The Greedier the Better?
Addition of Greedy Aspect on Popular Strategies in Prisoner's Dilemma

Seungjae Moon, Mingzhong Deng, Simon Shih

March 8, 2018

Abstract

The Prisoner's Dilemma is a widely applicable evolutionary game where two players independently strategize either to cooperate or defect to increase their payoff. The standard argument is that Tit-For-Tat (TFT), a strategy of mimicking the opponent's previous move, is the most effective at maximizing one's payoff for a repeated game. We attempt to modify this TFT strategy by adding an aspect of greediness, a common human characteristic, by incorporating intervals and probabilities. We analyze the new strategies, Fixed Greediness (FG) and Greedy Tit-For-Tat (GrTFT) by creating payoff matrices and analyzing the stability and the stochastic matrix. Our results show FG and GrTFT's dominance or lack thereof and expected payoff when matched up against popular strategies. Therefore, we are able to explore the evolutionary dynamics of our new strategies and conditions at which they are achieved.

1 Introduction

In the game of Prisoner's Dilemma, two players, who cannot communicate with each other, can choose to cooperate with the other or defect the other. If both players choose to cooperate, they both get payoff as reward, R. If one of the player chooses to defect while the other chooses to cooperate, the defecting player get a payoff as temptation, T, and the cooperating player gets a payoff as sucker, S. If both player choose to defect, they both get a payoff as punishment, P. One of the rules of the game is $T > R > P > S$. There is one more requirement of this game: $T + P < 2R$ to make sure that the total payoff of two players cooperating is higher than one person cooperating and one person defecting.

	C	D
C	R	S
D	T	P

Figure 1: Diagram of a RSTP payoff matrix [6]

In a one-round Prisoner's Dilemma, both player will decide to defect assuming they are both rational players; by the nature of the game, defecting is the better strategy for both of them because defecting gives you a larger payoff no matter what the other player chooses to do. In a one-shot game, there is no other round for revenge, so memory will not be taken into consideration. Even in a short-term game or a finitely iterative game, the rational strategy is to always defect (ALLD) [6]. If both players know the number of iteration, there is no incentive for them to be cooperative at the last round. By induction, there is no incentive for them to cooperate in the second-to-last round and all the rounds before that.

In real life, Prisoner's Dilemma occurs in various situations including in economics, sports, and personal relationship problems including one's reputation. Moreover, almost all social interactions with others occur repeatedly and infinitely. Oil price negotiation between countries, product prices between companies in the same industry, and national defense are examples of Prisoner's Dilemma. One simple realistic example is women wearing makeup. Both A and B wear makeup to look better. If both A and B use normal makeups, they both use a minimal cost while they both present the similar and good-looking image towards the public; in this case they both have a reward (R). However, if one of them use more expensive and fancy makeup, that particular woman presents a better image and has an advantage over the other; in this case she gets the temptation (T), and the other gets the sucker (S). If both of them used expensive make up, they do not have advantage over the other, but the cost increases; in this case, they have the punishment (P). This makeup situation happens every day, and it is an instance of infinite round of Prisoner's Dilemma game. There are various other real life situations that can be presented as an infinite round of Prisoner's Dilemma games. Then, what is the most useful strategy to use in the infinite rounds of games? Should people cooperate more because of the reward of reciprocity relationship or defect more because of the nature of the game that the temptation is the highest payoff in the long run?

In this article, we try to test a solution that is more aggressive and greedy against the other existing popular strategies. Since the temptation, T, is the largest element in the payoff matrix, choosing defection gives you a chance to exploit your opponent while not getting exploited yourself in a single round game. We want to examine if this holds true even in the iterative Prisoner's Dilemma. Since Prisoner's Dilemma happens so subtly and naturally, people might not even realize the payoff-maximizing strategy in real life.

In 1984, Robert Axelrod, professor of political science at the University of Michigan, made an online competition of iterative Prisoner's Dilemma game [1]. The winning strategy came out to be Tit-For-Tat (TFT). In order to maximize the payoff, TFT cooperates at the first time, and does what the other player did in the previous round. When facing a player who always cooperate, TFT will always cooperate. When facing a player who always defect, TFT will always defect [5]. However, there are few flaws in the TFT strategy.

1. TFT strategy fails to exploit the always-cooperative person; a rational person should choose defect to maximize its own payoff, but TFT model does not take into account such factor.

2. TFT cannot correct mistakes. Between two TFT players, if occasional mistakes happen by both players, they will choose to defect forever, and that decreases the overall payoff, rather than both choose to cooperate, like how they are supposed to act.

In fact, TFT stays cooperating if your opponent stays cooperating. In the real world, people will decide to defect in order to get the temptation. To model the frequent defection strategy, we played the game of iterative Prisoner's Dilemma with two players where one of the player uses the frequent defection strategy we created, and the other player uses a few of the popular strategies that previous scholars have created, such as ALLC (always cooperate), ALLD (always defect), and TFT. After deciding a concrete strategy, we will come up with the specific payoff matrix that is appropriate for both strategies and analyze the stability and the expected payoff. Based on the analysis, we will be able conclude whether strategies with the greedy aspect maximize the payoff.

2 Simplifications

To add an aspect of greediness, we chose two popular Prisoner's Dilemma strategies to modify: ALLC and TFT. Knowing the number of rounds to be played affects interactive strategies [6], so we begin by assuming a setting with infinite rounds.

Since ALLC strategy assumes that the player will cooperate indefinitely, we decided to analyze a strategy in which a player would become greedy and defect once every few turns and cooperate otherwise. Since the player never changes, it is assumed that the player gets greedy and defects at a fixed interval, α. We named it the FG (Fixed Greediness) strategy.

Round#	1	2	3	4	$\cdots$
$\alpha = 2$	C	D	C	D	$\cdots$

,

Round#	1	2	3	4	5	6	$\cdots$
$\alpha = 3$	C	C	D	C	C	D	$\cdots$

, $\cdots$

Figure 2: The tables illustrate how a player using the FG strategy would choose depending on their greediness level. Greedier player has a lower fixed interval, α, of defection. The same pattern continues for $\alpha > 1$. Note that $\alpha = 1$ would represent ALLD.

When both players use the FG strategy, we assumed that the interval is the same. This assumption imposes a limitation to the model because the model will start to lose accuracy as the gap between the level of greediness widen for the two players.

We also created the GrTFT (Greedy Tit-For-Tat) strategy inspired from TFT. For this strategy, we consider the probability of a player cooperating when the opponent cooperates, β. Hence, a lower β-value resembles a greedier player since one is more likely to defect even when the opponent cooperated in the previous round. Note that $0 < \beta < 1$ since $\beta = 0$ would behave like ALLD, and $\beta = 1$ would behave like TFT. GrTFT is interactive since the decision to cooperate or defect depends on the opponent; however, when the opponent defects, GrTFT player would defect without taking probability into consideration. When GrTFT plays against GrTFT, it is assumed by nature of greediness that if one defects, the other is quick to defect as well and lead to an infinite defection.

The reactive strategies of ALLC, ALLD, TFT, and GrTFT are defined by two parameters. Let p be the probability of cooperating after opponent's cooperation in the previous round and q be the probability of cooperating after opponent's defection in the previous round. Then ALLC, ALLD, TFT, and GrTFT can be denoted by the pair (p,q).

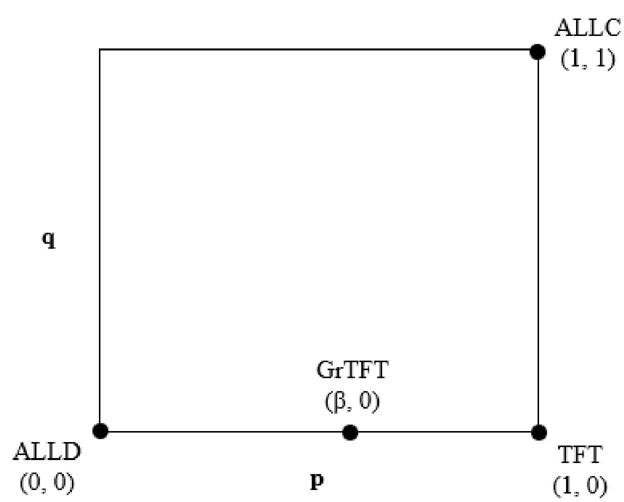

Figure 3: Specific reactive strategies and their corresponding parameters are ALLC $= (1, 1)$, ALLD $= (0, 0)$, TFT$= (1, 0)$, and GrTFT$= (\beta, 0)$.

3 Mathematical Model

A 2x2 payoff matrix is often used to study the match-up between two strategies:

	A	B
A	a	b
B	c	d

Figure 4: A diagram of a standard two-player game payoff matrix; a = payoff strategy A gets playing against A, b = payoff A gets playing against B, c = payoff B gets playing against A, and d = payoff B gets playing against B)

Each element is represented by summing the payoff (R,S,T, and P) of each iteration or round between two strategies. From the simplifications, the payoff matrix can be completed by

1. analyzing the first few rounds out of infinite of them until a general pattern establishes, then

2. algebraically representing the total payoff for any given number of rounds, m. Let α be the fixed interval at which FG strategy defects. Let β be the probability of cooperating when opponent cooperates.

3.1 ALLC vs. FG

For ALLC against ALLC, the two players always cooperate. For ALLC against FG, the players both start by cooperating. However, after every α round, FG would defect. For FG against FG, the two players defect every α rounds and cooperate rest of the time having assumed that α is the same for either player (e.g. $\alpha = 4$). Then, we have

$$\begin{array}{|c|cccc|}\hline AALC & \multicolumn{4}{c|}{\overbrace{C \quad C \quad C \quad \cdots}^{\text{m}}} \\ AALC & C & C & C & \cdots \\ \hline \end{array} \tag{i}$$

$$\begin{array}{|c|cccc:cccc:c|}\hline AALC & C & C & C & C & C & C & C & C & \cdots \\ FG & C & C & C & D & C & C & C & D & \cdots \\ \hline \end{array} \tag{ii}$$

$$\begin{array}{|c|cccc:cccc:c|}\hline FG & C & C & C & D & C & C & C & D & \cdots \\ FG & C & C & C & D & C & C & C & D & \cdots \\ \hline \end{array} \tag{iii}$$

Recall that, for each round, we can extract the payoff from the RSTP payoff matrix [figure 1]. Based on (i), ALLC gets a reward each round, so the payoff is **mR**. Based on (ii), ALLC receives a sucker's payoff once every α rounds (denoted by a dashed line), and FG receives a temptation payoff. Both get a reward on the rest of the rounds, so the payoff is $\boldsymbol{\frac{\alpha-1}{\alpha}\mathbf{mR}+\frac{1}{\alpha}\mathbf{mS}}$ for ALLC and $\boldsymbol{\frac{\alpha-1}{\alpha}\mathbf{mR}+\frac{1}{\alpha}\mathbf{mT}}$ for FG. Based on (iii), FG receives a punishment payoff once every α round and reward on the rest of the rounds, so the payoff is $\boldsymbol{\frac{\alpha-1}{\alpha}\mathbf{mR}+\frac{1}{\alpha}\mathbf{mP}}$. Therefore, we have the following payoff matrix:

$$\begin{array}{c|cc} & ALLC & FG \\ \hline AALC & mR & \frac{\alpha-1}{\alpha}mR+\frac{1}{\alpha}mS \\ FG & \frac{\alpha-1}{\alpha}mR+\frac{1}{\alpha}mT & \frac{\alpha-1}{\alpha}mR+\frac{1}{\alpha}mP \end{array} \tag{1}$$

3.2 ALLD vs. FG

Similarly, the ALLD vs. FG dynamical game yields the payoff matrix (2).

$$\begin{array}{|c|cccc|}\hline AALD & D & D & D & \cdots \\ AALD & D & D & D & \cdots \\ \hline \end{array}$$

$AALD$	D	D	D	D	D	D	D	D	$\cdots$
FG	C	C	C	D	C	C	C	D	$\cdots$

$$
\begin{array}{c|cc}
 & ALLD & FG \\
\hline
AALD & mP & \frac{\alpha-1}{\alpha}mT + \frac{1}{\alpha}mP \\
FG & \frac{\alpha-1}{\alpha}mS + \frac{1}{\alpha}mP & \frac{\alpha-1}{\alpha}mR + \frac{1}{\alpha}mP
\end{array}
\tag{2}
$$

3.3 TFT vs. FG

Likewise, when TFT competes with FG, they generate the payoff matrix (3).

TFT	C	C	C	$\cdots$
TFT	C	C	C	$\cdots$

TFT	C^*	C	C	C	D	C	C	C	D	$\cdots$
FG	C^*	C	C	D	C	C	C	D	C	$\cdots$

*Even though the first round does not fit into the pattern, the effect of this irregularity is relatively small when number of round grows ($m \to \infty$), thus negligible to the overall payoff.

$$
\begin{array}{c|cc}
 & TFT & FG \\
\hline
TFT & mR & \frac{T+S+(\alpha-2)R}{\alpha}\cdot m \\
FG & \frac{T+S+(\alpha-2)R}{\alpha}\cdot m & \frac{\alpha-1}{\alpha}mR + \frac{1}{\alpha}mP
\end{array}
\tag{3}
$$

3.4 ALLC vs. GrTFT

For ALLC against GrTFT, GrTFT has the probability $1-\beta$ to be greedy and defect each round, so the exact round when GrTFT player defects is unknown. Meanwhile, ALLC still always cooperates. For GrTFT against GrTFT, once one of the GrTFT player defects under the given probability, we assumed that the its opponent, who is also playing GrTFT, tends to defect immediately. Then, we have

$$\begin{array}{|c|cccccccccccc|}\hline ALLC & \cdots & C & C & C & C & C & \cdots & C & C & C & C & C & \cdots \\ GrTFT & \cdots & C & C & D & C & C & \cdots & C & C & D & C & C & \cdots \\ \hline\end{array} \tag{iv}$$

$$\begin{array}{|c|cccccccc|}\hline GrTFT & \cdots & C & C & D & C & D & D & \cdots \\ GrTFT & \cdots & C & C & C & D & D & D & \cdots \\ \hline\end{array} \tag{v}$$

Dashed line in the tables now represents a particular round, m_i, when GrTFT defects.

Based on (iv), when ALLC plays against GrTFT, ALLC and GrTFT cooperate and get rewards for the first βm times. The other $(1-\beta)m$ times, ALLC receives a sucker, while GrTFT receives a temptation. Hence, ALLC's payoff against GrTFT is $\mathbf{(1-\boldsymbol{\beta})mS + m\boldsymbol{\beta}R}$ and GrTFT's payoff against ALLC is $\mathbf{(1-\boldsymbol{\beta})mT + m\boldsymbol{\beta}R}$. Based on (v), when GrTFT plays against GrTFT, one gets a reward for the initial $\beta^2 m$ times because the probability for each GrTFT players is independent. Then, GrTFT receives a punishment for the rest of the $(1-\beta^2)m$ times. Since we assumed that the effect of rounds between first player's defection and the second player's defection is negligible, the payoff is $\mathbf{\boldsymbol{\beta}^2 mR + (1-\boldsymbol{\beta}^2)mP}$.

Therefore, we have the following payoff matrix:

$$\begin{array}{c|cc} & ALLC & GrTFT \\ \hline AALC & mR & (1-\beta)mS + m\beta R \\ GrTFT & (1-\beta)mT + m\beta R & \beta^2 mR + (1-\beta^2)mP \end{array} \tag{4}$$

3.5 ALLD vs. GrTFT

Same method from ALLC vs. GrTFT was used for completing the ALLD vs. GrTFT payoff matrix.

ALLD	$\cdots$	*D*	*D*	*D*	*D*	*D*	*D*	$\cdots$
GrTFT	$\cdots$	*C*	*D*	*D*	*D*	*D*	*D*	$\cdots$

$$\begin{array}{c|cc} & ALLD & GrTFT \\ \hline AALD & mP & T+(m-1)P \\ GrTFT & S+(m-1)P & \beta^2 mR+(1-\beta^2)mP \end{array} \tag{5}$$

3.6 TFT vs. GrTFT

Analogous to the derivation of ALLC vs. GrTFT (3.5), TFT vs. GrTFT has the following payoff matrix.

TFT	$\cdots$	*C*	*C*	*C*	*D*	*C*	*D*	*D*	$\cdots$
GrTFT	$\cdots$	*C*	*C*	*C*	*C*	*D*	*D*	*D*	$\cdots$

$$\begin{array}{c|cc} & TFT & GrTFT \\ \hline TFT & mR & m\beta R+(1-\beta)mP \\ GrTFT & m\beta R+(1-\beta)mP & \beta^2 mR+(1-\beta^2)mP \end{array} \tag{6}$$

4 Solution of the Mathematical Problem

Comparing the stability between two strategies allows us to analyze which strategy maximizes the payoff. Since Prisoner's Dilemma is an evolutionary game, the game follows the selection dynamics. Therefore, we can derive the following differential equation for the two strategies A and B using the procedure in [Appendix A]:

$$\frac{dx}{dt} = x(1-x)\big(f_A(\vec{x}) - f_B(\vec{x})\big) \tag{7}$$

where x is the frequency of strategy A, $\vec{x} = (x, 1-x)$, and $f_A(\vec{x})$ and $f_B(\vec{x})$ are the fitness of the given strategy.

Then, for two-player game with a payoff matrix [figure 4], $f_A = x_A a + x_B b$ and $f_B = x_A c + x_B d$. Therefore, we get

$$\frac{dx}{dt} = x(1-x)\big[(a-b-c+d) + (b-d)\big].$$

Two of the equilibria are $x^* = 0$ and $x^* = 1$, and their stabilities are dependent on $\big[(a-b-c+d) + (b-d)\big]$. That said, the stability analysis shows that the frequency of strategy A would become 0, 1, or behave in other predictable ways against strategy B after infinite rounds. In other words, the stability analysis under this selection dynamics indicates which strategy is more dominant. Specifically,

- if $a > c$ and $b > d$, A dominates B.
- if $a < c$ and $b < d$, B dominates A.
- if $a > c$ and $b < d$, A and B are bistable.
- if $a < c$ and $b > d$, A and B can stably coexist.
- if $a = c$ and $b = d$, A and B are mutual.

Comparing $a, b, c,$ and d requires the comparison of the 4 payoff types (R, S, T, P), so it is important to remember that $T > R > S > P$ and $2R > T + P$.

Since GrTFT is a reactive strategy, we can use a stochastic matrix to further predict the expected payoff for two strategies playing each other.

4.1 Stability Analysis

It is natural to classify the stabilities of respective payoff matrices (1-6) using the differential equation (7) for selection dynamics.

4.1.1 ALLC vs. FG

$$
\begin{aligned}
& a \stackrel{?}{=} c && b \stackrel{?}{=} d \\
& mR \stackrel{?}{=} \frac{\alpha-1}{\alpha} mR + \frac{1}{\alpha} mT && \frac{\alpha-1}{\alpha} mR + \frac{1}{\alpha} mS \stackrel{?}{=} \frac{\alpha-1}{\alpha} mR + \frac{1}{\alpha} mP \\
& \equiv 0 < \frac{1}{\alpha}(T-R) && \equiv 0 < \frac{1}{\alpha}(P-S) \\
& \equiv a < c && \equiv b < d
\end{aligned}
$$

which means FG dominates ALLC.

4.1.2 ALLD vs. FG

$$
\begin{aligned}
& a \stackrel{?}{=} c && b \stackrel{?}{=} d \\
& mP \stackrel{?}{=} \frac{\alpha-1}{\alpha} mS + \frac{1}{\alpha} mP && \frac{\alpha-1}{\alpha} mT + \frac{1}{\alpha} mP \stackrel{?}{=} \frac{\alpha-1}{\alpha} mR + \frac{1}{\alpha} mP \\
& \equiv P - S > \frac{1}{\alpha}(P-S) && \equiv T - R > \frac{1}{\alpha}(T-R) \\
& \equiv a > c && \equiv b > d
\end{aligned}
$$

which means ALLD dominates FG.

4.1.3 TFT vs. FG

$$a \stackrel{?}{=} c \qquad\qquad b \stackrel{?}{=} d$$

$$mR \stackrel{?}{=} \frac{T+S+(\alpha-2)R}{\alpha} \cdot m \qquad\qquad \frac{T+S+(\alpha-2)R}{\alpha} \cdot m \stackrel{?}{=} \frac{\alpha-1}{\alpha}mR + \frac{1}{\alpha}mP$$

$$\equiv 2R > T+S \qquad\qquad T+S \stackrel{?}{=} R+P$$

$$\equiv a > c$$

which means the stability depends on $T+S$ and $R+P$. There are two cases to be discussed.

1. If $T+S > R+P$, then $b > d$, thus TFT dominates FG.

2. If $T+S < R+P$, then $b < d$, thus TFT and FG are bistable.

4.1.4 ALLC vs. GrTFT

$$a \stackrel{?}{=} c \qquad\qquad b \stackrel{?}{=} d$$

$$mR \stackrel{?}{=} (1-\beta)T + m\beta R \qquad\qquad m(1-\beta)S + m\beta R \stackrel{?}{=} \beta^2 mR + (1-\beta^2)mP$$

$$\equiv \beta(T-R) < T-R$$

$$\equiv a < c$$

which means stability depends on β. From [figure 5], if $\beta < \beta_{crit}$, then GrTFT dominates ALLC. If $\beta > \beta_{crit}$, GrTFT and ALLC can stably coexist.

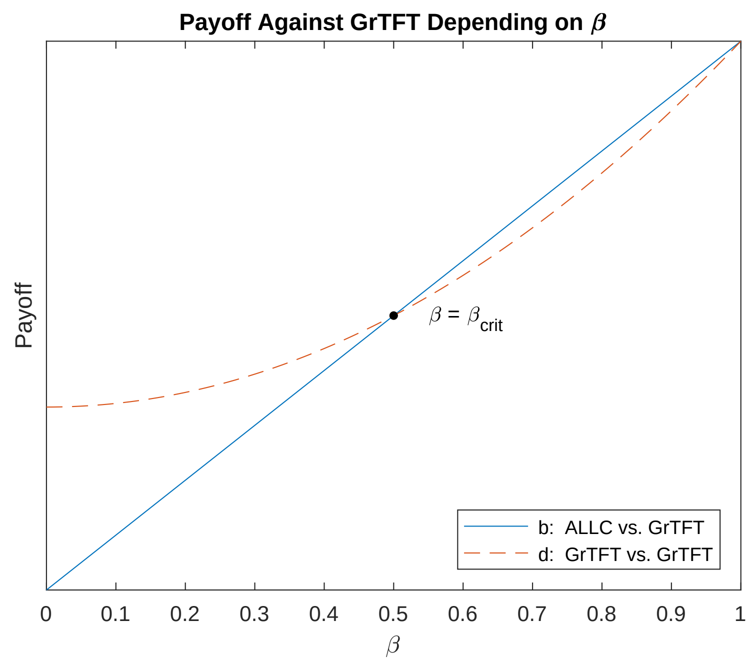

Figure 5: The graph plots the b and d components of the payoff matrix. The intersection point, β_{crit} is marked, which is where the change in stability happens. When $\beta < \beta_{crit}$, $b < d$; when $\beta > \beta_{crit}$, $b > d$. The constants (R, S, T, P) are set to $(3, 0, 5,1)$ which is a generic payoff matrix for the Prisoner's Dilemma [6] that follows the rule $T > R > S > P$ and $2R > T+P$. Any constants that follow these rules would result in a similar plot.

4.1.5 ALLD vs. GrTFT

$$
\begin{aligned}
& a \stackrel{?}{=} c && b \stackrel{?}{=} d \\
& mP \stackrel{?}{=} S + (m-1)P && T + (m-1)P \stackrel{?}{=} \beta^2 mR + (1-\beta^2)mP \\
& \equiv P > S && \equiv T - P < m\beta^2(R - P) \\
& \equiv a > c && \equiv b < d
\end{aligned}
$$

which means ALLD and GrTFT are bistable.

4.1.6 TFT vs. GrTFT

$$
\begin{aligned}
& a \stackrel{?}{=} c && b \stackrel{?}{=} d \\
& mR \stackrel{?}{=} m\beta R + (1-\beta)mP && \beta R + (1-\beta)P \stackrel{?}{=} \beta^2 R + (1-\beta^2)P \\
& \equiv 0 > (R-P)(\beta-1) && \equiv R-P > \beta(R-P) \\
& \equiv a > c && \equiv b > d
\end{aligned}
$$

which means TFT dominates GrTFT.

4.2 Expected Payoff for Strategies Against GrTFT

The expected value of payoff when strategy 1 plays strategy 2 is

$$E(S_1, S_2) = Rs_1s_2 + Ss_1(1-s_2) + T(1-s_1)s_2 + P(1-s_1)(1-s_2),$$

which is derived from the stochastic model [Appendix B]. From [figure 3], we have ALLC = $(1,1)$, ALLD = $(0,0)$, TFT = $(1,0)$, and GrTFT = $(\beta, 0)$.

When GrTFT plays ALLC, $r_1 = \alpha$, $r_2 = 0$, $s_1 = \alpha$, $s_2 = 0$.

$$E(S_1, S_2) = R\alpha + T(1-\alpha)$$

The expected value of GrTFT dominates ALLC and it is higher than all rewards for all the iteration.

When GrTFT plays ALLD, $r_1 = \beta$, $r_2 = 0$, $s_1 = 0$, $s_2 = 0$.

$$E(S_1, S_2) = P$$

The equilibrium state when GrTFT plays ALLD is punishment for all the rounds and it is lower than all rewards for all the iteration.

When GrTFT plays TFT, $r_1 = \alpha$, $r_2 = 1$, $s_1 = 0$, $s_2 = 0$.

$$E(S_1, S_2) = P$$

The equilibrium state when GrTFT plays TFT is punishment for all the rounds and it is lower than all rewards for all the iteration.

5 Results and Discussion

With the motivation to champion in the Prisoner's Dilemma, we have tried two strategies, namely Fixed Greediness (FG) and Greedy Tit-For-Tat (GrTFT), to see if FG and/or GrTFT can win other established strategies. For one-shot or fixed-round game, the approaches we invented are not applicable at all, as proven by induction, the only reasonable choice is to defect. The composition of two best responses of playing defect drives the game to a lose-lose ending, given that all players are rational. Alternatively speaking, (Defect, Defect) is the Nash equilibrium, a state in which all players have no incentive to deviate from the selection. However, under this equilibrium, both are worse off. To see why this is true, we verify that both players have a higher payoff if they were to cooperate.

Then, for the iterative game, we should remark that fitness and payoff have a positive correlation (i.e. if a strategy has a higher overall payoff, then it has a better fitness). When we pose FG against ALLC, ALLD, and TFT, corresponding to first three analyses in section 4.1, we learned the following facts:

Analysis (1) shows that when playing against FG, ALLC is a costly plan to implement, because FG defects occasionally. ALLC simply lets go of its own profits. ALLC in the real world mimics someone that has a stronger preference in sustainable growth and never retaliates when its opponent exploit the trust among players. It is doubtful that we see an abundant population of this kind of individuals.

Analysis (2) states that ALLD is a better strategy than FG; intuitively ALLD imitates a stubborn competitor that will never cooperate. FG has no chance to overturn. Suppose two companies are competing, and the one with greater capitals is aiming to boost a higher market share. To push the smaller firm out of the market, the giant enterprise could offer a cut-throat price and begin a price war. Eventually, the smaller company will be kicked out from the market. We affirm that FG is unlikely to endure the waves caused by ALLD.

Analysis (3), the stability of TFT and FG depends on the payoff matrix. Two cases are to be discussed.

1. In (i) we observe that if $T + S > R + P$, then TFT dominates FG. For example, two hotel chains are trying to open a new branch at the same location. Then, the payoff matrix will have a smaller sum of R and P, compared to T and S. R would be smaller as two hotels share the profits if they operate adjacently, and P is that neither of them will gain monetary benefits for not opening. That is, they share an

incentive to cooperatively respond to each other's choice, so that at the end they could maximize their payoff. Note that this game is very similar to the Hawk-Dove game [3].

2. The (ii) condition reads that $T+S < R+P$, meaning the players actually have better payoffs if they play the same strategy at each iteration. Battle of the sexes [2] illustrates why this could happen. Imagine a couple wants to have a date, and they have two options: either watch a basketball game or a ballet performance. Playing different strategies means that the game ends at (Cooperate, Defect) or (Defect, Cooperate), so there is no dating at all (the worst consequences). Picking the same activity means they go out for a date, absolutely a better outcome. Cooper et al. has also shown that there exists a tendency for the member in the game to periodically switch between the activities, which is a matching of the bistable equilibrium.

Next, we will take a look of what happens if deploying GrTFT against ALLC, ALLD, and TFT. These are echoing last three investigations in section 4.1.

Analysis (4) indicates that the β value decides the dynamics of the game. We see that with an extremely low β approaching 0 from the right, a low probability of cooperation when its opponent cooperates, meaning a high chance of defect. GrTFT is essentially ALLD. That is, the results of GrTFT vs. ALLC follows ALLD vs. ALLC (ALLD will always get temptations). Conversely, with a higher β, there exists a bistability between GrTFT and ALLC. The argument is the same as in analysis (5).

Analysis (5) points out that ALLD and GrTFT are bistable. We can easily see why ALLD is stable by thinking through what happens when one-shot game repeats multiple times. Recall that the payoff matrix has a $T > R > P > S$ lineup, and a prudent stakeholder normally will choose T rather than anything else, but this mindset initiates the irreversible steps of reaching Nash equilibrium, and it ends up with choosing defect for both players every time they match-up. On the other hand, GrTFT would also be a stable equilibrium as all players take turns to enjoy the temptations.

Analysis (6): TFT dominates GrTFT. From the reactive discussion in Evolutionary Dynamics [6], we know that in a mixed population (ideally, the size is large) is skewed toward TFT, implying most people prefer ALLD to TFT, this is a corollary of the Guess $\frac{2}{3}$ of the average experiment [4]. Now, if we force that participants cannot always defect, then all participants will have the inclination to do TFT (remember that TFT is also an stable equilibrium) to maximize their payoffs, GrTFT will not be stable because whenever one player does something unplanned, the earning system will collapse and all intelligent players will work really hard to avoid such thing from occurring.

Our models do not address the issue of discount rate, which is a problem when playing infinite rounds of game. The payoff in the later stage economically worth less, or the phenomena of inflation. In simple words, the same amount of money given today and ten years later are appreciated variously. To fix this problem, we will have to revise the expected payoff. Then, the next question would be why we can be certain that the discount rate is constant across the game period. Apparently, the interest rate is never fixed (equivalently saying discount rate is never steady), but this is a problem of time series, which is out of the scope of this study.

From the reasoning demonstrated above, our attempts to surpass the popular strategies are not successful. In short, FG is too weak to combat with ALLD (2), yet it takes advantage of ALLC (1), and TFT is a prioritized choice based on the two cases we assessed in (3). GrTFT has a more complicated result, GrTFT and ALLD are bistable (4). Although GrTFT dominates ALLC when β is small, it becomes stably coexist when β is larger (i.e. $\beta > \beta_{crit}$) (5). GrTFT is an inferior version of the TFT, as TFT dominant in the game GrTFT versus TFT (6). Therefore, we do not have an improvement with the new strategies, we find neither FG nor GrTFT completely exceeds any existing strategy.

6 Improvement

The winning strategy in an iterative Prisoner's Dilemma game has two important properties: NICENESS and PROVOCABILITY [1]. Niceness means that a particular strategy cannot be the first player to defect. Provocability means that the strategy incorporates the ability to retaliate in a subsequent time. The strategy that implemented both properties was TFT. However, TFT cannot tolerate any kind of error, which is unrealistic. An example of an error would be not reacting correctly to the opponent's choice. If there is an error in TFT, it leads to continuous defection for both parties, which then is no longer the most beneficial strategy. Hence, the best strategy must

1. be a nice strategy,
2. have the ability to provoke after getting exploited,
3. and have a property of forgiveness that is willing to sacrifice to return to a cooperating state.

As an attempt for improvement, we implemented FG, which has forgiveness but not niceness, and it resulted in a worse overall payoff compared to TFT. FG also was dominated by ALLD because FG cannot resist invasion of ALLD. Also, FG is not a reactive strategy. It cannot respond to others no matter what strategy the opponent is playing. For example, ALLC and ALLD are not reactive strategies either. FG follows its own pattern of choices between cooperation and defection. It cannot punish other player if they get a sucker. It does not have the ability to keep the previous choice if it gets a temptation. Most importantly, it is not a rational strategy that a human would choose. Human has the ability to react to the outcome of the previous game and FG does not consider that.

Then, we implemented a more realistic version of TFT, GrTFT. In contrast to FG, GrTFT is also a reactive strategy, and it responds to the opponent's cooperation and defection with a certain probability. However, GrTFT lacked niceness, and was dominated by TFT.

We confirmed that only having few features of the best strategy is not enough. Therefore, we can improve the model by having all three properties (niceness, provocability, and forgiveness). However, there are two major difficulties in implementing such model. The first is balancing forgiveness and one's own payoff. Forgiveness requires the player to get a sucker and make the system to be trustworthy again, but too much forgiveness will make the model vulnerable to defective strategies. Secondly, measuring when the error occurs is impossible. When the error occurs, the continuity of the payoff would no longer hold true. Modeling an instance of error requires recalculation each time, which cannot be generalized by a single model.

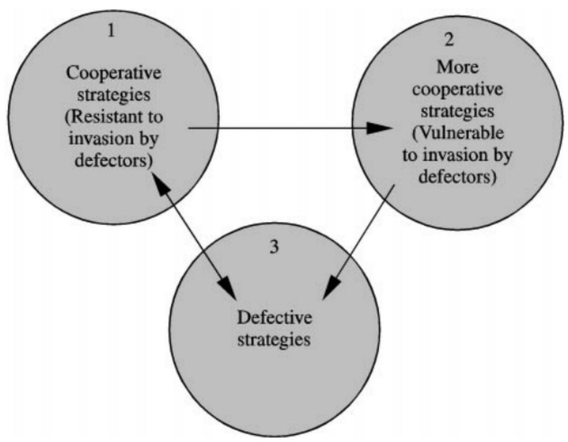

Figure 6: The dynamics of cooperation [7]

Alternatively, Wahl et.al suggests another point of view to the Game Theory, which confirms how complex modeling the Prisoner's Dilemma can get [7]. The cycle in [figure 6] represents the gradual change in strategies based on realistic rational decisions. For instance, let a population start by playing the "cooperative" strategies. When error occurs and lead to continuous defection, the population will either adapt the "defective strategies" or starting forgiving and adapt the "more cooperative" strategies to maximize their payoff. Soon after, part of the population will play defective strategies since they will result in a higher payoff against "more cooperative" strategies. Then, the remaining population will provoke and switch back to "cooperative" strategies. Assuming each strategy interacts with multiple different strategies within the population, the best strategy depends on the current state of the population, which we cannot estimate generally.

7 Conclusions

Game theory is a topic of mathematics, psychology and economics. In this paper, we emphasize the mathematical and economical perspectives of a major branch in Game Theory: Prisoner's Dilemma. Since one-shot game has been explored exhaustively, and there exists a dominant strategy to play, we move our focus on multiple rounds of games. We question the statement that Tit-for-Tat (TFT) is the best strategy in playing infinite rounds of games, so we invented two strategies, Fixed Greediness (FG) and Greedy Tit-for-Tat (GrTFT), inspired by the imperfections TFT has. FG defects periodically, as TFT fails to exploit the resources playing against always cooperative individual (ALLC). GrTFT adds a probability for correction so that mistakes appear in the system will not lead to an unfavorable consequence (all players then play defect). Our hope was that at least one of these proposed strategies could replace TFT to become a better strategy in Prisoner's Dilemma, but our analysis shows that FG is roughly as good as TFT, and GrTFT does not do a better job comparing to TFT. Furthermore, after reviewing several scholars' articles we found that to build a strategy better than TFT, one has to fix all the flaws at once. So far, with the assumption that errors do not randomly occur, we confirm that TFT is still the dominant strategy in Prisoner's Dilemma.

Appendix

A Derivation for the Selection Dynamics Differential Equation

Given selection dynamics,

$$\frac{dx_A}{dt} = x_A(f_A(\vec{x}) - \phi),$$

$$\frac{dx_B}{dt} = x_B(f_B(\vec{x}) - \phi)$$

where x_A and x_B are the frequency of two given strategies, $\vec{x} = (x_A, x_B)$, $f_A(\vec{x})$ and $f_B(\vec{x})$ are the fitness of these strategies, and ϕ, the average fitness, is defined as $x_A f_A(\vec{x}) + x_B f_B(\vec{x})$.

Since $x_A + x_B = 1$, we have $x = x_A$ and $\vec{x} = (x, 1-x)$ where x_B is represented in terms of x_A

$$\begin{aligned}\frac{dx}{dt} &= x(f_A(\vec{x}) - \phi), \quad \phi = x_A f_A(x) + (1-x) f_B(x) \\ &= x(1-x)\big(f_A(\vec{x}) - f_B(\vec{x})\big)\end{aligned}$$

B Method for Predicting the Expected Payoff

Let $S_1 = (p_1, q_1)$ and $S_2 = (p_2, q_2)$ and M_{ij} is the probability of change from i state to j state.

$$M = \begin{bmatrix} p_1 p_2 & p_1(1-p_2) & (1-p_1)p_2 & (1-p_1)(1-p_2) \\ q_1 p_2 & q_1(1-p_2) & (1-q_1)p_2 & (1-q_1)(1-p_2) \\ p_1 q_2 & p_1(1-q_2) & (1-p_1)q_2 & (1-p_1)(1-q_2) \\ q_1 q_2 & q_1(1-q_2) & (1-q_1)q_2 & (1-q_1)(1-q_2) \end{bmatrix}$$

Let $r_1 = p_1 - q_1$ and $r_2 = p_2 - q_2$ and the matrix M is regular if $abs(r_1 r_2) < 1$. We have the equality

$$x_{t+1} = M x_t$$

and the eigenvector is

$$x = \begin{bmatrix} s_1 s_2 \\ s_1(1-s_2) \\ (1-s_1)s_2 \\ (1-s_1)(1-s_2) \end{bmatrix}$$

where

$$s_1 = \frac{q_2 r_1 + q_1}{1 - r_1 r_2}$$

and

$$s_2 = \frac{q_1 r_2 + q_2}{1 - r_1 r_2}.$$

C Matlab Code for Producing Plots

```
beta=0:0.01:1;
R = 3; S = 0; T = 5; P = 1;           % generic payoff matrix
f = @(beta) (1-beta)*S+beta*R;        % payoff for ALLC against GrTFT
g = @(beta) beta.^2*R+(1-beta.^2)*P;  % payoff for GrTFT against GrTFT

% solves for the intersection point with an inital guess
beta_0 = 0.5;
beta_crit = fzero(@(beta) f(beta)-g(beta), beta_0);
fg = f(beta_crit);
plot(beta,f(beta),beta,g(beta),'--')
hold on
plot(beta_crit,fg,'k.','MarkerSize', 12)
title('Payoff Against GrTFT Depending on \beta')
xlabel('\beta')
ylabel('Payoff')
set(gca,'ytick',[])
legend(' b:  ALLC vs. GrTFT', ' d:  GrTFT vs. GrTFT','Location','southeast')
text(beta_crit + 0.05, fg, '\beta = \beta_{crit}')
```

References

[1] Axelrod, R. (1980). More Effective Choice in the Prisoner's Dilemma. *The Journal of Conflict Resolution.,24,* 379-403. Retrieved March 6, 2018, from http://www.jstor.org/stable/173638

[2] Cooper, R., DeJong, D., Forsythe, R., & Ross, T. (1989). Communication in the Battle of the Sexes Game: Some Experimental Results. *The RAND Journal of Economics, 20*(4), 568-587. Retrieved from http://www.jstor.org/stable/2555734

[3] Grafen, A. (1979). The hawk-dove game played between relatives. *Animal Behaviour,* 27, 905-907.

[4] Nagel, R. (1995). Unraveling in Guessing Games: An Experimental Study. *The American Economic Review,* 85(5), 1313-1326. Retrieved from http://www.jstor.org/stable/2950991

[5] Nowak, M., & Sigmund, K. (1993). A strategy of win-stay, lose-shift that outperforms tit-for-tat in the Prisoner's Dilemma game. *Nature,364*(6432), 56-58. doi:10.1038/364056a0

[6] Nowak, M. A. (2006). *Evolutionary dynamics: exploring the equations of life.* Cambridge: Harvard University Press.

[7] Wahl, L. M., & Nowak, M. A. (1999). The Continuous Prisoner's Dilemma: II. Linear Reactive Strategies with Noise. *Journal of Theoretical Biology,*200(3), 323-338. doi:10.1006/jtbi.1999.0997