Iterated Prisoner's Dilemma with Partial Imitation in Noisy Environments

Andre Amend, Degang Wu and K. Y. Szeto
Department of Physics, The Hong Kong University of Science and Technology, Clear Water Bay, Hong Kong, Hong Kong

Keywords: Adaptation, Prisoner's Dilemma, Memory, Noisy Environment.

Abstract: Players with one-step memory in an iterated Prisoner's Dilemma game can adaptively change their strategies after playing some games with their opponent. The probability of change of strategies depends on noise levels, the players' patience (or reaction time), and initial strategies. Players perform partial imitation, since, realistically, they can only imitate what they observe. Patience determines the frequency of a player's possible strategies changes. In this paper, we focus on the evolution of strategies between two major categories of players whose innate characters belong either to cheaters (traitors) or nice (benevolent) players. We consider them as agents whose characters are fixed, but their detailed genetic makeup can still vary among several types, so that, for example, the cheaters can evolve among different types of cheaters. We observe their evolutions by means of their degree of cooperation, where the variables are initial strategies, noise, and patience. Here, noise is incorporated in a sigmoid function that accounts for errors in learning. The numerical results show interesting features that we can explain heuristically: in the iterated games between an adaptive cheater against a patient nice player in a noisy environment, we observe a minimum degree of cooperation at a specific noise level.

1 INTRODUCTION

Game Theory describes mathematical models of conflict and cooperation between decision-makers (Axelrod, 1984; Smith, 1982; Antony, 2011). The field of evolutionary game theory was first established when Maynard Smith and Prince introduced the concept of evolutionary stable strategies in 1973 (Smith, 1982), and the emergence of cooperation in a population of selfish individuals has been studied by many researchers since. In evolutionary game theory the success of an individual (or a species) is determined by how it interacts with other individuals (or species). In this paper, we focus on the Prisoner's Dilemma (PD) (Poundstone, 1992), one of the simplest models of the interaction of two decision-makers. Using the PD game we investigate the effect of patience on the partial imitation (Wu, 2010) process of agents (or players). In this work we investigate a prisoner's dilemma game in which players have memory and can adapt their strategy. Players are modeled as being variants of two basic strategy categories: a strategy that will look for a short-term gain via betrayal or a strategy that will look for a long-term gain via cooperation, both will be in effect unless otherwise influenced by the actions of the opponent, where the player remembers his and his opponent's last move. Additionally, noisy environments in which player strategy executions are not perfect are also investigated. We also introduce the concept of "patience" of a player, which models his willingness to keep a strategy when it is not yet successful for some time. The observations of the cheater's degree of cooperation under different conditions might be used to maximize a system's efficiency that contains such individuals. Each player has two choices: to cooperate (C), which is better for the overall payoff, or to defect (D), which may be better for the individual's payoff. For example, a PD game can be represented by the following notation: CD, where one player cooperated (C), while the other player defected (D). In the model a certain value, representing a payoff, is assigned to each of the four possible outcomes of a PD game:

$$P_{ij} = \begin{bmatrix} Payoff\ CC & Payoff\ CD \\ Payoff\ DC & Payoff\ DD \end{bmatrix} = \begin{bmatrix} R & S \\ T & P \end{bmatrix} \quad (1)$$

Here P_{ij} is the payoff for an i-strategist against a j-strategist, R is called the reward for mutual cooperation, S is called the sucker's payoff, T is called the temptation for a player to defect, and P is called the punishment for mutual defection. For two players,

Amend A., Wu D. and Szeto K..
Iterated Prisoner's Dilemma with Partial Imitation in Noisy Environments.
DOI: 10.5220/0005075402280235
In *Proceedings of the International Conference on Evolutionary Computation Theory and Applications* (ECTA-2014), pages 228-235
ISBN: 978-989-758-052-9

Alice and Bob, in a Prisoner's Dilemma, the best possible outcome for Alice is to defect when Bob is cooperating. When both she and Bob cooperate they will both receive the reward R, this is only the second best outcome for Alice. If both defect both will receive the punishment P that is worse than the reward. However, the worst possible outcome for Alice occurs when she cooperates while Bob defects, in this case she will receive the sucker's payoff S. From this we can see that a PD game requires that $T > R > P > S$. An additional restriction $2R > T + S$ is often used when the PD game is played repeatedly; this ensures that mutual cooperation yields the highest total payoff of both players. In this paper the same payoff parameters were used as in Axelrod's famous PD computer tournament (Antony, 1992): S=0, R=3, P=1, T=5. The main conflict addressed by a Prisoner's Dilemma game is the best strategy for a selfish player is the worst strategy for the society that benefits from the total payoff from all players. Nowak et al. summarized five rules for the emergence of cooperation (Nowak, 2006): kin selection, direct reciprocity (Lindgren, 1994), indirect reciprocity, network reciprocity (Nowak, 1993) and group selection. In this paper it is direct reciprocity that influences the players. The same players play the PD game repeatedly and the players are given memory; this means that they can remember a certain number of past PD games and their results. Each Player also possesses a set of responses to every possible outcome of the previous games that are remembered, we call this set of responses a strategy. A result of this is that when a player defects, although he may gain a greater payoff in that round, his opponent might defect more in future rounds. This can result in a lower payoff for him, which discourages defection. In this paper the players are additionally given a certain patience, which determines the frequency of a player's strategy changes, as well as the past payoff he considers at the time of the strategy change. We also examine the effect of noise on the degree of cooperation of a cheater (who cannot adapt a strategy that maintains mutual cooperation).

2 METHOD

2.1 Partial Imitation of Players

A two-player PD game yields one of the four possible outcomes because each of the two independent players has two possible moves, cooperate (C) or defect (D). To an agent i, the outcome of playing a PD game with his opponent, agent j, can be represented by an ordered pair of responses S_iS_j. Here S_i can be either C for cooperate or D for defect. Thus, there are four possible histories for any one game between them: S_iS_j takes on one of these four outcomes (CC, CD, DC, DD). In general, for n games, there will be a total of 4^n possible scenarios. A particular pattern of these n games will be one of these 4^n scenarios, and can be described by an ordered sequence of the form $S_{i1}S_{j1}\cdots S_{in}S_{jn}$. This particular ordered sequence of outcomes for these n games is called a history of games between these two players. In a PD-game with a fixed memory-length m, the players can get access to the outcomes of the past m games and decide their next move. We use only players with one-step memory mainly for the sake of simplicity. There is no contradiction with the concept of patience, which is measured by the number of games played before the player changes his strategy. The reason is that the cumulative effect of unsatisfactory performance of a strategy over n games can be accumulated by the payoff of each game once at a time, which requires only a simple registry of the player with one-step memory (this will be discussed in more detail in section 2.3). Extension to players with longer memory will make the present problem very complex and will be investigated in future works.

2.2 Player Types

The number of moves in a strategy, given that the agent can memorize the outcomes of the last m games, is $\sum 4^m$ (Baek, 2008; Antony, 2013). In this paper we consider only one-step memory players ($m = 1$) for their simplicity, the one-step memory-encoding scheme is now described in detail. We allow our agents to play moves based on their own last move and the last move of their opponent. Thus we need 4 responses S_P, S_T, S_S and S_R for the DD, DC, CD and CC histories of the last game. The agents also need to know how to start playing if there is no history. We add an additional first move S_0. This adds up to a total of 5 moves for a one-step memory strategy. A strategy in one-step memory is then denoted as $S_0|S_P S_T S_S S_R$, where S_0 is the first move. There are $2^5 = 32$ possible strategies. In this paper, we consider a random player, whose every move in a PD game is completely random, and players with one-step memory. A one-step memory player can use any of the 32 one-step strategies, which can be classified into types as shown in Table 1.

Our definition of "nice" players are those who start the game with C and when the last game both players use C, he will also use C. As to the "cheaters", they are players who start with D, and when the last game both players use C, the cheater will use D. Note

Table 1: Classification of one-step strategies, as taken from (Antony, 2013). A square (❑) indicates that the category applies no matter what move is chosen at that point. Thus, for each square, one can either choose C or D.

History		DD	DC	CD	CC
Move	S_0	S_P	S_T	S_S	S_R
Grim Trigger	C	D	D	D	C
Tit For Tat	C	D	C	D	C
Pavlov	C	C	D	D	C
always defect	D	D	D	❑	❑
always cooperate	C	❑	❑	C	C
nice	C	❑	❑	❑	C
trusting (cheating)	❑	❑	❑	❑	$C(D)$
sucker (retaliating)	❑	❑	❑	$C(D)$	❑
contrite (exploiting)	❑	❑	$C(D)$	❑	❑
repentant (spiteful)	❑	$C(D)$	❑	❑	❑

that there are in total 32 kinds of players, and there are only 8 kinds of nice players, and 16 kinds of cheaters, while the remaining 8 kinds are neither classified as cheaters or nice players. We expect that this classification of players can capture some interesting behaviors of real players through this rather simple model.

2.3 Meta-strategies for Strategy Switching

A player with memory can change his strategy according to the results of the previous games in the iPD (iterated Prisoner's Dilemma). To do this, a mechanism, which we called meta-strategy, must be introduced for the player to switch his strategy intelligently. The mechanisms we used consist of a condition and a switching-rule. The condition applies to past payoffs: the player remembers his own and his opponent's cumulative payoffs for a certain number of rounds n. After n rounds he compares his and his opponent's cumulative payoff of the last n rounds and may apply the switching rule. Having done this, the player waits for another n rounds before repeating the process again. When his cumulative payoff of n rounds is less than his opponent's, then the condition for switching strategy is fulfilled and the player applies a rule to switch his strategy. This method may seem strange, since we are considering a one-step memory player and he can know the payoff of more than just the last round. However, the player doesn't remember any past games or their outcomes beyond the last game, all he knows is the payoff (which could be modeling wealth, status, etc.) that he and his counterpart accumulated over the course of n games, after n games this payoff accumulation is reset. Since the player does not remember the last n games (when $n > 1$), but only accesses information available to him now, he can have one-step memory. In this paper the rule used is called partial imitation, (Antony, 2011; Wu, 2010) by using partial imitation the player imitates the opponent's last move. After imitating his opponent's last move the player will use the same move (C or D) that his opponent used in this game, when he encounters the same situation his opponent encountered. The partial imitation rule is different from the commonly used traditional imitation rule (Wu, 2010; Antony, 2013), which allows players to imitate (i.e. copy) the entire strategy of their opponents with all possible moves. This is rather unrealistic as in each game, only part of the strategy can be observed (only one of the five responses in $S_0|S_PS_TS_SS_R$, is known to a one-step memory player). The partial imitation rule makes the more realistic assumption that players can only imitate what they observed before; the players do not imitate anything which they have not seen and thus this is not a copying process of the opponents entire strategy. Players that use meta-strategies can be further characterized. A players who waits longer before deciding to make a switch, uses a meta-strategy with a greater n, is named a 'patient' player. A player who makes the decision of changing their strategy faster, thus uses a meta-strategy with smaller n, is called 'impatient'. We will explore the difference in cumulative playoff and degree of cooperation of patient players versus impatient ones.

3 ONE-STEP MEMORY PLAYER VERSUS RANDOM PLAYER

3.1 Introduction

To compare the success of different meta-strategies we let a one-step memory player play against an opponent that chooses each move randomly. The simulation is run for many iPDs, where each iPD contains a certain number of games. The one-step memory player can change his strategy between games within each iPD, but in the next iPD he will start with a certain initial strategy. Each iPD is repeated an equal amount of times for each of the 32 possible one-step initial strategies.

Parameters

- The number of PD games or rounds played in one iPD.
- The level of patience (n) of a player; n applies to the player's meta-strategy.

Observables

- The payoff quotient, it is the quotient of the player's and his opponent's total payoff of all games in the iPD, or, equivalently, $payoff\ quotient\ (t) = C_f(t)/C_o(t)$, where $C_f(t)$ is the cumulative payoff of the focal player from the beginning of an iPD to time t, measured by the number of games played, and $C_o(t)$ is the cumulative payoff of its opponent.

The payoff quotient represents the player's success (in terms of payoff) relative to the opponent's success, thus different iPDs with different number of games can be compared directly; the average payoff quotient is the average over all initial strategies. Since the opponent in this game is random, the dominant strategy, that yields the highest payoff, is *always defect*. Adopting the strategy *always defect* yields an average payoff of $(P+T)/2 = (1+5)/2 = 3$ per game, the random opponent will then receive an average payoff of $(P+S)/2 = (1+0)/2 = 0.5$ per game, thus the highest possible average payoff quotient is $3/0.5 = 6$.

3.2 Results

We show in Fig. 1 the results of a one-step memory player using a meta-strategy with n=1 up to n=5 (n numbers of rounds before each strategy switch) against a random opponent. The results of up to 20.000 games per iPD are shown. The meta-strategies that do not use n=3 eventually all reach an average payoff quotient of 6, which is the highest-possible payoff quotient. The meta-strategy using n=3, however, only tends to an average payoff quotient of about 4.4.

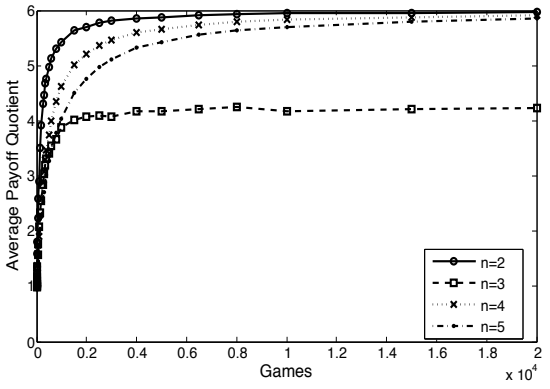

Figure 1: The average payoff quotient of all initial strategies versus the PD games per iPD of a one-step player using Partial Imitation against a random player for various patience levels (n). Shown are the averages of 1000 repetitions, for iPDs with less than 1000 PD games, and 100 repetitions, for iPDs with more than or exactly 1000 PD games.

It can be seen that, in general, impatient players increase their payoff more quickly (within fewer games) than patient players. For sufficiently few played games and the same switching rule, an impatient player is always more successful than a patient player, although the differences grow smaller when n becomes larger. A smaller n causes a higher frequency of switching, thus the player can adapt a more successful strategy faster. We see that meta-strategies, using smaller n, are always faster to change, but while the average results of most meta-strategies eventually tend to the highest possible payoff quotient 6, the meta-strategy using n=3 only reaches an average payoff quotient of about 4.4. To understand why a certain average payoff quotient can be reached it is necessary to look at the strategy-switching process under each meta-strategy in detail.

3.3 Explanation

Before analyzing the results of each meta-strategy, let us first consider what strategies can achieve the highest possible payoff quotient 6. For discussion purpose, we use the notation $DC \to C$ to denote a partial strategy. What it means is that the partial strategy $DC \to C$ is to respond with a C to a previous game where the player used D and the opponent used C. The interpretation of this notation can be easily extended to other partial strategies, e.g. $CC \to C$, $CD \to D$, etc. The player does not have to respond to all previous outcomes by using D to achieve the payoff quotient 6. After playing several games, the player will invariably use D most of the time (since D yields a higher average payoff than C). Thus, after several games, the crucial element for achieving a high payoff is that the parts of the player's strategy, which react to the player using D in the previous game, make him use D again ($DC \to D$ and $DD \to D$). If he can adopt such a strategy, then his behavior will be the same as that of an 'all D' player and his average payoff quotient, after sufficiently many games, will be 6. In the results of partial imitation (Fig. 1), meta-strategies n=2, 4, 5 all reach an average payoff quotient of 6 eventually, but n=3 is less successful, even for large numbers of games per iPD.
The player using n=3 can only reach an average payoff quotient of about 4.4. This must mean that, for certain initial strategies, he is unable to adopt the strategies $DC \to D$ or $DD \to D$. Of the two partial strategies, $DC \to D$ is the more difficult to adopt, since a game with outcome DC yields a high payoff for the player compared to his opponent. Recall that the necessary condition for a switch of strategy to occur is that the player's payoff is lower than the opponent's.

This condition is more difficult to be satisfied when DC yields a higher payoff than the opponent's. We see that using n=3 results in a relatively unstable average payoff quotient of about 4.4. We expect that this lower value of the payoff quotient for n=3 is due to the fact that some strategies cannot realize the switch from $DC \to C$ to $DC \to D$. The part of the strategy responding to a previous game DC (the player used D, opponent used C) can only be switched from $DC \to C$ to $DC \to D$ by partial imitation, when, first of all, the opponent uses D after a previous game CD (player used C, opponent used D) and, second of all, the opponent got a higher payoff after this subsequent game. For n=3 the player cannot fulfill these two necessary conditions to make the switch when using certain initial strategies (for example when his strategy includes $DC \to C$ and $CC \to D$, this will be discussed in in detail later). Other initial strategies do not necessarily enable the player to make this switch and he can end up with different final strategies. Since the opponent is random some of those strategies can sometimes adopt $DC \to D$ and achieve the payoff quotient 6, but sometimes they get stuck with the strategy $DC \to C$ and then they only achieve the quotient 1.5; this uncertainty makes the results somewhat unstable, compared to other n. For example: when the player's strategy includes $CC \to C$ and $CD \to D$, then the switch to $DC \to D$ is possible as follows:

Table 2: Adapting $DC \to D$ using 'partial imitation' and n=3.

Strategy:		Cumulative Payoff:	
player	opponent	player	opponent
C	C	3	3
C	D	3	8
D	D	4	9

If these three games were to happen at the right time, then the player using n=3 would imitate the opponent's strategy of the last game: $DC \to D$. This is possible when the player's strategy includes the responses: $DC \to C$ and $CC \to C$. Now, if all responses in the player's strategy are D except $DC \to C$, then there is no possibility for the player to adopt $DC \to D$. Thus, a player with the partial strategies $CC \to C$, $DC \to C$ adopts $CC \to D$ first, he cannot adopt $DC \to D$ in any subsequent game and he will only receive an average payoff quotient of 1.5.
When taking a look at the individual results of each initial strategy we can see which initial strategies are liable to only receive a payoff quotient of 1.5: Initial strategies that have all D ($CC \to D$, $CD \to D$, $DD \to D$) except for $DC \to C$ (and with a first move C or D) always receive the payoff quotient 1.5. Initial strategies that include $DC \to C$ and $CD \to D$ sometimes do and sometimes do not make the switch to $DC \to D$ and their average payoff quotient is thus either 1.5 or 6 (thus the unstable results). A player with patience characterized by n=3 yields lower payoff than the other n; because this player will face situations in which either the crucial change (adopting $DC \to D$) is never available across the entire iPD when the switching condition is fulfilled, or the switching condition cannot be fulfilled (due to payoff differences) once the crucial change is available. In both cases it is impossible for certain initial strategies to adopt $DC \to C$ to $DC \to D$. For more patient players (using larger n) this is avoided; when the cumulative payoffs of more games are considered, there is always a possibility that their payoff is lower when the crucial change of strategy is available. This is so because, when the payoffs of more games are considered, the payoffs of the last two games do not impact as much on the cumulative payoff. We therefore can argue that patient players are more flexible in their strategy changes. Now, how about the very impatient player with n=2? He can still reach the maximum payoff quotient value of 6 at long time, while n=2 player considers even less games than n=3. The previous explanation for n=3 cannot be applied for the n=2 case. For n=3 the average payoff quotient could not reach 6 because some of the default strategies could not adopt $DC \to D$ (instead of $DC \to C$), but if we look carefully at the process of partial strategy imitation with n=2, we see that this is not the case here. When we imagine that the player adapted a strategy that responds to everything by using D, except for $DC \to C$ (the strategy the player could not change with n=3), then the change can still occur for n=2, as shown in Table 3 below.

Table 3: Adapting $DC \to D$ using 'partial imitation' and n=2.

Strategy:		Cumulative Payoff:	
player	opponent	player	opponent
C	D	0	5
D	D	1	6

After such two games the impatient player (n=2) will adopt the strategy $DC \to D$ (which is not possible for a player with the same strategy using n=3). Thus the player using n=2 is able to adopt both $DD \to D$ and $DC \to D$, therefore achieving an average payoff quotient of 6.

4 CHEATING ONE-STEP MEMORY PLAYER VERSUS NICE ONE-STEP MEMORY PLAYER

4.1 Introduction

After we obtain the interesting features associated with one-step memory players with zero memory players, we like to consider the competition between two one-step memory players, but with different characteristics. For social interest, we like to investigate the evolution of the degree of cooperation between a cheater with a nice guy. Will the world be better off by the nice guys playing against the cheaters? We also allow possible mistakes made by each player so that the effect of noise can be included in our investigation. Our simulation is based on the following setup. The first player is a cheater (refer to Table 1 for the definition of cheater) that can change his strategy using a meta-strategy. Because of his nature of a cheater, we only allow him to change to one of the 16 cheating strategies. A cheater always has the partial strategy of $CC \rightarrow D$. The second player is a nice player that has infinite patience. By his nature, he always starts with C and responds to CC with C. Since we assume that the nice player do not change his strategy (infinite patience), his initial strategy is therefore one of the 8 nice strategies listed in Table 1. For simplicity we introduce a *noise* variable that that applies to both players. The noise in our simulation is modeled by a probability for players to use the opposite of what his strategy prescribes. The noise is represented by a number from 0 to 50 in percentage. Thus when the player intends to use C, according to his strategy, then he will use D with a probability that is equal to the noise level. *Noise* = 0% is deterministic and *noise* = 50% is completely random as the player use C or D equally likely.

Parameters

- The *noise*, which represents the probability that a player will do the opposite of what he intends to do. The noise applies to both players independently.
- The level of patience (n) of a player; n applies to the player's meta-strategy.

Observables

- Degree of Cooperation (DoC): the DoC shows the level of cooperation of the player. The DoC is the number of times C (cooperate) was used by the player (because of strategy or mistake) divided by the total number of games played in an iPD, $DoC = NC(t)/T$, where $NC(t)$ is the number of C used by the player up to time t, and T is the number of games in an iPD. Thus DoC = 0.0 (1.0) suggests the player is least (most) cooperative. Here, we only look at the cheater's DoC since he is the only player that can change his strategy during an iPD in the setup of our simulation, so that the DoC we shown in Fig. 2 is the DoC for a particular cheater strategy playing against an opponent with a particular nice strategy. Note however that the noise can change the strategies of both cheater and nice players, though not their fundamental natures as defined in Table 1.

4.2 Results

When we plot the DoC versus the *noise* of a cheater with patience n that plays against a fixed nice player for 1.000 games per iPD, we observe that for all values of the cheater's patience (n). The curves start at DoC approximately at 0.32 for zero *noise* and reach a minimum point at a *noise* between 1% and 3%, after which the curves increase linearly to DoC = 0.5 (at *noise* = 50% both players make completely random moves). In Fig. 2 we show the mean of the DoC for patience n=1 to n=5.

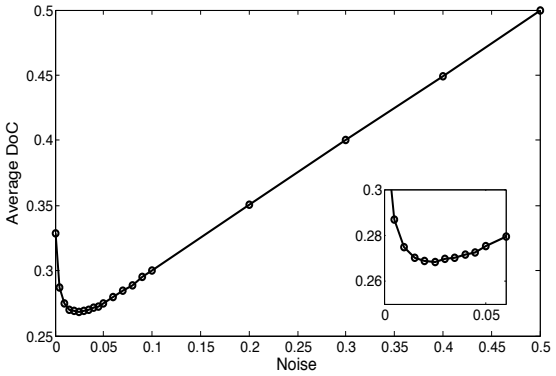

Figure 2: Average Degree of Cooperation (DoC) of $n = 1,...,5$ versus the *noise*. The insert shows a more detailed plot of the minimum (*noise* = 0 to 0.05). A cheater with fixed patience n plays against an infinitely patient nice player for 1000 games in an iPD with 100 realizations for each combination of initial strategies.

4.3 Explanation

In the game between a cheater and a nice player, both players possess some fixed moves (see Table 1) which define their characteristics, but the rest of their strategy can be considered random since we average over

all possible initial strategies of both players. For the cheater the strategy $CC \rightarrow D$ is fixed, even in the presence of noise, thus the cheater will defect after a game of mutual cooperation and a stable cooperation between the players is impossible. When the noise level is raised to about 2.5% the cheater's DoC strictly increases. This higher DoC at higher noise can be explained by the fact that the cheater's strategy includes more D than C (his DoC at *noise*=0 is around 0.3), thus most of the mistakes due to noise result in the using C instead of D and so the DoC increases. This reasoning can also provide a heuristic argument for the continued increase of DoC for a *noise* above 10%. However, for a *noise* less than approximately 2.5%, the DoC decreases with increasing *noise*. In Fig. 2, a non-linear dependency between *noise* and DoC can be seen when the noise level is below 6%. This suggests that there is a second mechanism, stronger than the first (mentioned above) at *noise* smaller than 2.5% that counteract the increase in DoC due to mistake that changes D to C.

This second mechanism, that decreases the DoC with increasing *noise*, could be connected to the *noise*'s effect on the nice player. The nice player is fixed to use C as his initial move and has a fixed response $CC \rightarrow C$, thus, by the same argument as for the cheater, the nice player will use more D when *noise* is introduced. When the nice player uses more D, the cheater's payoff decreases, since a PD game against an opponent that uses D always yields less payoff than against an opponent that uses C. Receiving less payoff puts more pressure on the cheater to change his strategy, since he is now more likely to have received less payoff than his opponent in the last n games. Strategy changes generally tend to be changes to D and only seldom to C, because the player only imitates the opponent when the opponent achieved a higher payoff, to receive a higher payoff it is necessary to use D at least once and normally several times, thus there is a greater likelihood that the opponent's last move before the player decides to imitate him was D. Therefore, when the cheater is under more pressure to make strategy changes, he will more likely change his response to D, and consequently his DoC decreases. This second mechanism (caused by the nice player defecting more) is stronger than the first mechanism (caused by the cheater cooperating more) for *noise* smaller than 2.5%, as a result there is a minima in the DoC with an initial rapid decrease on the lower side, and a linear increase larger side in the curve DoC vs. *noise*.

5 CONCLUSION AND DISCUSSION

We investigated one-step memory players, that used meta-strategies to change their strategy, and compared their relative success when playing against a random opponent in an iPD. We observed 'patient' and 'impatient' players in particular, who make decisions on strategy changes more and less frequently and base their decision partially on the cumulative payoff that they documented. We found that impatient players are faster to adopt more successful strategies, and thus they are always more successful than patient players for sufficiently few games. However, we also found that impatient players are less flexible than patient players when switching strategy. Impatient players are not able to imitate certain moves (since imitation is only possible when a payoff lower than the opponent's was achieved during the last n games and the right strategies were used in the most recent game). In some situations an impatient player cannot fulfill both of the conditions necessary for a certain switch and thus he may be unable to adopt a more successful strategy. In certain cases this makes the impatient player less successful than the patient player in the long run.

We also investigated iPD games between one-step memory players who use only cheating and nice strategies respectively. The nice player was very patient and thus didn't change his initial strategy, the cheater however used partial imitation to change his strategy and we observed his degree of cooperation after a large number of game. We found that the cheater's patience only makes a small difference in his degree of cooperation and that the cheater always cooperated about 33% (DoC = 0.33) of the times. This leads us to think about possible mistakes the player can make in their decision processes. We handled this situation simply by studying the effect of noise on the cheater's degree of cooperation. We took the average results at different patience levels and varied the noise from zero to a maximum *noise* that leads to completely random decisions of C and D. At a certain small noise level the DoC reached a minimum value, after that, it steadily increased for higher noise levels. At the maximum noise level the cheater cooperated 50% of the times, since at that noise he chooses defection and cooperation with an equal probability. This is the first effect we discussed in section 4 where we argue that the DoC will increase as noise increases. This effect is dominant at high levels of noise. There is also a second effect at small noise levels. By committing mistakes the nice player defects more, thus the cheater receives a lower payoff and is under more

pressure to adopt a more successful strategy. The best strategy against a defecting opponent is to defect, and thus the cheater's DoC decreases. This effect is dominant at low levels of noise, so that we expect a minimum DoC to emerge as a result of the competition between these two effects that both influence the cheater's DoC. Indeed, after the minimum, when the noise increases further, all the cheater's moves become increasingly random and thus his DoC tends to 0.5. Because the first effect is more prevalent at small and the second effect is more prevalent at large noise levels, the cheater becomes most uncooperative at a certain noise level. In future works, we aim at an analytical calculation of these interesting phenomena, especially the noise level when least cooperation occurs, as this may be useful for optimization problems.

ACKNOWLEDGEMENTS

K. Y. Szeto acknowledges the support of grant FS-GRF13SC25 and FS-GRF14SC28.

REFERENCES

Axelrod, R., 1984. The Evolution of Cooperation. *Basic Books*.

Smith, J. M., 1982. Evolution and the Theory of Games. *Cambridge University Press*.

Antony, M., 2011. Partial Information, Noise and Network Topology in 2x2 Games with Memory, MPhil Thesis, Department of Physics, the Hong *Kong University of Science and Technology*.

Poundstone, W., 1992. Prisoner's Dilemma: John von Neumann, Game Theory and the Puzzle of the Bomb. Doubleday, New York.

Wu, D., Antony, M., Szeto, K.Y., 2010. Evolution of Grim Trigger in Prisoner Dilemma Game with Partial Imitation, in: Di Chio, C., Cagnoni, S., Cotta, C., Ebner, M., Ekrt, A., Esparcia-Alczar, A.I., Goh, C.-K., Merelo, J.J., Neri, F., Preuß, M., Togelius, J., Yannakakis, G.N. (eds.) *EvoApplications 2010*, Part I. LNCS, vol. 6024, pp. 151-160. Springer, Heidelberg.

Nowak, M.A., 2006. Five rules for the evolution of cooperation. Science 314(5805), 1560-1563.

Lindgren, K., Nordahl, M.G., 1994. Evolutionary dynamics of spatial games. *Physica D Nonlinear Phenomena* 75, 292-309

Nowak, M.A., May, R.M., 1993. The spatial dilemmas of evolution. Int. J. of Bifurcation and Chaos 3(1), 35-78.

Baek, S.K., Kim, B.J., 2008. Intelligent tit-for-tat in the iterated prisoner's dilemma game. *Physical Review E (Statistical, Nonlinear, and Soft Matter Physics)* 78(1), 011125.

Antony, M., Wu, D., Szeto, K.Y., 2013. Partial Imitation Hinders Emergence of Cooperation in the Iterated Prisoner's Dilemma with Direct Reciprocity. In Esparcia-Alczar, A.I. (eds.) *EvoApplications 2013*, LNCS, vol. 7835, pp. 92-101. Springer, Heidelberg.