thus, players have difficulty in having a common understanding about which market their opponent should be punished in and when punishment should be started and ended. We first theoretically show that an extension of 1-period mutual punishment (1MP) for an arbitrary number of markets can be an equilibrium. Second, by applying a verification method, we identify a simple equilibrium strategy called "locally cautioning (LC)" that restores collusion after observation error or deviation. We then numerically reveal that LC significantly outperforms 1MP and achieves the highest degree of collusion.

Introduction

This paper analyzes repeated multimarket contact with observation errors where two players operate in multiple markets simultaneously. A company usually sells a variety of goods across multiple markets and often supplies an identical good to areas that are geographically apart, e.g., U.S. and Japan. In such a multimarket contact situation, it is pointed out that tacit collusion among companies is likely to occur (Chellappaw, Sambamurthy, and Saraf 2010). Therefore, multimarket contact has received much attention from the literature of economics, management, and information systems. However, despite vast empirical studies (Yu and Albert A. Cannella 2013) that have examined whether multimarket contact fosters cooperation/collusion, little is theoretically known as to how players behave in an equilibrium when each player receives a noisy observation of other firms' actions.

Without noisy observation, i.e., under *perfect* monitoring where each player can observe his opponent's action, the theoretical results in the single market case can be applied (Bernheim and Whinston 1990). When a strategy for a single market is an equilibrium, the joint strategy for multiple markets is an equilibrium. However, there exists no equilibrium strategy such that players can increase their profit aggregating the information from the multiple markets. With noisy observation, one exception is a case where players do share common information; all players always observe a noisy, but common signal (*public* monitoring). A generalization of the trigger strategy is an equilibrium and optimal among the possible strategies (Kobayashi and Ohta 2012). However, the trigger strategy is no longer an equilibrium in a case where players do not share common information on each other's past history; each player may observe a different signal (*private* monitoring).

Analytical studies on this class of games have not been very successful. This is because finding (pure) strategy equilibria in such games has been considered to be extremely hard. Indeed, it requires very complicated statistical inferences to estimate the history a player reaches at a period and to compute the continuation payoff from the period on (Kandori 2010). As a result, such analysis has been possible for only very restricted distributions of signals such as *belief-free* equilibria where an optimal strategy is constructed so that a player's belief (about her opponent) does not matter (Ely, Horner, and Olszewski 2005).

Recently, Kandori and Obara (2010) establish a connection between the partially observable Markov decision process (POMDP) (Kaelbling, Littman, and Cassandra 1998) and private monitoring, by employing *belief-based* strategy representation. Iwasaki et al. (2014) further develop a computationally feasible algorithm to verify whether a strategy profile can constitute an equilibrium. Accordingly, this enables us to analyze equilibria for arbitrary distributions of signals.

Building upon those studies, we first examine a finite state automaton (FSA) that follows a strategy called J-1MP, which is the joint version of 1MP in Fig. 1[1] for an arbitrary number of markets. Each player acts according to 1MP in

[1] Here, g or b is a private signal, (*good* or *bad*), which is a noisy observation of opponent action C or D (*cooperate* or *defect*).

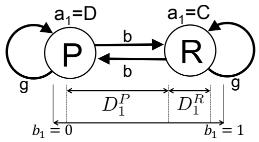


Figure 1: 1MP

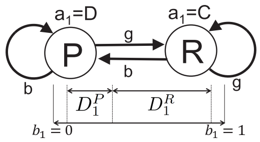


Figure 2: TFT

each market. It is quite natural that if 1MP is an equilibrium in a single market, J-1MP is in multiple markets. However, J-1MP is still a very important benchmark to consider the multimarket contact situation. In addition, to the best of our knowledge, we are the first to analytically characterize a condition where 1MP is an equilibrium under private monitoring. By observing this result, even in a single market, we illustrate the difficulty of analyzing equilibria under private monitoring.

This paper aims to find an FSA that can form an equilibrium and that achieves a better payoff than J-1MP. To this end, we exhaustively search for small-sized FSAs in the case of two markets and find a novel FSA called "locally cautioning (LC)" depicted in Fig. 5a. Furthermore, the exhaustive search confirms that no other FSA that can be an equilibrium is more efficient than LC.

Under this FSA, a player first cooperates in each market. If her opponent defects in both markets, she also defects only in either market. However, after one period of mutual cooperation in the market where she continues to cooperate, she returns cooperation in both markets. Requiring such a mutual cooperation is beneficial in establishing a robust coordination among players. Such LC behavior resembles tit-for-tat (TFT): she returns to cooperation after observing her opponent's cooperation, ignoring the outcome in the market she is punishing (Fig. 2). The surprising feature is that LC can be an equilibrium across substantially different private monitoring structures.

Related literature

Let us briefly explore the literature of repeated games in economics and multi-agent systems. In economics, the main body of repeated games is understanding how cooperation among self-interested players is sustained in long-term relationship (Mailath and Samuelson 2006; Kandori 2010). Perfect monitoring has been considerably investigated: *folk theorems* and *one-shot deviation principle*. Those results can be extended to imperfect public monitoring. In contrast, imperfect private monitoring remains considerably less explored, though it is closely related to multi-agent systems in an environment with uncertainty such that each agent (player) partially observes the state of the environment.

In multi-agent systems, there are many streams associated with repeated games (Burkov and Chaib-draa 2013): the complexity of equilibrium computation (Littman and Stone 2005; Borgs et al. 2008) multi-agent learning (Blum and Monsour 2007; Shoham and Leyton-Brown 2008), POMDP (Hansen, Bernstein, and Zilberstein 2004; Doshi and Gmytrasiewicz 2006; Wu, Zilberstein, and Chen. 2011; Wunder et al. 2011), and so on.

Traditionally, POMDP imposes partial observability on an opponent's strategy (behavior rule) and not for opponent's past actions. The main objective is guessing an optimal (best reply) strategy against an unknown strategy (not always fixed) from perfectly observable actions (perfect monitoring). On the other hand, our objective is checking whether a given strategy profile is a *mutual* best reply after any history, i.e., finding an equilibrium, with partially observable actions (private monitoring). Therefore, the private monitoring problem is significant and remains difficult to solve, though it has been missed in the POMDP literature.

POMDP also focuses mainly on "cooperative" environments where multiple agents share a common goal, i.e., no conflict of interest such as an ad-hoc team setting (Hansen, Bernstein, and Zilberstein 2004; Wu, Zilberstein, and Chen. 2011). In contrast, this paper applies the POMDP technique to "competitive" environments where they have some conflict of interest such as prisoners' dilemma and the lemonade-stand game (Doshi and Gmytrasiewicz 2006; Wunder et al. 2011).

To the best of our knowledge, very few existing works have addressed such environments and focused on finding equilibria under private monitoring. That being said, partially observable stochastic games (POSGs) are a generalization of private monitoring if the opponent's actions are partially observable (they often are not, though). As one exception, Hansen, Bernstein, and Zilberstein (2004) develop an algorithm that iteratively eliminates dominated strategies. However, this algorithm can not identify a cooperative equilibrium in a competitive environment, e.g., in prisoners' dilemma, a cooperative action is dominated. Doshi and Gmytrasiewicz (2006) propose an algorithm for general POSGs that converges to an equilibrium, which is different from ours. However, they do not successfully develop a new concrete strategy in any game. show only the concept and the algorithm This paper raises multimarket contact as an important issue, elaborates a theoretical structure of the private monitoring problem as a POMDP, and finds an entirely novel strategy. Accordingly, we believe that it casts light on a new research frontier for POMDP.

Model

Let us consider $N = \{1, 2\}$ to be the set of two players (firms) competing in two markets $Z = \{A, B\}$ simultaneously. Each player repeatedly plays the same stage games over infinite horizon $t = 0, 1, \ldots$. In each period, player i in market z takes some action a_i^z from a finite set $A_i^z = \{C, D\}$. We denote actions for player i in both markets by $a_i = (a_i^A, a_i^B) \in A_i = \Pi_{z \in Z} A_i^z$ and an action profile by $\boldsymbol{a} = (a_1, a_2) \in \Pi_{i \in N} A_i$. Player i's expected payoff in market z in a period is given by a stage game payoff function $g_i^z(a^z)$ for z, where $a^z = (a_1^z, a_2^z) \in \Pi_{i \in N} A_i^z$ is the action profile in that period. A stage game payoff $g_i(\boldsymbol{a})$ is defined as the sum of the stage game payoffs for each market: $g_i(\boldsymbol{a}) = \sum_{z \in Z} g_i^z(a^z)$.

Within each period, each player's private signal for market z is $\omega_i^z \in \Omega_i^z$, which is a noisy observation of the opponent's action for that market. Let us introduce the joint distribution of private signals $o^z(\omega^z \mid a^z)$ for z. Let

$\omega_i = (\omega_i^A, \omega_i^B) \in \Omega_i = \Pi_{z\in Z}\Omega_i^z$ denote an observation for player i. We assume that the private signals of each market are independently distributed and are unaffected by the actions taken in the other market. Formally, $\boldsymbol{o}(\boldsymbol{\omega}|\boldsymbol{a}) = \Pi_{z\in Z} o^z(\omega^z \mid a^z)$ where $\boldsymbol{\omega} = (\omega_1, \omega_2)$. We also assume that no player can infer which action was taken (or not taken) by another player for sure; to this end, we assume that each signal profile $\boldsymbol{\omega}$ occurs with a positive probability for any $\boldsymbol{a}$ (*full support assumption*).

Player i's expected discounted payoff V_i from a sequence of action profiles $\boldsymbol{a}(0), \ldots, \boldsymbol{a}(t), \ldots$ is $\sum_{t=0}^{\infty} \delta^t g_i(\boldsymbol{a}(t))$, with discount factor $\delta \in (0,1)$. The discounted *average payoff* (payoff per period and market) is defined as $(1-\delta)V_i/|Z|$.

Strategy representation and equilibrium concept

A private history for player i at the end of time t is the record of her past actions and signals,

$$h_i^t = (a_i(0), \omega_i(0), \ldots, a_i(t), \omega_i(t)) \in H_i^t := (A_i \times \Omega_i)^{t+1}.$$

To determine the initial action of each player, we introduce a dummy initial history (a null history) h_i^0 and let H_i^0 be a singleton set $\{h_i^0\}$. A pure strategy s_i for i is a function specifying an action after any history, or, formally, $s_i : H_i \to A_i$, where $H_i = \bigcup_{t\geq 0} H_i^t$.

A finite state automaton is a popular approach for compactly representing the behavior of a player. An FSA M is defined by $\langle \Theta, \hat{\theta}, f, T\rangle$, where Θ is a set of states, $\hat{\theta} \in \Theta$ is an initial state, $f : \Theta \to A_i$ determines the action choice for each state, and $T : \Theta \times \Omega_i \to \Theta$ specifies a deterministic state transition. $T(\theta(t), \omega_i(t))$ returns next state $\theta(t+1)$ when the current state is $\theta(t)$ and the private signal is $\omega_i(t)$. We call an FSA without the specification of the initial state, i.e., $m = \langle \Theta, f, T\rangle$, as a finite state *preautomaton* (pre-FSA). $\boldsymbol{m}$ denotes a profile of pre-FSAs,

Assume that player $j (\neq i)$ acts based on m_j. A *belief* of i is a probability distribution over the current state of j, which is represented as $b_i \in \Delta(\Theta_j)$, where $\Theta_j = \Theta$ due to symmetry and $\Delta(\Theta)$ represents the probability distribution on Θ. Let $\boldsymbol{\theta}$ denote the state profile (θ_i, θ_j). $b_i(\theta_j)$ denotes the probability that the states of j are θ_j. For example, if two players act based on 1MP in Fig. 1, b_i is represented as a vector of two elements: $(b_i(R), b_i(P))$. However, since $b_i(R) + b_i(P) = 1$, we have only one degree of freedom. Thus, we represent b_i by the value of $b_i(R)$.

We adopt a *belief-based strategy* representation (Bhaskar and Obara 2002), and represent player i's strategy as tuple $(m_i, D_i, \hat{\theta}_i)$. Game theory requires that a strategy be a complete contingent action plan. In other words, a strategy of a player not only specifies her equilibrium behavior (i.e., on-path plan), but it must also describe what she should do after deviating from it. Mathematically, a strategy in game theory is represented by a mapping from a private history to a current action.

Now, we formally define a belief division. as well as several related conditions. Belief division D_i of player i is an array $(D_i^1, \ldots, D_i^{k_i})$, such that $\forall D_i^l \in D_i, D_i^l \subseteq \Delta(\Theta)$. We assume each D_i^l is associated with each state $\theta_i^l \in \Theta_i$ of player i's pre-FSA m_i. D_i is *covering* if $\bigcup_{D_i^l \in D_i} D_i^l = \Delta(\Theta)$. We also describe the profile of belief divisions as $\boldsymbol{D}$.

Let $\chi_i[(a_i, \omega_i), b_i]$ denote the posterior belief for i where the current belief is b_i, the current action is a_i, and the obtained observation is ω_i, assuming other players act based on m_j. For history h_i^t, $\chi_i[h_i^t, b_i]$ denotes a posterior belief after observing h_i^t starting from b_i. These beliefs can be calculated by Bayes' rule for any h_i^t and b_i because of the full support assumption.

We analyze a (symmetric) *pure-strategy equilibrium*, where both players have an incentive to start at the common state R assuming that the opponent starts the same state. Initial joint state $\hat{\boldsymbol{\theta}} = (\hat{\theta}_1, \hat{\theta}_2)$ is (R, R) and thus the initial belief b_i is one.

Definition 1 (Resilient finite state equilibrium) *A tuple $(\boldsymbol{m}, \boldsymbol{D}, \hat{\boldsymbol{\theta}})$ forms a pure resilient finite state equilibrium (RFSE) if the following conditions hold:*

(i) *$\boldsymbol{D}$ accommodates $\hat{\boldsymbol{\theta}}$, i.e., for any history starting at $\hat{\boldsymbol{\theta}}$ no posterior is outside belief division $\boldsymbol{D}$.*

(ii) *for any $i \in N$, for any belief b_i after any history starting at $\hat{\boldsymbol{\theta}}$, and for all l with $b_i \in D_i^l$, an FSA $M_i = (m_i, \theta_i^l)$ is a plan or behavior that maximizes her expected discounted payoff given i's belief b_i.*

This definition implies that RFSE is a *sequential equilibrium* with initial joint state $\hat{\boldsymbol{\theta}}$. Iwasaki et al. (2014) call the equilibrium *resilient*, because a player always goes back to a particular state of pre-FSA m_i after any deviation. It must be emphasized here that we are not restricting the possible strategy spaces of players (i.e., we are *not* assuming that players can only use FSAs).

Verification method

Assuming player j acts based on m_j, player i confronts POMDP, where the state of the world is represented by the state of j's pre-FSA. An optimal *policy* in this POMDP corresponds to an optimal belief-based strategy. Let $v_i^{\boldsymbol{\theta}}$, where $\boldsymbol{\theta} = (\theta_i, \theta_j)$, be player i's expected discounted payoff associated with (m_i, θ_i). Based on the joint pre-FSA, we obtain $v_i^{\boldsymbol{\theta}}$ by solving a system of linear equations defined as follows, where $\boldsymbol{\theta'} = T(\boldsymbol{\theta}, \boldsymbol{\omega})$, i.e., $\boldsymbol{\theta'}$ is the next joint state:

$$v_i^{\boldsymbol{\theta}} = g_i(f(\boldsymbol{\theta})) + \delta \sum_{\boldsymbol{\omega}\in \Pi_{i\in N}\Omega_i} v_i^{\boldsymbol{\theta'}} \cdot o(\boldsymbol{\omega} \mid f(\boldsymbol{\theta})). \quad (1)$$

$V_i^{M_i}(b_i)$ denotes the expected discounted payoff (or the continuation payoff) of player i, when her belief of j's state is b_i, where i acts based on an FSA M_i and j acts based on m_j. In particular, $V_i^{(m_i,\theta_i)}(b_i)$ is $\sum_{\theta_j\in\Theta_j} v^{(\theta_i,\theta_j)} b_i(\theta_j)$. Note that $V_i^{(m_i,\theta_i)}(b_i)$ is linear in belief b_i. The upper envelop of these expected discounted payoffs is called m_i's *value function* in the POMDP literature.

The concept of a one-shot extension (Kandori and Obara 2010) (also known as a backup operator) is convenient to prove the optimality of an FSA.

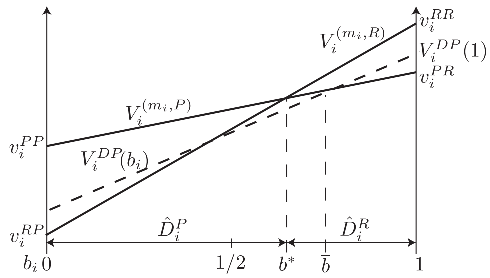


Figure 3: Expected discounted payoffs where Eqs. 2 and 3 hold.

Definition 2 (One-shot extension) *For pre-FSA m_i, its one-shot extension, denoted as $(a_i, ((\omega_1, \theta_{\omega_1}), \ldots))$, is a new FSA created as follows: (1) it has new initial state θ'_i, where $a_i \in A_i$, is played, (2) after ω_i is observed, FSA (m_i, θ_{ω_i}) is played, where $\theta_{\omega_i} \in \Theta$.*

Next we describe the set of all one-shot extensions of m_i as $Ex(m_i)$. Note that $Ex(m_i)$ has a finite number of elements ($= |A_i| \cdot k_i^{|\Omega_i|}$, where k_i is the number of states in m_i). We define a belief division such that no profitable one-shot extension exists, which we call a *target belief division*.

Definition 3 (Target belief division) *Target belief division $\widehat{D}_i$ for player i is a belief division, where each $\widehat{D}_i^l$ is chosen so that the following condition holds: $\forall b_i \in \widehat{D}_i^l$*

$$V_i^{(m_i, \theta_i^l)}(b_i) \geq V_i^{M_i}(b_i), \forall M_i \in Ex(m_i).$$

Figure 3 illustrates an example when two players act according to 1MP. The x-axis shows the belief of player i, and the y-axis shows the expected discounted payoffs of 1MP and its one-shot extension. For instance, $V_i^{(m_i,R)}$ is the payoff when player i starts at R at a period and follows 1MP. Given current belief b_i, $V_i^{DP}(b_i)$ is the payoff of a one-shot extension where i plays D at a period and, regardless of the current observation, follows 1MP from state P from the next period on. The target belief division $\hat{D}_i$ consists of $\hat{D}_i^P = [0, b^*]$ and $\hat{D}_i^R = [b^*, 1]$. Because $\hat{D}_i$ is covering, the optimal continuation plan given b_i is always included in m_i. If the current belief is $\bar{b}$, starting at R is the optimal plan and the continuation payoff is $V_i^{(m_i,R)}(\bar{b})$. However, as we show in Fig. 4, it may be a case that the optimal continuation plan is not included in a pre-FSA.

The target belief division is useful in verifying RFSE because the following theorem holds.

Proposition 1 (Theorem 1 (Iwasaki et al. 2014)) *A profile of pre-FSAs $\boldsymbol{m}$ and $\hat{\boldsymbol{\theta}}$ form a pure RFSE iff $\hat{\boldsymbol{D}}$ accommodates $\hat{\boldsymbol{\theta}}$.*

This proposition assures that unless for any player i and for any history h_i^t starting from $\hat{\theta}_i$, her belief b_i (updated by Bayes' rule) goes outside the target belief division $\hat{D}_i$, pre-FSA m_i gives an optimal continuation plan, i.e., the best reply for all i. Thus, ($\boldsymbol{m}$, $\hat{\boldsymbol{D}}$, $\hat{\boldsymbol{\theta}}$) is an equilibrium.

	$a_2^z = C$	$a_2^z = D$
$a_1^z = C$	$1, 1$	$-y, 1 + x$
$a_1^z = D$	$1 + x, -y$	$0, 0$

Table 1: Prisoners' dilemma in market z.

	$\omega_2^z = g$	$\omega_2^z = b$
$\omega_1^z = g$	p	q
$\omega_1^z = b$	q	s

(a) $(a_1^z, a_2^z) = (C, C)$

	$\omega_2^z = g$	$\omega_2^z = b$
$\omega_1^z = g$	q	s
$\omega_1^z = b$	p	q

(b) $(a_1^z, a_2^z) = (C, D)$

Table 2: Joint signal distributions under nearly-perfect monitoring.

Stage game and monitoring structures

The stage game payoff is the sum of $g_i^z(a^z)$ for each market. Given $x > 0$ and $y > 0$, the payoff for each market z is defined in Tab. 1. We consider this prisoners' dilemma (PD) throughout this paper. Each player's private signal $\omega_i \in \{(g, g), (g, b), (b, g), (b, b)\}$. For example, when the opponent chooses C in each market, player i is more likely to receive correct signal $\omega_i = (g, g)$ in both markets, but sometimes observation error provides a wrong signal: $\omega_i = (g, b)$, (b, g), or (b, b).

In what follows, we describe some monitoring structures concentrating on a single market case without loss of generality. There are two representative classes of private monitoring: *nearly-perfect* and *almost-public*. Monitoring is nearly-perfect if each player can identify the action taken by her opponent with errors, and it is almost-public if all players receives the same public signal with errors. Due to the space limitation, let us focus on nearly-perfect monitoring

Table 2 shows the joint distribution of private signals $o^z(\omega^z \mid a^z)$ for PD under nearly-perfect monitoring. When the action profile is (C, C), the joint distribution is given in Tab. 2a (when the action profile is (D, D), p and s are exchanged). Notice that the probability that players 1 and 2 observe (g, g) is p, and the probability that they observe (g, b) is q. Similarly, when the action profile is (C, D), the joint distribution of the private signals is given in Tab. 2b When the action profile is (D, C), p and s are exchanged. These joint distributions of private signals require the constraints of $p + 2q + s = 1$. In addition, p is required to be much larger than q or s. Private monitoring is a generalization of perfect monitoring. In fact, if $q = 0$ and $s = 0$, Table 2 is equivalent to perfect monitoring.

In the theory and simulation of repeated games, TFT is much more popular than 1MP. However, since it does not prescribe mutual best replies after a deviation, it is *not* a subgame perfect Nash equilibrium (SPNE), but 1MP is under perfect monitoring. Under imperfect monitoring, whether an FSA can be a sequential equilibrium depends on the monitoring structure. Important lessons are that under almost-public monitoring, TFT can be an equilibrium, but not 1MP and that under nearly-perfect monitoring, 1MP can be an equilibrium, but not TFT (Iwasaki et al. 2014).

1-period mutual punishment under nearly-perfect monitoring

This section considers the 1-period mutual punishment (1MP) shown in Fig. 1. This FSA is also known as "win-stay, lose-shift (Nowak and Sigmund 1993)." According to this FSA, a player first cooperates. If her opponent defects, she also defects, but after one period of mutual defection, she returns to cooperation. We extend 1MP and consider a pre-FSA where each player behaves according to the 1MP for each market and let the FSA be the joint 1MP (J-1MP). This FSA is an important benchmark to consider multimarket contact.

Theorem 1 *Under nearly-perfect monitoring where p, q, $s = 1-p-2q$, and $p > s > q$, J-1MP forms a pure RFSE starting from (R,R) for each market if and only if*

$$\delta \geq \frac{(p-s+X)x+(p-s-X)y}{2(p-s)X} \quad \text{and} \tag{2}$$

$$\delta \geq \frac{Y+\sqrt{Y^2+4(p-s)(s-q)x}}{2(p-s)(s-q)} \tag{3}$$

both hold where $X = p+q-\sqrt{(q-s)^2+4pq}$ and $Y = -p+qy+(1+x)s$.

Theorem 1 is straightforwardly derived from the following lemma.

Lemma 1 *Under the same setting as Theorem 1, 1MP forms a pure RFSE starting from (R,R) if and only if Equations 2 and 3 hold.*

To the best of our knowledge, the necessary and sufficient condition under which 1MP is an equilibrium under private monitoring has not been identified, even in the case of a single market. Equations 2 and 3 are both related to a one-shot extension (deviation), which plays D in the current period and follows 1MP from state P from the next period on. Equation 2 requires that this deviation does not improve the payoff at a history where a player has observed a very long string of (D,b), and Equation 3 requires that it does not improve the payoff at the initial history. It is possible that one of the RHSs of Equations 2 and 3 is greater than or equal to one. If so, 1MP never forms an equilibrium.

Proof 1 ***If part:*** *Suppose Equations 2 and 3 hold. First, derive the expected discounted payoffs (we use continuation payoffs interchangeably) for each joint state of 1MP, which is either RR, RP, PR, or PP. Each is a solution of Eq. 1: v_i^{RR}, v_i^{RP}, v_i^{PR}, and v_i^{PP}.*

For given belief b_i, the continuation payoffs of the following 1MP starting at R and P are given by

$$V_i^{(m_i,R)}(b_i) = b_i v_i^{RR} + (1-b_i)v_i^{RP} \text{ and}$$
$$V_i^{(m_i,P)}(b_i) = b_i v_i^{PR} + (1-b_i)v_i^{PP}$$

which are illustrated in Fig. 3. Next, let b^* be a belief such that $V_i^{(m_i,R)}(b^*) = V_i^{(m_i,P)}(b^*)$: a player is indifferent as to which state to start, as long as she conforms to 1MP. Formally, b^* is uniquely derived so that

$$b^* = \frac{v_i^{PP}-v_i^{RP}}{v_i^{RR}-v_i^{RP}-v_i^{PR}+v_i^{PP}} = \frac{y+\delta(p-s)}{y-x+2\delta(p-s)}.$$

Clearly, $1/2 < b^*$ holds from $x > 0$ and $y > 0$. By transforming Eq. 2, we obtain

$$\delta(p-s)-x \geq \frac{\{\sqrt{(q-s)^2+4pq}-q-s\}(x+y)}{2\{p+q-\sqrt{(q-s)^2+4pq}\}}.$$

Since the denominator and the numerator are positive, $\delta(p-s) > x$ holds and thus $b^* < 1$ holds. Therefore, $b^* \in (1/2, 1)$.

Now we show that under any belief b_i and each of eight possible one-shot extensions $M_i \in Ex(m_i)$,

$$\max\{V_i^{(m_i,R)}(b_i), V_i^{(m_i,P)}(b_i)\} \geq V_i^{M_i}(b_i). \tag{4}$$

This is sufficient to prove the "if" part, because $V_i^{(m_i,R)}(b_i) \geq V_i^{(m_i,P)}(b_i)$ holds if and only if $b_i \geq b^*$. Hence, the belief divisions such that $\hat{D}_i^P = [0,b^*]$ and $\hat{D}_i^R = [b^*,1]$ for each $i = 1,2$ are target belief divisions, and they are covering and therefore accommodate $\boldsymbol{\theta} = (R,R)$. Accordingly, the 1MP profile forms an RFSE.

Note first that Equation 4 clearly holds if M_i is equivalent to (m_i,R) or (m_i,P): M_i plays C and shifts to state R after observing g and shifts to P after b, or M_i plays D and shifts to R after b and P after g.

Next, note that for any b_i, Bayes' rule implies

$$\chi_i\big[(C,b),b_i\big] = \frac{q}{b_i(s+q)+(1-b_i)(p+q)} < \frac{1}{2},$$

where the inequality follows from $p > s > q$. Similarly, we have $\chi_i\big[(D,g),b_i\big] < 1/2$. Since $1/2 < b^*$, this implies that any one-shot extension which shifts to state R when the current outcome is either (C,b) or (D,g) is dominated by one that instead shifts to state P after (C,b) or (D,g).

Consequently, it suffices to prove Eq. 4 for the following two one-shot extensions; playing D and always moving to state P (DP), and playing C and always moving to state P (CP).

If the continuation payoff of DP given belief b_i, $V_i^{DP}(b_i)$, satisfies $V_i^{(m_i,P)}(b_i) \geq V_i^{DP}(b_i)$ for any b_i, Equation 4 holds for DP. Suppose otherwise and $\bar{b} < 1$ exists such that $V_i^{(m_i,P)}(\bar{b})$ is equal to $V_i^{DP}(\bar{b})$. We have $\chi_i\big[(D,b),\bar{b}\big] = b^*$:

$$\frac{\bar{b}s+(1-\bar{b})p}{\bar{b}(q+s)+(1-\bar{b})(p+q)} = b^*.$$

This is because DP and (m_i,P) differ only after (D,b) and are indifferent if $\chi_i\big[(D,b),\bar{b}\big] = b^*$. Solving this, we obtain

$$\bar{b} = \frac{p-b^*(p+q)}{(1-b^*)(p-s)}.$$

Since $\chi_i\big[(D,b),b_i\big]$ is decreasing in b_i, $V_i^{(m_i,P)}(b_i)$ is smaller than or equal to $V_i^{DP}(b_i)$ if and only if $b_i \geq \bar{b}$. Hence Equation 4 holds if both $V_i^{DP}(1) \leq V_i^{(m_i,R)}(1)$ and $b^* \leq \bar{b}$ hold. Some calculations show that $V_i^{DP}(1) \leq V_i^{(m_i,R)}(1)$ is equivalent to Eq. 3, though we omit the details due to space limitations. $b^* \leq \bar{b}$ is equivalent to

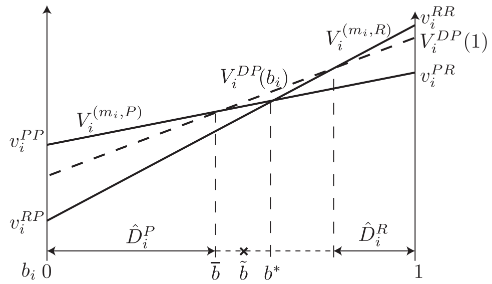


Figure 4: Expected discounted payoffs where Equation 2 is violated.

$(p-s)(b^*)^2 - (2p+q-s)b^* + p \geq 0$. Solving this, we obtain

$$b^* \leq \frac{2p+q-s-\sqrt{(q-s)^2+4pq}}{2(p-s)}.$$

Using the definition of b^* and solving for δ, we see that this is equivalent to Eq. 2.

Finally, for CP, due to the symmetry of the signal distributions, by applying a similar argument to DP, we can prove that Equation 4 holds if Equations 2 and 3 hold.

Only if part: *Suppose first that Equation 2 fails. Then $b^* > \bar{b}$ follows, which is shown in Fig. 4. Now no target belief division $\hat{D}_i$ is covering; $V_i^{DP}(b_i)$ exceeds $V_i^{(m_i,P)}$ on $(\bar{b}, b^*)$. We show that player i's belief is in the range under some history, i.e, no $\hat{D}_i$ accommodates $\boldsymbol{\theta}$.*

To this end, let $\tilde{b}$ be a belief such that $\chi_i\big[(D,b),\tilde{b}\big] = \tilde{b}$. Since $\chi_i\big[(D,b),b_i\big]$ is decreasing in b_i, $\tilde{b}$ uniquely exists. For any b_i and $b_i' \neq b_i$, we have

$$(b_i - b_i')\Big\{\chi_i\big[(D,b),b_i\big] - \chi_i\big[(D,b),b_i'\big]\Big\} < 0.$$

Letting $b_i = \bar{b}$ and $b_i' = \tilde{b}$ and recalling $\chi_i[(D,b),\bar{b}] = b^*$, we obtain $(\bar{b}-\tilde{b})(b^*-\tilde{b}) < 0$. Since $b^* > \bar{b}$, $\bar{b} < \tilde{b} < b^*$ follows.

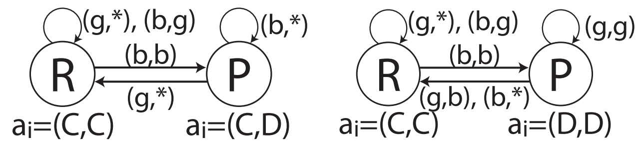
Let $b(t)$ be player i's belief when she observed (D,b) in all past t periods. For any t, we have

$$\begin{aligned} b(t+2) - b(t) &= \frac{b(t+1)s + \big\{1-b(t+1)\big\}p}{b(t+1)(q+s) + \big\{1-b(t+1)\big\}(p+q)} - b(t) \\ &= \frac{(p-s)\big\{b(t)\big\}^2 - (2p+q-s)b(t) + p}{\big\{b(t)s + \big(1-b(t)\big)p\big\}(q+s) + q(p+q)}(q+s). \end{aligned}$$

Since this equation is decreasing in $b(t)$ and zero at $b(t) = \tilde{b}$, it is nonnegative if $b(t) \leq \tilde{b}$ and nonpositive if $b(t) \geq \tilde{b}$. Therefore, $b(t)$ converges to $\tilde{b}$ as $t \to \infty$. This proves that player i's belief is in $(\bar{b}, b^*)$, as is desired.

Finally, suppose Equation 3 fails. Then DP clearly attains a greater payoff at the initial history, i.e., $b_i = 1$. Thus, this contradicts that 1MP forms an RFSE.

Note that the posterior belief that followed from $\bar{b}$ is a crucial role for 1MP to be an equilibrium. Indeed, when 1MP is not an equilibrium, at the posterior (in this case b^*), the corresponding one-shot extension is profitable. From this feature, we would like to consider a more efficient verification algorithm in future work.



(a) Locally-cautioning (b) Globally-cautioning

Figure 5: New FSAs for two markets: $*$ indicates g or b.

Locally cautioning

This section investigates novel FSAs in Figs. 5a and 5b, which we call *locally-cautioning* (LC) and *globally-cautioning* (GC). Unfortunately, it is extremely difficult to analyze these FSAs with equal clarity as in Theorem 1. Indeed, we have to find out which one-shot extensions are most relevant from the $4 \cdot 2^4 = 64$ one-shot extensions even when the number of markets is two (1MP has only eight) and verify that they do not improve the continuation payoff. Accordingly, we entrust the verification task to a numerical analysis.

In what follows, the stage game payoff is set from $x = y = 0.12$ and the discount factor δ is fixed at 0.9, while the parameters of the joint signal distributions p, q, and if necessary p' vary. s and s' are derived from those parameters. With these settings, using the verification method (Iwasaki et al. 2014) we exhaustively search for small-sized FSAs that are an equilibrium. We enumerate all possible FSAs with two or fewer states, i.e., $|A|^{|\Theta|} \cdot |\Theta|^{|\Theta|\cdot|\Omega|}$=4096 FSAs, and check whether they form an RFSE. For FSAs with three states, it is impossible to enumerate all possible FSAs because we require about 34 millions of them ($4^3 \times 3^{3\times 4} = 34{,}012{,}224$). Thus, an exhaustive search is currently intractable. However, we examined three state FSAs that we infer from LC and GC and will mention later. As a result of an exhaustive search with two state FSAs, we found that LC is the most efficient (achieves the highest average payoff) among the FSAs that are RFSE in a reasonably wide range of signal parameters and that GC is the most efficient among such FSAs that a player at state P defects in both markets.

In LC, a player first cooperates in both markets. If her opponent defects in both of them, she also defects in either of them. However, after one period of mutual cooperation in the market that she continues to cooperate, she returns cooperation. We say that she *locally* cautions (defects in either of markets) her opponent when she punishes him. In GC, a player at state R behaves similarly to LC. However, when a player punishes her opponent, she *globally* cautions her opponent (defects in all markets). After punishment, she continues punishment only if she observes cooperation in both markets. Otherwise, she returns to cooperation.

Under nearly-perfect monitoring, in addition to J-1MP, both LC and GC can be sequential equilibria (RFSE). Figure 6 illustrates the range of signal parameters over which LC is RFSE. For comparison, we show the ranges where J-1MP and GC are RFSE. We can see that LC and GC are RFSE in a reasonably wide range of signal parameters, though the range is smaller than J-1MP. When the proba-

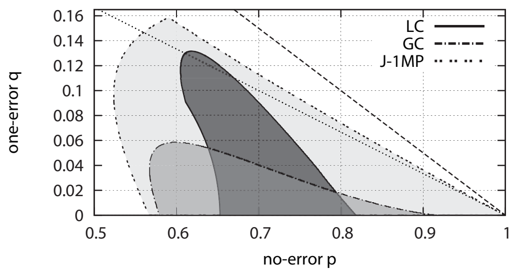


Figure 6: Range of signal parameters over which J-1MP/LC/GC is RFSE under nearly-perfect monitoring. Note that the feasible parameter space is $p + 2q \leq 1$.

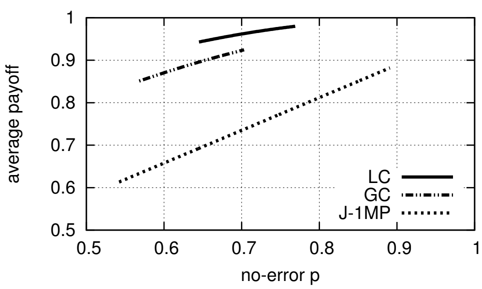


Figure 7: Average payoffs under nearly-perfect monitoring ($q = 0.04$).

bility that either agent receives the wrong signal (one-error) is quite small ($q \fallingdotseq 0$), LC is RFSE in the range of signal of no-error $p \in [0.65, 0.82]$. As the one-error probability becomes larger ($q > 0.13$), LC is no longer RFSE. GC is RFSE in a wider range of no-error probabilities than LC ($p \in [0.58, 0.91]$). However, GC is not RFSE for $q > 0.06$, which is smaller than the upper bound of LC. J-1MP is the most robust against noisy observation and completely overlays the regions of LC and GC. Note that J-1MP's region includes the case of $q > s$ which corresponds to the bottom-left side of $q = 1/3 - p/3$ in Fig. 6. The numerical verification also shows that Theorem 1 covers most of the J-1MP's region.

Now, let us examine the average payoffs. In Fig. 7, the x-axis indicates no-error probability p, while one-error probability q is fixed at 0.04. The y-axis indicates the average payoff per period and market. Note that an average payoff is 1 if mutual cooperation is always achieved. It is clear that LC significantly outperforms GC and GC outperforms J-1MP, regardless of the no-error probability. For example, when $p = 0.7$, these FSAs are RFSE simultaneously. LC achieves a payoff of 0.96, while GC and J-1MP achieve 0.92 and 0.74.

Let us next examine LC and GC with respect to Proposition 1. When LC is RFSE, the target belief division is almost covering and no profitable one-shot extension exists. Thus, the target belief division accommodates initial state (R, R); each player's belief is not outside the target belief division for any history starting at (R, R). The GC's target belief division is not covering. For example, for $p = 0.7$ and $q = 0.04$, the target belief division $\hat{D}_i$ is $\{\hat{D}_i^P, \hat{D}_i^R\} = \{[0, 0.41], [0.85, 1]\}$. Let us consider a case where a player with $b_i \in \hat{D}_i^R$ deviates to (C, D) and observes (b, g). If her opponent is at R, since she observes the bad signal in market A, her opponent observes b with higher probability $\frac{s}{q+s}$ than g with $\frac{q}{q+s}$. Likewise, in market B, her opponent is likely to observe b. She suspects that her opponent is moving to P in the next period. If her opponent is at P, the same reasoning leads to the fact that her opponent observes (g, g) and moves to P with high probability. The player has the belief that her opponent is at P in the next period for sure. Accordingly, after any deviation and observation, a player is sure which state her opponent is. The GC's target belief division accommodates (R, R) under nearly-perfect monitoring.

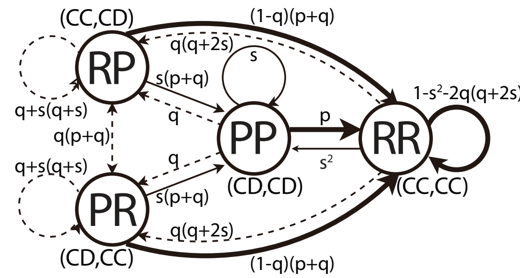


Figure 8: Joint FSA for LC under nearly-perfect monitoring

Discussion

LC is an equilibrium and attains a high level of payoffs even when the monitoring is nearly-perfect, since there exists no such a small-sized FSA that can be an equilibrium except the trigger strategy for the single market case.

Under perfect monitoring, neither LC nor GC is SPNE. For behavior specified by an FSA to be an equilibrium in repeated games, Equation 4 ensures that the current gain from each one-shot extension (deviation) never exceeds the future loss caused by that deviation. Hence, no future loss arises and she can safely exploit her opponent by deviating to (C, D) or (D, C) when each player is at R.

Under nearly-perfect monitoring, that consequence drastically changes. Figure 8 shows the joint FSA of LC. Under any joint state, the players cooperate in market A, and hence (g, g) is the most likely signal profile in that market. Since each player will be at R in the next period whenever his signal in market A is g, they are most likely to be at RR in the next period irrespective of the current joint state (the thick arrows in Figure 8). This tendency to quickly return to RR explains why LC attains a large payoff.

In LC, any deviation is punished by a cautioning in market B only. The large difference between p and s ensures that the punishment is sufficiently severe. This is true despite that, as we argued above, the cautioning in market B is likely to last only for one period. A primary reason is that the deviation gain x is small, which is just offset by one-period punishment.

Things are quite different for GC, which is an equilibrium only under near-perfect monitoring. Note that GC is similar to 1MP, in the sense that the transition of the states given

(g, g) or (b, b) exactly corresponds to that of 1MP given g or b, respectively. This correspondence explains why GC works under near-perfect monitoring, where 1MP also works well.

As we have observed, LC and GC devise state transitions by the combination of signals inherent in multimarket contact and realize high average payoffs in equilibria. One might think that a more complicated FSA improves the payoff, i.e., FSAs with more than three states are more efficient than LC. However, this is not true. Indeed, adding a new rewarding state, e.g., R' where a player chooses (C, C), is ineffective for achieving an equilibrium since it gives a player a chance to exploit her opponent and simply increases her incentive to defect. Conversely, adding a punishment state P' may lead to an equilibrium. However, such an FSA decreases the payoff because a player is likely to punish. Thus, more complicated FSAs would not perform better.

One natural question is that LC works under more than two markets. In the three market case, we confirm that some extensions of LC are an equilibrium in a range of signal parameters. For example, consider an extension such that a player chooses (C, C, C) at R and (C, C, D) at P. This extended FSA with appropriate state transitions achieves a significantly higher payoff than J-1MP. Another extension such that she chooses (C, D, D) at P is more likely to be an equilibrium than the first extension, but it lowers the payoff even though it remains higher than J-1MP.

Conclusions

This paper identifies simple pure strategy equilibria in repeated multimarket contact with a noisy signal about the opponent's action, which is one instance of private monitoring. Our contribution is twofold. First, we fully characterize a condition where J-1MP is an equilibrium by introducing several assumptions. Second, we propose LC, which incorporates observable signals and decides the actions in each market and show that it is an equilibrium. Its efficiency is the most efficient than any other FSAs through an exhaustive search. In future works, we would like to characterize the equilibrium condition for LC and to generalize its concept to an arbitrary number of markets.

References

Bernheim, B. D., and Whinston, M. D. 1990. Multimarket Contact and Collusive Behavior. *RAND Journal of Economics* 21(1):1–26.

Bhaskar, V., and Obara, I. 2002. Belief-based equilibria in the repeated prisoners' dilemma with private monitoring. *Journal of Economic Theory* 102(1):40 – 69.

Blum, A., and Monsour, Y. 2007. Learning, regret minimization, and equilibria. In *Algorithmic game theory*. Cambridge University Press. 79–101.

Borgs, C.; Chayes, J.; Immorlica, N.; Kalai, A.; Mirrokni, V.; and Papadimitriou, C. 2008. The myth of the folk theorem. In *Proceedings of the 40th Annual ACM Symposium on Theory of Computing (STOC)*, 365–372.

Burkov, A., and Chaib-draa, B. 2013. Repeated games for multiagent systems: A survey. *The Knowledge Engineering Review* 1–30.

Chellappaw, R. K.; Sambamurthy, V.; and Saraf, N. 2010. Competing in crowded markets: Multimarket contact and the nature of competition in the enterprise systems software industry. *Information Systems Research* 21(3):614–630.

Doshi, P., and Gmytrasiewicz, P. J. 2006. On the Difficulty of Achieving Equilibrium in Interactive POMDPs. In *Proceedings of the 21st National Conference on Artificial Intelligence (AAAI)*, 1131–1136.

Ely, J.; Horner, J.; and Olszewski, W. 2005. Belief-free equilibria. *Econometrica* 73(2):377–415.

Hansen, E. A.; Bernstein, D. S.; and Zilberstein, S. 2004. Dynamic programming for partially observable stochastic games. In *Proceedings of the Nineteenth National Conference on Artificial Intelligence (AAAI)*, 709–715.

Iwasaki, A.; Kandori, M.; Obara, I.; and Yokoo, M. 2014. A computationally feasible method for verifying equilibria in repeated games with private monitoring. mimeo, https://sites.google.com/site/a2ciwasaki/home/resilient-2014.pdf.

Kaelbling, L. P.; Littman, M. L.; and Cassandra, A. R. 1998. Planning and acting in partially observable stochastic domains. *Artificial Intelligence* 101:99–134.

Kandori, M., and Obara, I. 2010. Towards a Belief-Based Theory of Repeated Games with Private Monitoring: An Application of POMDP. mimeo, http://mkandori.web.fc2.com/papers/KObb10June4.pdf.

Kandori, M. 2010. Repeated games. In Durlauf, S. N., and Blume, L. E., eds., *Game theory*. Palgrave Macmillan. 286–299.

Kobayashi, H., and Ohta, K. 2012. Optimal collusion under imperfect monitoring in multimarket contact. *Games and Economic Behavior* 76(2):636–647.

Littman, M., and Stone, P. 2005. A polynomial-time Nash equilibrium algorithm for repeated games. *Decision Support Systems* 39(1):55–66.

Mailath, G., and Samuelson, L. 2006. *Repeated Games and Reputation*. Oxford University Press.

Nowak, M., and Sigmund, K. 1993. A strategy of win-stay, lose-shift that outperforms tit for tat in prisoner's dilemma. *Nature* 364:56–58.

Shoham, Y., and Leyton-Brown, K. 2008. Learning and teaching. In *Multiagent systems: Algorithmic, Game-Theoretic, and Logical Foundations*. Cambridge University Press. 189–222.

Wu, F.; Zilberstein, S.; and Chen., X. 2011. Online planning for ad hoc autonomous agent teams. In *Proceedings of the 22th International Joint Conference on Artificial Intelligence (IJCAI)*, 439–445.

Wunder, M.; Kaisers, M.; Yaros, J. R.; and Littman, M. 2011. Using iterated reasoning to predict opponent strategies. In *The 10th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, 593–600.

Yu, T., and Albert A. Cannella, J. 2013. A comprehensive review of multimarket competition research. *Journal of Management* 39:76–109.