CARF Working Paper

CARF-F-452

Behavioral Theory of Repeated Prisoner's Dilemma: Generous Tit-For-Tat Strategy

Hitoshi Matsushima
University of Tokyo

February 22, 2019
This Version: November 2, 2019

CARF is presently supported by The Dai-ichi Life Insurance Company, Limited, Nomura Holdings, Inc., Sumitomo Mitsui Banking Corporation, MUFG Bank, The Norinchukin Bank, The University of Tokyo Edge Capital Co., Ltd., Finatext 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.

Behavioral Theory of Repeated Prisoner's Dilemma: Generous Tit-For-Tat Strategy[1]

Hitoshi Matsushima[2]
University of Tokyo

February 22, 2019
This Version: November 2, 2019

Abstract

This study investigates infinitely repeated games of a prisoner's dilemma with additive separability in which the monitoring technology is imperfect and private. Behavioral incentives indicate that a player is not only motivated by pure self-interest but also by social preference such as reciprocity, and that a player often becomes naïve and selects an action randomly due to her cognitive limitation and uncertain psychological mood as well as the strategic complexity caused by monitoring imperfection and private observation. By focusing on generous tit-for-tat strategies, we characterize a behavioral version of Nash equilibrium termed behavioral equilibrium in an accuracy-contingent manner. By eliminating the gap between theory and evidence, we show that not pure self-interest but reciprocity plays a substantial role in motivating a player to make decisions in a sophisticated manner.

JEL Classification Numbers: C70, C71, C72, C73, D03.

Keywords: Repeated Prisoner's Dilemma, Imperfect Private Monitoring, Generous Tit-for-Tat Strategy, Reciprocity, Naïveté, Behavioral Equilibrium.

---

[1] This research 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 Yutaka Kayaba and Tomohisa Toyama for their support in drafting an earlier version of this study. All errors are mine.

[2] Department of Economics, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan. E-mail: hitoshi@e.u-tokyo.ac.jp

1. Introduction

This study examines the impact of behavioral incentives on equilibrium outcomes in infinitely repeated games of a prisoner's dilemma with the additive separability of payoff matrix. We assume that the monitoring technology is imperfect and private. Each player cannot directly observe whether the opponent selects the cooperative action or the defective action. However, each player can imperfectly monitor the opponent's action choice through private observation of a noisy signal. The signal is either "good" or "bad," and a player is more likely to observe the good signal when the opponent selects the cooperative action. In this setting, this study investigates the role of noisy signal observations on implicit collusion.

Previous works in the repeated game literature have investigated the impact of monitoring accuracy on the degree to which the history-dependence of players' strategic behaviors facilitates their implicit collusion.[3] Since the monitoring technology is imperfect, it is not certain that each player will receive the good signal when the opponent makes the cooperative action choice. Therefore, monitoring imperfection inevitably interferes with the full achievement of implicit collusion.[4]

When the monitoring accuracy is high enough, a player can avoid the welfare loss caused by 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. Hence, the more accurate the monitoring technology, the more effectively a player can retaliate against the opponent.[5] A prediction from standard equilibrium theory indicates that a player retaliates *less* in the high accuracy scenario than in the low accuracy scenario.

In contrast with this prediction, however, Kayaba, Matsushima, and Toyama (2019) reports experimental results indicating that subjects in laboratory experiments tend to retaliate *more* in the high accuracy scenario than in the low accuracy scenario. In

[3] For a survey, see Mailath and Samuelson (2006) for instance.
[4] See Green and Porter (1984) and Abreu, Pearce, and Stacchetti (1990), for example.
[5] The folk theorem holds even with such imperfect monitoring, thus indicating that if the discount factor is close to unity, a wide variety of allocations including approximate efficiency can be attained by subgame perfect equilibria (e.g., Fudenberg, Levine, and Maskin, 1994; Matsushima, 2004; Sugaya, 2019).

laboratory experiments, the expected payoff to an individual from the cooperative action choice tends to be higher than that from the defective action choice when the monitoring is accurate, while the expected payoff from the cooperative action choice tends to be lower than that from the defective action choice when the monitoring is inaccurate. Hence, these experimental findings suggest the presence of incentives of real people to pursue retaliation and cooperation beyond the simple maximization of pure self-interest.

Based on these observations, this study sheds light on *behavioral* aspects of players such as social preference and bounded rationality. We assume that noisy signal observation influences the observer's psychological state, thus motivating her social preferences and cognitive limitation.

A player is often motivated not only by pure self-interest but also by social preference such as *reciprocity*; a player feels guilty when she selects the defective action even though she observed the good signal, while a player is annoyed when she selects the cooperative action even if she observed the bad signal. Moreover, a player often becomes *naïve* and selects an action randomly, independently of her pure self-interest and reciprocal motives, due to her cognitive limitation and uncertain psychological mood as well as the strategic complexity caused by monitoring imperfection and private observation.

By incorporating such reciprocity and naïveté into players' incentives, we define *behavioral equilibrium* as an extension of standard Nash equilibrium notion. To simplify strategic interaction, this study will focus on generous tit-for-tat (g-TFT) strategies, which are straightforward stochastic extensions of the tit-for-tat (TFT) strategy (e.g., Molander, 1985; Nowak and Sigmund, 1992; Takahashi, 2010; Matsushima, 2013). This study also assumes that in each period, each player does not consider the impact of her current action choice on her future behavioral attitude of reciprocity and naïveté. We propose a characterization of g-TFT behavioral equilibria in an accuracy-contingent fashion in this setting.

G-TFT is the most concise manner to describe cooperation, retaliation, and forgiveness in repeated interactions. In a g-TFT strategy, a player retaliates against the opponent by selecting the defective action more often when she observes the bad signal than when she observes the good signal. In line with this argument, importantly, the experimental studies of Kayaba, Matsushima, and Toyama (2019) indicates that among a wide variety of strategies, a significant proportion of experimental subjects adopts a g-TFT strategy even if they employ heterogeneous g-TFT.

G-TFT has a substantial advantage over deterministic TFT, which generally fails to be an equilibrium, while, given a high enough discount factor, g-TFT equilibria (in the standard sense) exist irrespective of the level of monitoring accuracy. In any (accuracy-contingent) g-TFT equilibrium, the more accurate the monitoring technology is, the less intensively a player retaliates against the opponent. This view, however, contradicts the experimental results given by Kayaba, Matsushima, and Toyama (2019).

By incorporating reciprocity and naïveté into equilibrium theory, this study demonstrates a characterization result implying that in an accuracy-contingent g-TFT behavioral equilibrium, the more accurate the monitoring technology is, the more severely each player retaliates against the opponent. This result contradicts the prediction of standard equilibrium theory that does not account for behavioral incentives but is more consistent with the experimental evidences.

Our characterization result also indicates that the more often a player behaves naïvely, the less motivated she is by reciprocity. Hence, reciprocity motivates a player to behave in the more sophisticated manner. We further show that the more accurate the monitoring technology is, the less *kind* a player is against the opponent; given low enough levels of monitoring accuracy, the less accurate the monitoring technology is, the more positively reciprocal the player is. Given high enough levels of monitoring accuracy, the more accurate the monitoring technology is, the more negatively reciprocal the player is. At a medium level of monitoring accuracy, a player is neither positively nor negatively reciprocal; she becomes the most naïve at this medium level of monitoring accuracy.

This study should be regarded as the first systematic attempt in the repeated game literature to propose a behavioral theory that reconciles with experimental results. The literature of experimental repeated games has examined the relevance of theoretical predictions without behavioral incentives and the prevalence of various strategies by employing the Strategy Frequency Estimation Method (e.g., Dal Bò and Fréchette, 2011; Fudenberg, Rand, and Dreber, 2012; Aoyagi, Bhaskar, and Fréchette, 2019; Kayaba, Matsushima, and Toyama, 2019). These works commonly supported the predictions that subjects are more likely to collude as the monitoring technology is more accurate, and also indicated that subjects tend to employ heterogeneous strategies. Importantly, Kayaba, Matsushima, and Toyama (2019) experimentally showed that a large proportion of subjects employ heterogeneous g-TFT strategies, and that their retaliation is severer as the monitoring technology is more accurate. Since the latter experimental observation is inconsistent with the theoretical prediction without behavioral incentives, it should be regarded as an important research to develop a new behavioral theory that can describe this observation as equilibrium behavior. This is exactly what this study attempts to do.

Previous studies in the behavioral economics literature show that social preferences facilitate cooperation (e.g., Güth, Schmittberger, and Schwarze, 1982; Berg, Dickhaut, and McCabe, 1995; Fehr and Gächter, 2000), and preferences depend on various contexts (e.g., Rabin 1993; Charness and Rabin, 2002; Dufwenberg and Kirchsteiger, 2004; Falk and Fishbacher, 2005). This study parameterizes the relevant contexts simply by the level of monitoring accuracy. Duffy and Muñoz-García (2012) demonstrate that social preference facilitates collusion when the discount factor is insufficient. In our study, the monitoring technology is a crucial determinant of whether social preferences aid collusion. Social preferences facilitate collusion when monitoring is inaccurate, while they prevent people from colluding when monitoring is accurate.

There exists a literature of bounded rationality in economics and game theory such as limited attention, limited cognitive power, and limited awareness.[6] This study makes an extreme assumption that in each period, a player is either ideally sophisticated or perfectly naïve (unaware), and whether she is ideally sophisticated is exogenously determined in a history-dependent, stochastic, and unconscious manner.

The remainder of this study is organized as follows. Section 2 defines the repeated prisoner's dilemma with additive separability and with imperfect private monitoring. Section 3 introduces the g-TFT strategy and behavioral equilibrium. We then demonstrate our characterization result. Section 4 investigates accuracy-contingent symmetric models. Section 5 concludes.

2. Prisoner's Dilemma with Additive Separability

This study investigates an infinitely repeated game of *prisoners' dilemma with*

---

[6] See Spiegler (2014) for instance.

additive separability, the component game of which is described by Figure 1.

		player 2 C	player 2 D
player 1	C	$1 \quad 1$	$-g_1 \quad 1+g_2$
	D	$1+g_1 \quad -g_2$	$0 \quad 0$

Figure 1: Prisoners' Dilemma with Additive Separability

Let us call C and D the *cooperative* action and *defective* action, respectively. In each period, each player selects either C or D. When an action profile $a \equiv (a_1, a_2) \in \{C, D\}^2$ is selected, each player $i \in \{1,2\}$ receives the (expected value of) instantaneous payoff $u_i(a_1, a_2) \in R$, where

$$u_1(C,C) = u_2(C,C) = 1,\ u_1(D,D) = u_2(D,D) = 0,$$

$$u_1(C,D) = -g_1,\ u_2(C,D) = 1 + g_2,$$

$$u_1(D,C) = 1 + g_1, \text{ and } u_2(D,C) = -g_2.$$

Due to the additive separability of payoff matrix, irrespective of the opponent's action choice, each player i generates a cost g_i by selecting the cooperative action C instead of the defective action D, but provides the opponent $j \neq i$ a benefit equal to $1 + g_j$. We assume that $g_i > 0$ for each $i \in \{1,2\}$, and $|g_1 - g_2| < 1$. Hence, the cooperative action profile (C,C) maximizes their total welfare, while the defective action profile (D,D) is the dominant strategy profile, and it is Pareto-inferior to (C,C).

We assume that monitoring is *imperfect* and *private*. Each player i cannot directly observe the actions that the opponent $j \neq i$ has selected in the current and previous periods. However, at the end of each period, she privately observes a noisy signal denoted by $\omega_j \in \{c, d\}$ for the opponent $j's$ action choice. Let us call c and d the *good* and *bad* signals, respectively.

We define the level of *monitoring accuracy* for each player $i's$ action choice as a probability index $p_i \in (1/2, 1)$; the opponent $j \neq i$ observes the good signal c (the bad signal d) with probability p_i when player i selects the cooperative action C (the defective action D, respectively). Since $p_i > 1 - p_i$, the probability of the opponent $j's$ receiving the good signal c is higher when player i selects C than when she selects D. The greater p_i, the more accurately the opponent $j \neq i$ can monitor player $i's$ action choice. Note that by cooperating, player i can increase the chance that the opponent $j \neq i$ observes the good signal c, making this opponent likely to respond cooperating in the next period.

To measure the level of monitoring accuracy as a single value for simplicity, this study assumes that the probability of the good signal when the cooperative action is selected and the probability of the bad signal when the defective action is selected are equivalent. (This assumption is irrelevant to the outcome of this study.)

3. Generous Tit-For-Tat Strategy

Let $\delta_i \in (0,1)$ denote the *discount factor* of player i. This study allows players to have different discount factors. Each player $i's$ payoff in the infinitely repeated game is given by the discounted sum of her instantaneous payoffs $\sum_{t=1}^{\infty} \delta_i^{t-1} u_i(a(t))$, where $a(t)$ denotes the action profile selected in period t. We define a strategy for each player i as a mapping from histories of her previous action choices and private signal observations to her mixed actions, which is denoted by s_i. Each player i selects a strategy s_i to maximize the associated expected payoff $E[\sum_{t=1}^{\infty} \delta_i^{t-1} u_i(a(t)) \mid s]$, provided that she expects her opponent $j \neq i$ to behave according to s_j, where $s = (s_1, s_2)$ and $E[\cdot \mid s]$ denotes the expectation operator conditional on the strategy profile s.

This study focuses on equilibrium play according to a *generous tit-for-tat (g-TFT) strategy*, and also on the player's incentive to make a signal-contingent action choice by ignoring incentive issues in the first period 1. Hence, we denote a g-TFT strategy for each player $i \in \{1, 2\}$ simply by

$$s_i = (r_i(c), r_i(d)) \in (0,1)^2,$$

according to which, in each period $t \geq 2$, player i makes the cooperative action choice C with probability $r_i(\omega_j) \in (0,1)$ whenever she observes signal $\omega_j \in \{c, d\}$ in period $t-1$. To eliminate trivial cases, we assume $r_i(\omega_j) > 0$ for each $i \in \{1,2\}$ and $\omega_j \in \{c, d\}$ throughout this study.

Associated with a g-TFT strategy s_i, we define the *retaliation intensity* of player i as the difference in cooperation rate between the good and bad signal scenarios, that is, $r_i(c) - r_i(d)$. The retaliation intensity measures the degree to which a player punishes the opponent when she observes the bad signal.

As Matsushima (2013) has shown, due to the additive separability of payoff matrix, the incentive constraint in each period $t \geq 2$ is simply described by the following conditions: for each $i \in \{1,2\}$,

(1) $$[\, g_i > \delta_i (1+g_i)(2p_i - 1)\{r_j(c) - r_j(d)\}\,] \Rightarrow [\, r_i(c) = r_i(d) = 0\,],$$

and

(2) $$[\, g_i < \delta_i (1+g_i)(2p_i - 1)\{r_j(c) - r_j(d)\}\,] \Rightarrow [\, 1 - r_i(c) = 1 - r_i(d) = 0\,].$$

Each player $i's$ choice of C instead of D generates a cost g_i in period t, while, in period $t+1$, this choice generates a gain $1+g_i$ from the cooperative response of the opponent j with probability $p_i r_j(c) + (1-p_i) r_j(d)$ rather than probability $(1-p_i) r_j(c) + p_i r_j(d)$. Note that the opponent j observes the good signal c with probability p_i $(1-p_i)$ when player i selects the cooperative action C (the defective action D, respectively). It is important to note that due to the additive-separability of payoff matrix, the instantaneous gain from her action choice in period $t+1$ plus the discounted sum of payoffs after period $t+2$ is independent of her action choice in the current period $t+2$, provided that the opponent follows a g-TFT strategy after period $t+2$. This is why we can express the incentive constraints simply by (1) and (2).

Due to the belief-free nature (e.g., Ely and Välimäki, 2002; Piccione, 2002), we can derive the equilibrium constraints of g-TFT strategy profile, (1) and (2), simply from the equivalence between the instantaneous gain from selecting C rather than D, given by g_i, and the resultant loss in the next period, given by

$$\delta_i(1+g_i)\{p_i r_j(c)+(1-p_i)r_j(d)\}-\delta_i(1+g_i)\{(1-p_i)r_j(c)+p_i r_j(d)\}$$
$$=\delta_i(1+g_i)(2p_i-1)\{r_j(c)-r_j(d)\},$$

that is,

(3) $$g_i=\delta_i(1+g_i)(2p_i-1)\{r_j(c)-r_j(d)\}.$$

Let us define

$$R_i=R_i(p_j)\equiv\frac{g_j}{\delta_j(1+g_j)(2p_j-1)}.$$

Note that the equality (3) is equivalent to

(4) $$r_i(c)-r_i(d)=R_i(p_j)\quad\text{for each}\quad i\in\{1,2\}.$$

Hence, a g-TFT strategy profile s is a Nash equilibrium in this section's sense if and only if the associated retaliation intensity of player i is equivalent to $R_i(p_j)$ for each $i\in\{1,2\}$. It is clear from (4) and $0<r_i(c)-r_i(d)\leq 1$ that there exists a (non-trivial) g-TFT Nash equilibrium if and only if for each $i\in\{1,2\}$, $0<R_i(p_j)\leq 1$, that is,

(5) $$\delta_j\geq\frac{g_j}{(1+g_j)(2p_j-1)}.$$

Note that $R_i(p_j)$ is *decreasing* in the level of monitoring accuracy p_j for the opponent, thus contradicting the experimental evidences reported by Kayaba, Matsushima, and Toyama (2019).

4. Behavioral Equilibrium

This section introduces a *behavioral* version of g-TFT equilibrium and shows the possibility that the retaliation intensity is *increasing* in the level of monitoring accuracy. We define the notion of behavioral equilibrium as a modification of the conditions (1) and (2) in the following manner. Each player $i\in\{1,2\}$ is motivated not only by pure self-interest but also by cognitive limitation termed *naïveté*, which is denoted by $\varepsilon_i\in[0,\frac{1}{2})$, and by social preference termed *reciprocity*, which is denoted by $(w_i(c),w_i(d))\in R_+^2$.

In every period, with probability $2\varepsilon_i$, player i becomes naïve and randomly selects between actions C and D. The player has cognitive limitation the degree of which depends on her uncertain psychological mood as well as various parameters such as the level of monitoring accuracy. For simplicity, we assume that her psychological mood is either ideally sophisticated or totally unaware, and that whether she is ideally sophisticated or not is randomly determined. Due to this naïveté, a g-TFT strategy for player i, $s_i = (r_i(c), r_i(d))$, must satisfy

(6) $$\min[r_i(c), 1 - r_i(c), r_i(d), 1 - r_i(d)] \geq \varepsilon_i .$$ [7]

We assume that the above-mentioned randomness is drawn independently across players.

With the remaining probability $1 - 2\varepsilon_i$, player i makes her action choice in the sophisticated manner. In addition to pure self-interest, we introduce a motive of social preference termed *reciprocity* $(w_i(c), w_i(d))$ as follows. Suppose that player i observes the good signal c; she feels guilty when she selects the defective action D despite observing the good signal. In this case, she can avoid a psychological cost $w_i(c) \geq 0$ by selecting the cooperative action C. Next, suppose that player i observes the bad signal d; she is annoyed when she selects the cooperative action C despite observing the bad signal d. In this case, she can avoid a psychological cost $w_i(d) \geq 0$ by selecting the defective action D.

Whenever player i observes the good signal c, her instantaneous gain from selecting action D is given by $g_i - w_i(c)$, while her resultant future loss is given by $\delta_i(1 + g_i)(2p_i - 1)\{r_j(c) - r_j(d)\}$. We assume that each player i does not consider the impact of her current action choice on her future behavioral aspects. From this assumption, player i is willing to select the cooperative action C if

$$g_i - w_i(c) < \delta_i(1 + g_i)(2p_i - 1)\{r_j(c) - r_j(d)\},$$

while she is willing to select the defective action D if

---

[7] The introduction of naïveté generally restricts the class of g-TFT equilibria discussed in Section 3, but it still leaves this class non-empty whenever

$$R_i(p_j) \leq 1 - 2\varepsilon_i \quad \text{for each } i \in \{1, 2\}.$$

$$g_i - w_i(c) > \delta_i(1+g_i)(2p_i - 1)\{r_j(c) - r_j(d)\}.$$

Similarly, whenever player i observes the bad signal d, her instantaneous gain from selecting action D is given by $g_i + w_i(d)$, while her resultant future loss is given by $\delta_i(1+g_i)(2p_i - 1)\{r_j(c) - r_j(d)\}$. Hence, player i is willing to select the cooperative action C if

$$g_i + w_i(d) < \delta_i(1+g_i)(2p_i - 1)\{r_j(c) - r_j(d)\},$$

while she is willing to select the defective action D if

$$g_i + w_i(d) > \delta_i(1+g_i)(2p_i - 1)\{r_j(c) - r_j(d)\}.$$

Based on these arguments, we define *behavioral equilibrium* as an extension of the conditions (1) and (2) as follows.

Definition 1: A g-TFT strategy profile $s \equiv (s_1, s_2)$ is said to be a *behavioral equilibrium, or, shortly, an equilibrium, with respect to* $(g_i, \delta_i, p_i, \varepsilon_i, w_i(c), w_i(d))_{i\in\{1,2\}}$ if for each $i \in \{1,2\}$,

(7) $$[g_i - w_i(c) > \delta_i(1+g_i)(2p_i - 1)\{r_j(c) - r_j(d)\}] \Rightarrow [r_i(c) = \varepsilon_i],$$

(8) $$[g_i - w_i(c) < \delta_i(1+g_i)(2p_i - 1)\{r_j(c) - r_j(d)\}] \Rightarrow [1 - r_i(c) = \varepsilon_i],$$

(9) $$[g_i + w_i(d) > \delta_i(1+g_i)(2p_i - 1)\{r_j(c) - r_j(d)\}] \Rightarrow [r_i(d) = \varepsilon_i],$$

and

(10) $$[g_i + w_i(d) < \delta_i(1+g_i)(2p_i - 1)\{r_j(c) - r_j(d)\}] \Rightarrow [1 - r_i(d) = \varepsilon_i].$$

This study assumes that for each $i \in \{1,2\}$:

either $w_i(c) = 0$ or $w_i(d) = 0$.

A player i is said to be *positively (negatively) reciprocal* if $w_i(c) > 0$ ($w_i(d) > 0$, respectively). The following theorem demonstrates an important characterization result and clarifies the relation between behavioral aspects and equilibrium.

Theorem 1: *Consider an arbitrary g-TFT strategy profile* s*, where for each* $i \in \{1,2\}$*,* $r_i(c)$*,* $r_i(d)$*,* $1 - r_i(c)$*, and* $1 - r_i(d)$ *are all different, and equality (4) does not hold.*

Then, s *is an equilibrium if and only if for each* $i \in N$*, equality (1) and the following properties hold; if*

$$r_j(c) - r_j(d) < R_j(p_i),$$

then:

(11) $$w_i(c) = g_i - \delta_i(1+g_i)(2p_i - 1)\{r_j(c) - r_j(d)\} > 0,$$

$$w_i(d) = 0, \text{ and } \varepsilon_i = r_i(d).$$

If

$$r_j(c) - r_j(d) > R_j(p_i),$$

then:

(12) $$w_i(d) = \delta_i(1+g_i)(2p_i - 1)\{r_j(c) - r_j(d)\} - g_i > 0,$$

$$w_i(c) = 0, \text{ and } \varepsilon_i = 1 - r_i(c).$$

Proof: The proof of the "if" part is a direct consequence of Definition 1. The proof of the "only if" part is as follows. Suppose that

$$r_j(c) - r_j(d) < R_j(p_i).$$

Then, the left-hand side of (9) holds; as a result,

$$\varepsilon_i = r_i(d).$$

Since $r_i(c)$ and $1 - r_i(c)$ are different from ε_i, it follows from (7) and (8) that

$$g_i - w_i(c) = \delta_i(1+g_i)(2p_i - 1)\{r_j(c) - r_j(d)\},$$

thus implying (11). Since $w_i(c)$ is positive, player i is positively reciprocal, meaning that $w_i(d) = 0$.

Next, suppose that:

$$r_j(c) - r_j(d) > R_j(p_i).$$

Then, the left-hand side of (8) holds; as a result,

$$\varepsilon_i = 1 - r_i(c).$$

Since $r_i(d)$ and $1 - r_i(d)$ are different from ε_i, it follows from (9) and (10) that

$$g_i + w_i(d) = \delta_i(1+g_i)(2p_i - 1)\{r_j(c) - r_j(d)\},$$

thus implying (12). Since $w_i(d)$ is positive, player i is negatively reciprocal, meaning that $w_i(c)=0$.

Q.E.D.

From Theorem 1, player i is positively reciprocal if the retaliation intensity is lower than R_i, while she is negatively reciprocal if the retaliation intensity is higher than R_i. The more negatively reciprocal or the less positively reciprocal a player is, the greater the opponent's retaliation intensity is. From Theorem 1, we can automatically derive the *uniqueness* of equilibrium as follows.

Theorem 2: *A g-TFT strategy profile s is a behavioral equilibrium if and only if the following property holds for each player $i\in\{1,2\}$; if the opponent $j\neq i$ is positively reciprocal ($w_i(c)>0$), then*

$$r_i(c)=\frac{g_j-w_j(c)}{\delta_j(1+g_j)(2p_j-1)}+\varepsilon_i \quad \text{and} \quad r_i(d)=\varepsilon_i,$$

while, if opponent j is negatively reciprocal ($w_i(d)>0$), then

$$r_i(c)=1-\varepsilon_i \quad \text{and} \quad r_i(d)=1-\varepsilon_i-\frac{g_j+w_j(d)}{\delta_j(1+g_j)(2p_j-1)}.$$

The uniqueness implied by Theorem 2 is in contrast with the null-reciprocal (standard) case, which has multiplicity of g-TFT Nash equilibrium whenever the strict inequality of (5) holds for each $i\in\{1,2\}$.

From Theorem 2, we can automatically derive the *existence* of equilibrium; that is, there exists a behavioral equilibrium if and only if for each $i\in\{1,2\}$, either

$$w_i(c)>0,\ \ g_j\geq w_j(c),\ \text{and}\ \ \delta_j\geq\frac{g_j-w_j(c)}{(1-2\varepsilon_i)(1+g_j)(2p_j-1)},$$

or

$$w_i(d)>0\ \ \text{and}\ \ \delta_j\geq\frac{g_j+w_j(d)}{(1-2\varepsilon_i)(1+g_j)(2p_j-1)}.$$ [8]

---

[8] From Theorem 2, it generically holds that for each $i\in\{1,2\}$, $r_i(c)$, $r_i(d)$, $1-r_i(c)$, and

Hence, from the comparison with the condition (5), the condition for the existence is more restrictive when a player is negatively reciprocal than when she is null-reciprocal (purely self-interested), while it is less restrictive when a player is positively reciprocal than when she is null-reciprocal.

Moreover, from Theorem 2, we can automatically derive the retaliation intensity $r_i(c)-r_i(d)$ associated with the (unique) behavioral equilibrium; if the opponent $j \neq i$ is positively reciprocal, then

$$r_i(c)-r_i(d)=\frac{g_j-w_j(c)}{\delta_j(1+g_j)(2p_j-1)},$$

while, if the opponent j is negatively reciprocal, then

$$r_i(c)-r_i(d)=\frac{g_j+w_j(d)}{\delta_j(1+g_j)(2p_j-1)}.$$

5. Symmetry

This section considers a *symmetric* model in which there exists (g,δ,p) such that

$$(g_i,\delta_i,p_i)=(g,\delta,p) \text{ for each } i\in\{1,2\}.$$

We allow players' behavioral aspects to be heterogeneous and contingent on a common level of monitoring accuracy p; therefore, $(\varepsilon_i(p),w_i(c;p),w_i(d;p))$ instead of $(\varepsilon_i,w_i(c),w_i(d))$. In line with Kayaba, Matsushima, and Toyama (2019), we allow players' g-TFT strategies to be heterogeneous. Moreover, an equilibrium should be contingent on the level of monitoring accuracy p; therefore, $s_i(p)=(r_i(c;p),r_i(d;p))$ instead of $s_i=(r_i(c),r_i(d))$.

Let us set an arbitrary level $\underline{p}\in(\frac{1}{2},1)$, which we call the minimum level of monitoring accuracy. This section assumes that both the accuracy-contingent behavioral aspect $(\varepsilon_i(p),w_i(c;p),w_i(d;p))$ and the accuracy-contingent equilibrium strategy $s_i(p)=(r_i(c;p),r_i(d;p))$ are *continuous* in $p\in[\underline{p},1]$.

$1-r_i(d)$ are all different.

5.1. Kindness and Retaliation Intensity

As a measure of the degree to which the accuracy-contingent reciprocity $(w_i(c;p), w_i(d;p))$ motivates player i to make the cooperative action choice, this section considers a notion of *kindness* as follows.

Definition 2: Player i is said to be *more kind at* p *than at* $\tilde{p}$ if:

either $w(c;p) > w(c;\tilde{p})$ or $w(d;p) < w(d;\tilde{p})$.

Definition 2 implies that a player is more kind at p than at $\tilde{p}$ if she is positively reciprocal at p and is negatively reciprocal at $\tilde{p}$. The experimental work by Kayaba, Matsushima, and Toyama (2019) reported that the retaliation intensities observed in laboratories are *increasing* in the level of monitoring accuracy. The following theorem indicates that the introduction of behavioral aspects plays the substantial role in explaining this experimental observation.

Theorem 3: *Suppose that* $s(p)$ *is an accuracy-continent equilibrium. If the retaliation intensity of player* i, $r_i(c;p) - r_i(d;p)$, *is increasing in* $p \in [\underline{p}, 1]$, *then, the higher the monitoring accuracy* p *is, the less kind the opponent* j *is.*

Proof: See Appendix A.

Theorem 3 implies that in the high accuracy scenario, a player tends to be more negatively reciprocal as the level of monitoring accuracy increases. This tendency makes the retaliation intensity more severe and works against the success in cooperation induced by the improvement of monitoring technology. In the low accuracy scenario, a player tends to be more positively reciprocal as the level of monitoring accuracy decreases. This tendency makes the retaliation intensity milder and mitigates the lack of cooperation caused by the deterioration of the monitoring technology.

5.2. Naïveté and Reciprocity

Kayaba, Matsushima, and Toyama (2019) also reported experimental results showing that the more likely experimental subjects are to make the cooperative action choice, the more accurate the monitoring technology is. Based on this finding, this section considers an accuracy-contingent equilibrium $s(p)$ such that for each $i \in \{1,2\}$, $r_i(c;p)$, $r_i(d;p)$, and $r_i(c;p) - r_i(d;p)$ are all continuous and increasing in p.

Theorem 4: *Consider an arbitrary accuracy-contingent equilibrium $s(p)$. Suppose that for each $i \in \{1,2\}$, $r_i(c;p)$, $r_i(d;p)$, and $r_i(c;p) - r_i(d;p)$ are continuous and increasing in p. Then, for each $i \in N$, $\varepsilon_i(p)$ is increasing, $w_i(c;p)$ is decreasing, and $w_i(d;p) = 0$ in $p \in [\underline{p}, \hat{p}_i]$, while $\varepsilon_i(p)$ is decreasing, $w_i(d;p)$ is increasing, and $w_i(c;p) = 0$ in $p \in [\hat{p}_i, 1]$.*

Proof: See Appendix B.

In the high enough accuracy scenario, a player tends to be negatively reciprocal and is likely to be sophisticated as unintended results. In the low enough accuracy scenario, a player tends to be positively reciprocal and is likely to be sophisticated. In the medium accuracy scenario, a player tends to be purely self-interested and is likely to be naïve. Theorem 4 shows the presence of the unintended trade-off between naïveté and reciprocity; the more likely a player is to be naïve, the less reciprocal she tends to be. Hence, we can conclude that a player's sophisticated decision-making is motivated not by her pure self-interest but by social preference concerning reciprocity.

6. Concluding Remarks

This study incorporated reciprocity and naïveté into infinitely repeated prisoner's dilemma with imperfect private monitoring, and described strategic behavior as behavioral equilibrium in a consistent manner with experimental evidences. This study is the first systematic analysis of repeated games that fills the gap between theory and evidence.

We have several issues that are left unsolved as possible future research. For instance, this study focused on g-TFT strategies; however, other types of strategies such as grim-trigger, lenience, and long-term punishment, which are prominent from theoretical and empirical viewpoints, should also be investigated.

This study should be extended to more general games beyond prisoner's dilemma. Without any substantial difficulty, we can extend this study to a three-or-more-player social dilemma with additive separability. Since this study's analysis crucially depends on the additive separability assumption, it would be a challenging problem to consider games without this assumption.

This study assumes that each player does not consider the impact of her current action choice on her future behavioral aspects. Due to the backward induction technique, without any substantial change of this study's arguments, we can replace this assumption with the assumption that each player does not consider the impact of her current action choice on her future behavioral aspects after a fixed *finite* time interval. On the other hand, it would be a substantial and difficult open problem to consider the case in which a player considers the impact of her current action choice on her behavioral aspects *throughout the future*.

The impact of reciprocity and naïveté on implicit collusion should be investigated more directly through creating new experimental designs. For example, to clarify the impact of reciprocity, we should conduct experiments in which subjects play repeated games against not real people (natural intelligence) but machines (artificial intelligence). To clarify the impact of naïveté, we should conduct experiments in which each subject has a plenty of time to think before decision and she is even permitted to decide in consultation with multiple neutral people.

All the research directions described above are important but beyond the purpose of this study.

References

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

Aoyagi, M., V. Bhaskar, and G. Fréchette (2019): "The Impact of Monitoring in Infinitely Repeated Games: Perfect, Public, and Private," *American Economic Journal: Microeconomics* 11(1), 1-43.

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

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.

Dal Bó, P. and G. Fréchette (2011): "The Evolution of Cooperation in Infinitely Repeated Games: Experimental Evidence," *American Economic Review* 101(1), 411–429.

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.

Fudenberg, D., D.G. Rand, and A. Dreber (2012): "Slow to Anger and Fast to Forgive: Cooperation in an Uncertain World," *American Economic Review* 102(2), 720–749.

Green, E. and R. Porter (1984): "Noncooperative Collusion under Imperfect Price Information," *Econometrica* 51, 87-100.

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 (2019): "Accuracy and Retaliation in Repeated Games with Imperfect Private Monitoring: Experiments," Discussion Paper CARF-F-466, University of Tokyo.

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

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.

Spiegler, R. (2014): *Bounded Rationality and Industrial Organization*, Oxford University Press, USA.

Sugaya, T. (2019): *The Folk Theorem in Repeated Games with Private Monitoring*, mimeo. https://docs.google.com/viewer?a=v&pid=sites&srcid=ZGVmYXVsdGRvbWFpbnx0YWt1b3N1Z2F5YXxneDozNjE1YTM2OGQwZTM3M2Ex

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

Appendix A: Proof of Theorem 3

Since $r_i(c;p)-r_i(d;p)$ is increasing and $R(p)$ is decreasing in p, it follows from Theorem 1 that there is a unique level of monitoring accuracy $\hat{p}_j \in [\underline{p},1]$ that satisfies the following properties for each $p \in [\underline{p},1]$; if $p < \hat{p}_j$, then:

$$r_i(c;p)-r_i(d;p) < R(p),$$

and, therefore, $w_j(c;p)$ is decreasing in $p \in [\underline{p},\hat{p}_i]$. If $p > \hat{p}_j$, then:

$$r_i(c;p)-r_i(d;p) > R(p),$$

and, therefore, $w_j(d;p)$ is increasing in $p \in [\hat{p}_i,1]$. These properties imply that the higher p is, the less kind opponent j is.

Q.E.D.

Appendix B: Proof of Theorem 4

From Theorem 2, if $p < \hat{p}_i$, then,

$$\varepsilon_i(p) = r_i(d;p) .$$

Since $r_i(d;p)$ is increasing in p, $\varepsilon_i(p)$ is increasing in $p \in [\underline{p}, \hat{p}_i]$. If $p > \hat{p}_i$, then,

$$\varepsilon_i(p) = 1 - r_i(c;p) .$$

Since $r_i(c;p)$ is increasing in p, $\varepsilon_i(p)$ is decreasing in $p \in [\hat{p}_i, 1]$. From these observations and Theorem 1, we obtain the proof of Theorem 4.

Q.E.D.