CENTER FOR ADVANCED RESEARCH IN FINANCE

CARF Working Paper

CARF-F-432

Psychological Aspect of Monitoring Accuracy in Repeated Prisoners' Dilemma

Yutaka Kayaba
University of Tokyo

Hitoshi Matsushima
University of Tokyo

Tomohisa Toyama
International Christian University

April 24, 2018

CARF is presently supported by Dai-ichi Mutual Life Insurance Company, Nomura Holdings, Inc., Sumitomo Mitsui Banking Corporation, MUFG Bank, Finatext Ltd., The University of Tokyo Edge Capital Co., Ltd., Sompo Holdings, Inc., Tokio Marine & Nichido Fire Insurance Co., Ltd. and Nippon Life Insurance Company. This financial support enables us to issue CARF Working Papers.

Psychological Aspect of Monitoring Accuracy in Repeated Prisoners' Dilemma[1]

Yutaka Kayaba[2]
University of Tokyo

Hitoshi Matsushima[3]
University of Tokyo

Tomohisa Toyama[4]
International Christian University

April 24, 2018

Abstract

This study theoretically investigates an infinitely repeated prisoners' dilemma in which the monitoring technology is imperfect and private. In contrast to previous works, we shed light on the psychological aspect of monitoring imperfection rather than its informational aspect. We demonstrate a behavioral model in which a player is motivated not only by pure self-interest but also by social preferences such as reciprocity and naïveté. We then focus on the possibility that a generous tit-for-tat strategy, a simple Markovian stochastic strategy, satisfies equilibrium properties. We show that the prediction from the behavioral model is opposed to, but much more compatible with, daily experiences and existing experimental evidence than the prediction from the standard model with pure self-interest.

JEL Classification Numbers: C70, C71, C72, C73, D03.
Keywords: Repeated Prisoner's Dilemma, Imperfect Private Monitoring, Generous Tit-for-Tat, Retaliation Intensity, Reciprocity, Naïveté.

---

[1] This study is a modification of Section 9 in Kayaba, Matsushima, and Toyama (2017). Research for this study was financially supported by a grant-in-aid for scientific research (KAKENHI 25285059) from the Japan Society for the Promotion of Science (JSPS) and the Ministry of Education, Culture, Sports, Science and Technology (MEXT) of the Japanese government, as well as by the Center of Advanced Research in Finance at the University of Tokyo. We are grateful to the participants of seminars at the SAET conference, Hitotsubashi University, and the University of Tokyo for their useful comments. All errors are our own.
[2] Department of Economics, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan. Email: ykayaba@e.u-tokyo.ac.jp
[3] Department of Economics, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan. E-mail: hitoshi@e.u-tokyo.ac.jp
[4] International Christian University, College of Liberal Arts Division of Arts and Sciences, 3-10-2 Osawa, Mitaka-shi, Tokyo 181-8585, Japan. E-mail: tomohisa.toyama@gmail.com

1. Introduction

This study examines an infinitely repeated prisoners' dilemma with imperfect private monitoring; in each period of the infinite time horizon, each player cannot directly observe the opponent's action choice but can indirectly and privately observe it through a noisy signal. We investigate the impact of such noisy signal observations on players' strategic behavior.

Previous works in the repeated game literature have intensively investigated the *informational aspect* of monitoring imperfection. In contrast, this study will shed light on another aspect, i.e., the *psychological aspect*. We then argue that this psychological aspect influences players' strategic behavior in a manner opposite to that of the informational aspect. Importantly, the impact of noisy signal observations with respect to the psychological aspect is much more compatible with the daily experiences and existing experimental evidence than the impact with respect to the informational aspect. This study should be regarded as the first theoretical attempt to systematically analyze the psychological aspect of informational imperfection in the repeated game literature.

In our prisoners' dilemma model, each player can observe the *good* signal instead of the *bad* signal with a higher probability when the opponent makes the *cooperative* action choice than when the opponent makes the *defective* action choice. However, since the monitoring technology is imperfect, it is not certain that each player will observe the good signal even if the opponent makes the cooperative action choice. This monitoring imperfection inevitably interferes with the achievement of implicit collusion.

Despite this, theoretical studies have shown that sufficiently patient players can employ cooperative strategies as an equilibrium even if the monitoring is imperfect; the folk theorem holds even with imperfect monitoring, indicating that if the discount factor is close to unity, a wide variety of allocations can be attained by subgame perfect equilibrium (see Fudenberg, Levine, and Maskin, 1994; Matsushima, 2004; and Sugaya, 2012). In the respective proofs, however, equilibrium strategies were tailored to the fine details of signal histories in a complicated manner.

Even if players utilize only simple Markovian strategies, they can still collude with each other, not fully but partially, in imperfect private monitoring. In fact, partially collusive behaviors can be attained by the *generous tit-for-tat (g-TFT)* strategy equilibrium, a straightforward stochastic extension of the well-known tit-for-tat (TFT) strategy (e.g., Molander, 1985; Nowak and Sigmund, 1992; Takahashi, 2010; Matsushima, 2013). According to a g-TFT, a player retaliates against the opponent by selecting the defective action more often when she (or he) observes the bad signal than when she observes the good signal. This study will intensively consider g-TFT strategies that satisfy some equilibrium properties.

G-TFT should be regarded as the briefest manner of reciprocal behavior pattern that describes cooperation, retaliation, and forgiveness in repeated interactions. G-TFT has a great advantage over TFT; TFT generally fails to be a subgame perfect equilibrium, while, provided that the discount factor is sufficient, g-TFT equilibria always exists irrespective of the level of monitoring accuracy. Furthermore, TFT cannot escape the death spirals; once it falls into this spiral, players who follow TFT repeat the alternating play of cooperation and defection endlessly. G-TFT can avoid such spirals because it permits each player to change her action choice on a trial basis.

It is reasonable to expect that actual individual human beings and populations of animals randomly conduct such experimentations. In evolutionary biology, it was reported that animals maintain peaceful coexistence instead of the weak coast by adopting g-TFT (e.g., Molander, 1985; Nowak and Sigmund, 1992). In societies of human beings, TFT is expected to provide the opportunity to avoid the crisis of nuclear war and build a peaceful relationship. Importantly, the experimental studies by the companion paper, Kayaba, Matsushima, and Toyama (2018), reported that amongst a wide variety of strategies such as grim-trigger, lenience, long-term punishment strategies, and their stochastic variants, a significant proportion of experimental subjects follow g-TFT.

The purpose of this study is to theoretically examine what kind of g-TFT strategies players employ as equilibrium behavior. Specifically, we investigate how the level of monitoring accuracy influences their equilibrium behavior.

Previous works have investigated the standard model in which players behave as maximizers of pure self-interest. These works have focused on the informational aspect of monitoring imperfection; the more accurate the monitoring technology is, the more convinced a player, who observes the bad (good) signal, is that the opponent made the defective (cooperative) action choice. This implies that the more accurate the monitoring technology is, the more effectively a player retaliates against the opponent. In other words, with a higher monitoring accuracy, a player can more effectively penalize the deviant even if the retaliation that she employs is not very intensive (i.e., even if the retaliation intensity, the difference between the good signal observation and the bad signal observation in the cooperation rate, is not very large). According to the belief-free nature in which a player is always indifferent between the cooperative action choice and the defective action choice in equilibrium (e.g., Ely and Valimaki, 2002; Bhaskar and Obara, 2002; Piccione, 2002), the more accurate the monitoring technology is, the less severely each player retaliates against the opponent.

In contrast to these works, this study will shed light on the psychological aspect of monitoring imperfection in addition to the informational aspect; a noisy signal observation influences the observer's psychological state, motivating her (or his) social preferences. In particular, each player is often motivated not only by pure self-interest but also by *reciprocity*; a player feels guilty when she selects the defective action even though she observed the good signal, while she is annoyed when she selects the cooperative action even though she observed the bad signal. We further assume that each player often becomes *naïve* enough to select an action randomly.

By incorporating social preferences such as reciprocity and naïveté into players' incentives in addition to pure self-interest, this study introduces a *behavioral model* as an alternative to the standard model and then introduces a new equilibrium notion, *behavioral equilibrium*, as an alternative to the standard equilibrium notion. We argue that, in an accuracy-contingent behavioral equilibrium, the more accurate the monitoring technology is, the more severely each player retaliates against the opponent. This implies that the behavioral model describes a players' behavior pattern that is opposite to what the standard model describes because, in the standard model, the more accurate the monitoring technology is, the less severely each player retaliates against the opponent.

It is important to note that the behavioral equilibrium in the behavioral model is much more compatible with the daily experiences and existing experimental evidence than the equilibrium in the standard model. The companion paper, Kayaba, Matsushima, and Toyama (2018), experimentally reported that subjects tend to retaliate more in the high accuracy treatment than in the low accuracy treatment. This experimental indication contradicts the prediction from the standard model, while it is more consistent with the prediction from the behavioral model.

Furthermore, Kayaba, Matsushima, and Toyama (2018) experimentally reported that subjects retaliate more in the high accuracy treatment, while they retaliate less in the low accuracy treatment than the g-TFT equilibria predict in the standard model. Hence, the expected payoff to an individual subject from cooperation tends to be greater than that from defection when the monitoring is accurate, while the expected payoff from cooperation tends to be less than that from defection when the monitoring is inaccurate. Their experimental work also reported that subjects tend to progress in their learning process quite slowly. These experimental findings suggest to us a new incentive issue that encompasses motivations for retaliation beyond just maximizing pure self-interest. This study will fill the gap between theory and reality by considering social preferences in addition to pure self-interest.

According to our main theorem, we can see that a player is well-motivated by pure self-interest when the level of monitoring accuracy is medium, while she is well-motivated by reciprocity when the level of monitoring accuracy is either poor or rich. Importantly, a player tends to behave consciously when she is well-motivated by reciprocity, while she is likely to be naïve, or unconscious when she is motivated by pure self-interest. This implies that it is not pure self-interest, but the psychological aspect, that plays the central role in a player being conscious in decision making.

The experimental literature has noted that social preferences facilitate cooperation (e.g., Güth, Schmittberger, and Schwarze, 1982; Berg, Dickhaut, and McCabe, 1995; Fehr and Gächter, 2000). The literature assumed that preferences depend on various contexts (e.g., Rabin 1993; Charness and Rabin, 2002; Dufwenberg and Kirchsteiger, 2004; Falk and Fishbacher, 2005). In this respect, this study makes relevant contexts parameterized by the level of monitoring accuracy.

The behavioral model of this study and the model in Duffy and Muñoz-García (2012) share a common characteristic in that social preferences play an important role in people undertaking collusion in repeated games. The implications of social preferences, however, differ. Duffy and Muñoz-García (2012) demonstrated that social preference facilitates collusion when the discount factor is insufficient. By contrast, in our study, the monitoring technology is a crucial determinant of whether social preferences aid collusion; social preferences serve to facilitate collusion when monitoring is inaccurate, while they inhibit people from colluding when monitoring is accurate.

This study does not assert that the actual behavior is literally an equilibrium. This study does not assert that TFT, or its variants such as g-TFT, survives as the most successful strategy like the round-robin tournament experiments and evolutionary simulations like Axelrod (1984). The motivation of this study is that it is inevitably necessary to consider a new viewpoint of players' incentives based on the fact that the behavior of the experimental subject is far from the standard theory and it is not only because they are merely making wrong predictions.

The rest of this study is organized as follows. Section 2 presents the standard model and introduces g-TFT strategy. Section 3 presents the behavioral model and introduces behavioral equilibrium. Section 4 characterizes the behavioral model that is consistent with an accuracy-contingent g-TFT strategy. Section 5 provides discussion of the characterization theorem in Section 4. Section 6 concludes.

2. Standard Model and Generous Tit-For-Tat

This study investigates an infinitely repeated game played by two players (i.e., players 1 and 2). Specifically, we consider a *prisoners' dilemma with symmetry and additive separability* as the component game, which is described in Figure 1. We assume $g > 0$ in this figure.

Figure 1

		Player 2: C	Player 2: D
Player 1	C	$1 \quad 1$	$-g \quad 1+g$
	D	$1+g \quad -g$	$0 \quad 0$

Let us call C and D the *cooperative* action and *defective* action, respectively. Additive separability implies that, irrespective of the opponent's action choice, selecting the cooperative action C instead of the defective action D costs g, but gives the opponent the benefit $1+g$. Note that the cooperative action profile (C,C) maximizes their welfare, while the defective action profile (D,D) is a dominant strategy profile and Pareto-inferior to (C,C) (i.e., $(0,0)<(1,1)$).

We assume that monitoring is *imperfect* and *private*. Each player i cannot directly observe the action that the opponent $j\neq i$ has selected. Instead, she (or he) privately observes a noisy signal for the opponent $j's$ action choice, which is denoted by $\omega_j\in\{c,d\}$. Let us call c and d the *good* and *bad* signals, respectively. We define the level of *monitoring accuracy* as the probability index $p\in(\frac{1}{2},1)$; player i observes the good signal c (the bad signal d) with probability p when the opponent j selects the cooperative action C (the defective action D). The greater p is, the more accurately each player can monitor the opponent's action choice. Hence, inequality $p>\frac{1}{2}$ implies that the probability of the good signal c occurring for the corresponding player's action choice is greater when this player selects C than when she selects D.

Let us denote by $\delta\in(0,1)$ the *discount factor*. The solution concept that this section employs for the infinitely repeated prisoners' dilemma is the standard notion of subgame perfect equilibrium (*equilibrium*), where this section assumes the standard model in which each player is solely motivated by her pure self-interest. (The next sections replace the standard model with a behavioral model where each player is motivated not only by pure self-interest but by social preferences.)

We introduce the *generous tit-for-tat (g-TFT) strategy*, which is denoted by $(r(c),r(d))\in[0,1]^2$; at each period $t\geq 2$, player i makes the cooperative action choice C with probability $r(\omega_j)$ when she observes signal $\omega_j\in\{c,d\}$ in the previous period $t-1$. (To eliminate irrelevant complexity and focus on the incentive to make signal-dependent action choices, we ignore the incentive issue in the first period.) We then focus on g-TFT equilibria. Consider an arbitrary period $t\geq 2$. Suppose that both players employ the same g-TFT strategy $(r(c),r(d))$ from the next period $t+1$. Then, a player $i's$ selecting C instead of D costs her g in the current period t, whereas in the next period $t+1$, she can gain $1+g$ from the response of the opponent $j\neq i$ with probability $pr(c)+(1-p)r(d)$ instead of probability $(1-p)r(c)+pr(d)$.

Since she must be incentivized to select both actions C and D at once (because of the belief-free nature), the indifference between these action choices at all times must be a necessary and sufficient condition for equilibrium:

$$-g+\delta(1+g)\{pr(c)+(1-p)r(d)\}=\delta(1+g)\{(1-p)r(c)+pr(d)\},$$

that is,

$$r(c)-r(d)=\frac{g}{\delta(2p-1)(1+g)}.$$

Because $\delta<1$ and $r(c)-r(d)\leq 1$, the following inequality must hold:

$$\delta>\frac{g}{(2p-1)(1+g)}.$$

We define the *retaliation intensity* of a g-TFT strategy $(r(c),r(d))$ as the difference in the cooperation rate between the good and bad signals:

$$r(c)-r(d).$$

The greater the retaliation intensity $r(c)-r(d)$, the more severely players retaliate against their opponents. We further define

$$w(p)\equiv\frac{g}{\delta(2p-1)(1+g)}.$$

Note that the retaliation intensity of a g-TFT equilibrium (i.e., the equilibrium retaliation intensity $r(c)-r(d)$), is equal to $w(p)$. It is important to note that the equilibrium retaliation intensity $w(p)$ is *decreasing in* p ; the less accurate the monitoring technology is, the more severely players retaliate against their opponents.

This decreasing property is essential for understanding how players overcome the difficulty of achieving cooperation under imperfect private monitoring. To incentivize a player to make the cooperative action choice, it is necessary that her opponent makes the defective action choice when observing the bad signal more often than when observing the good signal. When monitoring is less accurate, it is more difficult for her opponent to detect whether the player makes the cooperative action choice or the defective action choice. In this case, the enhancement in retaliation intensity is necessary to incentivize the player. Hence, the equilibrium retaliation intensity in the standard model $w(p)$ must be decreasing at the level of monitoring accuracy p.

3. Behavioral Model and Accuracy-Contingent G-TFT

This section introduces a *behavioral model* of the infinitely repeated prisoners' dilemma by incorporating psychological aspects (i.e., social preferences such as reciprocity and naïveté) into the standard model; each player is motivated not only by pure self-interest but also by *reciprocity*. Each player often becomes *naïve* enough to select actions randomly. By incorporating reciprocity and naïveté into players' incentives in addition to pure self-interest, we will introduce the notion of *behavioral equilibrium* as an alternative to the standard equilibrium notion.

Fix an arbitrary level of monitoring accuracy $\underline{p} \geq \frac{1}{2}$ as the minimum level. We then define an *accuracy-contingent g-TFT strategy* as

$$(r(c;p), r(d;p))_{p\in(\underline{p},1)},$$

where we assume that, irrespective of monitoring accuracy $p \in (\underline{p},1)$, a player selects both actions C and D with positive probabilities:

(1) $$0 < r(c;p) < 1 \text{ and } 0 < r(d;p) < 1.$$

For each level of monitoring accuracy $p \in (\underline{p},1)$, a player makes stochastic action choices according to the corresponding g-TFT strategy $(r(c), r(d)) = (r(c;p), r(d;p))$. We permit the retaliation intensity $r(c;p) - r(d;p)$ dependent on the monitoring accuracy p.

We introduce *naïveté* as a psychological aspect; at every period, with probability $2\varepsilon(p) \in [0, \frac{1}{2}]$, a player who becomes naive randomly selects between actions C and D:

$$\min[r(c;p), 1-r(c;p), r(d;p), 1-r(d;p)] \geq \varepsilon(p).$$

With the remaining probability $1-2\varepsilon(p)$, the player makes the action choice in a conscious manner.[5]

[5] To calm the tense relationship between rationality and empirical data, economic theory and empirics have used stochastic choice models, such as logit and probit models, that incorporate random error into the equilibrium analysis. In the model of quantal response equilibrium, it is assumed that the deviation errors from the optimal action choice are negatively correlated with the resultant payoffs (e.g., Goerree, Holt, and Palfrey, 2008). By contrast, we assume that the deviation errors induced by naïveté are independent of either the resultant payoff or the observed signal but depend on the level of

We further introduce *reciprocity* as another psychological aspect. Suppose that a player observes the good signal c. She feels guilty when she selects the defective action D despite the observation of the good signal c. In this case, she can save the psychological cost $s(c;p) \geq 0$ by selecting the cooperative action C. Hence, the instantaneous gain from selecting action D should be equal to $g - s(c;p)$, while the resultant future loss is equal to $\delta(1+g)(2p-1)\{r(c;p) - r(d;p)\}$, the same value as in the standard model.[6]

Next, suppose that a player observes the bad signal d. She is annoyed when she selects the cooperative action C despite the observation of the bad signal d. In this case, she can save the psychological cost $s(d;p) \geq 0$ by selecting the defective action D. Hence, the instantaneous gain from D should be equal to $g + s(d;p)$, while the resultant future loss is equal to $\delta(1+g)(2p-1)\{r(c;p) - r(d;p)\}$, the same value as in the standard model.

Based on the above arguments, we define a *behavioral model* as

$$(\varepsilon(p), s(c,p), s(d;p))_{p \in (\underline{p},1)},$$

where we assume that $s(c;p)$ and $s(d;p)$ are continuous in p. In the behavioral model $(\varepsilon(p), s(c,p), s(d;p))_{p \in (\underline{p},1)}$, a player is said to be *positively reciprocal* for monitoring accuracy p if

$$s(c;p) > 0.$$

She is said to be *negatively reciprocal* for p if

$$s(d;p) > 0.$$

She is said to be *null-reciprocal* for p if

$$s(c;p) = 0 \text{ and } s(d;p) = 0.$$

We assume that

$$\text{either } s(c;p) = 0 \text{ or } s(d;p) = 0.$$

With this assumption, we can divide players into three categories: *positively reciprocal*

---

monitoring accuracy.

6 To eliminate irrelevant complexity, this study excludes the impact of the current action choice on the psychological mode in the future from the calculation of the future loss.

players, negatively reciprocal players, and *null-reciprocal players*.

When the instantaneous gain is greater (less) than the future loss, any player who is not naïve has an incentive to select the defective action D (the cooperative action C). Based on this, we shall call an accuracy-contingent g-TFT strategy $(r(a;p), r(b;p))_{p\in(\underline{p},1)}$ a *behavioral equilibrium* in the behavioral model $(\varepsilon(p), s(c,p), s(d;p))_{p\in(\underline{p},1)}$ whenever the following four properties hold for all $p\in(\underline{p},1)$:

(2) $$[\,g - s(c;p) > \delta(1+g)(2p-1)\{r(c;p)-r(d;p)\}\,] \Rightarrow [\,r(c;p)=\varepsilon(p)\,],$$

(3) $$[\,g - s(c;p) < \delta(1+g)(2p-1)\{r(c;p)-r(d;p)\}\,] \Rightarrow [\,r(c;p)=1-\varepsilon(p)\,],$$

(4) $$[\,g + s(d;p) > \delta(1+g)(2p-1)\{r(c;p)-r(d;p)\}\,] \Rightarrow [\,r(d;p)=\varepsilon(p)\,],$$

and

(5) $$[\,g + s(d;p) < \delta(1+g)(2p-1)\{r(c;p)-r(d;p)\}\,] \Rightarrow [\,r(d;p)=1-\varepsilon(p)\,].$$ [7]

These properties imply that if the instantaneous gain is greater (less) than the future loss, then the player selects the cooperative action (the defective action) only with the minimal probability $\varepsilon(p)$. Clearly, the notion of behavioral equilibrium is the same as the standard equilibrium notion when the behavioral model is degenerate:

$$(\varepsilon(p), s(c,p), s(d;p)) = (0,0,0) \quad \text{for all} \quad p\in(\underline{p},1).$$

4. Characterization of Behavioral Model

According to the experimental evidence of Kayaba, Matsushima, and Toyama (2018) and our daily experiences, let us consider an accuracy-contingent g-TFT strategy $(r(c,p), r(d;p))_{p\in(\underline{p},1)}$ that satisfies the following properties that are supported by their experimental work:

(i) both $r(c;p)$ and $r(d;p)$ are increasing and continuous in p

and

(ii) the retaliation intensity $r(c;p)-r(d;p)$ is increasing in p.

---

[7] Without any substantial change, we can interpret $s(c;p)$ as the psychological benefit, whereby the player feels better by selecting the cooperative action instead of the defective action after observing the good signal. The same interpretation can apply to $s(d;p)$.

Property (ii) implies that the more accurate the monitoring technology is, the more often each player makes the cooperative action choice. Property (ii) implies that the more accurate the monitoring technology is, the more severely each player retaliates against the opponent's defection. Any g-TFT equilibrium in the standard model fails to satisfy property (ii) because $w(p)$ is decreasing in p.

Since $w(p)$ is decreasing in p, there is a critical level $\hat{p} \in [\underline{p}, 1]$ such that

$$r(c;p) - r(d;p) > w(p) \qquad \text{if} \quad p > \hat{p},$$

and

$$r(c;p) - r(d;p) < w(p) \qquad \text{if} \quad p < \hat{p}.$$

The retaliation intensity $r(c;p) - r(d;p)$ is greater than the equilibrium retaliation intensity $w(p)$ if the monitoring accuracy p is greater than the critical level $\hat{p}$, while the retaliation intensity is less than the equilibrium retaliation intensity if the level of monitoring accuracy is worse than the critical level.

The following theorem shows that the abovementioned behavioral equilibrium constraints, (i.e., (2), (3), (4), and (5)) *uniquely* determine the underlying behavioral model.

The Theorem: *The accuracy-contingent g-TFT strategy* $(r(c;p), r(d;p))_{p \in (\underline{p},1)}$ *is a behavioral equilibrium in the behavioral model* $(\varepsilon(p), s(c,p), s(d;p))_{p \in (\underline{p},1)}$ *if and only if*

(6) $$r(c;p) = 1 - \varepsilon(p) \quad \textit{and}$$

$$r(c;p) = 1 - \varepsilon(p) - w(p) - \frac{s(c;p)}{\delta(1+g)(2p-1)} \quad \textit{for all} \quad p > \hat{p}$$

and

(7) $$r(c;p) = \varepsilon(p) + w(p) - \frac{s(c;p)}{\delta(1+g)(2p-1)} \quad \textit{and}$$

$$r(d;p) = \varepsilon(p) \quad \textit{for all} \quad p < \hat{p}.$$

Proof: The proof of the "if" part is straightforward from (2), (3), (4), (5), (6), and (7).

Fix an arbitrary $p \in (\underline{p},1)$. From continuity, we can assume without loss of generality that $1 - r(c;p) \neq r(d;p)$.

Suppose $p > \hat{p}$:

$$r(c;p) - r(d;p) > w(p) > 0.$$

This, along with $w(p) \equiv \frac{g}{\delta(1+g)(2p-1)}$ and $s(c;p) \geq 0$, implies inequality (3):

$$r(c;p) = 1 - \varepsilon(p) \text{ and } r(d;p) < 1 - \varepsilon(p).$$

Hence, from (4) and (5), either $r(d;p) = \varepsilon(p)$ or

$$g + c(d;p) = \delta(1+g)(2p-1)\{r(c;p) - r(d;p)\},$$

which implies (6). Note that $r(b;p) \neq \varepsilon(p)$ holds because $r(c;p) = 1 - \varepsilon(p)$ and $1 - r(d;p) \neq r(d;p)$.

Next, suppose $p < \hat{p}$:

$$r(c;p) - r(d;p) < w(p).$$

This, along with $w(p) \equiv \frac{g}{\delta(1+g)(2p-1)}$ and $s(d;p) \geq 0$, implies inequality (4), and therefore,

$$r(d;p) = \varepsilon(p) \text{ and } r(c;p) > \varepsilon(p).$$

Hence, from (2) and (3), either $r(c;p) = 1 - \varepsilon(p)$ or

$$g - s(a;p) = \delta(1+g)(2p-1)\{r(c;p) - r(d;p)\},$$

which implies (7). Note that $r(c;p) \neq 1 - \varepsilon(p)$ holds because $r(d;p) = \varepsilon(p)$ and $1 - r(c;p) \neq r(d;p)$.

Q.E.D.

From the theorem, the behavioral model $(\varepsilon(p), s(c,p), s(d;p))_{p \in (\underline{p},1)}$ is *uniquely* determined; for every $p > \hat{p}$

$$\varepsilon(p) = 1 - r(c;p),$$

$$s(c;p) = 0$$

and

$$s(d;p) = \delta(1+g)(2p-1)\{r(c;p) - r(d;p) - w(p)\},$$

and for every $p < \hat{p}$,

$$\varepsilon(p) = r(d;p),$$

$$s(c;p) = \delta(1+g)(2p-1)\{w(p) - r(c;p) + r(d;p)\},$$

and

$$s(d;p) = 0.$$

The uniquely derived behavioral model has the following properties:

(iii) The player is null-reciprocal for the critical level $\hat{p}$:

$$s(c;\hat{p}) = s(d;\hat{p}) = 0.$$

She is positively reciprocal for any monitoring accuracy that is greater than the critical level $\hat{p}$:

$$s(c;p) > 0 \text{ and } s(d;p) = 0 \text{ for all } p > \hat{p}.$$

She is negatively reciprocal for any monitoring accuracy that is less than the critical level $\hat{p}$:

$$s(c;p) = 0 \text{ and } s(d;p) > 0 \text{ for all } p < \hat{p}.$$

(iv) The parameter of *naïveté* $\varepsilon(p)$ is *single-peaked* with the peak at the critical level $\hat{p}$; $\varepsilon(p)$ is increasing in $p \in (\underline{p}, \hat{p})$ and decreasing in $p \in (\hat{p}, 1)$.

(v) The psychological cost of feeling guilty $s(c;p)$ is decreasing in $p \in (\underline{p}, \hat{p})$. The more accurate the monitoring technology is, the less positively the player is reciprocal.

(vi) The psychological cost of being annoyed $s(d;p)$ is increasing in $p \in (\hat{p}, 1)$. The more accurate the monitoring technology is, the more negatively the player is reciprocal.

In the next section, we will discuss the implication of the Theorem in more details.

5. Discussion

5.1. Accuracy and Kindness

A behavioral model $(\varepsilon(p), s(c,p), s(d;p))_{p\in(\underline{p},1)}$ is said to be *more kind* in p than in p' if either

it is more positively reciprocal in p than in p':

$s(c; p) > s(d; p')$,

or

it is less negatively reciprocal in p than in p':

$s(d; p') > s(d; p)$.

In other words, a player who is kind in our terminology rarely gets angry at the opponent by observing the bad signal and often praises the opponent by observing the good signal. From the Theorem, the behavioral model has the following trade-off between accuracy and kindness:

(vii) The less kind a player is, the more accurate the monitoring technology is.

Given a sufficient level of monitoring accuracy, a player tends to be more negatively reciprocal as monitoring is more accurate. This tendency makes the retaliation intensity severer, and, therefore, works against the better success in cooperation caused by the improvement of monitoring technology.

Given an insufficient level of monitoring accuracy, a player tends to be more positively reciprocal as monitoring is less accurate. This tendency makes the retaliation intensity milder and thereby mitigates the worse success in cooperation caused by the deterioration of monitoring technology.

5.2. Naïveté and Reciprocity

From the Theorem, the behavioral model has the following trade-off between naïveté and reciprocity:

(viii) The more likely a player makes the action choice randomly, or unconsciously, the less reciprocal she is.

When a player is more negatively reciprocal, she tends to be more conscious; she is less likely to mistakenly select the defective action despite observing the good signal as she is more negatively reciprocal. When a player is more positively reciprocal, she tends to be more conscious; she is less likely to mistakenly select the cooperative action despite observing the bad signal as she is more positively reciprocal.

In other words, a player tends to behave more consciously as she is more motivated by reciprocity. A player tends to behave consciously as the level of monitoring accuracy is either poor or rich; she tends to be naive, or unconscious, as the level of monitoring accuracy is medium. Hence, a player's conscious decision making is motivated not by her pure self-interest but by her reciprocal social preference.

5.3. Uniqueness of Behavioral Equilibrium

This subsection shows that the accuracy-contingent g-TFT strategy $(r(c;p),r(d;p))_{p\in(\underline{p},1)}$ is the *only plausible* accuracy-contingent g-TFT behavioral equilibrium in the corresponding behavioral model $(\varepsilon(p),s(c,p),s(d;p))_{p\in(\underline{p},1)}$. Note that we have shown the uniqueness of the behavioral model in Section 4, while this subsection will demonstrate the uniqueness of the behavioral equilibrium.

If $\underline{p}<\hat{p}$, there is no other accuracy-contingent g-TFT behavioral equilibrium strategy. If $\underline{p}=\hat{p}$, however, there is another accuracy-contingent g-TFT behavioral equilibrium, which is specified as $(\tilde{r}(c;p),\tilde{r}(d;p))_{p\in(\underline{p},1)}$ where

$$\tilde{r}(c;p)=\varepsilon(p)+w(p) \text{ and } \tilde{r}(d;p)=\varepsilon(p) \text{ for all } p\in(\underline{p},1).$$

(Note that $(\tilde{r}(c;p),\tilde{r}(d;p))_{p\in(\underline{p},1)}$ and $(r(c;p),r(d;p))_{p\in(\underline{p},1)}$ are the only behavioral g-TFT equilibria.) However, $(\tilde{r}(c;p),\tilde{r}(d;p))_{p\in(\underline{p},1)}$ is implausible, because it is inconsistent with the existing experimental evidence; it does not satisfy properties (i) and (ii). Both $\tilde{r}(c;p)$ and $\tilde{r}(d;p)$ are decreasing, and the retaliation intensity $\tilde{r}(c;p)-\tilde{r}(d;p)$ is decreasing, in monitoring accuracy p. Note also that $(\tilde{r}(c;p),\tilde{r}(d;p))_{p\in(\underline{p},1)}$ is less efficient than $(r(c;p),r(d;p))_{p\in(\underline{p},1)}$ irrespective of $p\in(\underline{p},1)$. Hence, we can safely ignore $(\tilde{r}(c;p),\tilde{r}(d;p))_{p\in(\underline{p},1)}$.

The above-mentioned uniqueness of (plausible) behavioral equilibrium is in contrast with the standard model; the standard model has numerous g-TFT equilibria.

6. Conclusion

This study investigated the psychological aspect of monitoring imperfection in a repeated prisoners' dilemma. We demonstrated a behavioral model by incorporating social preferences such as reciprocity and naïveté into the standard model with pure self-interest. According to the existing experimental evidence and daily experiences, we intensively studied the possibility of a simple Markovian stochastic strategy termed g-TFT strategy being a behavioral equilibrium; to make our theoretical analysis consistent with experimental evidence, we focused on accuracy-contingent g-TFT strategies such that the more accurate the monitoring technology is, the more severely a player retaliates against the opponent. We showed that the standard model fails to support such a plausible g-TFT strategy as an equilibrium, while we can uniquely characterize the behavioral model that can support this g-TFT strategy as the unique plausible behavioral g-TFT equilibrium. The behavioral model derived in our characterization implies that the more accurate the monitoring technology is, the more kindly a player behaves; the more conscious a player behaves, the more a player is motivated by reciprocal social preference; and a player tends to be conscious (naïve) as the level of monitoring accuracy is either poor or rich (medium).

This study should be regarded as the first theoretical attempt to systematically analyze the psychological aspect of monitoring imperfection in the repeated game literature. Therefore, there are many issues remaining after this study. For example, the existing experimental evidence notes the diversity of strategies among subjects, even if subjects' strategies are commonly described as g-TFT. It is important to question how to theoretically explain this diversity. It is also important to extend this study to a more general class of strategic interactions, rather than the prisoners' dilemma. These investigations are beyond the scope of this study.

References

Axelrod, R. (1984): *The Evolution of Cooperation*, Basic Books: New York.

Berg, J., J. Dickhaut, and K. McCabe (1995): "Trust, Reciprocity, and Social History," *Games and Economic Behavior* 10, 122–142.

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

Charness, G., and M. Rabin (2002): "Understanding Social Preferences with Simple Tests," *Quarterly Journal of Economics* 117(3), 817-869.

Duffy, J., and F. Muñoz-García (2012): "Patience or Fairness? Analyzing Social Preferences in Repeated Games," *Games* 3(1), 56–77.

Dufwenberg, M., and G. Kirchsteiger (2004): "A Theory of Sequential Reciprocity," *Games and Economic Behavior* 47(2), 268-298.

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

Falk, A., and U. Fischbacher (2005): "Modeling Fairness and Reciprocity," in Moral Sentiments and Material Interests: *The Foundations of Cooperation in Economic Life*. H. Gintis, S. Bowles, R. Boyd, and E. Fehr, eds. Cambridge: MIT Press.

Fehr, E., and S. Gächter (2000): "Fairness and Retaliation: The Economics of Reciprocity," *Journal of Economic Perspectives* 14(3), 159-181.

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

Goeree, J., C. Holt, and T. Palfrey (2008): "Quantal Response Equilibrium," *New Palgrave Dictionary of Economics*, Second Edition, Edited by S. Durlauf and L. Blume, Palgrave Macmillan.

Güth, W., R. Schmittberger, and B. Schwarze (1982): "An Experimental Analysis of Ultimatum Bargaining," *Journal of Economic Behavior and Organization* 3, 367–388.

Kayaba, Y., H. Matsushima, and T. Toyama (2018): "Accuracy and Retaliation in Repeated Games with Imperfect Private Monitoring: Experiments and Theory," Discussion Paper CARF-F-433, University of Tokyo (Revise-and-Resubmit to *Game and Economic Behavior*).

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

Matsushima, H. (2013): "Interlinkage and Generous Tit-For-Tat Strategy," *Japanese Economic Review* 65, 116–121.

Molander, P. (1985): "The Optimal Level of Generosity in a Selfish Uncertain

Environment," *Journal of Conflict Resolution* 29, 611-618.

Nowak, M. and K. Sigmund (1992): "Tit-For-Tat in Heterogeneous Populations," *Nature* 355, 250–253.

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

Rabin, M. (1993): "Incorporating Fairness into Game Theory and Economics," *American Economic Review* 83(5), 1281-1302.

Sugaya, T. (2012): *The Folk Theorem in Repeated Games with Private Monitoring*, Ph.D. dissertation.

Takahashi, S. (2010): "Community Enforcement when Players Observe Partners' Past Play," *Journal of Economic Theory* 145, 42–64.