論文 / 著書情報
Article / Book Information

題目(和文)	
Title(English)	Aspiration-Based Learning Shaped by Sharing Mechanism and Its Applications
著者(和文)	SiallaganManahan
Author(English)	Manahan Siallagan
出典(和文)	学位:博士(工学), 学位授与機関:東京工業大学, 報告番号:甲第9366号, 授与年月日:2013年12月31日, 学位の種別:課程博士, 審査員:出口 弘,寺野 隆雄,新田 克己,野田 五十樹,小野 功
Citation(English)	Degree:Doctor (Engineering), Conferring organization: Tokyo Institute of Technology, Report number:甲第9366号, Conferred date:2013/12/31, Degree Type:Course doctor, Examiner:,,,,
学位種別(和文)	博士論文
Type(English)	Doctoral Thesis

2013-Year, Department of Computational Intelligence and Systems Science, Doctoral Thesis

Aspiration-Based Learning Shaped by Sharing Mechanism and Its Applications

Manahan Parlindungan Saragih Siallagan

Supervisor: Hiroshi DEGUCHI

TOKYO INSTITUTE OF TECHNOLOGY

Aspiration-Based Learning Shaped by Sharing Mechanism and Its Applications

by

Manahan Parlindungan Saragih Siallagan

A thesis submitted in partial fulfillment for the
degree of Doctor of Engineering

in the
Department of Computational Intelligence and Systems Science
Interdisciplinary Graduate School of Science and Engineering

December 2013

TOKYO INSTITUTE OF TECHNOLOGY

Abstract

Department of Computational Intelligence and Systems Science
Interdisciplinary Graduate School of Science and Engineering

Doctor of Engineering

by Manahan Parlindungan Saragih Siallagan

Learning process and the availability of information on how decision makers behave in environments that exhibit strategic interdependence is a crucial factor in theory of interdependent decision-making. Modeled agents as hyper rational have been shifted to simple adaptive leaner (bounded rational behavior) agents. These behavioral models have cognitively less demands and more plausible descriptions of real decision-making processes. In this thesis we develop such kind of behavioral model of learning that we called aspiration-based learning shaped by sharing mechanism to investigate individuals' behavior in some problems, such as social dilemmas, learning organizational, and duopoly markets. Our results suggest that within the learning method shaped by sharing mechanism the individuals tend to coordinate their action to equilibrium state. The performance of the system can be improved in some cases, e.g., social dilemmas and organizational learning, and can prove two existing equilibriums, i.e., Nash equilibrium and collusive equilibrium in duopoly markets.

Acknowledgements

First and foremost I offer my sincerest gratitude to my supervisor, Prof. Hiroshi Deguchi, who has supported me throughout my thesis with his patience, motivation, enthusiasm, and immense knowledge. You give me many inspirations not only in my research but also in my life. I would like to express my deep gratitude and respect to Dr. Manabu Ichikawa and Mrs. Masami Hayashibara whose advices and helps was invaluable to me. For helping and supporting me in my research.

In addition, I would like to thank my doctoral defense committee, Prof. Takao Terano, Prof. Katsumi Nitta, Prof. Itsuki Noda, and Prof. Isao Ono for their important comments and suggestions through the process of finishing this dissertation.

I would like to thank students in Deguchi lab, Mr. Nguyen, Miss. Chang, Miss. Xue, Mr. Aizawa, Mr. Ozaki, Mr. Watanabe, Mr. Yamagishi, Mr. Takahashi, Mr. Hashimoto, Mr. Hyoudou, Mr. Kawamura, Mr. Sakaematsu, and Mr. Atsumi for our discussion and for all the fun we have had in the Deguchi lab.

Finally, I thank my parents and families for supporting me throughout all my studies.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Learning process and the availability of information on how decision makers behave in environments that exhibit strategic interdependence is a crucial factor in theory of interdependent decision-making. In this thesis we develop a new method of learning based on aspiration level and a new information searching mechanism to investigate individuals' behavior in some problems, such as social dilemmas, learning organizational, and duopoly markets. Our results suggest that within the learning method shaped by sharing mechanism the individuals tend to coordinate their action to equilibrium state. The performance of the system can be improved in some cases, e.g., social dilemmas and organizational learning and can prove two existing equilibriums, i.e., Nash equilibrium and collusive equilibrium in duopoly markets.

1.1 Motivation

The need of learning in strategic environment has become an important theory in social science and economic (Borgers, 1996; Macy and Flache, 2002; Brenner, 2004; Izquierdo et al., 2008). The assumption of rational agents has been shifted to bounded rational agents (Simon, 1955). The reasons of this shifting can be described as follows:

1. Strong assumptions concerning rationality and common knowledge appear implausible as descriptions of the behavior of real world agents (Erev and Roth, 1998).

2. In relatively complex environments, expected payoff maximization requires too much knowledge about the environment. This notion supposes that every agent understands the environment well enough to precisely estimate payoff functions, formulate beliefs concerning the actions of others, and subsequently compute the solution to an optimization problem. If one is allowed to assume that in face of

such complexities, rational behavior entails a cost of implementation (Dziubinski and Roy, 2007; Bendor et al., 2001b).

These reasons create a space for behavioral models that are cognitively less demanding and more plausible descriptions of real decision-making processes. In these behavioral models, the agents are modeled as simple adaptive leaners. The agents do not necessarily use best responses when deciding about their strategies. The agents follow a simple rule, a satisfactory action tends to be repeated if it is given a satisfactory payoff, and explores the others' action if it give unsatisfactory payoff. This view originated in the behavioral psychology literature as stimulus-response learning or reinforcement learning. The implications of reinforcement learning in a strategic context have received much attention especially in social dilemmas.

To investigate the social dilemmas, game theory has formalized the issue as cooperation problems. The problems were represented as a mixed-motive two-person game, i.e., prisoner's dilemma game and n-person dilemma game, i.e., public goods game. The Bush-Mosteller Stochastic learning model (Bush and Mosteller, 1955) is also known as reinforcement learning model, which is designed to capture the "Law of Effect" (Thorndike, 1911). Positive reinforcement increases the tendency to play an action, while negative reinforcement decreases it. Positive or negative reinforcement is judged by a cognitive factor to stimulate their action. The standard cognitive factor is aspiration level. The difference between payoff and aspiration level will generate a stimulus (Flache and Macy, 2002; Macy and Flache, 2002; Izquierdo et al., 2008). This aspiration level is not static but evolves slowly as a player gains experience. Within these models, the behavior of the agents to reach the equilibrium states can be analyzed.

The variant of this model, i.e., payoff matching model, has successfully described human behavior in experimental studies of social dilemmas (Roth and Erev, 1995; Erev and Roth, 1998; Erev and Rapoport, 1998). This model predicts that players will learn to cooperate depending on the payoff structure. In the theoretical analysis and simulation approach, a large number of researches have been examined to solve the prisoner's dilemma game. The results showed that cooperative behavior could emerge and survive in the long run. However, the emergence of cooperative behavior depends on certain payoff conditions (Palomino and Vega-Redondo, 1999), sufficiently slow speed of updating the aspiration level (Bendor et al., 2001a), or a combination of these two factors (Flache and Macy, 2002). Flache and Macy (2002) and Macy and Flache (2002) have shown the emergence of cooperative behavior depends on learning rate and initial value of aspiration level. Besides, all previous researches use self-play environment, i.e., all agents use the same learning model.

The aspiration level plays an important factor in reinforcement learning models. The aspiration level is linearly adjusted in the direction of outcome experienced via learning rate within the model. However, this concept is also important in behavioral theory of firms (March and Simon, 1958; Cyert and March, 1963). They noted that firm management tends to express a preference for a particular performance level and this performance level seems to be persistently greater than zero. In short, management desires projects that are not just positive net present value, but projects that are significantly greater than zero net present value. The level they wish to obtain is a sociological comfort level of profits referred to as the firm's aspiration level. Aspiration levels are the borderline between perceived success and failure and denote the starting point of doubt and conflict in decision making (Greve, 1998). The difference between realized performance and the aspiration level is attainment discrepancy (Lant, 1992). This attainment discrepancy is determined by the performance history of the firm and performance feedback governs the direction of aspiration.

In the economic field, such as duopoly markets, the aspiration level has also attracted the investigation of collusive behavior in the duopoly markets. The information on the industries' average profitability might induce more collusive outcomes. In this sense, the firms perceive the industries average profitability as aspiration levels. The firms will try new strategies anytime their profits fall below the industry's average profitability (Dixon, 2000; Dixon et al., 2006). However, those researches assume that all firms have the same aspiration level which is represented by overall average profit as a reference point. Another assumption is that the information on the overall average profit is provided.

From the above explanations, we can summarize the motivation of this thesis into three parts as follows:

1. The use of aspiration level in reinforcement learning to investigate the social dilemmas still have some problems related to emergence of cooperative behavior, i.e., initial value of aspiration level, learning rate, and self-play environment.

2. The model of organizational learning purposed by March (1991) does not use the aspiration level in mutual learning between organization and the members of the organization. However, the concept of aspiration level is also important in behavioral theory of the firm as claimed by March and Simon (1958) and Cyert and March (1963).

3. The use of industries' average profitability as an aspiration level to all firms in duopoly markets is quite naive because in real world the firms may have their own aspiration level.

1.2 Aim and Contributions

The overall aim of this thesis is to advance the investigation of the impact of learning process and availability of information on how decision makers behave in environments that exhibit strategic interdependence. To fulfill this aim, firstly, we list the limitation of aspiration-based reinforcement learning from previous researches as follows:

1. In case of social dilemma: dependence of solution to some model's parameters, e.g., initial aspiration level, learning rate, and habituation.
2. Homogeneity assumption: use the same parameters and the model of learning.
3. Individual learning: no interaction.

According to our motivation, we proposed a research question as follows: If the agents follow the aspiration learning and the information about agents' aspiration level will be shared through interaction, what kind of behavior would emerged in the macro level?

To answer the research question we proposed research objective as follows:

1. To build aspiration-based learning that handles the effect of learning rate and habituation parameters.
2. To build a mechanism of sharing information (aspiration level) that handles heterogeneity.
3. To applied the learning model and the mechanism in different fields, i.e., social dilemma, organizational learning and economic (duopoly market).

The specific contributions of this thesis are to build a new aspiration-based learning that uses the dynamics learning rate. Within this method the heterogeneity aspects, i.e., different habituation parameter, different learning model, and different initial aspiration level can be handled. Besides, we also proposed a new model of sharing information to change individual learning to social learning through interaction. This method can make the agents coordinate their action and also coordinate their aspiration level. We also implemented the proposed models to different area of investigation, i.e., social dilemmas, organizational learning, and duopoly markets. This investigation is important to generalize the proposed models.

Contributions of this thesis can be described as follows:

- Improving the availability of Aspiration-Based Reinforcement Learning in terms of:

  - The number of agents involved in the problem.
  - The heterogeneity of the agents.
  - The application of the model.
  - The interaction among agents.

- Complementing the existing methodologies about human learning such as, mathematical approach, experimental approach, and numerical approach.

1.3 Related Literature in Reinforcement Learning

The first use of reinforcement learning models appear in the mathematical psychology literature with studies by Estes (1954) and Bush and Mosteller (1955). After these studies, the researches were extended to some field, i.e., social science, economics, and computer science and engineering. In social science and economics, reinforcements learning is focused on repeated game to investigate how the players learn to coordinate on efficient outcomes. Macy and Flache (2002) explored the dynamics of 2x2 (2-player 2-strategy) social dilemma games. They studied a variant of Bush and Mosteller (1955) linear stochastic model of reinforcement learning. This variant is a particular type of a wider class of aspiration-based reinforcement learning models Bendor et al. (2001a). Reinforcement learners interact with their environment and use their experience to choose or avoid certain actions based on their consequences. Actions that lead to satisfactory outcomes (i.e. outcomes that met or exceeded aspirations) in the past tend to be repeated in the future, whereas choices that lead to unsatisfactory experiences are avoided. In line with this work, Izquierdo et al. (2008) advanced this model and formalized the solution concepts of Macy and Flache reinforcement learning.

In the context of experimental game theory with human subjects, several authors have used simple models of reinforcement learning to successfully explain and predict behavior in a wide range of games (Roth and Erev, 1995; Erev and Roth, 1998; Erev and Rapoport, 1998). The purpose of these researches to fit experimental data assumed that players can only learn over immediate actions but not over a strategy set including repeated-game strategies. Theoretical works have also been done (Karandikar et al., 1998; Pazgal, 1997; Kim, 1999; Palomino and Vega-Redondo, 1999; Bendor et al., 2001a,b) showing mutual cooperation is a common long-run outcome in the social dilemma games.

The computer science and engineering literature has also used such models representing various natures of automata learning as in Narendra and Thathachar (1989). Besides, in area of multi-agent system, machine learning reinforcement learning model is also used to investigate cooperation and coordination problems (Claus and Boutilier, 1998; Dipyaman

and Sen, 2007; Bowling and Veloso, 2002; Tuomas and Crites, 1995). Basically, most of the reinforcement learning model in machine learning do not use aspiration level as reference point to make decision. The reinforcement learning model in machine learning use action-value function to find an optimal action-selection policy for any given (finite) Markov decision process (MDP) (Sutton and Barto, 1998). The well-known model is Q-learning and the variant of it.

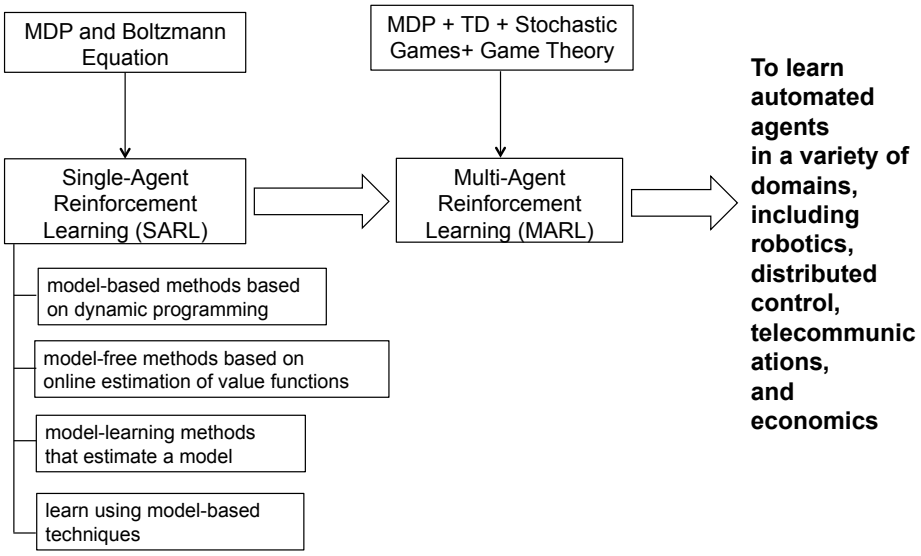


FIGURE 1.1: Multi-Agents Reinforcement Learning

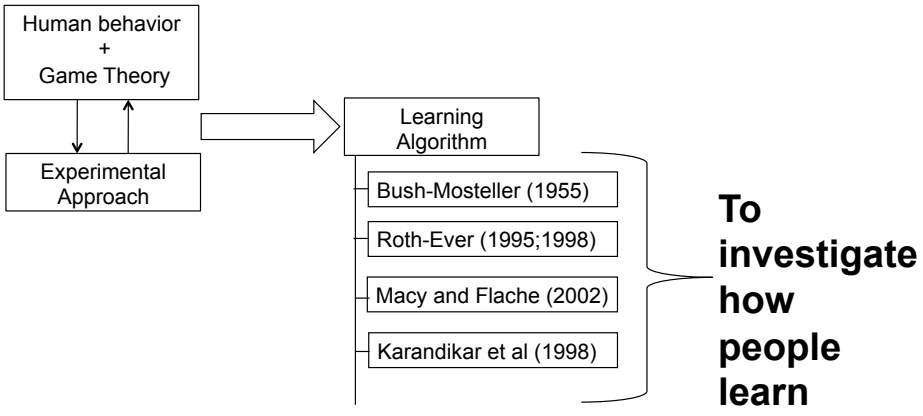


FIGURE 1.2: Aspiration-Based Reinforcement Learning

Figure 1.1 and Figure 1.2 show the investigation of Multi-Agents Reinforcement Learning (MARL) and Aspiration-Based Reinforcement Learning (ABRL). MARL is focused on machine learning or artificial intelligence fields which use Markov Decision Process

(MDP) and Boltzmann equation to investigate learning process in automated agents. This investigation also uses stochastic game and game theory to include the aspect of interactive decision making. On the other hand, Aspiration-Based Reinforcement Learning (ABRL) focuses on human behavior in terms of how behavior involved in learning process. The investigation also includes experimental approach with human subject.

1.3.1 Investigation of Previous Research on ABRL

The previous research on ABRL focused on experimental approach, a few of them have used mathematical analysis or simulation approach. Figure 1.3 shows the previous research on ABRL and their limitations.

	Methodology	**Field of investigation**	**Number of agent/agent environment**	**Result**
Estes (1954);Bush and Mosteller (1955).	Experimental + Mathematical Analysis	How people learn in game	<=2; homogeneous; no interaction; self-play learning	Dependence on parameters
Roth and Erev, 1995; Erev and Roth, 1998; Erev and Rapoport, 1998	Experimental + Mathematical Analysis	How people learn in game (non-cooperative and cooperative game)	<=2; homogeneous; no interaction; self-play learning	Dependence on parameters
Karandikar et al., 1998; Pazgal, 1997; Kim, 1999; Palomino and Vega-Redondo, 1999; Bendor et al., 2001a,b, Izquierdo et al. (2008)	Experimental + Mathematical Analysis + Evolutionary Analysis	How people learn in game (cooperative game: PD game)	<=2; homogeneous; no interaction; self-play learning	Dependence on parameters
Macy and Flache (2002)	Agent-Based Simulation	How people learn in game (cooperative game: PD game)	<=2; homogeneous; no interaction; self-play learning	Dependence on parameters; convergence time

FIGURE 1.3: Previous Research on ABRL and their Limatation

In these research, ABRL combines experimental and mathematical analysis to investigate how people learn. The main field of investigation is how people learn to play a game. From these previous research, some limitations is observed:

1. Number of agent = 2: only investigate a simple form of environment.
2. Homogeneous agent: agents have the same model of parameters.
3. Parameters dependence: the results or the solution of the model depends on parameters.
4. No interaction: there is no information which be changed.

5. Self-play learning: all players have the same learning model.

6. Models application: Application is limited to game such as Prisoner's Dilemma game

Related to these limitations, this thesis wants to overcome these limitations.

1.4 Concept of Aspiration Level

The first concept of aspiration level is an integrated scheme of level of aspiration experiment about goal-setting behavior, achievement motivation theory and social comparison theory (Quaglia and Casey, 1996). From the experiment of level of aspiration, aspiration level can be described as the level of future performance in a familiar task which an individual, knowing his level of past performance in that task, explicitly undertakes to reach. The achievement motivation can be defined as the conscious or unconscious drive to do well in an achievement-oriented activity. This achievement motivation affects the level of aspiration and the theory of social comparison suggests that aspiration levels are determined by the performance of similar others (Festinger, 1954). Through these three theories Quaglia and Casey (1996) defined aspiration level as an individual's ability to identify and set goals for the future, while being inspired in the present to work toward those goals. This construct of aspirations has two major underpinnings: inspiration and ambition. Inspiration reflects that an activity is exciting and enjoyable to the individual and the awareness of being fully and richly involved in life here and now. It is depicted by an individual who becomes involved in an activity for its intrinsic value and enjoyment. An individual with a high level of inspiration is one who believes an activity is useful and enjoyable. Ambition represents the perception that an activity is important as a means to future goals. It reflects individuals' perceptions that it is both possible and desirable to think in future and to plan for the future.

The second concept of aspiration level comes from the theory of firm's behavior (Cyert and March, 1963; Levinthal and March, 1981). The basic principle of organization behavior is that failure and success always depend on the context in which they are considered. When making evaluative judgments about success and failure, people have specific reference points they use to decide whether an outcome was a success or a failure. How people feel about success and where they set the explicit cutoff point for what determines failure depends on the information they consider important when they are making the decision. Aspiration level is determined by the performance history of the firm itself. The recent performance history of the organization can be used to set an aspiration level that holds differences among organizations. The historical

aspiration level gradually accommodates to the current performance of the firm. This is done by specifying it as an exponentially weighted moving average of experienced performance (Levinthal and March, 1981; Lant, 1992; Greve, 1998). This concept is also related to the theory of decision making. An aspiration level in individual decision making has been described as a reference point that is psychologically neutral or as the smallest outcome that would be deemed satisfactory by the decision maker (Greve, 1998). This definition is similar to a reservation value in bargaining. However, this result is very difficult to evaluate in continues measurement because of bounded rationality Simon (1955). Therefore, decision makers try to simplify evaluation by transforming a continuous measure of performance in to a discrete measure of success or failure.

1.5 Outline of the Thesis

The structure of this thesis is as follows: in Chapter 2 we build a model of aspiration-based by using dynamics learning rate. We use this model to investigate cooperative behavior in social dilemmas. The purpose of the model is to handle some problems that arose in previous aspiration-based model, i.e., initial aspiration, learning rate parameter, and habituation. In Chapter 3 we build a sharing mechanism to investigate cooperative behavior in social dilemmas. The purpose of the model is to introduce interaction among the agents. Through the interaction, agents share their aspiration levels. We use three kinds of aspiration-based learning models to describe different learning model. In Chapters 4 we use the aspiration-based learning model and the sharing mechanism to the problems of organizational learning. In the organizational learning, balancing between exploitation and exploration is an important factors to improve the knowledge earned by the organization and it members as well. To much exploitation means that the organizational may trapped in local maxima. To much exploration means that the organization may lose much money without gain any advantages. In Chapter 5 we use the aspiration-based learning and the sharing mechanism to investigate the firms' behavior in duopoly markets. Some researches state that if the firms perceive the industries' average profitability as aspiration levels, then the markets will converge to the collusive behavior. However, those researches assume that all firms have the same aspiration level which is represented by overall average profit as a reference point. Another assumption is that the information on the overall average profit is provided. To handle these disadvantages, we proposed the aspiration-based learning in which each firm has its aspiration level and using the sharing mechanism to search the information about the other's aspiration level. Chapter 6 summarizes the main conclusions of this work and identifies areas for further research.

Chapter 2

Learning to Cooperate in Heterogeneous Agents

Application of learning models to the problem of cooperation in social dilemmas has been used to analyses the convergence of cooperative solution. The reinforcement-learning model such as Bush-Mosteller predicts stochastic collusion as a backward-looking solution to the problems based on random walk. However, lock-in a cooperative equilibrium solution depends heavily on learning rate and aspiration level. This research examines dependence of learning rate and aspiration level by agent-based simulation approach. We modify the model to handle heterogeneous agents, i.e., different initial learning rate, initial aspiration level, and habituation parameter, and elaborate the model as an N-way Prisoner's Dilemma. We also consider the structure of interaction among the players. By varying the learning rate, the model overcomes heterogeneity of agents and leads to cooperative solution.

2.1 Bush-Mosteller Reinforcement Learning and Its Disadvantage

Consider a game in which the player has limited information about the environment and receives little information with each action. In this situation, a simple behavior of the player is to select an action based on past experience by playing the same game. This process will develop, consciously or unconsciously, a standard pattern of response. Specifically, the player will learn during sequences of the game. Experimental studies have been conducted to investigate the learning process in the incomplete information case (Roth and Erev, 1995; Erev and Roth, 1998).

To develop a standard pattern of response, the players adapt their behavior through simple trial-and-error learning. A well-known model is reinforcement learning such as Bush and Mosteller (BM) model (Bush and Mosteller, 1955). In this model, a successful action is more likely to be repeated, while unsuccessful ones are less likely. Each player has an aspiration level to evaluate his/her outcomes. Based on current probabilities of players, each player makes a decision and plays the game. Each of player receives an outcome and evaluates the outcome as satisfactory or unsatisfactory relative to their aspiration level. Satisfactory choices become more likely to be repeated and increase the probability of the associated choice, while unsatisfactory choices become less likely and decrease the probability of the associated choice. This process is also known as The Law of Effect. Through the learning process, each individual has a learning rate, which is interpreted as a magnitude to the individual to perceive his/her experienced such as interest in the outcome. Low learning rate means the individual perceive his/her experienced slowly through the learning process, and vice versa.

Flache and Macy have investigated reinforcement-learning model in social dilemma games (Flache and Macy, 2002; Macy, 1991). They also define two learning-theoretic equilibriums in the Prisoner's Dilemma (PD), i.e., self-reinforcing equilibrium (SRE) and self-correcting equilibrium (SCE). On the other hand, the SRE obtains when a strategy pair yields payoffs that are mutually regarding. The SCE is characterized by dissatisfying behavior. Dissatisfying means that both players will try to avoid an outcome that is better than their worst possible payoff.

Macy and Flache argue that the BM model identifies stochastic collusion as a backward looking solution for social dilemma games such as Prisoner's Dilemma Game, Chicken Dilemma Game, and Stag Hunt Game. However, this solution depends on the aspiration level of players. The individual may not learn completely to reach the solution if his/her aspiration level is too low or too high from all the available outcomes. Moreover, lock-in a mutually cooperative equilibrium is also depends on the learning rate. The lower the learning rate, the larger the number of steps that must be needed to lock-in a mutually cooperative equilibrium. It is clear that the learning rate must be same to all players so that the players can be synchronized they move.

There are some interesting questions that may arise in reinforcement learning model. This research considers two of them. The first one is how the model should be modified to overcome the effect of learning rate. Consider a social dilemma in a certain society. Each individual may come with different capacity to perceive the problem. Some of them come with low capacity (low learning rate) and others with medium or high capacity (medium or high learning level). According to the model, learning rate must be the same and thus providing synchronizes moves to reach desirable result. However, in

the beginning of the problem, individual may have different learning rate and hence learn how to synchronize the learning rate. The second one is how the model should be modified in order to overcome the effect of aspiration level. Again, individual may come with different aspiration level and may change over time. Macy and Flache model focuses most of their analysis on the case of constant aspiration level. In the case of dynamic aspiration level, they have found that habituation destabilizes stochastic collusion. Moreover, the parameter of habituation is assumed same to all players and the value is set to sufficient small to guarantee the convergence of the model (Pazgal, 1997; Oechssler, 2002; Macy and Flache, 2002; Izquierdo et al., 2008) .

With these considerations, we try to construct an environment with heterogeneous agents. By heterogeneous agents, we refer to the agents with different initial learning rate, initial aspiration level and parameter of habituation. The model we propose in this research is based on the capability of the agents to use the strength of reinforcement to update their probability of an action. During the game, the agent accumulates the stimulus (the difference between receives payoff and aspiration) for an action that has been selected. The average of the accumulate stimulus for each action will determine the strength of reinforcement via a response function. Output of the response function will be used to update probability of the associated action.

2.2 The Bush-Mosteller Reinforcement Learning Model

In this paper, we focus on the Prisoner's Dilemma (PD) game. Structure of payoffs in this game is described as in Table 1, where $T = 4 > R = 3 > P = 1 > S = 0$.

TABLE 2.1: Prisoner's Dilemma Payoff Structure

P1/P2	C	D
C	R,R	S,T
D	T,S	P,P

In general, the Bush-Mosteller model implements a stochastic decision process. In this model, individuals are modeled as stimulus-response mechanism shaped by learning forces. A player takes an action based on the propensity (probability) of the action. Aspiration and reward will generate stimulus of the action. Positive outcomes increase the probability that the associated action will be repeated, while negative outcomes reduce it.

Let P be a set of players, $P = \{1, 2\}$, $\rho_i(t)$ be the aspiration level for player i at time t, $pr_i(a, t)$ be probability of an action $a \in A = \{C, D\}$ for player i at time t, and $R_i(t)$ be a payoff for player i at time t. A stimulus associated with payoff $R_i(t)$ and aspiration level $\rho_i(t)$ for taken an action $a \in A$ for player i is $S_i(a, t) = R_i(t) - \rho_i(t)$. The stimuli for players i is defines as follows:

$$S_i(a, t) = \frac{R_i(t) - \rho_i(t)}{Z}, a \in A = \{C, D\} \tag{2.1}$$

where $Z = sup[|R_i(t) - \rho_i(t)|]$, represent the upper value of the set of possible differences between payoff and aspiration. This scaling factor will make the stimulus in the range $|S_i(a, t)| < 1$.

The BM model updates probabilities after an action (cooperation or defection) as follows:

$$pr_i(a, t + 1) = \begin{cases} pr_i(a, t) + LS_i(a, t)(t)(1 - pr_i(a, t)) & \text{if } S_i(a, t) \geq 0 \\ pr_i(a, t) + LS_i(a, t)(t)pr_i(a, t) & \text{if } S_i(a, t) < 0 \end{cases} \tag{2.2}$$

$L =$, $0 < L < 1$ is the learning rate.

The updated aspiration level $\rho_i(t + 1)$ is a weighted mean at the prior aspiration level at time t and the payoff $R_i(t)$ receives at t.

$$\rho_i(t + 1) = (1 - h)\rho_i(t) + hR_i(t) \tag{2.3}$$

where h indicated habituation, i.e., the degree to which aspiration level of floats toward the payoff. If $h = 0$, the aspiration is constant, that is, recent payoffs are ignored and the initial aspiration level ρ_{io} is preserved throughout the game. If $h = 1$, aspirations float immediately to the payoff that was received in the previous iteration.

By introducing the strength of reinforcement, we modify updating rule for probability in Eq. (2.2) as follows: Let $\beta_i(a, t)$ be the strength of reinforcement of an action $a \in A$ for player i at time t,

$$pr_i(a, t + 1) = \begin{cases} pr_i(a, t) + \beta_i(a, t)(1 - pr_i(a, t)) & \text{if } S_i(a, t) \geq 0 \\ pr_i(a, t) - \beta_i(a, t)pr_i(a, t) & \text{if } S_i(a, t) < 0 \end{cases} \tag{2.4}$$

In this equation, the strength of reinforcement is interpreted as how much interest of an player to the stimulus he/she receives by playing action a. We assume the value of this strength lies on interval (0,1]. This value has a role as learning rate, however, the updating rule of probability in Eq.(2.4) has a condition depending on the stimulus the agents receive. One interesting result here related to dynamics of the PD game can be explained by Eq.(2.4). If both players use only their stimulus to perceive positive or

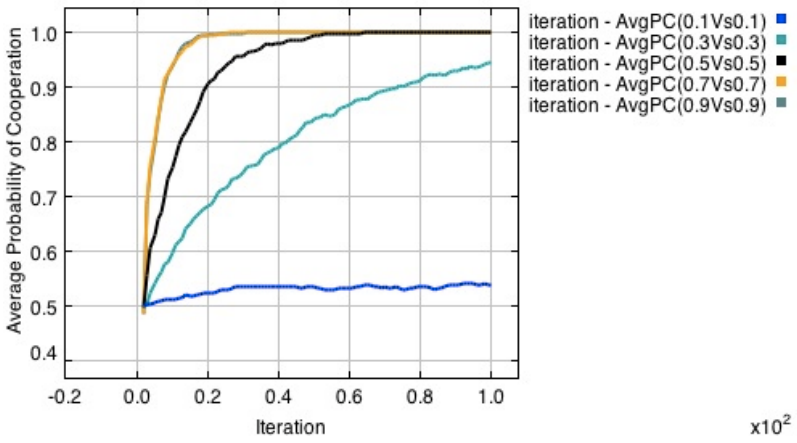


FIGURE 2.1: Same value of beta

negative reinforcement and apply Eq.(2.4) to update their probability of an action, the dynamic of the PD game will change drastically.

Figure 2.1 shows the dynamic of the PD game by using Eq.(2.4). In this simulation, we use only the stimulus $S_i(a,t)$ to perceive positive or negative reinforcement and set the same value of β for both players, i.e., $\beta_i(a,t) = [0.1, 0.3, 0.5, 0.7, 0.9]$ for all $i \in P = 1, 2$ and all t. We use $\rho_1(0) = \rho_2(0) = 2$ and habituation level $h = 0.1$ for both players. In this figure, the $x - axis$ is the number of iterations, 100 iteration, and $y - axis$ is the average of cooperative action of both players in 1000 running.

As we can see, the dynamics of cooperative action depends on the value of β. Cooperative behavior can be maintained if both players have high value of β. However, Figure 2.2 shows the dynamics of the game by using different value of β for each player. We set $\beta_1(t) = [0.1, 0.2, 0.3, 0.4]$ and $\beta_2 = [0.5, 0.6, 0.7, 0.8]$ for all t. In this figure, only one combination of β that yield full cooperative pattern, i.e., 0.4 Vs 0.8. In other cases, cooperative behavior cannot be maintained. The first player has relatively low value of β. He failed to adjust his probability of cooperation against an player who has relatively high value of β.

From the above results, we have found that the strength of reinforcement is important to improve cooperative behavior in PD game. In contrast with Eq.(2.2), Flachey and Macy have shown that the cooperative behavior also depends on the learning rate. The players should have the same and sufficient high level of learning rate in order to establish cooperative result. The initial value of aspiration level must lie on the range between maximin and the payoff of mutual cooperation (R) to produce optimal results in PD game. Moreover, by using habituation in term of parameter h, the habituation destabilized mutual cooperation.

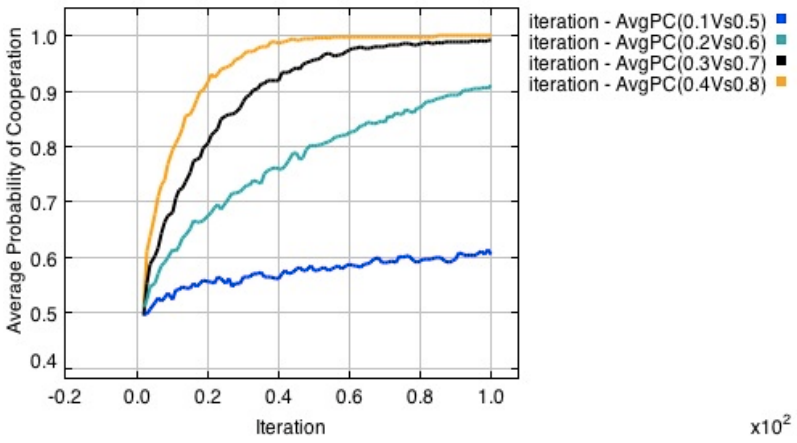


FIGURE 2.2: Different value of beta

The difference between updating probability in Eq.(2.2) and Eq.(2.4) can be explained as follows. In Eq.(2.2), the value of stimulus $S_i(a,t)$ is used both in perceiving the reinforcement effects (condition of updating probability) and in updating probability. However, in Eq.(2.4), the value of stimulus is only used in perceiving the reinforcement effects (condition of updating probability). Because the stimulus is used in updating probability in Eq.(2.2), the strength of reinforcement in term of $LS_i(a,t)$ is effected by the learning rate L. Suppose at time t player i payoff $R_i(t) = 3$ by playing action $a = C$ and $\rho_i(t) = 2$, the stimulus $S_i(C,t) = 0.5$ after scaled this values. Therefore, the strength of reinforcement is $0.5L$. We can see this strength relatively small (weak) except for L close to 1. Moreover, this strength evolves over time depend on R_i and ρ_i. This causes the value of the probability fluctuate over time. On the other hand, Eq.(2.4) do not use R_i and ρ_i to update the probability.

However, as we have already shown in Figure 2.1 and Figure 2.2, the dynamics of cooperative behavior is not guaranteed mutual cooperation regarding to the value of β. Moreover, Eq.(2.4) uses only R_i and ρ_i to determine updating condition. In our proposes model, we want to define β as a function of R_i and ρ_i.

2.3 Proposed Model

As we have explained, the value of β is considered as learning rate with respect to updating probability in Eq.(2.4). By using Eq.(2.4), we have shown the convergence to cooperative behavior can be completely established if both players have the highest value of the learning rate or β. However, in the situation in which the players have low and high value of beta, the cooperative behavior cannot completely establishes.

This is caused by a constant value of the learning rate (β), so that the players updated their probability with the same strength. Because the effect of reinforcement in term of stimulus is used only in the condition of updating probability, the highest value of the learning rate will strongly increases the probability of an action and vice versa. In this case, both players are failed to synchronize their updating process of probability.

Our proposed model can be described as follows: during the game each player will accumulates the value of stimulus his/her receives by playing an action $a \in A = \{C, D\}$. This accumulate value will be averaged by time. After that, the average value of stimulus will pass a response function to determine the value of β or learning rate to update the probability (Figure 3.1).

Let $\omega_i(t, a)$ be the total average of stimulus of action a at time t for player i:

$$\omega_i(a,t) = \begin{cases} \frac{1}{t}[\omega_i(a,t-1) + S_i(a,t)] & \text{if a is chosen} \\ \omega_i(a,t-1) & otherwise \end{cases} \tag{2.5}$$

The value of β will be updated according to:

$$\beta_i(a,t) = \frac{e^{\omega_i(a,t)}}{\sum_{a^- \in A} e^{\omega_i(a^-,t)}} \text{ if a is chosen} \tag{2.6}$$

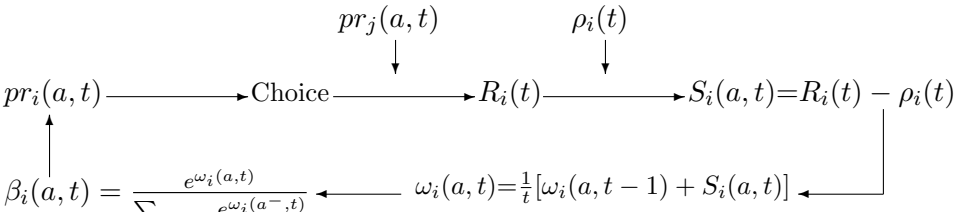


FIGURE 2.3: Schematic of the model

By using the above updating function for β, the value of β can be guaranteed in (0,1]. An action, which played frequently, will lead β to zero if the action has negative reinforcement, i.e., e^{-x} as a numerator. Therefore, the learning rate (strength of reinforcement) will be weak, so that the probability of the action has a small change. Conversely, an action, which played frequently, will lead β to one if the action has positive reinforcement, i.e., e^{x} as a numerator. Therefore, the learning rate (strength of reinforcement) will be sufficient strong, so that the probability of the action has a big change. However, the total average of stimulus will tend to zero as t increase, therefore, beta will tends to 0.5. In this process, the players should give more interest to positive stimulus and synchronize their learning rate to 0.5. We use Eq.(2.4) to update the probability.

2.4 Simulation and Results

In order to include heterogeneity of agents, we simulate the model based on several scenarios as follows:

2.4.1 The first scenario

In this simulation we want to show the effect of different value of initial aspiration and parameter of habituation. Initial aspiration level will be set differently and one player has initial aspiration below the maximin, i.e., $\rho_1(0) = 0.8$ and $\rho_2(0) = 2$. Parameter of the habituation h will be set differently, i.e., $h_1 = 0.1$ and $h_2 = 0.3$ (Figure 2.4). We also compare the propose model with BM model and Eq.(2.4) by using constant learning rate, i.e., 0.1 for the first player and 0.5 for the second player. The iteration time will be set 100 and the simulation will be replicate 1000 time under the same condition. The initial value of $\omega_i(0) = 0$ and $pr_i(C, 0) = pr_i(D, 0) = 0.5\ \forall i$ for each simulation.

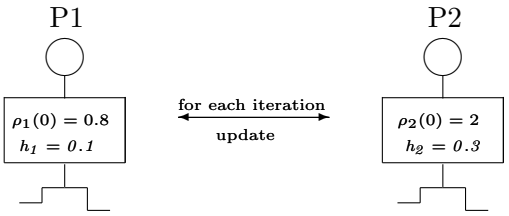


FIGURE 2.4: Structure of interaction of the first scenario

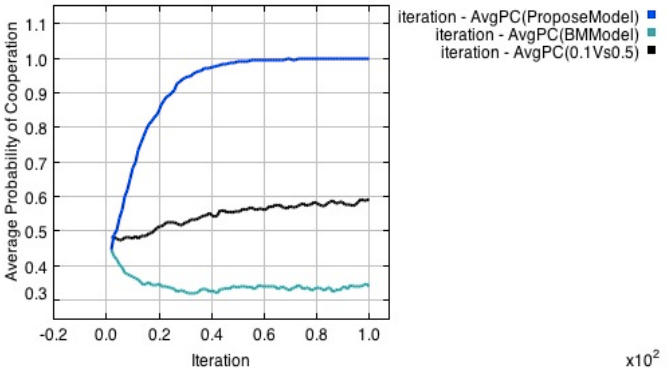


FIGURE 2.5: Two-person PD Game: Dynamics of model

Figure 2.5 shows the two-person PD game. The model improves and maintains cooperative behavior even in the situation in which the players have different initial aspiration level and parameter of habituation. The initial aspiration level of the first player is below the maximin value ($0.8 < 1.0$), therefore, he/she receives positive reinforcement

in case of DD, DC, or CC outcome and negative reinforcement in case of CD. For the second player, he/she receive positive reinforcement in case of CC or CD outcome and negative reinforcement in case of DC or DD outcome.

Table 2.2 shows the action taken for both players of one running of the simulation. As we can see, the first outcome is CD, therefore, the first player receives negative reinforcement (decreasing β value), while the second player receives positive reinforcement (increasing β value). Both players tend to play D in the next iteration because the probability of action C for the first player decreases, while the probability of action D increases for the second player. The outcome DD for the second iteration gives positive reinforcement for the first player but negative reinforcement for the second player. The first player keeps the action D until his/her aspiration level exceeded the payoff of DD outcome (the sixth iteration, Figure 2.6).

TABLE 2.2: Actions taken for both players

Player/Iteration	1	2	3	4	5	6	7	8	9	10	.	.	.	100
FirstPlayer	C	D	D	D	D	D	C	D	C	C	.	.	.	C
SecondPlayer	D	D	C	C	C	D	D	D	C	C	.	.	.	C

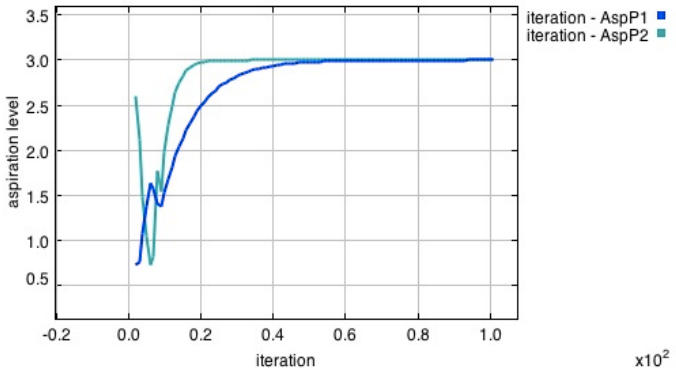


FIGURE 2.6: Dynamics of aspiration level

Through this situation, the aspiration level of the second player decreases as long as the first player played D. After that, the outcome DD is unfavorable for both players, they try to play C to make their aspiration level above the DD outcome. At this point, the outcome CC increases their aspiration level and gives positive reinforcement to the value of beta. The dynamics of beta value can be seen in Figure 2.7. As we can see, the beta value of the second player decreases as he/she receives negative reinforcement, while increases for the first player as he/she receives positive reinforcement. After that, they synchronize the value close to 0.5.

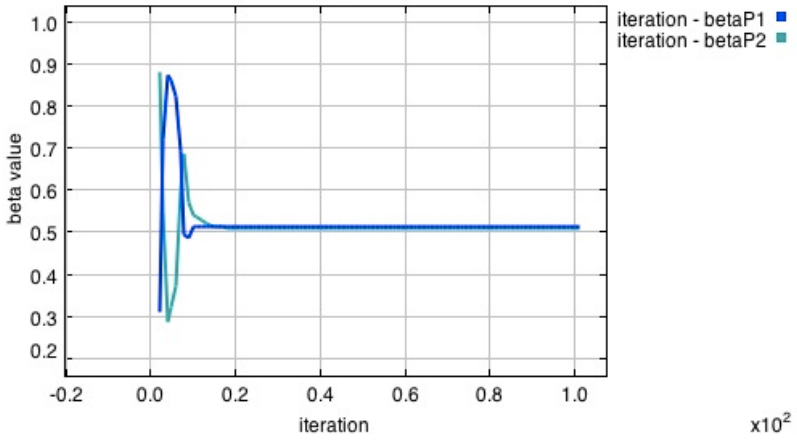


FIGURE 2.7: Dynamics of beta value

2.4.2 The second scenario

We varying initial aspiration level for the first player, i.e., $\rho_1(0) = [0...4] = 0.1 * inc$, where $\in inc = \{0, 1, 2, ..., 40\}$ and $\rho_2(0) \in CR = \{0.1, 0.5, 1.0, 2.5\}$. Parameter of the habituation h will be set differently, i.e., $h = 0.1$ for the first player and $h = 0.3$ for the second player (Figure 2.9). We want to determine the range of initial aspiration level that exhibit cooperative behavior in two-player case. The iteration time will be set 100 and the simulation will be replicate 1000 time under the same condition. The initial value of $\omega_i(0) = 0$ and $pr_i(C, 0) = pr_i(D, 0) = 0.5\ \forall i$ for each simulation.

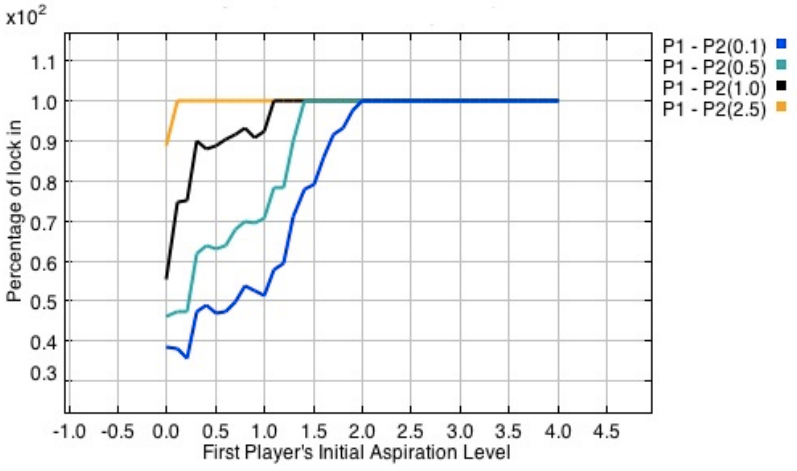


FIGURE 2.8: Varying the initial aspiration level

Figure 2.8 shows the range for which the cooperative behavior can be achieved. In this figure, the $x - axis$ is the initial value of aspiration level for the first player which vary between zero and four. The $y - axis$ is the percentage of lock-in mutual cooperation.

As we can see, the range depends on the initial aspiration level of the second player. The highest the initial aspiration level of the second player, the larger the range that exhibit cooperative behavior. The second player with $\rho_2(0) = 2.5$ exhibited the larger range for the initial aspiration level of the first player. The initial aspiration level for the first player can be made in the range between 0.1 and 4.0 and performed full mutual cooperation.

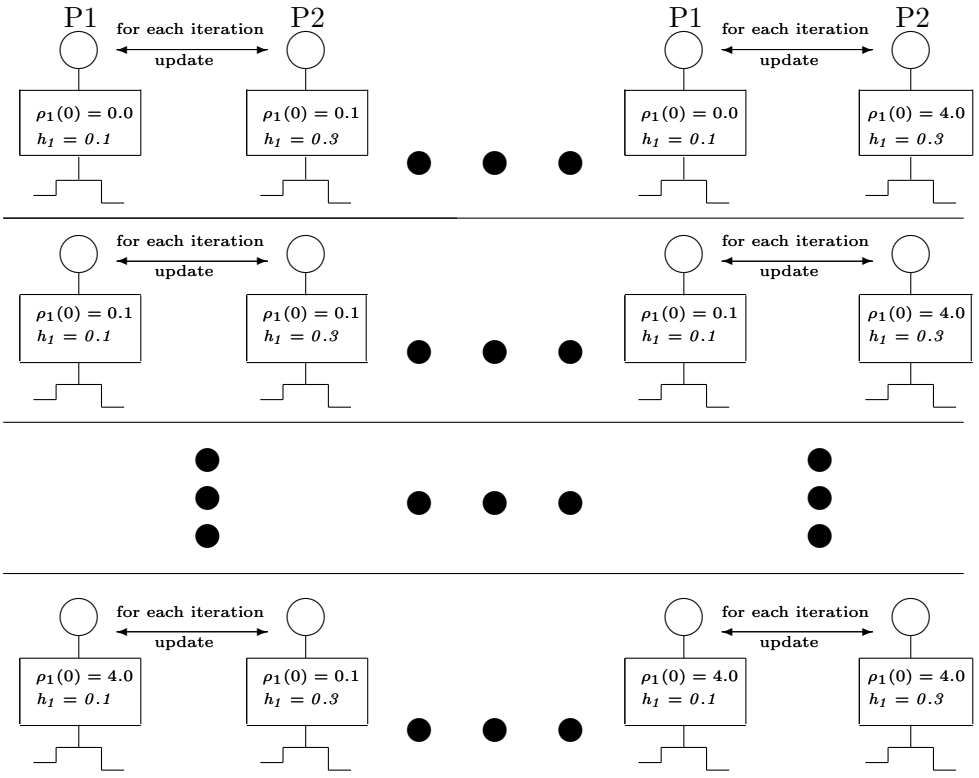


FIGURE 2.9: Structure of interaction of the second scenario

2.4.3 The third scenario

In this simulation we want to observe the capability of the model by increasing the number of players. We set the number of players $N = \{9, 16, 25, 36\}$ and for each iteration two players will be selected randomly to play the PD game for three times (Figure 2.10). For each game, selected players update their aspiration and probabilities, while not selected players will set previous value of their aspiration and probabilities. Initial aspiration level for each player $\rho_i(0) = 1.0 + \Delta A$ with $\Delta A = i * 0.1$, $i \in \{1, 2, ..., N\}$ and habituation for each player $h_i = 0.01 + \Delta H$ with $\Delta H = i * 0.01$, $i \in \{1, 2, ..., N\}$. The iteration time will be set 5000 and the simulation will be replicate 1000 time under

the same condition. The initial value of $\omega_i(0) = 0$ and $pr_i(C, 0) = pr_i(D, 0) = 0.5\ \forall i$ for each simulation.

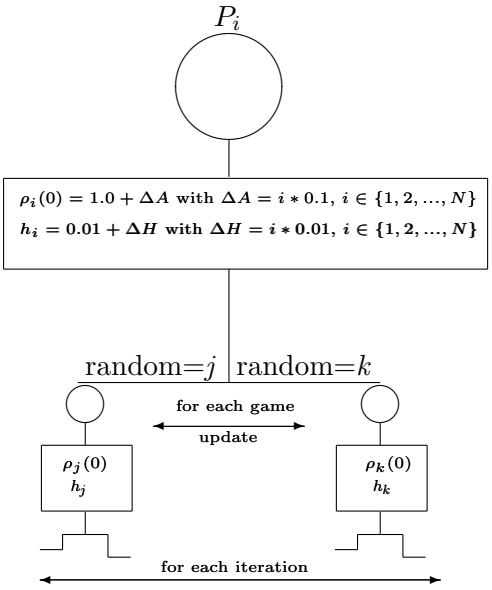


FIGURE 2.10: Structure of interaction of the third scenario

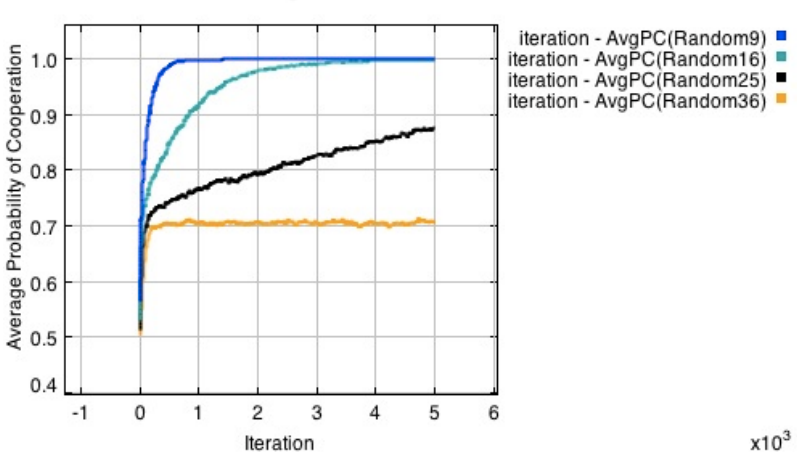


FIGURE 2.11: Random matching

In random matching structure of interaction, the more players involve in the game, the hardest cooperative behavior can be achieved. As we can see in Figure 2.11, as the number of players increase, it is difficult for the players to synchronize their heterogeneity. The players interact randomly, therefore, it is difficult for them to synchronize their learning process.

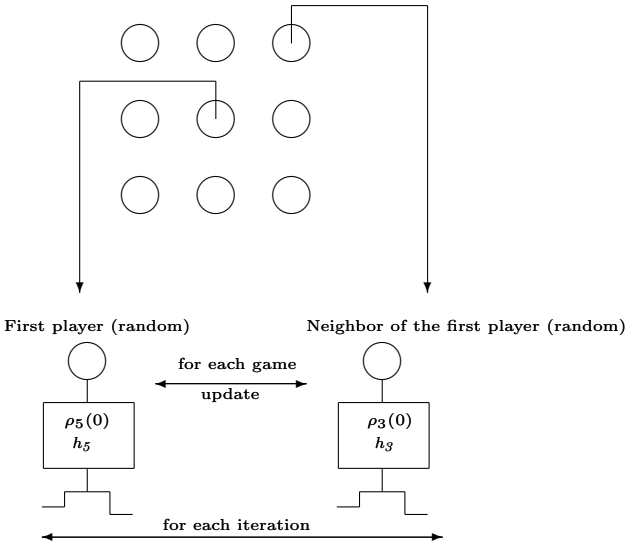


FIGURE 2.12: Structure of interaction, (rectangular $3x3$)

2.4.4 The forth scenario

In contrast to the third scenario, in this simulation each player has regular interaction between his/her neighbor. We set a rectangular regular grid with torus. We set 3x3, 4x4, 5x5, and 6x6 dimension and use Von-Neumann 4-Neighborhood. For each iteration, we select one player randomly and one neighbor will be selected randomly from the selected player to play the PD game for three times (Figure 2.12). Initial aspiration level ρ_i and habituation h_i for each player will be set as in (3). The iteration time will be set 5000 and the simulation will be replicate 1000 time under the same condition.

Besides that, we change the structure of interaction. We set a circle with $N = \{9, 16, 25, 36\}$ players or nodes and 4-Neighborhood for each player. For each iteration, we select one player randomly and one neighbor will be selected randomly from the selected player to play the PD game for three times (Figure 2.13). Initial aspiration level ρ_i and habituation h_i for each player will be set as in (3). The iteration time will be set 5000 and the simulation will be replicate 1000 time under the same condition. The initial value of $\omega_i(0) = 0$ and $pr_i(C, 0) = pr_i(D, 0) = 0.5\ \forall i$ for each simulation.

In contrast to regular interaction, i.e., rectangular grid and circle, the cooperative behavior can be achieved by increasing the number of players until 25 (Figure 2.14 and 2.15). The circle structure of interaction perform well in term of speed of convergence. Moreover, in case of the number of players increase until 36, the circle structure can achieve about 0.91 of cooperative probability. The model do better to establish full cooperative behavior if the players can interact regularly. The model can also handle variety of initial aspiration and parameter habituation in regular interaction.

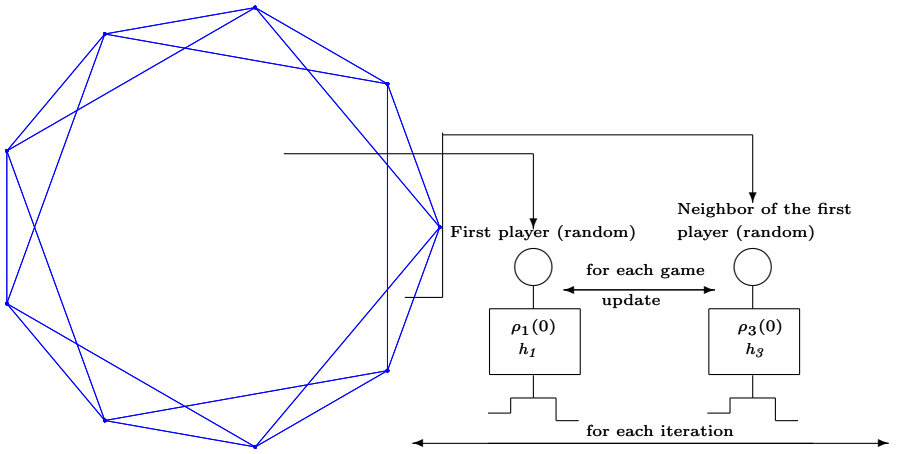


FIGURE 2.13: Structure of interaction, (circle $N = 9$)

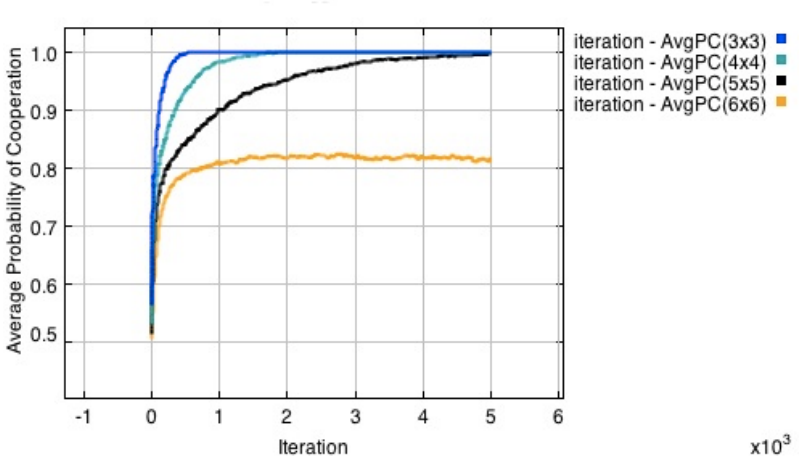


FIGURE 2.14: Rectangular Structure

2.4.5 The fifth scenario

In this simulation we want to observe a threshold of the number of players for which the cooperative behavior is maintained. We increase the number of players $N = \{9, 16, 25, ...\}$ until we found the threshold. For each iteration two players will be selected randomly to play the PD game for three times (Figure 2.10). We divided this scenario to two types. The first type, we use the same initial aspiration for all players, i.e., $\rho_i(0) = 2.5$ and the same habituation level for all players, i.e., $h_i = 0.01$. The iteration time will be set 150000 and the simulation will be replicate 1000 time under the same condition. The initial value of $\omega_i(0) = 0$ and $pr_i(C, 0) = pr_i(D, 0) = 0.5\ \forall i$ for each simulation. In the second type, we set the initial aspiration level for each player $\rho_i(0) = 1.0 + \Delta A$ with $\Delta A = i * 0.1$, $i \in \{1, 2, ..., N\}$ and habituation for each player $h_i = 0.01 + \Delta H$ with $\Delta H = i * 0.01$, $i \in \{1, 2, ..., N\}$. The iteration time will be set

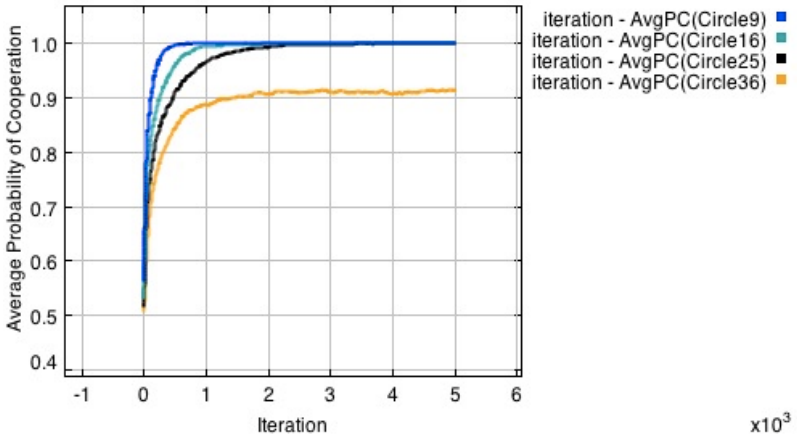


FIGURE 2.15: Circle Structure

50000 and the simulation will be replicate 1000 time under the same condition. The initial value of $\omega_i(0) = 0$ and $pr_i(C, 0) = pr_i(D, 0) = 0.5\ \forall i$ for each simulation.

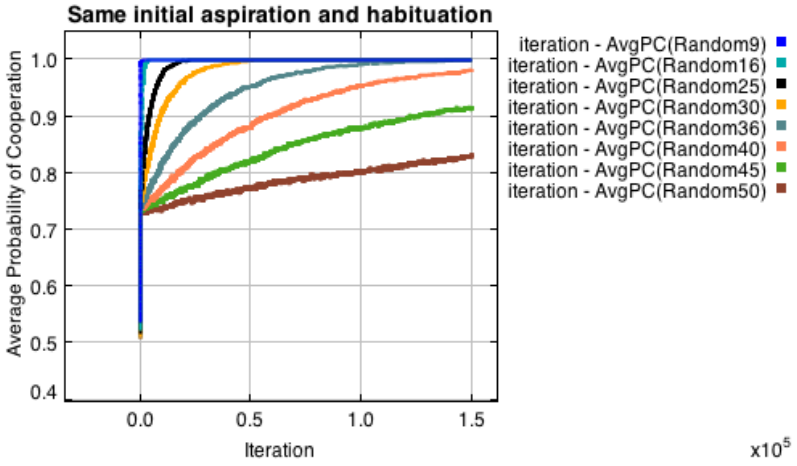


FIGURE 2.16: Same initial aspiration level and habituation

As we can see in Figure 2.16, the threshold is about 40 players. In this case we use 150000 iterations to observe the convergence to the cooperative behavior. As the number of players increase, the more time is needed to reach the convergence to the cooperative behavior. However, within 150000 iterations, the threshold is about 40 players. On the other hand, with different initial aspiration and habituation level the threshold number is 26 players. Above this threshold the dynamics remind oscillated around 0.72. The coordination process is hard in this environment compare with the case in which all players come with same initial aspiration level and habituation.

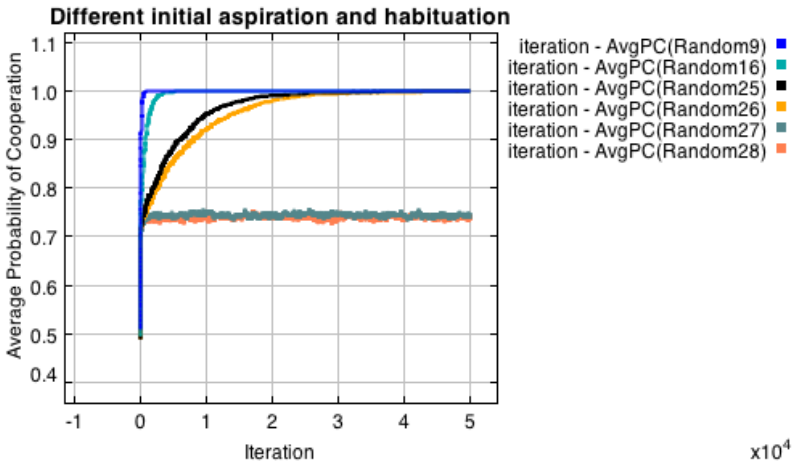


FIGURE 2.17: Different initial aspiration level and habituation

2.4.6 Trembling hands process

In this simulation we want to study the robustness of the model where players suffer from"trembling hands" (Selten, 1975): after having decided which action to undertake, each player i may select the wrong action with some probability $\epsilon_i > 0$ in each iteration. Initial aspiration level will be set differently and one player has initial aspiration below the maximin, i.e., $\rho_1(0) = 0.8$ and $\rho_2(0) = 2$. Parameter of the habituation h will be set differently, i.e., $h_1 = 0.1$ and $h_2 = 0.3$. We also compare the propose model with BM model and Eq.(2.4) by using constant learning rate, i.e., 0.1 for the first player and 0.5 for the second player. The iteration time will be set 100 and the simulation will be replicate 1000 time under the same condition. The initial value of $\omega_i(0) = 0$ and $pr_i(C, 0) = pr_i(D, 0) = 0.5\ \forall i$ for each simulation. We set the $\epsilon_i = 0.001$

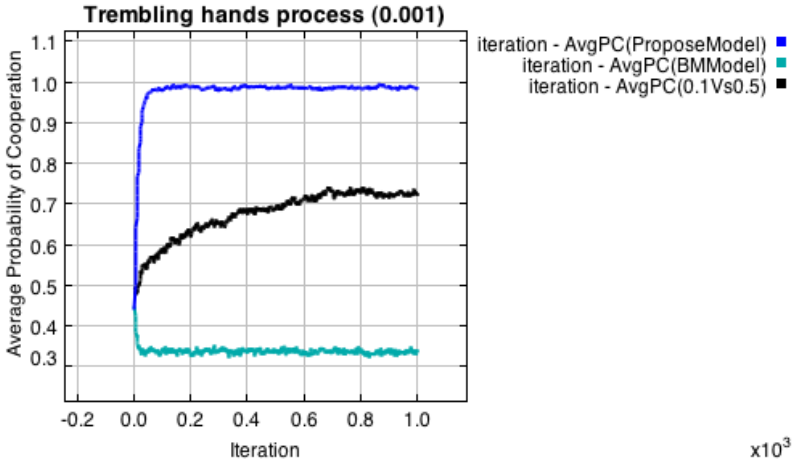


FIGURE 2.18: Trembling hands process

As we can see in Figure 2.17, the proposed model is quite robust to small value of trembling hand process, i.e., $\epsilon_i = 0.001$ compare with BM model or the model by using constant learning rate, i.e., Eq.(2.4).

2.5 Discussion and Conclusion

We have shown the model improve and maintain cooperative behavior in heterogeneous agents. Varying the learning rate can eliminate the heterogeneity of agents. In this model, the stimulus is used for condition to update the probability of an action, therefore, the effect of aspiration level not directly involve in updating process. This situation makes the value of probability not fluctuate in case of the players (agents) have habituation in term of the parameter h. Besides that, the model uses the aggregate value of the stimulus to vary the learning rate. This process makes proportional updating value of probability regarding to the aggregate value of the stimulus received by players (agents).

The model also describes a characteristic of learning to reach an optimal outcome. The agents should slowly increase their learning rate after they receive negative stimulus and slowly decrease their learning rate after they receive positive stimulus. Therefore, the process of exploration and exploitation of the available actions can be made balanced until they found an optimal action.

The model has a threshold about 40 players in which all players have the same initial aspiration level and habituation. In the case all players have different initial aspiration level and habituation, the threshold is 26 players. The model improve the coordination among players. The model is quite robust dealing with the trembling hands process for small value of probability $\epsilon > 0$.

The structure of interaction can effects dynamic of cooperative behavior by using the model. The model successfully establishes cooperative behavior in regular interaction in which all players continuously interact with others. However, the more agents involve in the game, the more difficult to reach cooperative result. In the simulation, we select one player randomly and one neighbor will be selected randomly from the selected player to play the PD game for three times. Therefore, there is a chance for the players will not be selected if the number of players increases. This is only one possibility based on our simulation. The future work will be investigated the other possibilities.

Chapter 3

A Mechanism of Sharing Aspiration to Promote Cooperative Behavior in a Group

Many experimental studies have been built to investigate social dilemmas. There are some clues in these studies that can be used to create mechanisms to overcome social dilemmas. This research deals with a simulation model that uses a mechanism of sharing aspiration based on some of such clues, i.e., individuals' expectations, information seeking and communication, obtained from previous experimental studies on social dilemmas. A mechanism of sharing aspiration is combined with a learning process to promote cooperative behavior in a group. Simulation results show that the mechanism of sharing aspiration shaped by learning can promote cooperative behavior in a group.

3.1 Social Dilemmas in Psychology View

The tension between individual interests and collective interests is a primary concern in social dilemmas. Individuals in social dilemmas are better off by contributing nothing or very little of one's own resources to the public goods. However, if all group members act in this way, public goods will not be provided. Also, self-interested individuals can take advantage of cooperative individuals (the individuals who contribute to the provision of public goods) (Dawes and Messick, 2000).

Laboratory experimentation has been conducted to investigate this issue. On one hand, there are some experimental studies related to the role of a leader in a group. The results suggest that the effectiveness of the leader's role to improve the cooperative behavior in

a group not only depends on the leader's characteristics but also depends on individuals expectations (Cremer and Vugt, 2002; Vugt et al., 2004). On the other hand, information seeking and communication are also related to solving social dilemmas. Accessibility of information refers to the individuals' strategies, that is, people have clear preferences of particular pieces of information and that information preferences vary systematically across individuals as a function of their contribution strategies (Kurzban and Descioli, 2008). When the information about aggregated contribution to the public is provided, the cooperative behavior of individuals would vary from strong co-operators to strong free riders (Kurzban and Houser, 2001).

The interaction of individual opinions in group discussion can improve the cooperative behavior (Hopthrow and Hulbert, 2005). This interaction is also known as communication effect on cooperation (Bicchieri and Lev-On, 2007). As long as communication persists, cooperation rates are high and stable. Conversely, without communication, cooperation rates gradually decline. The communication is achieved in subsequent periods, which allowed the individuals to share information and reach a better understanding of their tasks (Kerr and Kaufmann-Gilliland, 1994).

As we can see in the above findings of experimental studies on social dilemmas, individuals' expectations, information, and communication have an important role in order to spur cooperative behavior in social dilemmas. However, this role also depends on the way that individuals identify, discuss, and commit to make cooperation possible (Kerr and Kaufmann-Gilliland, 1994). The identification is related to the way that individuals perceive others in terms of similarities and closeness. The more similar and closer the individuals are, the easier the communication happens. The discussion is about the cues on the character and motives of individuals that are exchanged during the communication process. If the cues of the character and motives are about "the dilemmas", the cooperative behavior could be improved. Conversely, the cooperative behavior could be decreased if the cues of the character and motives are not about "the dilemmas". The commitment is a promise to keep cooperative behavior during the interaction. All these factors would determine the level of cooperation in the social dilemmas.

In line with laboratory experimentations to investigate the social dilemmas, game theory has formalized the issue as cooperation problems. The problems are represented as a mixed-motive two-person game, i.e., prisoner's dilemma game or n-person game, i.e., public goods game. In recent studies, the analysis of the games has been shifted from high game theory to low game theory. In high game theory the players are modeled as hyper rational. In low game theory the players are modeled as simple adaptive learners (Roth and Erev, 1995). The interpretation of the player as a learner has resulted in a number of learning models such as Bush-Mosteller Stochastic learning model (Bush

and Mosteller, 1955). This model is also known as reinforcement learning model, which is designed to capture the "Law of Effect" (Thorndike, 1911). Positive reinforcement increases the tendency to play an action, while negative reinforcement decreases it. Positive or negative reinforcement is judged by a cognitive factor to stimulate their action. The standard cognitive factor is aspiration level. The difference between payoff and aspiration level will generate a stimulus. This aspiration level is not static but evolves slowly as a player gains experience.

The variant of this model, i.e., payoff matching model, has successfully described human behavior in experimental studies of social dilemmas (Roth and Erev, 1995; Erev and Roth, 1998; Erev and Rapoport, 1998). This model predicts that players will learn to cooperate depending on the payoff structure. In the theoretical analysis and simulation approach, a large number of researches have been examined to solve the prisoner's dilemma game. The results showed that cooperative behavior could emerge and survive in the long run. However, the emergence of cooperative behavior depends on certain payoff conditions (Palomino and Vega-Redondo, 1999), sufficiently slow speed of updating the aspiration level (Bendor et al., 2001a), or a combination of these two factors Flache and Macy (2002).

Based on the above information, we claim that there are three crucial factors to overcome social dilemmas, i.e., individuals' expectation, information and communication. We use these three factors to develop a mechanism of sharing information, i.e., sharing the level of aspiration. In a group, people may have different aspirations toward their relationship. We can consider this aspiration as a goal or an expectation which is what the people are willing to achieve. We assume the members of the group can interact (communicate) with each other to share their aspiration level (goal or expectation). The information that one person would use depends on the closeness of this person to other person. We adapt a social comparison theory (Festinger, 1954) to build a mechanism for sharing information. The information that has been received by one person is used to update his/her aspiration level by comparing with his/her current aspiration level. This concept reflects a process that involves identification, discussion and commitment within the interaction.

We combine the sharing aspiration process with the learning process. The discrepancy between the current payoff and the aspiration level will generate a stimulus, which would be used to update the probabilities of the actions. We use three models of learning, i.e., Roth-Erev, Borgers-Sarin, and Satisfying, which are based on stimulus-response mechanism shaped by a learning force. This situation reflects the fact that people may learn with different models of learning. Within this framework, we want to promote the cooperative behavior in a group.

3.2 Public Goods Game

A conflict situation in a group can be modeled as Prisoner's Dilemma (PD) game. In contrast to the more familiar PD game in which the strategy space of each player is binary (cooperate/defect), the strategy space in-group conflict that we consider here is discrete. Consider a group consisting of N players. Assume each of the N players has the same endowment that we denote by e. At each time t, every player is faced with a decision of allocating a units of his/her endowment. We assume the strategy space is any discrete number that does not exceed the endowment. This assumption can be made more general if the strategy space is continuous, i.e. any fraction that does not exceed the endowment (Anna and Rapoport, 2006).

Let a_i be the amount contributed by player i, where $a_i \in A = \{0, 1, ...,\text{e}\}$ and let the total contribution of N players by $X = \sum_{i=1}^{N} a_i$. The payoff of player i at time t is given by,

$$R_i(t) = (e - a_i) + g\frac{X}{Ne} \tag{3.1}$$

where, g is public good that all of N players generates if each of player contribute his entire endowment e. The second term in equation (3.1) is equal to zero if each player contributes nothing, g if each contributes his/her entire endowment e, and some intermediate value between 0 and g if $a_i \in A \setminus \{0, e\}$. The game has the PD property if $0 < g < Ne$, the equilibrium solutions for player i is to contribute nothing, i.e., $a_i = 0$ (Rapoport and Amaldoss, 1999).

3.3 Aspiration-Based Learning Models

In this section we describe three models of learning based on aspiration. In these models, players' behaviors are based on two properties. First, they have a stimulus, i.e., the discrepancy between payoff and aspiration level which divides outcome into two subsets, i.e., satisfactory and unsatisfactory. Second, players learn via trial-and-error, and become more inclined to try actions that satisfy their aspiration level and less likely to try those that do not. Aspiration level is endogenous, i.e., they adjust to a players experience (payoff).

In this section we described three models of learning based on aspiration. The models use discrepancy between payoff and aspiration level as a stimulus to update the probability of an action. Figure 3.1 shows a scheme of the learning models (Flache and Macy, 2002). A player takes an action based on the probability of the action. Aspiration and the payoff will generate a stimulus of the action. Positive outcomes increase the probability that

the associated action will be repeated, while negative outcomes reduce it. In this sense, the stimulus of a player will encourage him/her to find a "good" action (in this game is to contribute to public goods). In real life situations, an individual will receive a stimulus from a discrepancy between achievement (outcome) of his/her works and what he/she expects from the work (aspiration). However, the real life situations involve a random exposure to environmental variables and humans are sometimes inertial, i.e., they do not invariably adapt or learn as well. To accommodate these circumstances, we can introduce randomness to extend the models. Thus with probability ϵ_1, a player does not adjust his/her probability in a current period. With probability ϵ_2, a player may not adjust his/her aspiration level and after having decided which action to undertake, he/she may select the wrong action with probability ϵ_3. However, the current research does not consider this extended model and it will be considered in future research.

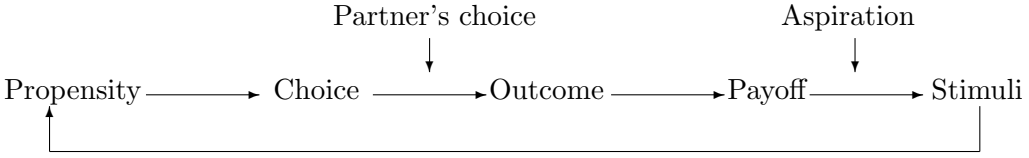


FIGURE 3.1: Schematic of Learning Models

Let, $\rho_i(t)$ denotes the aspiration level for player i at time t, let $pr_i(a,t)$ be probability of an action $a \in A = \{0, 1, ..., e\}$ for player i at time t, and $R_i(t)$ be a payoff for player i at time t. A stimuli associated with payoff $R_i(t)$ and aspiration level $\rho_i(t)$ for taken an action $a \in A$ for player i is $S_i(t) = R_i(t) - \rho_i(t)$. We described the three learning models and differences among them as follows.

3.3.1 Roth-Erev Model

In this model, we denote $q_i(a,t)$ as the propensity of an action $a \in A$ at time t. The propensity to take an action $a \in A$ for player i is updated by setting:
if action a was chosen at t,

$$q_i(a, t+1) = \max\{\nu, (1-\phi)q_i(a,t) + (1-\epsilon)S_i(t)\} \tag{3.2}$$

otherwise,

$$q_i(a, t+1) = \max\{\nu, (1-\phi)q_i(a,t) + \epsilon S_i(t)\} \tag{3.3}$$

Parameter $\nu > 0$ is a technical parameter to ensure that propensities remain positive. Parameter ϕ slowly reduces the importance of past experience and parameter ϵ prevents the probability of choosing any action from going to zero. The probability of choosing

action $a \in A$ at time t for player i is proportional to past average propensities,

$$pr_i(a,t) = \frac{q_i(a,t)}{\sum_{a \in A} q_i(a,t)} \forall a \in A \tag{3.4}$$

The aspiration level is also linearly adjusted in the direction of outcome experienced, if $R_i(t) < \rho_i(t)$,

$$\rho_i(t+1) = (1-\omega^-)\rho_i(t) + \omega^- R_i(t) \tag{3.5}$$

if $R_i(t) \geq \rho_i(t)$,

$$\rho_i(t+1) = (1-\omega^+)\rho_i(t) + \omega^+ R_i(t) \tag{3.6}$$

The parameters ω^- and ω^+, control the adjustment of the aspiration level following negative and positive rewards.

3.3.2 Borgers-Sarin Model

The probability of choosing action $a \in A$ at time t for player i is given as follows:
If $S_i(t) \geq 0$,
if a was chosen at t,

$$pr_i(a,t+1) = (1-\lambda S_i(t))pr_i(t) + \lambda S_i(t) \tag{3.7}$$

otherwise,

$$pr_i(a,t+1) = (1-\lambda S_i(t))pr_i(t) \tag{3.8}$$

If $S_i(t) < 0$,
if a was chosen at t,

$$pr_i(a,t+1) = (1+\lambda S_i(t))pr_i(t) \tag{3.9}$$

otherwise,

$$pr_i(a,t+1) = (1+\lambda S_i(t))pr_i(t) - \lambda S_i(t) \tag{3.10}$$

The parameter λ controls the effect of rewards in $pr_i(a,t+1)$, and can assume any value guaranteeing that the absolute value of $\lambda S_i(t)$ always lies (strictly) between zero and one (learning rate). The greater value of λ, the faster the adaption. The aspiration level is also linearly adjusted in the direction of outcome experienced,

$$\rho_i(t+1) = (1-\beta)\rho_i(t) + \beta R_i(t) \tag{3.11}$$

where, $0 \leq \beta < 1$.

3.3.3 Satisfying Model

Each player i will make decision based on following criteria: if $R_i(t) \geq \rho_i(t)$, $a_i(t+1) = a_i(t)$, otherwise select an action $a \in A$ randomly. The aspiration level is updated as the convex combination of the old aspiration and the current reward via learning rate λ,

$$\rho_i(t+1) = (1-\beta)R_i(t) + \beta\rho_i(t) \tag{3.12}$$

3.3.4 Differences of The Three Learning Models

All three learning models use a stimulus, i.e., the discrepancy between payoff and aspiration levels, to determine an action to be chosen. Moreover, the three models also use the same formulation to update the aspiration level, i.e., linearly adjusted in the direction of outcome experienced via learning rate. The main difference is the way that a stimulus is used to determine the action that will be chosen. The Roth-Erev model uses the stimulus to calculate the propensity of the action and uses this propensity to update the probability. In the Borgers-Sarin model, the stimulus is used directly to update the probability of an action and also as the condition to update the probability. In the satisfying model, the stimulus is only used as a condition to change the action. This model is similar to satisfying theories that state that decision makers search for new alternatives if and only if today's action is unsatisfactory, i.e., yields a payoff that falls below the decision maker's aspiration level (Bendor et al., 2004). On the other hand, Roth-Erev and Borgers-Sarin models are reinforcement learning models that use a probability of an action. If such action produces a satisfactory payoff in the current period, then the player will not decrease his/her probability of that action. Therefore, the state space of reinforcement learning models is more complex in terms of the probability over all available actions.

3.4 Mechanism of Sharing Aspiration

The theory of social comparison states that a person tends to make a self-evaluation based on comparison with other persons. In this situation, the information of others would determine the behavior of the person in the future. Therefore, the competitive environment may be occurring in this process and there is a pressure toward uniformity (Festinger, 1954).

We assume each player updates his/her aspiration level by sharing the information of other players' aspiration level and then compares this information with his/her current

aspiration level. The sharing process is based on the interaction scheme that is given in the beginning of the simulation. Within the interaction, a player will obtain the information about the other players' aspiration level depending on the closeness of the player in the given interaction scheme. The closeness is represented by weights.

The model that we purpose assume players can interact each other. Within the interaction, players change information that is their aspiration. Each player interacts randomly at all time. Let $\alpha_i(t)$ be a level of sharing aspiration for player i at time t, and let T_i be a set of player that interact with player i. Let n_i be a number of player who interact with player i, and n_k be a number of player who interact with player k, we calculate $\alpha_i(t)$ as follow:

$$\alpha_i(t) = w_i\rho_i(t) + \sum_{k\in T_i} w_k\rho_k(t) \tag{3.13}$$

where, $w_k = \frac{1}{1+\max\{n_i,n_k\}}$ $\forall k \in T_i$ and $w_i = 1 - w_k$. Equation (4.27) look similar with distributed algorithm for distributed averaging problem (Xiao and Boyd, 2004), but differ in term of what information will be communicated. We explain the model as

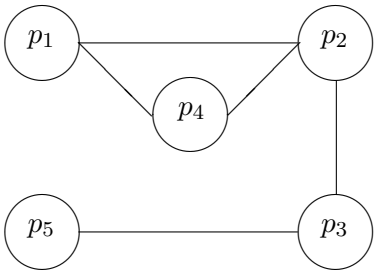


FIGURE 3.2: Structure of Interaction at Specific Time t

follow: Let at time t the interaction happens as in Figure 3.2.

TABLE 3.1: Matrix Representation of Interaction

	P_1	P_2	P_3	P_4	P_5
P_1	0	1	0	1	0
P_2	1	0	1	1	0
P_3	0	1	0	0	1
P_4	1	1	0	0	0
P_5	0	0	1	0	0

We can describe this interaction in form of matrix as shown in Table 5.1. From Table 5.1 we can count n_i $\forall i \in \{1,2,3,4,5\}$, i.e. $n_1 = 2$, $n_2 = 3$, $n_3 = 2$, $n_4 = 2$, and $n_5 = 1$. Level of sharing aspiration for player p_1 can be calculated as follow: p_1 interact with p_2 and p_4:

$w_2 = \frac{1}{1+\max\{2,3\}} = \frac{1}{4}$, $w_4 = \frac{1}{1+\max\{2,2\}} = \frac{1}{3}$, so we get $w_1 = 1-(w_2+w_4) = 1-(\frac{1}{4}+\frac{1}{3}) = \frac{5}{12}$. After that we can calculate $\alpha_1(t) = w_1\rho_1(t)+w_2\rho_2(t)+w_4\rho_4(t) = \frac{5}{12}\rho_1(t)+\frac{1}{4}\rho_2(t)+\frac{1}{3}\rho_4(t)$.

We update aspiration level of player i based on a model of learning that player i uses.
Roth-Erev Model:
If $\alpha_i(t) \geq \rho_i(t)$,

$$\rho_i(t+1) = (1-\omega^+)\alpha_i(t) + \omega^+ R_i(t) \tag{3.14}$$

otherwise,

$$\rho_i(t+1) = (1-\omega^-)\rho_i(t) + \omega^- R_i(t) \tag{3.15}$$

Borgres-Sarin Model:
If $\alpha_i(t) \geq \rho_i(t)$,

$$\rho_i(t+1) = (1-\beta)\alpha_i(t) + \beta R_i(t) \tag{3.16}$$

otherwise,

$$\rho_i(t+1) = (1-\beta)\rho_i(t) + \beta R_i(t) \tag{3.17}$$

Satisfying Model:
If $\alpha_i(t) \geq \rho_i(t)$,

$$\rho_i(t+1) = (1-\beta)\alpha_i(t) + \beta R_i(t) \tag{3.18}$$

otherwise,

$$\rho_i(t+1) = (1-\beta)\rho_i(t) + \beta R_i(t) \tag{3.19}$$

A player adjusts his/her level of aspiration according to the level of sharing aspiration (information) that he/she get from interaction, and uses it if the level higher or equal to his/her aspiration level. It is mean that a player must increases his/her aspiration level to group's aspiration level.

3.5 Simulation and Results

In this simulation a set of players play the public goods game and repeat the game for a number of times. We define a set of players as a group. We do not consider a group as a dynamic group that involves the process of group formation, group membership, and group cohesion. We assume each player in a group follows a learning model that is embedded to him/her without the capacity to interpret what one learns. In this sense, the players are only aware of what actions are successful (give a satisfaction) through learning. Under this circumstance, we consider two scenarios. The main purpose of the first scenario is to compare a group in which players interact with a group in which

players do not interact. We define the interact group as a group in which the players share information about their aspiration level through the mechanism of sharing aspiration. Through this mechanism we expect that the players share information about their aspiration level, which in turn influences their aspiration toward cooperative behavior. Therefore, we can assert that the mechanism of sharing aspiration spurs cooperative behavior in interact group under such circumstance. Conversely, non- interact group is a group in which the players do not share information about their aspiration level. The scheme of interaction for this setting is given in Figure 3.2.

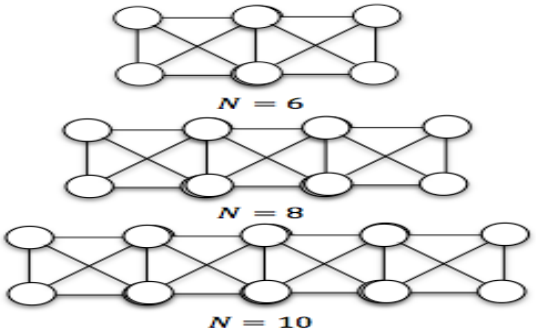


FIGURE 3.3: Scheme of Interaction for Strong Connectivity

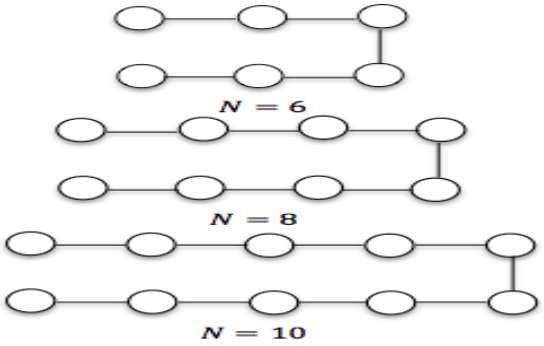


FIGURE 3.4: Scheme of Interaction for Weak Connectivity

The main purpose of the second scenario is to investigate the effectiveness of sharing aspiration mechanism in relations with the number of players in a group and the strength of connectivity. We vary the number of players in a group, i.e., 6, 8, and 10 and two strengths of connectivity, i.e., strong connectivity and weak connectivity. The scheme of interaction for strong connectivity and the number of players can be seen in Figure 3.3. Figure 3.4 shows the weak connectivity and the number of players.

Each player in a group has a learning model, i.e., REM stands for Roth-Erev Model, BSM stands for Borgers-Sarin Model, and SM stands for Satisfying Model. Table 3.2 shows the parameters of the learning models.

TABLE 3.2: Parameters of Learning Models

REM	$\nu = 10^{-9}$
	$\phi = 10^{-5}$
	$\epsilon = 10^{-5}$
	$\omega^{-} = 0.02$
	$\omega^{+} = 0.01$
BSM	$\lambda = 0.5$
	$\beta = h = 0.01$
SM	$\lambda = 0.99$

At each time t, each player takes an action depending on the probability of the action. They will receive a reward from this game. After that, they update the probabilities of the actions based on their learning model. Finally, they interact to share the aspiration level (only for interact group) and update their aspiration level (for interact and non-interact group).

3.5.1 First Scenario

In this simulation we compare interact group with non-interact group. Both groups consist of $N = 5$ players. At each time t, every player plays the public goods game. We use endowment $e = 2$, so that each player can take an action from $A = \{0, 1, 2\}$ and the public good $g = 9.9$. This condition satisfies $0 < g < Ne = 10$. We set initial value of $q_i(a, t_0) = 1$ so that $pr_i(a, t_0) = \frac{1}{|A|} = \frac{1}{3}$ $\forall a \in A$ and $\forall p_i \in P$ with $LM = REM$. We also set initial value of $pr_i(a, t_0) = \frac{1}{|A|} = \frac{1}{3}$ $\forall a \in A$ and $\forall p_i \in P$ with $LM = BSM$, and $pr_i(a, t_0) = \frac{1}{|A|} = \frac{1}{3}$ $\forall a \in A$ and $\forall p_i \in P$ with $LM = SM$ to determined $R_i(t_0)$.

No information will be shared in non-interact group and the aspiration level will be updated by using Eq. (3.5) and Eq. (3.6) for REM, Eq. (3.11) for BSM, and Eq. (3.12) for SM. The information will be shared in interact group by using the mechanism of sharing aspiration. The aspiration level will be updated by using Eq. (3.14) and Eq. (3.15) for REM and Eq. (3.16) and Eq. (3.17) for BSM, and Eq. (3.18) and Eq. (3.19) for SM. We use the scheme of interaction as in Figure 3.2 in this scenario. The initial value of aspiration level and the learning model of each member in each group can be seen in Table 3.3.

We set the same initial aspiration level and the same learning models for each member in each group (for comparison purpose). The parameters of learning models are as shown in Table 3.2. We run the simulation for 50 trials in 10000 iterations.

TABLE 3.3: Parameters of the first scenario

Number of players (N)	Public good (g)	Initial aspiration level	Learning Model
Interact group (5)	$g = 9.9$	5.5;6.5;7.5;8.5;9.5	REM;BSM;SM;REM;BSM
Non-interact group (5)	$g = 9.9$	5.5;6.5;7.5;8.5;9.5	REM;BSM;SM;REM;BSM

Figure 3.5 shows the average reward for each group for 50 trials and 10000 iterations. The average value of reward in interact group can reach the public goods value, i.e., $g = 9.9$. It means that all players in the group contribute all of their endowment, i.e., $e = 2$. While in non-interact group, the average value of reward is only about 8, lower than the public goods value.

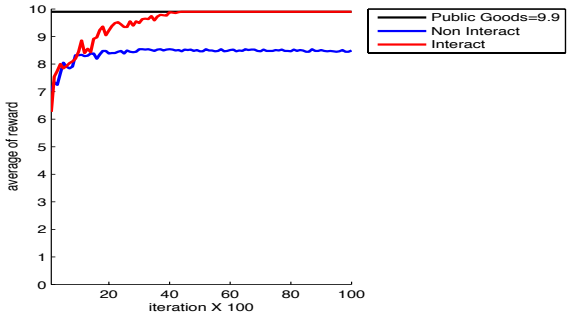


FIGURE 3.5: Comparison of Average Reward Between Interact Group and Non- Interact Group

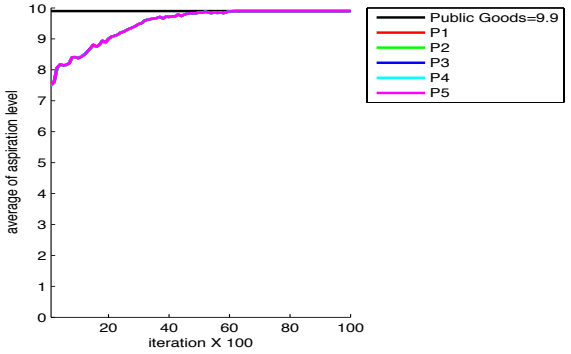


FIGURE 3.6: The Average of Aspiration Level for Interact Group

Figure 3.6 and Figure 3.7 also confirm this result. Figure 3.6 shows the average value of aspiration level for each member in interact group, while Figure 3.7 shows the average value of aspiration level for each member in non-interact group. As we can see in Figure 3.6, the average of aspiration level of all members almost has the same value from the beginning and converges to the public goods value, i.e., $g = 9.9$. This condition affects the behavior of each player in interact group to the cooperative behavior. All members

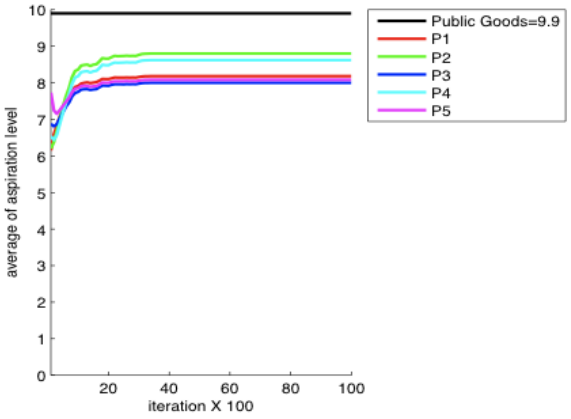


FIGURE 3.7: The Average of Aspiration Level for Non-Interact Group

in interact group contribute all of their endowment, i.e., $e = 2$ so that they get the reward of 9.9 which is equal to the public goods and his/her aspiration level as well. On the other hand, the average of aspiration level in non-interact group is not converging to the same value (Figure 3.7). Each player still has his/her own aspiration level so that they still have their own value to be contributed to the public goods. No interaction and communication in this group, so that no information can be shared.

3.5.2 Second Scenario

We use the scheme of interaction as in Figure 3.3 for strong connectivity and as in Figure 3.4 for weak connectivity. The initial value of aspiration level and the learning model of each member in each group can be seen in Table 3.4.

TABLE 3.4: Parameters of the second scenario

Number of players (N)	Public good (g)	Initial aspiration level	Learning Model
6	$g = 11.9$	0.5;1.5;2.5;3.5;4.5;5.5	REM;BSM;SM;REM;BSM;SM
8	$g = 15.9$	0.5;1.5;2.5;3.5;4.5;5.5;6.5;7.5	REM;BSM;SM;REM;BSM;SM;REM;BSM
10	$g = 19.9$	0.5;1.5;2.5;3.5;4.5;5.5;6.5;7.5;8.5;9.5	REM;BSM;SM;REM;BSM;SM;REM;BSM;SM;REM

We use endowment $e = 2$, so that each player can take an action from $A = \{0, 1, 2\}$. We set initial value of $q_i(a, t_0) = 1$ so that $pr_i(a, t_0) = \frac{1}{|A|} = \frac{1}{3}$ $\forall a \in A$ and $\forall p_i \in P$ with $LM = REM$. We also set initial value of $pr_i(a, t_0) = \frac{1}{|A|} = \frac{1}{3}$ $\forall a \in A$ and $\forall p_i \in P$ with $LM = BSM$, and $pr_i(a, t_0) = \frac{1}{|A|} = \frac{1}{3}$ $\forall a \in A$ and $\forall p_i \in P$ with $LM = SM$ to determined $R_i(t_0)$. All parameters of the models are shown in Table 3.2. We run the simulation for 50 trials in 100000 iterations to see the long-term dynamics of the model. The output of this scenario is the normalized average reward. Values of the normalized

average reward are in the range 0 and 1. The values close to 1 indicate full cooperative behavior. This normalization is needed because we have different values of g.

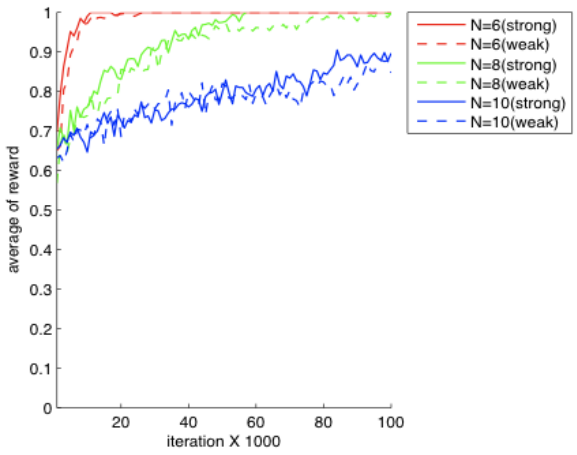


FIGURE 3.8: The Average of Reward for Strong and Weak Connectivity with Various Numbers of Players

As we can see in Figure 3.8, the strong connectivity is more quickly to converge to full cooperative behavior than the weak connectivity. As long as there is a connectivity that involves all players, the players will get the information that coordinate their strategy. As we can see in the weak connectivity with $N = 6$ (Figure 4), the first player can get the information from the second player, the second player can get the information from the third player, and so on. Because the connection is in two directions, there is a chain of information in the long run so that the players can coordinate their strategy. However, as the number of players increases, i.e., $N = 10$, full cooperative behavior is hard to achieve.

3.6 Discussion and Conclusion

Based on the simulation, we found two interesting results. The first one is that with interaction and sharing aspiration all players can improve their cooperative behaviors in a group. Players with different initial aspirations and learning models can adjust their aspirations through interaction and sharing aspiration process. In this situation, a player compares his/her aspiration with the group's aspiration level via sharing aspiration mechanism as long as his/her aspiration is below the group's aspiration level that he/she receives. Within this process, a player can share his/her aspiration with whom he/she interacts. By increasing aspiration of one player, then the aspirations of other players are also increased. This process makes the aspiration level of all members of the group converge to the cooperative results, i.e., all players contribute all of their endowment

to the public goods. This situation is also confirmed by social comparison theory, i.e., pressure towards uniformity (Festinger, 1954).

The second result is about the complexity of sharing aspiration in a group. A group with few numbers of players can improve their cooperative behavior more quickly in strong connectivity. A player in a group that consists of many players and in the weak connectivity needs more time to coordinate their action towards a goal. As long as there is communication (interaction/connectivity) in the group, cooperative behavior can be maintained. This result is in line with the previous experimental studies (Bicchieri and Lev-On, 2007; Kerr and Kaufmann-Gilliland, 1994). However, as the number of players increases, the level of cooperative behavior decreases.

The proposed mechanism of sharing aspiration uses some clues, i.e., individuals expectation, information, and communication obtained from experimental studies to investigate peoples behavior in social dilemmas. These three clues are important to spur cooperative behavior in a group Cremer and Vugt (2002); Vugt et al. (2004); Kurzban and Descioli (2008); Kurzban and Houser (2001); Hopthrow and Hulbert (2005); Bicchieri and Lev-On (2007); Kerr and Kaufmann-Gilliland (1994). However, in this research we are also linking this mechanism to aspiration-based reinforcement-learning models that can make players choose his/her action based on experience, i.e., aspiration level. As long as the members of the group are allowed to share their aspiration level, the cooperative behavior can be promoted.

One of the main assumptions of our model is that all players are willing to share their aspiration through interaction. This assumption is difficult to be fulfilled in reality. People might not want to share information about their aspiration. Our future research will investigate how to design a mechanism that can make people willing to share their aspiration.

The proposed model gives us insight into the aspiration level, which determines the behavior of the players in a group playing public goods game. However, in future research we would like to confirm our findings with experiment tests.

Chapter 4

Aspiration-Based Learning to Balance Exploration and Exploitation in Organizational Learning

This chapter considers organizational learning as mutual learning between an organization and the individuals working in it. The process of mutual learning has implications for understanding and managing the trade-off between exploration and exploitation. We propose an aspiration-based model to balance exploration and exploitation in organizational learning. The model is intended to improve the knowledge achieved by the members of an organization and the organization itself. In the proposed model, individuals in the organization are allowed to experiment with their believes according to their aspiration level and are allowed to interact to communicate their aspiration level. Simulation results show that the model improves the knowledge obtained by the members of the organization and the organization itself, and is able to deal with open systems.

4.1 The Concept of Mutual Learning in Organizational Learning

In behavioral studies of organization, organizations are seen as learning by encoding inference from history into routines that guide behavior (L and JG, 1988; J, 1996; HA, 1991). The routines include forms, rules, procedures, conventions, strategies, and technologies. Based on these routines organizations are constructed and operated. The

routines also include the structure of beliefs, frameworks, paradigms, codes, cultures, and knowledge that butters, elaborates, and contradicts the formal routines. Routines are independent of the individuals who execute them. Routines capture the experiential lessons of history in a way that makes the lessons accessible to organizations and organizational members who have not themselves experienced the history (Levinthal and March, 1993). Through socialization, education, imitation, professionalization, mergers, and acquisitions the routines are transmitted.

Based on this framework, March has built a model of mutual learning to investigate exploitation and exploration processes in organizational learning (March, 1991). In his model, individuals within the organization have diverse sets of beliefs about reality and the organization has a set of beliefs about reality in terms of organizational code. The code is continuously changed. This change is based on a group of superior individuals who have better interpretation about reality. At the same time, the code is socialized to individuals in the organization and the individuals learn from the code. Therefore, there is mutual learning between individuals and the organization. He argues that such mutual learning has implications for understanding and managing the trade-off between exploration and exploitation. Exploration refers to the search for new possibilities by creating variety in experience through experimentation and exploitation refers to creating reliability in experience through refinement. Exploitation yields more certain and immediate returns, however, it makes the discovery of new possibilities unlikely and can lead to suboptimal stable equilibrium. On the other hand, exploration can lead to the good solutions, however, it also causes a degradation of performance in the short run and greater risk in its process.

March's model has shown that slow learning in the side of individuals to adapt to the code improves the knowledge of organization. This adaption process should be followed by rapidly learning in the side of organization. However, this balancing process does not reflect reality completely, and therefore, there is some information about the reality that cannot be captured (or observed) by organization. This situation becomes worse if the environment is changing and if there is turnover in individuals. In March's model, the only source of learning is the organization code and in turn the individuals learn from the code. The individuals with more knowledge or competence improve the knowledge of the code. As long as diversity among the individuals knowledge exist, there is a chance for the code to be improved. However, if the individuals' knowledge converges to the organization's knowledge, the improvement will disappear even though reality has not been observed completely. Moreover, the single source of knowledge cannot handle the environmental turbulence and turnover.

In this research, we propose to use an aspiration-based model to balance exploration and exploitation in organizational learning. The model is intended to improve the knowledge achieved by the members of an organization and the organization itself. The individuals in the organization are allowed to make experimentation or exploration in their beliefs about reality. This process depends on how individuals' performance or competence deviates from their aspiration. The organization also has an aspiration level, by which the socialization and adapting process of the code is based. Besides that, the individuals are allowed to interact to share their aspiration. This process is used to make the aspiration level homogeneous.

4.2 The Aspiration-Based in Organizational Learning

We consider an organization as a complex adaptive system, where individuals interact with other individuals. The individuals and the organization develop mutual learning (L et al., 2000; Levinthal and March, 1981; March, 1991). In March's model, the organization stores its knowledge in form of procedures, norms, and rules. They accumulate such knowledge over time in organizational code, learning from their members. At the same time, individuals in the organization are socialized to organizational code or organizational beliefs. The original March's model consists of the following four key features: (1) There is an external reality that is independent of beliefs about it. Reality is described as having m dimensions, each of which has a value of 1 or -1. (2) The organization consists of n individuals. Each of them and an organizational code holds m beliefs about the corresponding elements of reality at each time step. Each belief for an individual has a value of 1, 0, or -1. A value of 0 means that an individual or the organization is not sure of whether 1 or -1 represents reality. This value may change over time. (3) Individuals modify their beliefs through socialization with probability p_1. (4) At the same time, with probability p_2, the organizational code will be adjusted to conform to the dominant beliefs within the superior group, i.e., those individuals whose beliefs correspond with reality on more dimension than does the code.

In the proposed model, we introduce several features of our model. Each individual has a target or an aspiration level as well as the organization (TK, 1992). The aspiration level of individuals and the organization is used as a stimulus to change their behavior. Our model has several features as follows:

4.2.1 Basic Elements

Like March's model, we consider reality as having m dimensions, each of which has a value of 1 or -1. The probability that any one dimension will have a value 1 (or -1) is 0.5. The organization consists of n individuals. Each of them and an organizational code holds m beliefs about the corresponding elements of reality at each time step. Each belief for an individual has a value of 1, 0, or -1. A value of 0 means that an individual or the organization is not sure of whether 1 or -1 represents reality.

4.2.2 Payoff

The individuals and the organization can observe the performance or knowledge level of their overall belief set, but they cannot directly observe how each element of the belief set contributes to this performance. We calculate the performance as the proportion of reality that is correctly represented in the belief set. Let c (c_{ind} for individuals and c_{org} for organization) be the number of correctly represented reality and nc (nc_{ind} for individuals and nc_{org} for organization) be the number of not correctly represented reality for non-zero elements,

$$IndPerf_i(t) = c_{ind} - nc_{ind} \tag{4.1}$$

$$OrgPerf(t) = c_{org} - nc_{org} \tag{4.2}$$

4.2.3 Experimentation Procedure

At each time step, an individual has a chance to change one element (choosing at random) of his/her belief set. The chance depends on the probability of experimentation. After that, the individual will get the stimulus $S_i(t) = IndPerf_i(t) - IndAsp_i(t)$, where $IndAsp_i(t)$ is aspiration level of individual i at time t. This value of stimulus will be averaging over time. Let $\omega_i^E(t)$ be the average value of stimulus for experimentation for individual i at time t and $\omega_i^{NE}(t)$ be the average value of stimulus for not experimentation for individual i at time t,
If experimentation:

$$\omega_i^E(t+1) = (\frac{1}{t}) * (\omega_i^E(t) + S_i(t)) \tag{4.3}$$

If not experimentation:

$$\omega_i^{NE}(t+1) = (\frac{1}{t}) * (\omega_i^{NE}(t) + S_i(t)) \tag{4.4}$$

The strength of reinforcement is determined as follows:
If experimentation:

$$\beta_i(t) = \frac{e^{\omega_i^E(t+1)}}{e^{\omega_i^E(t+1)} + e^{\omega_i^{NE}(t+1)}} \tag{4.5}$$

If not experimentation:

$$\beta_i(t) = \frac{e^{\omega_i^{NE}(t+1)}}{e^{\omega_i^E(t+1)} + e^{\omega_i^{NE}(t+1)}} \tag{4.6}$$

4.2.4 Updating Probability of Experimentation

The probability of experimentation will be updated as follows:
If $S_i(t) \geq 0$ and Experimentation

$$Pr_i^E(t+1) = Pr_i^E(t) + \beta_i(t) * (1 - Pr_i^E(t)) \tag{4.7}$$

If $S_i(t) < 0$ and Experimentation

$$Pr_i^E(t+1) = Pr_i^E(t) - \beta_i(t) * Pr_i^E(t) \tag{4.8}$$

If $S_i(t) \geq 0$ and Not Experimentation

$$Pr_i^{NE}(t+1) = Pr_i^{NE}(t) + \beta_i(t) * (1 - Pr_i^{NE}(t)) \tag{4.9}$$

If $S_i(t) < 0$ and Not Experimentation

$$Pr_i^{NE}(t+1) = Pr_i^{NE}(t) - \beta_i(t) * Pr_i^{NE}(t) \tag{4.10}$$

4.2.5 Updating the Organizational Code

The organizational code adapts to the beliefs of those individuals whose beliefs correspond with reality on more dimensions than does the code. At each time step, the organization has a chance to change each element of the organization code. The chance depends on the probability of learning by the code. We define the stimulus $S^{Org}(t) = AvgAsp^{sm}(t) - OrgAsp(t)$, where $AvgAsp^{sm}(t)$ is average of aspiration level of selected individual at time t and $OrgAsp(t)$ is organization aspiration at time t. This value of stimulus will be averaging over time. Let $\omega^{LBC}(t)$ be the average value of stimulus for learning by the code at time t and $\omega^{NLBC}(t)$ be the average value of stimulus for not learning by the code at time t then,
If learning

$$\omega^{LBC}(t+1) = (\frac{1}{t}) * (\omega^{LBC}(t) + S^{Org}(t)) \tag{4.11}$$

If not learning:

$$\omega^{NLBC}(t+1) = (\frac{1}{t}) * (\omega^{NLBC}(t) + S^{Org}(t)) \tag{4.12}$$

We define $\beta^{Org}(t)$ as the strength of reinforcement and is determined as follows:
If learning:

$$\beta^{Org}(t) = \frac{e^{\omega^{LBC}(t+1)}}{e^{\omega^{LBC}(t+1)} + e^{\omega^{NLBC}(t+1)}} \tag{4.13}$$

If not learning:

$$\beta^{Org}(t) = \frac{e^{\omega^{NLBC}(t+1)}}{e^{\omega^{LBC}(t+1)} + e^{\omega^{NLBC}(t+1)}} \tag{4.14}$$

The probability that the beliefs of the code will be adjusted to conform to the dominant belief within the superior group on any particular dimension is determined as follows:
If $S^{Org}(t) \geq 0$ and Learning

$$Pr^{LBC}(t+1) = Pr^{LBC}(t) + \beta^{Org}(t) * (1 - Pr^{LBC}(t)) \tag{4.15}$$

If $S^{Org}(t) < 0$ and Learning

$$Pr^{LBC}(t+1) = Pr^{LBC}(t) - \beta^{Org}(t) * Pr^{LBC}(t) \tag{4.16}$$

If $S^{Org}(t) \geq 0$ and Not Learning

$$Pr^{NLBC}(t+1) = Pr^{NLBC}(t) + \beta^{Org}(t) * (1 - Pr^{NLBC}(t)) \tag{4.17}$$

If $S^{Org}(t) < 0$ and Not Learning

$$Pr^{NLBC}(t+1) = Pr^{NLBC}(t) - \beta^{Org}(t) * Pr^{NLBC}(t) \tag{4.18}$$

where $Pr^{LBC}(t+1)$ is probability of learning by the code and $Pr^{NLBC}(t+1)$ is probability of not learning by the code, and $Pr^{LBC}(t+1) + Pr^{LBC}(t+1) = 1$.
The code is updated as follows:
Let δ_j be the sum of the individuals' belief on j dimension for non-zero values, i.e., $\sum_{i=1}^{n} b_j^i$, where n is the number of individuals in the superior group and b_j^i is the non-zero value of i's individual belief on j dimension, $e_j \rightarrow z_j$ (the old belief will be changed), where $z_j = 1$ if $\delta_j > 0$ and $z_j = -1$ if $\delta_j < 0$. In case of $\delta_j = 0$ then $z_j = 0$.

4.2.6 Updating the Individuals Beliefs

Individuals modify their beliefs continuously as a consequence of socialization into the organization code of beliefs. At each time step, the individual has a chance to learn from the code for each element of his/her belief. The chance depends on the probability

of learning from the code. We define $S_i^{LFC}(t) = IndPerf_i(t) - IndAsp_i(t)$ as stimulus for learning from the code at time t. This value of stimulus will be averaging over time. Let $\omega_i^{LFC}(t)$ be the average value of stimulus for learning from the code at time t and $\omega_i^{NLBC}(t)$ be the average value of stimulus for not learning from the code at time t then,
If learning

$$\omega_i^{LFC}(t+1) = (\frac{1}{t}) * (\omega^{LFC} i(t) + S_i^{LFC}(t)) \tag{4.19}$$

If not learning:

$$\omega_i^{NLFC}(t+1) = (\frac{1}{t}) * (\omega_i^{NLFC}(t) + S_i^{LFC}(t)) \tag{4.20}$$

We define $\beta_i^{LFC}(t)$ as the strength of reinforcement and is determined as follows:
If learning:

$$\beta_i^{LFC}(t) = \frac{e^{\omega_i^{LFC}(t+1)}}{e^{\omega_i^{LFC}(t+1)} + e^{\omega_i^{NLFC}(t+1)}} \tag{4.21}$$

If not learning:

$$\beta^{LFC}(t) = \frac{e^{\omega_i^{NLFC}(t+1)}}{e^{\omega_i^{LFC}(t+1)} + e^{\omega_i^{NLFC}(t+1)}} \tag{4.22}$$

The probability that the beliefs of the code will be adjusted to conform to the dominant belief within the superior group on any particular dimension is determined as follows:
If $S_i^{LFC}(t) \geq 0$ and Learning

$$Pr_i^{LFC}(t+1) = Pr_i^{LFC}(t) + \beta_i^{LFC}(t) * (1 - Pr_i^{LFC}(t)) \tag{4.23}$$

If $S_i^{LFC}(t) < 0$ and Learning

$$Pr_i^{LFC}(t+1) = Pr_i^{LFC}(t) - \beta_i^{LFC}(t) * Pr_i^{LFC}(t) \tag{4.24}$$

If $S_i^{Org}(t) \geq 0$ and Not Learning

$$Pr_i^{NLFC}(t+1) = Pr_i^{NLFC}(t) + \beta_i^{LFC}(t) * (1 - Pr_i^{NLFC}(t)) \tag{4.25}$$

If $S_i^{Org}(t) < 0$ and Not Learning

$$Pr_i^{NLFC}(t+1) = Pr_i^{NLFC}(t) - \beta_i^{LFC}(t) * Pr^{NLFC})i(t) \tag{4.26}$$

where $Pr_i^{LFC}(t)$ is probability of learning from the code and $Pr_i^{NLFC}(t)$ is probability of not learning from the code, and $Pr_i^{LFC}(t) + Pr_i^{NLFC}(t) = 1$. If the code is 0 on the particular dimension, individuals belief is not effected. In each time step in which the code differs on any particular dimension from the belief of an individual, individual belief changes to that of the code. In March's model, he uses a parameter in term of probability that reflects the effectiveness of socialization, i.e., learning from the code

(p_1). In this research we use the probability of learning from the code of individuals to determine the effectiveness of socialization.

4.2.7 Updating the Aspiration Level

At the end of each time step, the individuals interact with each other to share their aspiration level. Each individual interacts randomly at all time. Let $\alpha_i(t)$ be a level of sharing aspiration for individual i at time t, and let T_i be a set of individuals that interact with individual i. Let n_i be a number of individuals who interact with individual i, and n_k be a number of individuals who interact with individual k, we calculate $\alpha_i(t)$ as follows:

$$\alpha_i(t) = w_i * IndAsp_i(t) + \sum_{k \in T_i} w_k * IndAsp_k(t) \tag{4.27}$$

where, $w_k = \frac{1}{1+\max\{n_i,n_k\}}$ $\forall k \in T_i$ and $w_i = 1 - \sum_{k \in T_i} w_k$. We explain the model as

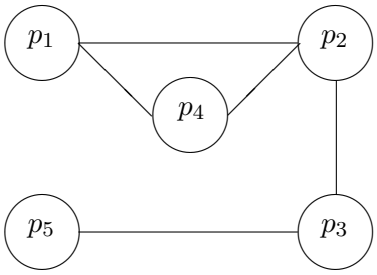


FIGURE 4.1: Structure of Interaction at Specific Time t

follows: At time t the interaction happens as in Figure 4.1.

TABLE 4.1: Matrix Representation of Interaction

	P_1	P_2	P_3	P_4	P_5
P_1	0	1	0	1	0
P_2	1	0	1	1	0
P_3	0	1	0	0	1
P_4	1	1	0	0	0
P_5	0	0	1	0	0

We can describe this interaction in form of matrix as shown in Table 5.1. From Table 5.1 we can count n_i $\forall i \in \{1,2,3,4,5\}$, i.e. $n_1 = 2$, $n_2 = 3$, $n_3 = 2$, $n_4 = 2$, and $n_5 = 1$. Level of sharing aspiration for individual p_1 can be calculated as follows: p_1 interact with p_2 and p_4:
$w_2 = \frac{1}{1+\max\{2,3\}} = \frac{1}{4}$, $w_4 = \frac{1}{1+\max\{2,2\}} = \frac{1}{3}$, so we get $w_1 = 1-(w_2+w_4) = 1-(\frac{1}{4}+\frac{1}{3}) =$

$\frac{5}{12}$. After that we can calculate $\alpha_1(t) = w_1 * IndAsp_1(t) + w_2 * IndAsp_2(t) + w_4 * IndAsp_4(t) = \frac{5}{12} * IndAsp_1(t) + \frac{1}{4} * IndAsp_2(t) + \frac{1}{3} * IndAsp_4(t)$.

We update aspiration level of individual i as follows.
If $\alpha_i(t) \geq IndAsp_i(t)$,

$$IndAsp_i(t+1) = (1 - h_i) * \alpha_i(t) + h_i * IndPerf_i(t) \tag{4.28}$$

otherwise,

$$IndAsp_i(t+1) = (1 - h_i) * IndAsp_i(t) + h_i * IndPerf_i(t) \tag{4.29}$$

where h_i is habituation parameter for individual i.

An individual adjusts his/her level of aspiration according to the level of sharing aspiration (information) that he/she gets from interaction, and uses it if the level higher or equal to his/her aspiration level. It means that an individual increases his/her aspiration level to group's aspiration level. Organization will update the aspiration level as follows:

$$OrgAsp_i(t+1) = (1 - ho) * OrgAsp_i(t) + ho * OrgPerf(t) \tag{4.30}$$

where ho is habituation parameter for organization.

4.2.8 Environmental Turbulence and Turnover

In this model, we also consider the effect of environmental turbulence. The value of any given dimension of reality shifts (from 1 to -1 or -1 to 1) with probability p_4 in a given time of time step. Turnover is the process of replacing some individuals by some new individuals in organization. At every time step each individual has a probability, p_3, of leaving the organization and being replaced by a new individual with a set of new beliefs described by an m-tuple, having values equal to 1, 0, or -1, with equal probabilities.

4.2.9 Simulation Cycle

The time step is set to 800 iterations and will be replicated by 80 runs. Reality is set to $m = 30$ dimensions and the number of individuals is set to $n = 50$. At the start of each run, every dimension in reality is set randomly to 1 or -1 ($m - dimensions$). The organizational code is initially 0, neutral beliefs on all dimensions. The individuals beliefs is set to 1, 0, or -1 with equal probabilities. The initial aspiration level for each individual and initial aspiration level of organization is set to a value in range between

$(0, 30)$ randomly. The initial probability of experimentation for each individual is set to random value between $(0, 1)$.

At each iteration, every individual has a chance to do experimentation and updates his/her performance, stimulus, average of stimulus, strength of reinforcement, and the probability of experimentation. The organization updates performance and selects the individuals to be a superior group. Based on the superior group, the organizational code will be updated and socialized to the individuals in the organization (updating the individuals beliefs). The individuals interact to share their aspiration level and update their aspiration level. The organization is also updated the aspiration level. The main output of this simulation is an equilibrium. The equilibrium is reached at which all individuals and the code share the same (not necessarily accurate) belief with respect to each dimension. The equilibrium is stable. We define the output as the knowledge level at equilibrium. Output were averaged over 80 runs.

4.3 Simulation and Results

The main part of our model is the existence of the experimentation process and interaction process to share the aspiration level. We would expect that the model can improve the knowledge level of organization at equilibrium. We conduct two scenarios in the simulation. The first scenario is the organization as in a closed system. In this settings, there are no environmental turbulence and turnover. The second scenario is organization as in an open system. Environmental turbulence and turnover are allowed in this setting.

4.3.1 First Scenario

We compare the result of our model with the March's model. Two set of parameters are used for March's model, i.e., $p_1 = 0.1; p_2 = 0.9$ and $p_1 = 0.5; p_2 = 0.5$. p_1 is probability of socialization and p_2 is probability of learning by the code in March's model. March's model has shown the greater knowledge level at equilibrium by using the first set of parameters ($p_1 = 0.1; p_2 = 0.9$). As we can see in Figure 4.2, the proposed model achieves the maximum values. The experimentation or exploration process by the individuals creates and preserves the variety of knowledge necessary for the organization. This process will continuously occur as long as the experimentation leads to positive stimulus for an individual and will stop if the process leads to negative stimulus. Based on the March's model, this result is caused by lower learning rate $p_1 = 0.1$ and high $p_2 = 0.9$. Moreover, in March's model the source to improve the code is only provided by slow

learning on the part of individuals that maintain diversity longer. As the individuals' beliefs converge to the code, there is no source to improve the knowledge for both sides. In Figure 4.3, we can see the average aspiration of both individuals and organization converge to the same value, i.e., the optimum value (accurately perceived the reality, $m = 30$). It means that the process of learning process and experimental process can be controlled. This idea is also confirmed by Figure 4.4, as we can see all probabilities goes to zero, which means the goal of individuals in the organization and the organization were achieved.

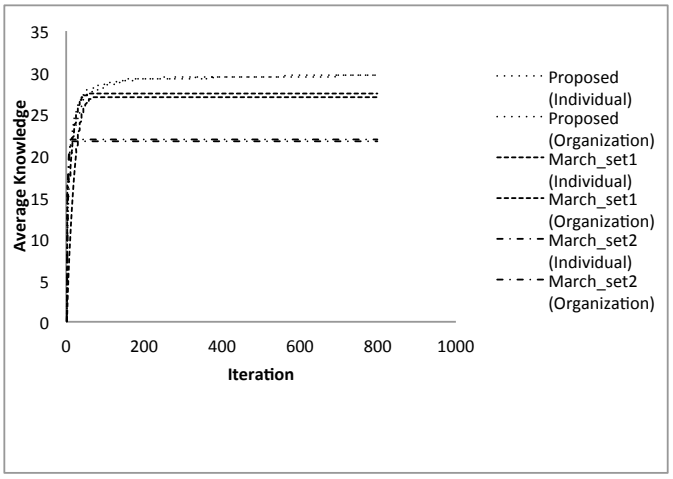


FIGURE 4.2: Average Knowledge in a Closed System

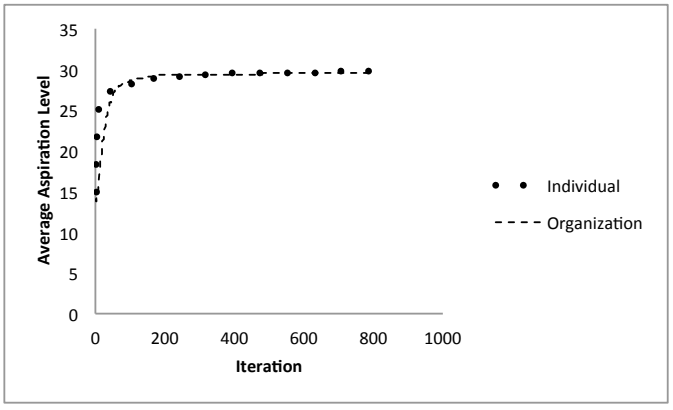


FIGURE 4.3: Dynamics of Average Aspiration Level in a Closed System

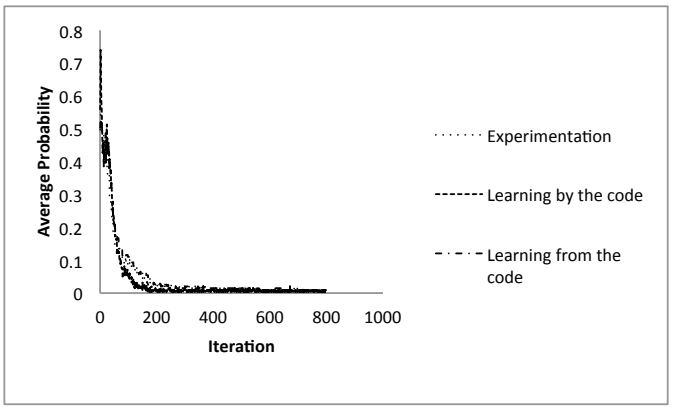


FIGURE 4.4: Dynamics of Average Probability in a Closed System

4.3.2 Second Scenario

In this simulation we consider an open system that involves environmental turbulence and turnover from individuals in organization. We set environmental turbulence probability $p_4 = 0.001$ and turnover probability $p_3 = 0.001$. We compare the proposed model to March's model by using $p_1 = 0.5; p_2 = 0.5$. According to the proposed model, the newcomer not only has a new set of beliefs, but also the variable of newcomer is also reseted, i.e., aspiration level, probability of experimentation, stimulus, average of stimulus, and strength of reinforcement. In the Figures 4.5 we can see the proposed model is more robust than the March's model in an open systems. The individuals in organization can adjust their experimental act and the learning processes by their aspiration level and organization's aspiration level. This is also shown in Figure 4.6 and Figure 4.7. Although the individuals are still doing experimentation (small probability of experimentation) and the individuals in organization and organization is still learning (small probabilities of learning from the code and learning by the code), this process is controlled by their aspiration level that already converges to the same value, i.e., optimal value.

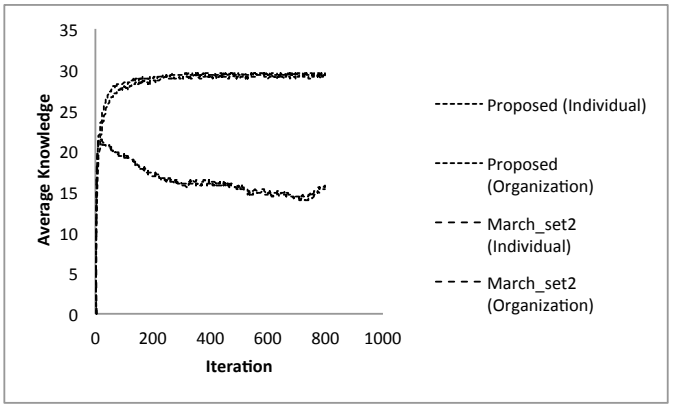


FIGURE 4.5: Average Knowledge in an Open System

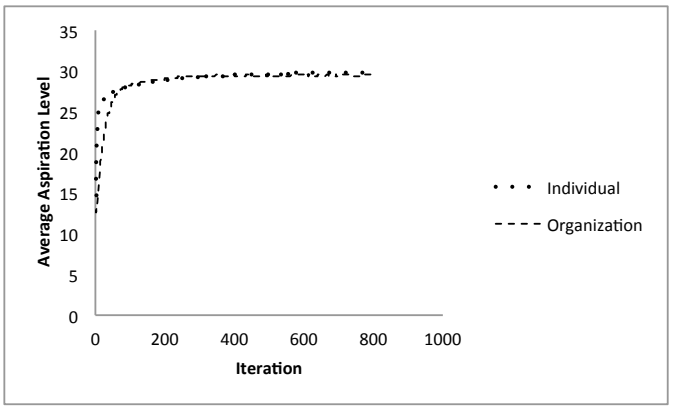


FIGURE 4.6: Dynamics of Average Aspiration Level in an Open System

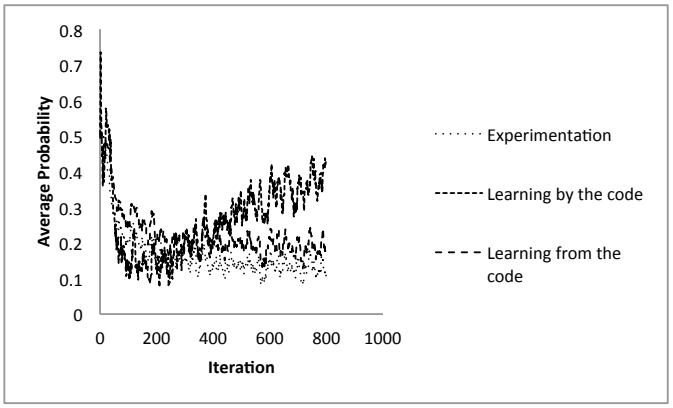


FIGURE 4.7: Dynamics of Average Probability in an Open System

4.4 Discussion and Conclusion

Balancing the process of exploration and exploitation in organizational learning can be achieved by individuals' experimentation and interaction to share their aspiration. The model suggests that exploration process via individuals' experimentation is allowed if the individuals feel their performance is not enough in terms of aspiration. As the individuals feel better with their performance, they will exploit their knowledge and contribute the knowledge to the organizational code. Because the selection process in the model is based on the individuals performance, the code will rapidly improve the performance of the organization.

Variability in beliefs perceived by individuals can be maintained via individuals' experimentation. Individuals can improve their knowledge. However, this experimentation is controlled by aspiration levels or individuals' goal so that they know when to experiment.

The individuals are rational. They tend to maximize their knowledge by experimentation (exploration) and by preventing the discarding of beliefs in which individuals have confidence (exploitation).

At the same time, the individuals in the organization interact with each other to share their aspiration. This process makes the heterogeneity in aspiration disappear. There is no inequality of aspiration levels. Each individual knows the other individuals' aspiration levels through sharing aspiration process. The model is robust in open systems, where there is turbulence and turnover. High knowledge level can be maintained.

Based on what we have found here, we can conclude that understanding about individuals' aspiration is important to balance exploration and exploitation process in organizational learning. Individuals' experimentation is not a risk as long as the heterogeneity in individuals' aspiration level can be eliminated.

Chapter 5

Aspiration-Based Learning in a Cournot Duopoly Model

This paper explores the implication of aspiration-based learning in a simple Cournot duopoly model. When the firms know the average industry-wide profit and perceive it as aspiration level, then the market leads to collusive outcome or collusive equilibrium. In this sense, all firms have the same reference point, i.e., the average industry-wide profit as their aspiration level. However, the firms may have their own aspiration level (e.g., a goal of profit) and will choose their strategy accordingly. Therefore, the firms will try to reach their own aspiration level. This aspiration level is not static and the firms will adjust their aspiration level. In this research we consider a market that consists of several firms with their own aspiration level. We propose an aspiration-based learning shaped by an information searching mechanism to examine the behavior of the firms in the market. A firm updates its aspiration level by searching the information of the other firms' aspiration level and then compares this information with its current aspiration level. Based on its aspiration level, the firm will choose the best strategy through learning. Simulation results show that the learning model and the information searching mechanism lead the market to competitive outcome, i.e., Nash equilibrium, if the firms have many strategies even if their initial aspiration level is low. However, if the firms have fewer strategies and start with high initial aspiration level, then collusive behavior will occur.

5.1 Investigation of Cournot Duopoly Game

The research on learning and the availability of information in duopoly market have attracted a great deal of attention. In the experimental research, the main purpose is

to examine the rules of learning on the equilibrium selection in oligopoly or duopoly market. As in Huck et al. (1999) and Lupi and Sbriglia (2003), the experiments were designed to test various learning theories in the context of a Cournot oligopoly. They analyzed the relationship between learning theories and the availability of information. The results showed that the availability of information and the rule of imitation could lead to the selection of competitive equilibrium. More information about demand and cost conditions yields less competitive behavior, while more information about the quantities and profits of other firms yields more competitive behavior. These results are also in line with Altavilla et al. (2006) and Vega-Redondo (1997) who state that imitation behavior will increase the level of competition in the market. However, information on the industries' average profitability might induce more collusive outcomes. In this sense, the firms perceive the industries' average profitability as aspiration levels. The firms will try new strategies anytime their profits fall below the industry's average profitability (Dixon, 2000; Dixon et al., 2006).

The aspiration learning model that has been analyzed in Oechssler (2002) and Palomino and Vega-Redondo (1999) is similar to satisfying theories that state that decision makers search for new alternatives if and only if today's action is unsatisfactory, i.e., yields a payoff that falls below the decision maker's aspiration level (Bendor et al., 2004). These satisfying theories have been applied as in Altavilla et al. (2006) and Dixon et al. (2006). The firms adopt a similar rule of behavior by considering the overall average profit as the individual aspiration level. These researches proved that convergence to collusive outcomes is likely to be observed in most experimental duopolies. These findings are also emphasized in Dixon (2000) by using simulation model. However, those researches assume that all firms have the same aspiration level which is represented by overall average profit as a reference point. Another assumption is that the information on the overall average profit is provided.

In this research we build an aspiration-based learning that uses reinforcement theory. Reinforcement learning model (Bush and Mosteller, 1955) is designed to capture the "Law of Effect" (Thorndike, 1911) where positive reinforcement increases the tendency to play an action, and negative reinforcement decreases it. In this sense, there is a probability of an action and if such action produces a satisfactory payoff in the current period, then the agent will not decrease his probability on that action. Therefore, the state space of reinforcement learning models is more complex in terms of the probability over all available actions.

In contrast to the previous studies, we assume the firms have their own aspiration level and the firms are encouraged to adjust their aspiration level by comparing with the other firms. We adapt a social comparison theory from the works of Festinger (1954) to build

an information searching mechanism. A firm updates its aspiration level by searching the information of the other firms' aspiration level and then compares this information with its current aspiration level. Discrepancy between the aspiration level and the current profit of the firm will generate a stimulus. The accumulation of stimuli i.e., strength of reinforcement, will affect the probability of the strategies that will be chosen in the next step. Through the learning model and the information searching mechanism we want to examine the behavior of the firms in the market, i.e., competitive behavior or collusive behavior.

5.1.1 Modeling Framework

The simulation model that we use in this research is inspired by experimental market in duopoly or laboratory duopoly markets (Dixon et al., 2006). In this laboratory duopoly markets, the experimental subjects were told that the experiment was divided into "days". At the beginning of each day, each subject had to choose his/her strategy for that days trading. The strategies set is the amount of quantity to be produced. During the day, each participant meets all of the other participants, and in each encounter his chosen trading strategy plays against the strategy chosen by the person he/she is playing against (play the Cournot duopoly game). Thus, if there are four experimental subjects, each one meets with the other three every day. At the end of the day, his/her earned profits as the total accumulated one throughout the day. We mimic this process to conduct the simulation model by using the Holt's experiment of duopoly model to represents the model of duopoly game (section 2). After that, we follow the same setting of experimental laboratory duopoly markets. We assume the strategies set is given and the total profit for a firm after meeting all the other firms is given in section 3. Within this setting, we can calculate payoff matrix or total profit for all combination of strategies. This model is similar with Dixon (2000). He considers an "economy" composed by several duopoly markets, where firms are matched to play a Cournot duopoly game. He uses the discrete time version of replicator dynamics to analysis competition and the evolution of collusion.

According to our model, the firms only have the strategies as decision variable. The market is quite simple which only involved the interaction among firms. There are no consumers involved in this model. We can consider the model as the game theory model (e.g., two-persons Prisoners Dilemma game), consists of players (the firms), a set of strategies (the amount of quantity) and the payoff (the total profit).

The main purpose of the experimental research in duopoly market is to investigate the impact of learning process and availability of information on equilibrium selection

in the market. Besides, the investigation of firms' behavior or individuals' behavior to play a duopoly game is important in behavioral economics. Therefore, it is inline with the purpose of our research. We mimic the laboratory duopoly markets setting to our simulation model. We develop a new method of learning based on aspiration level and a new information searching mechanism to investigate equilibrium selection in the market. We use repeated game theory approach by using a Cournot duopoly game to see long-term dynamics of equilibrium selection in the market. The long-term dynamics of equilibrium selection is hard to achieve in experimental research due to time limitation and condition of the subjects. We propose the simulation model to handle this disadvantage. In this sense, we use our simulation model as a laboratory duopoly markets.

5.2 The Standard Cournot Model

In this paper we consider a standard symmetric Cournot duopoly (see Kimbrough and Lu, 2003). There are two firms producing the same homogeneous commodity. The only decision variable for firm i is the quantity q_i to be produced. The inverse demand function is

$$P(Q) = a + bQ \tag{5.1}$$

where Q is the total quantity of the two firms i.e., $Q = q_1 + q_2$, $a > 0$ and $b < 0$. The total cost of firm i to produce quantity q_i is

$$C_i(q_i) = cq_i \tag{5.2}$$

In equation (5.2), there are no fixed costs and the marginal cost is constant at c. Given the market price P, the profit π_i for a firm is computed as follows:

$$\pi_i = Pq_i - C_i \tag{5.3}$$

We used Holt's experiment parameter i.e., $a = 12$, $b = -1/2$ and the variable costs are zero ($c = 0$). With this assumption there are three theoretical benchmarks as stated in Holt (1985), and the profit for a firms is:

$$\pi_i = (12 - 1/2Q)q_i \tag{5.4}$$

The first theoretical benchmark is Nash-equilibrium (NE) where each firm chooses the quantity that maximizes its profit given by the quantity of the other firm. The second benchmark is collusive equilibrium (CE) where all competitors act as if they were a single

monopolist to maximize their joint profits. The third benchmark is the competitive outcome (CO) where firms maximize their profits given the market-clearing price. At these benchmarks, the quantity and profit are given as follows:

$$q_1 = q_2 = 8 \text{ and } \pi_1 = \pi_2 = 32 \text{ for NE} \tag{5.5}$$

$$q_1 = q_2 = 6 \text{ and } \pi_1 = \pi_2 = 36 \text{ for CE} \tag{5.6}$$

$$q_1 = q_2 = 12 \text{ and } \pi_1 = \pi_2 = 0 \text{ for CO} \tag{5.7}$$

In the collusive equilibrium of a Cournot model, firms have an incentive to increase their quantity. The increase in quantity will bring more profit despite the fact that such increase drives down the market-clearing price.

5.3 Aspiration-Based Learning Model

We consider a market that consists of N firms. At each time t each firm meets all of the other firms to play a Cournot duopoly game. The firms choose a strategy (or quantity) from strategies space $QS = \{1, ..., n\}$ as an integer number. The total profit for firm i at time t after meeting all of the other firms is:

$$\pi_i(t) = \frac{1}{N-1} \sum_{i \neq j} (12 - 1/2(q_i(t) + q_j(t)))q_i(t) \tag{5.8}$$

Let $\rho_i(t)$ be the aspiration level for firm i at time t and let $pr_i(q, t)$ be the probability of a strategy $q \in QS = \{1, ..., n\}$ for firm i at time t. A stimulus associated with total profit $\pi_i(t)$ and aspiration level $\rho_i(t)$ for taking a strategy $q \in QS$ for player i is,

$$S_i(t) = \pi_i(t) - \rho_i(t) \tag{5.9}$$

Our proposed model can be described as follows: during the game each firm will accumulate the value of stimulus it receives by playing a strategy $q \in QS$. This accumulated value will be averaged by time. After that, the average value of accumulated stimulus will pass a response function to determine the value of $\beta_i(t)$ or strength of reinforcement to update the probability. Let $\omega_i(q, t)$ be the total average of stimulus of strategy q at time t for firm i:

$$\omega_i(q,t) = \begin{cases} \frac{1}{t}[\omega_i(q, t-1) + S_i(t)] & \text{if } q \text{ is chosen} \\ \omega_i(q, t-1) & \text{otherwise} \end{cases} \tag{5.10}$$

The value of $\beta_i(t)$ will be updated according to:

$$\beta_i(t) = \frac{e^{\omega_i(q,t)}}{\sum_{\tau \in QS} e^{\omega_i(\tau,t)}} \text{ if } q \text{ is chosen} \tag{5.11}$$

By using the above updating function for $\beta_i(t)$, the value of $\beta_i(t)$ can be guaranteed in (0,1].

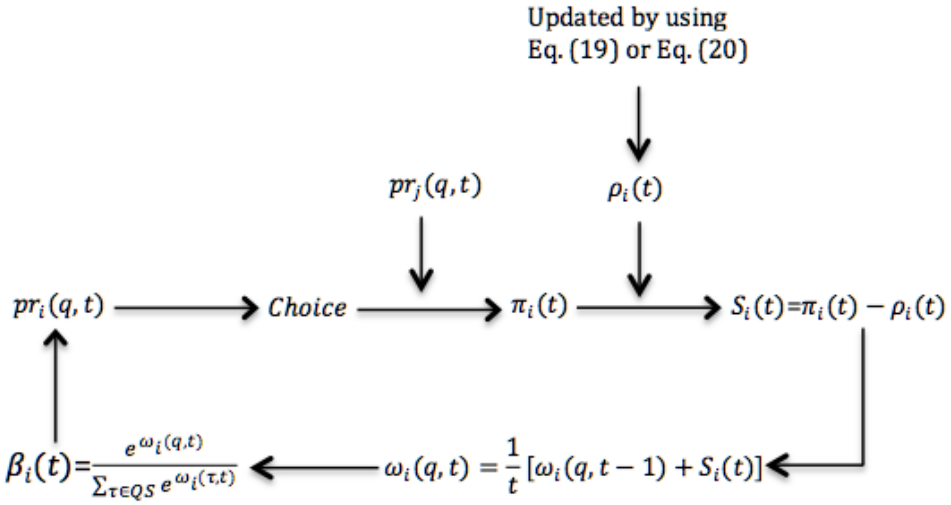


FIGURE 5.1: Schematic of the aspiration-based learning shaped by information searching mechanism.

Figure 5.1 shows a schematic of the model for firm i and $i \neq j$. Updating rules for aspiration level will be discussed in next section, i.e., equation (5.19) and equation (5.20) in Fig. 5.1.

The probability for a strategy will be updated as follows:
if $S_i(t) \geq 0$

$$pr_i(q,t+1) = \begin{cases} (1+\beta_i(t))pr_i(q,t) & \text{if } q \text{ is chosen} \\ (1-\beta_i(t))pr_i(q,t) & \text{otherwise} \end{cases} \tag{5.12}$$

if $S_i(t) < 0$

$$pr_i(q,t+1) = \begin{cases} pr_i(q,t) - \beta_i(t)(1-pr_i(q,t)) & \text{if } q \text{ is chosen} \\ \frac{\beta_i(t)}{n-1} + (1-\beta_i(t))pr_i(q,t) & \text{otherwise} \end{cases} \tag{5.13}$$

In this learning model, the value of aspiration level is not static. A firm updates its aspiration level based on the information it gets through interaction and then linearly adjusted in the direction of outcome (total profit) experienced.

5.4 Information Searching Mechanism

The theory of social comparison states that a person tends to make self-evaluation based on comparison with other persons. In this situation, the information of others would determine the behavior of the person in future. Therefore, a competitive environment may occur in this process and there is a pressure toward uniformity as stated by Festinger (1954).

In this paper, we assume each firm updates its aspiration level by searching the information of the other firms' aspiration level and then compares this information with its current aspiration level. The searching process is based on the interaction scheme that is given in the beginning of simulation. Within the interaction, a firm will obtain the information about other firms' aspiration level depending on the closeness of the firm in the given interaction scheme. The closeness is represented by weights.

Let $\alpha_i(t)$ be a level of information that will be received for a firm i at time t, and let T_i be a set of firms that interact with firm i. Let n_i be a number of firms who interact with firm i, and n_k be a number of firms who interact with firm k. We calculate $\alpha_i(t)$ as follows:

$$\alpha_i(t) = w_i\rho_i(t) + \sum_{k\in T_i} w_k\rho_k(t) \tag{5.14}$$

where, $w_k = \frac{1}{1+\max\{n_i,n_k\}}$ $\forall k \in T_i$ and $w_i = 1 - \sum_{k\in T_i} w_k$. w_k is a set of the weights that represented the closeness of a firm i in the given scheme. Equation (5.14) looks similar with distributed algorithm for distributed averaging problem (see Xiao and Boyd, 2004), but differs in term of what information will be communicated. We explain the model as follows:

Suppose the scheme of interaction is given in the beginning of simulation as shown in Fig. 5.2.

We can describe this interaction in form of matrix as shown in Table 5.1. The value of

TABLE 5.1: Matrix representation of interaction

	F_1	F_2	F_3	F_4	F_5
F_1	0	1	0	1	0
F_2	1	0	1	1	0
F_3	0	1	0	0	1
F_4	1	1	0	0	0
F_5	0	0	1	0	0

an entity $F_{ij} = 1$ if there is a connection between firm i and firm j, otherwise $F_{ij} = 0$.

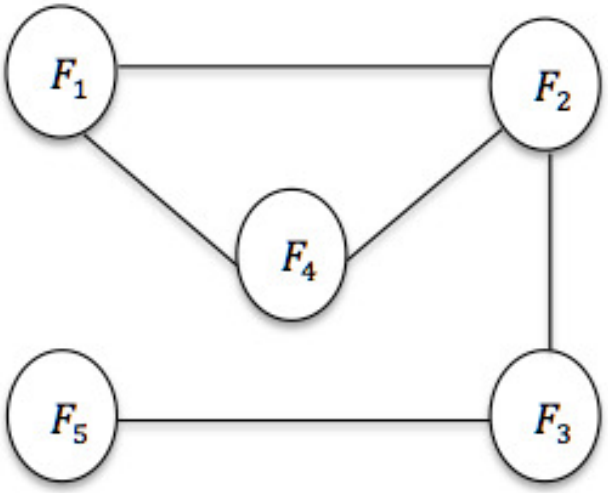


FIGURE 5.2: Scheme of interaction.

From Table 5.1 we can count n_i $\forall i \in \{1,2,3,4,5\}$, i.e. $n_1 = 2$, $n_2 = 3$, $n_3 = 2$, $n_4 = 2$, and $n_5 = 1$. Level of aspiration information for firm F_1 can be calculated as follows: F_1 interact with F_2 and F_4 so that,

$$w_2 = \frac{1}{1 + \max\{2,3\}} = \frac{1}{4} \tag{5.15}$$

$$w_4 = \frac{1}{1 + \max\{2,2\}} = \frac{1}{3} \tag{5.16}$$

so we get,

$$w_1 = 1 - (w_2 + w_4) = 1 - (\frac{1}{4} + \frac{1}{3}) = \frac{5}{12} \tag{5.17}$$

After that we can calculate,

$$\alpha_1(t) = w_1\rho_1(t) + w_2\rho_2(t) + w_4\rho_4(t) = \frac{5}{12}\rho_1(t) + \frac{1}{4}\rho_2(t) + \frac{1}{3}\rho_4(t) \tag{5.18}$$

As we can see, firm F_1 will get the aspiration information of firm F_2 and F_4 with different weight. We update aspiration level of firm i as follows:
if $\alpha_i(t) \geq \rho_i(t)$

$$\rho_i(t+1) = (1 - h_i)\alpha_i(t) + h_i\pi_i(t) \tag{5.19}$$

if $\alpha_i(t) < \rho_i(t)$

$$\rho_i(t+1) = (1 - h_i)\rho_i(t) + h_i\pi_i(t) \tag{5.20}$$

The formulations state that a firm adjusts its level of aspiration by comparing its level of aspiration with the level of information that it gets from interaction, and uses it if the level is higher or equal to its aspiration level. Parameter h_i is habituation for firm i, i.e., the degree to which the aspiration level floats toward the reward or total profit $\pi_i(t)$. If $h_i = 0$, then the second term in equation (5.19) and equation (5.20) will equal to zero. The firm i will update its aspiration level to the value of aspiration information it received, i.e., $\rho_i(t+1) = \alpha_i(t)$ in equation (5.19) or to the value of previous aspiration level, i.e., $\rho_i(t+1) = \rho_i(t)$ in equation (5.20). On the other hand, if $h_i = 1$, then the first term in equation (5.19) and equation (5.20) will equal to zero. The firm i will update its aspiration level to the value of reward (total profit) it received, i.e., $\rho_i(t+1) = \pi_i(t)$. This situation is similar to the situation in which no interaction, i.e., $\alpha_i(t) = 0$ in equation (5.19). If the scheme of interaction is fully connected, the value of $\alpha_i(t)$ will be similar to the average of aspiration level in the market.

By using equation (5.19) and equation (5.20) to update the aspiration level, we can see Fig. 5.1 as a complete model used in this research. The aspiration-based learning shaped by information searching mechanism is a learning process in which the firms update their aspiration level by comparing their aspiration level with the information of the other firms' aspiration level and then use the average value of accumulated stimulus as a strength of reinforcement to update the probability of strategies.

5.5 Simulation and Results

The main objective of the proposed model is to examine the relationship among initial aspiration level, interaction scheme, and the number of available strategies in order to reach an equilibrium, i.e., Nash or collusive equilibrium. We divide the initial aspiration into two categories. The first category is low initial aspiration level and the second is high initial aspiration level. Low initial aspiration level is the value of aspiration level which is lower than the profit of Nash equilibrium, i.e., $\rho_{low} < \pi^{NE} = 32$, and high initial aspiration level is the value of aspiration level which is higher than the profit of collusive equilibrium, i.e., $\rho_{high} > \pi^{CE} = 36$. We also divide the interaction scheme into two categories, i.e., fully connected scheme and not fully connected scheme as described in Fig. 5.3 and in Fig. 5.4.

We consider a universal set of strategies as $QS = \{1, 2, ..., 12\}$ which represent quantities to be produced by firm i. However, in this simulation we use two sets of strategies. The first set is the larger number of quantity, i.e., $QS7 = \{6, 7, 8, 9, 10, 11, 12\}$, we called it high quantity strategies. The second set is the benchmark strategies, i.e., $QS3 = \{6, 8, 12\}$, $q^{CE} = 6$ for collusive strategy, $q^{NE} = 8$ for Nash strategy, and

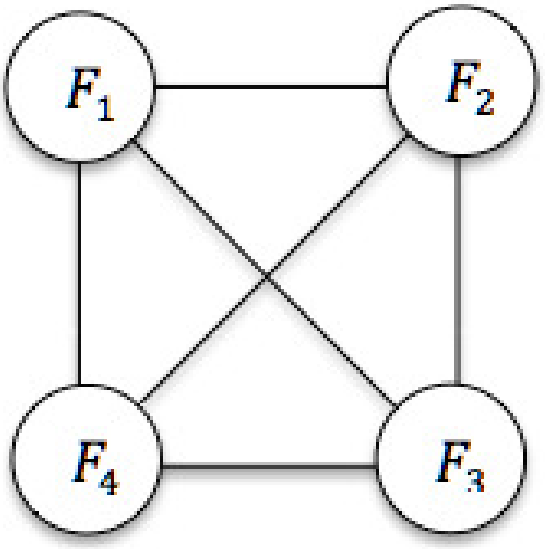


FIGURE 5.3: Fully connected scheme.

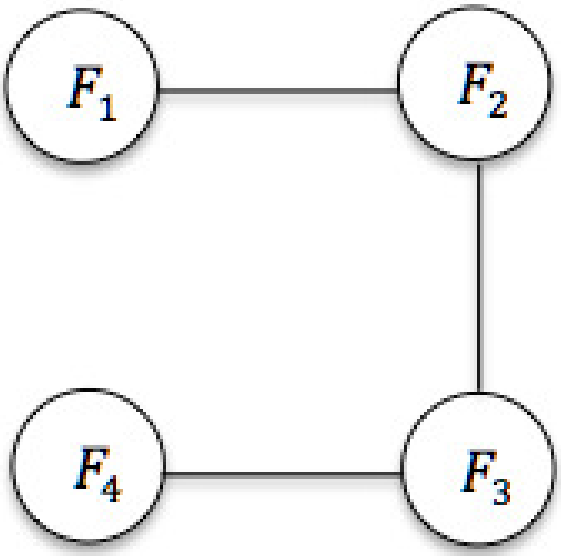


FIGURE 5.4: Not fully connected scheme.

$q^{CO} = 12$ for competitive strategy. We do not include the smaller number of quantities, i.e., $QSL = \{1, 2, 3, 4, 5\}$ because they give the same profit if both firms play the same strategy as we can see in Table 5.2. For example, if both firms play a strategy $q_1 = q_2 = 4$,

TABLE 5.2: Symmetric profit ($q_1 = q_2$)

$QS(Strategies/Quantities)$	$\pi_i = (12 - 1/2(q_1 + q_2)q_i)$
1 symmetry with 11	11
2 symmetry with 10	20
3 symmetry with 9	27
4 symmetry with 8	32
5 symmetry with 7	35

then they will get a profit $\pi_1 = \pi_2 = 32$, and this result is the same if both firms play a strategy $q_1 = q_2 = 8$, they will get a profit $\pi_1 = \pi_2 = 32$ as well. Because the main objective in this research is to examine the convergence to an equilibrium, we avoid the ambiguity. We choose the high quantity strategies which also include the strategies of the benchmark case.

The simulation will involve $N = 4$ firms. At each time t each firm meets all of other firms to play a Cournot duopoly game and the total reward or total profit is given in equation (5.8). We divide the simulation into two scenarios. In the first scenario we use the high quantity strategies, i.e., $QS7 = \{6, 7, 8, 9, 10, 11, 12\}$ and in the second scenario we use the benchmark strategies, i.e., $QS3 = \{6, 8, 12\}$. In each scenario we examine four conditions as follows:

1 . Fully connected vs. low initial aspiration level

2 . Fully connected vs. high initial aspiration level

3 . Not fully connected vs. low initial aspiration level

4 . Not fully connected vs. high initial aspiration level

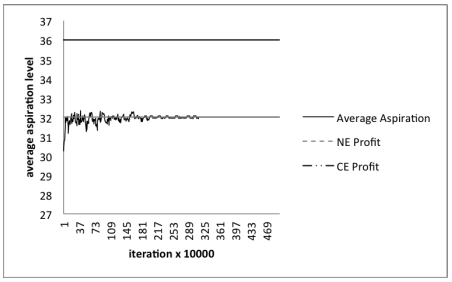
5.5.1 The first scenario: High quantity strategies

In this simulation we want to show the capability of the proposed model in the circumstance that involve many strategies. The strategies space are $QS7 = \{6, 7, 8, 9, 10, 11, 12\}$. We set the number of firms to $N = 4$. To investigate long-term dynamics of the simulation, we set the number of iterations to $5 * 10^6$ and run the simulation for 50 times. There are three outputs we want to show, i.e., the average profit in the market, the average aspiration level for all firms, and the average probabilities of the strategies. These output were averaging within 50 numbers of running.

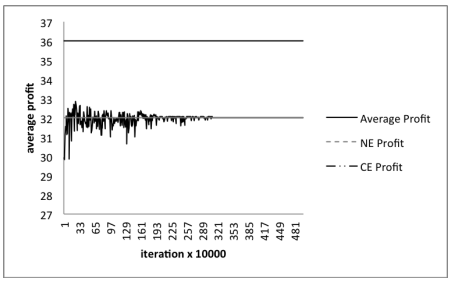
5.5.1.1 Fully connected vs. low initial aspiration level in high quantity strategies

We use scheme of interaction as described in Fig. 5.3. The initial aspiration level for each firm is $\{\rho_1(0) = 25, \rho_2(0) = 27, \rho_3(0) = 29, \rho_4(0) = 31\}$, all the initial aspiration level is lower than the profit of Nash equilibrium, i.e., $\rho_i(0) < \pi^{NE} = 32\ \forall i \in \{1, 2, 3, 4\}$. We also set the habituation $h_1 = h_2 = h_3 = h_4 = 4 * 10^{-4}$. The initial probability is $pr_1(q, 0) = pr_2(q, 0) = pr_3(q, 0) = pr_4(q, 0) = \frac{1}{7}$, $\forall q \in QS7$, and the initial of the total average of stimulus of strategy q is $\omega_1(q, 0) = \omega_2(q, 0) = \omega_3(q, 0) = \omega_4(q, 0) = 0$, $\forall q \in QS7$.

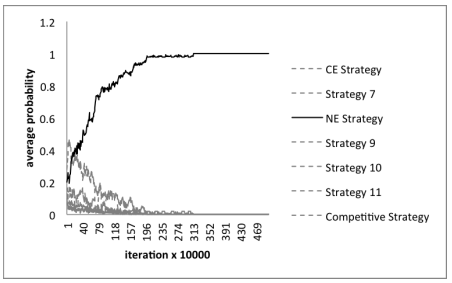
Within these conditions the market converges to Nash equilibrium. The overall average aspiration level goes to the Nash equilibrium profit, i.e., $\pi^{NE} = 32$ as we can see in Fig. 5.5(a). Also, Fig. 5.5(b) shows that the overall average profit goes to the Nash equilibrium profit, i.e., $\pi^{NE} = 32$. All firms tend to achieve their goal (aspiration level). Figure 5.5(c) emphasizes the results. All firms converge to Nash strategy, i.e., $q^{NE} = 8$, the probability of this strategy goes to 1. The convergence time is around $3.1 * 10^6$ and all firms learn to use the best reply strategy that leads the market to the Nash equilibrium.



(a) Dynamics of average aspiration level



(b) Dynamics of average profit

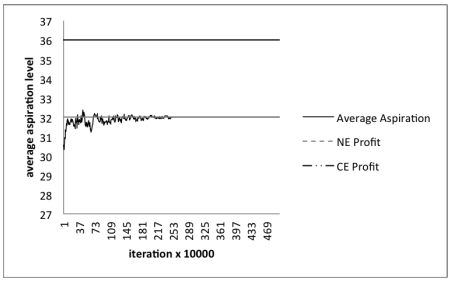


(c) Dynamics of average probability

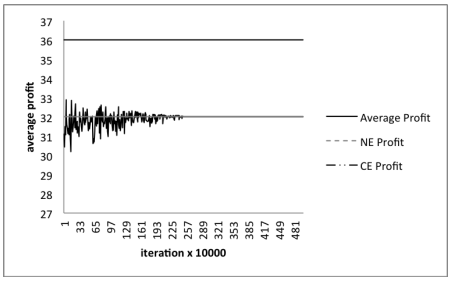
FIGURE 5.5: The outputs of fully connected vs. low initial aspiration level in high quantity strategies

5.5.1.2 Fully connected vs. high initial aspiration level in high quantity strategies

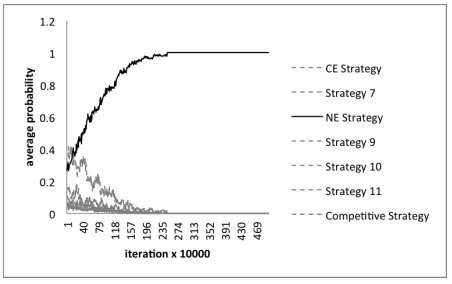
We use scheme of interaction as described in Fig. 5.3. The initial aspiration level for each firm is $\{\rho_1(0) = 40, \rho_2(0) = 42, \rho_3(0) = 45, \rho_4(0) = 47\}$, all the initial aspiration level is higher than the profit of collusive equilibrium, i.e., $\rho_i(0) > \pi^{CE} = 36\ \forall i \in \{1, 2, 3, 4\}$. We also set the habituation $h_1 = h_2 = h_3 = h_4 = 4 * 10^{-4}$. The initial probability is $pr_1(q, 0) = pr_2(q, 0) = pr_3(q, 0) = pr_4(q, 0) = \frac{1}{7}$, $\forall q \in QS7$, and the initial of the total average of stimulus of strategy q is $\omega_1(q, 0) = \omega_2(q, 0) = \omega_3(q, 0) = \omega_4(q, 0) = 0$, $\forall q \in QS7$.



(a) Dynamics of average aspiration level



(b) Dynamics of average profit



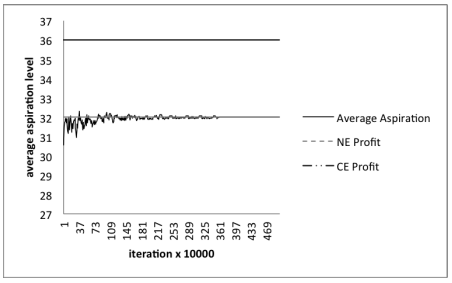
(c) Dynamics of average probability

FIGURE 5.6: The outputs of fully connected vs. high initial aspiration level in high quantity strategies

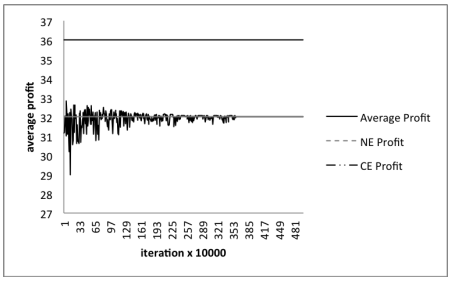
In the fully connected scheme with high initial aspiration level the market also converges to the Nash equilibrium as shown in Fig. 5.6(a), Fig. 5.6(b), and Fig. 5.6(c). All firms learn to use the best reply strategy that leads the market to the Nash equilibrium. However, the convergence time (around $2.5 * 10^6$) is more quickly than fully connected scheme with low initial aspiration level (around $3.1 * 10^6$). The firms take a long time to coordinate their aspiration level to reach the NE if they start with low aspiration level, i.e., below the Nash equilibrium profit $\pi^{NE} = 32$. There is no different result between low and high initial aspiration level in fully connected scheme except the time to converge.

5.5.1.3 Not fully connected vs. low initial aspiration level in high quantity strategies

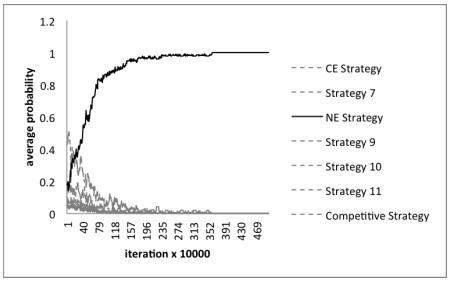
In this scenario we use scheme of interaction as described in Fig. 5.4. The initial aspiration level for each firm is $\{\rho_1(0) = 25, \rho_2(0) = 27, \rho_3(0) = 29, \rho_4(0) = 31\}$, all the initial aspiration level is lower than the profit of Nash equilibrium, i.e., $\rho_i(0) < \pi^{NE} = 32$ $\forall i \in \{1, 2, 3, 4\}$. We also set the habituation $h_1 = h_2 = h_3 = h_4 = 4 * 10^{-4}$. The initial probability is $pr_1(q, 0) = pr_2(q, 0) = pr_3(q, 0) = pr_4(q, 0) = \frac{1}{7}$, $\forall q \in QS7$, and the initial of the total average of stimulus of strategy q is $\omega_1(q, 0) = \omega_2(q, 0) = \omega_3(q, 0) = \omega_4(q, 0) = 0$, $\forall q \in QS7$.



(a) Dynamics of average aspiration level



(b) Dynamics of average profit



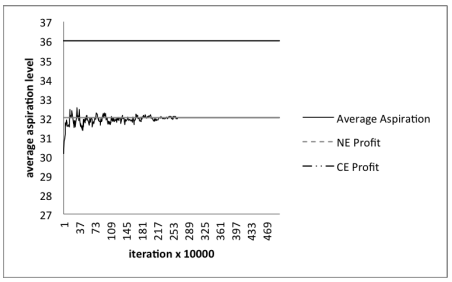
(c) Dynamics of average probability

FIGURE 5.7: The outputs of not fully connected vs. low initial aspiration level in high quantity strategies

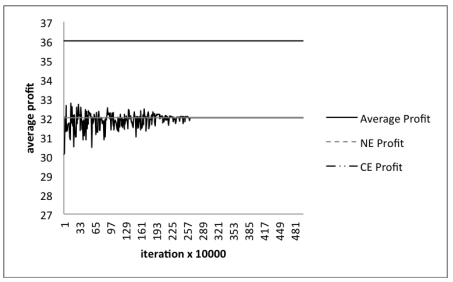
In the not fully connected scheme with low initial aspiration level the market also converges to the Nash equilibrium as shown in Fig. 5.7(a), Fig. 5.7(b), and Fig. 5.7(c). The convergence time is around $3.5 * 10^6$. Low initial aspiration level tends to take more time to reach the convergency to NE. The convergence time is almost similar with fully connected scheme with low initial aspiration (around $3.1 * 10^6$).

5.5.1.4 Not fully connected vs. high initial aspiration level in high quantity strategies

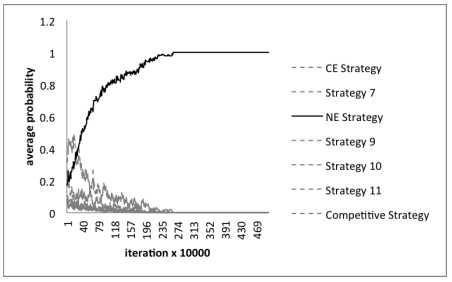
In this scenario we use scheme of interaction as described in Fig. 5.4. The initial aspiration level for each firm is $\{\rho_1(0) = 40, \rho_2(0) = 42, \rho_3(0) = 45, \rho_4(0) = 47\}$, all the initial aspiration level is higher than the profit of collusive equilibrium, i.e., $\rho_i(0) > \pi^{CE} = 36\ \forall i \in \{1, 2, 3, 4\}$. We also set the habituation $h_1 = h_2 = h_3 = h_4 = 4 * 10^{-4}$. The initial probability is $pr_1(q, 0) = pr_2(q, 0) = pr_3(q, 0) = pr_4(q, 0) = \frac{1}{7}$, $\forall q \in QS7$, and the initial of the total average of stimulus of strategy q is $\omega_1(q, 0) = \omega_2(q, 0) = \omega_3(q, 0) = \omega_4(q, 0) = 0$, $\forall q \in QS7$.



(a) Dynamics of average aspiration level



(b) Dynamics of average profit



(c) Dynamics of average probability

FIGURE 5.8: The outputs of not fully connected vs. high initial aspiration level in high quantity strategies

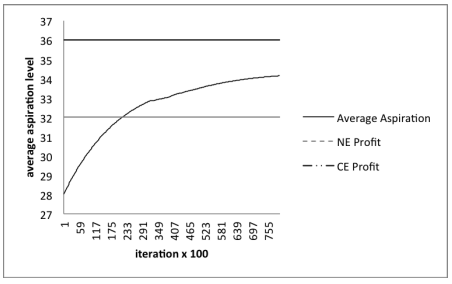
In the not fully connected scheme with high initial aspiration level the market converges to the Nash equilibrium as shown in Fig. 5.8(a), Fig. 5.8(b), and Fig. 5.8(c). All firms learn to use the best reply strategy that leads the market to the Nash equilibrium. The convergence time is around $2.7 * 10^6$ which is almost similar with the convergence time for fully connected scheme with high initial aspiration level (around $2.5 * 10^6$), but quicker than fully connected scheme with low initial aspiration level (around $3.1 * 10^6$). High initial aspiration level can accelerate the convergence time more quickly compare with low initial aspiration level.

5.5.2 The second scenario: Benchmark strategies

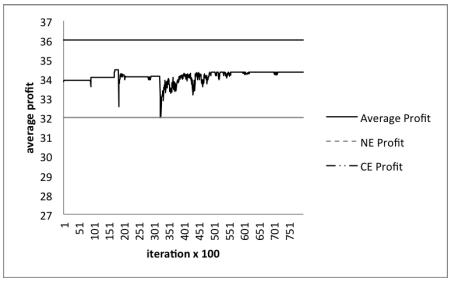
In this scenario we want to examine a circumstance in which the strategies space $QS3$ consists only of the benchmark strategies, i.e., $QS3 = \{6, 8, 12\}$. The number of firms are $N = 4$. We set the number of iterations to $8 * 10^4$ and run the simulation for 50 times. There are three outputs we want to show, i.e., the average profit in the market, the average aspiration level for all firms, and the average probabilities of the strategies. These output were averaging within 50 numbers of running.

5.5.2.1 Fully connected vs. low initial aspiration level in benchmark strategies

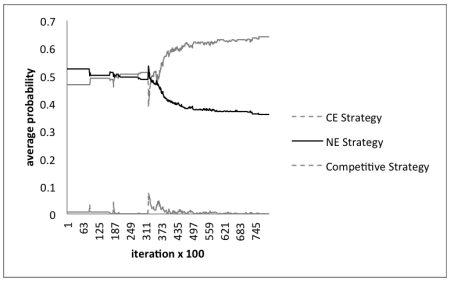
We use scheme of interaction as described in Fig. 5.3. The initial aspiration level for each firm is $\{\rho_1(0) = 25, \rho_2(0) = 27, \rho_3(0) = 29, \rho_4(0) = 31\}$, all the initial aspiration level is lower than the profit of Nash equilibrium, i.e., $\rho_i(0) < \pi^{NE} = 32\ \forall i \in \{1, 2, 3, 4\}$. We also set the habituation $h_1 = h_2 = h_3 = h_4 = 5 * 10^{-5}$. The initial probability is $pr_1(q, 0) = pr_2(q, 0) = pr_3(q, 0) = pr_4(q, 0) = \frac{1}{3}$, $\forall q \in QS3$, and the initial of the total average of stimulus of strategy q is $\omega_1(q, 0) = \omega_2(q, 0) = \omega_3(q, 0) = \omega_4(q, 0) = 0$, $\forall q \in QS3$.



(a) Dynamics of average aspiration level



(b) Dynamics of average profit



(c) Dynamics of average probability

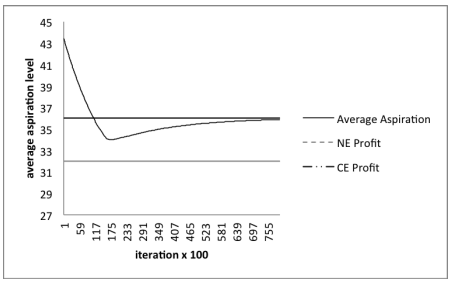
FIGURE 5.9: The outputs of fully connected vs. low initial aspiration level in benchmark strategies

There are two states of convergence in the fully connected scheme with low initial aspiration level, i.e., NE and CE. The proportion of the convergence is 64% for CE and 36% for NE. It means that in 50 numbers of running there are 32 times converge to CE and 18 times converge to NE. We can refer to this situation by looking at the dynamics of average probability in Fig. 5.9(c). The typical behavior of this result can be explained as follows. If a firm chooses Nash strategy ($q^{NE} = 8$), then the firm will get a profit $\pi^{NE} = 32$ (higher than its aspiration level) and will be satisfied with this strategy (increase the probability). If this strategy prevails, then in the long-run the firm will settle at this strategy. Also, if a firm chooses collusive strategy ($q^{CE} = 6$), then the firm will get a profit $\pi^{CE} = 36$ (higher than its aspiration level) and will be satisfied with this strategy (increase the probability) as well. If this strategy prevails, then in the long-run the firm will settle at this strategy. The competitive strategy $q^{CO} = 12$ will have a small chance to be chosen because this strategy will give zero profit ($\pi^{CO} = 0$) for both firms. As we can also see in Fig. 5.9(a), the aspiration level is below π^{NE} and π^{CE} at the beginning. If in the future the firms coordinate their strategies to $q^{NE} = 8$, each firm will get $\pi^{NE} = 32$. This value is higher than its aspiration level so that the firm satisfies with this state, i.e., converges to NE. On the other hand, if in the future the firms coordinate their strategies to $q^{CE} = 6$, each firm will get $\pi^{CE} = 36$. This value is also higher than its aspiration level so that the firm satisfies with this state, i.e, converge to CE. Therefore, we can see in Fig. 5.9(a) the overall average of aspiration level lies between NE and CE.

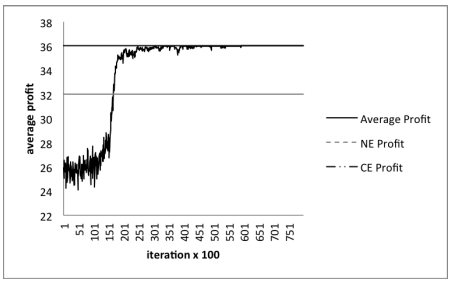
5.5.2.2 Fully connected vs. high initial aspiration level in benchmark strategies

We use scheme of interaction as described in Fig. 5.3. The initial aspiration level for each firm is $\{\rho_1(0) = 40, \rho_2(0) = 42, \rho_3(0) = 45, \rho_4(0) = 47\}$, all the initial aspiration level is higher than the profit of collusive equilibrium, i.e., $\rho_i(0) > \pi^{CE} = 36\ \forall i \in \{1, 2, 3, 4\}$. We also set the habituation $h_1 = h_2 = h_3 = h_4 = 5 * 10^{-5}$. The initial probability is $pr_1(q, 0) = pr_2(q, 0) = pr_3(q, 0) = pr_4(q, 0) = \frac{1}{3}$, $\forall q \in QS3$, and the initial of the total average of stimulus of strategy q is $\omega_1(q, 0) = \omega_2(q, 0) = \omega_3(q, 0) = \omega_4(q, 0) = 0$, $\forall q \in QS3$.

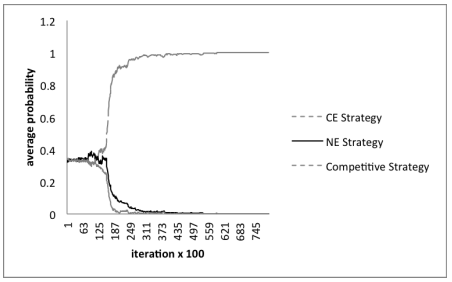
The market converges to CE in the fully connected scheme with high initial aspiration level. We can refer to this situation in Fig. 5.10(c) in which the probability for collusive strategy, i.e., $q^{CE} = 6$ goes to 1. As we can also see in Fig. 5.10(a), the aspiration level is higher than $\pi^{CE} = 36$ at the beginning. In the future the firms coordinate their strategies to $q^{CE} = 6$ and get $\pi^{CE} = 36$. If at one time some firms change their strategy to the other strategies (i.e., Nash strategy ($q^{NE} = 8$) with profit $\pi^{NE} = 32$ or



(a) Dynamics of average aspiration level



(b) Dynamics of average profit



(c) Dynamics of average probability

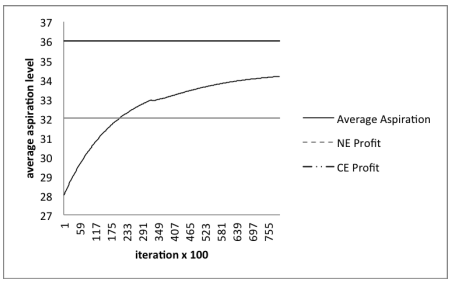
FIGURE 5.10: The outputs of fully connected vs. high initial aspiration level in benchmark strategies

competitive strategy ($q^{CO} = 12$) with profit $\pi^{CO} = 0$), they will get profit that is lower than $\pi^{CE} = 36$. Hence the firms will not be satisfied (decrease the probability). The firms should choose $q^{CE} = 6$ again and settle in this state. Also, we can see in Fig. 5.10(b) the firms' profit converge to $\pi^{CE} = 36$.

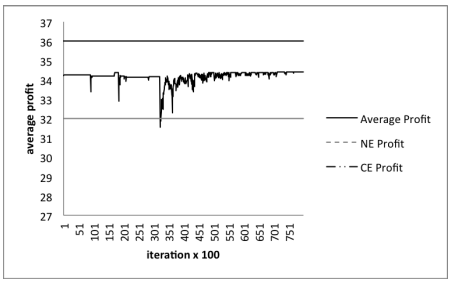
5.5.2.3 Not fully connected vs. low initial aspiration level in benchmark strategies

In this scenario we use scheme of interaction as described in Fig. 5.4. The initial aspiration level for each firm is $\{\rho_1(0) = 25, \rho_2(0) = 27, \rho_3(0) = 29, \rho_4(0) = 31\}$, all the initial aspiration level is lower than the profit of Nash equilibrium, i.e., $\rho_i(0) < \pi^{NE} = 32$ $\forall i \in \{1, 2, 3, 4\}$. We also set the habituation $h_1 = h_2 = h_3 = h_4 = 5 * 10^{-5}$. The initial probability is $pr_1(q, 0) = pr_2(q, 0) = pr_3(q, 0) = pr_4(q, 0) = \frac{1}{3}$, $\forall q \in QS3$, and the initial of the total average of stimulus of strategy q is $\omega_1(q, 0) = \omega_2(q, 0) = \omega_3(q, 0) = \omega_4(q, 0) = 0$, $\forall q \in QS3$.

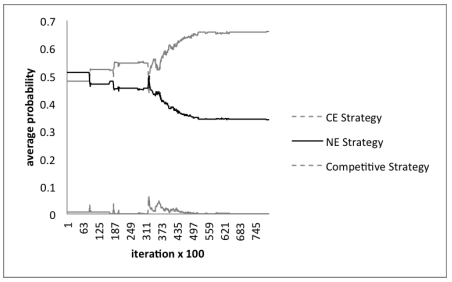
The results are similar to the fully connected scheme with low initial aspiration level. We can refer to Fig. 5.11(a), Fig. 5.11(b), and Fig. 5.11(c). If the firms' initial aspiration value is below the $\pi^{NE} = 32$, then the proportion of the convergence is 66% for CE and 34% for NE. It means that in 50 numbers of running there are 33 times converge to CE and 17 times converge to NE. We can refer to this situation by looking at the dynamics



(a) Dynamics of average aspiration level



(b) Dynamics of average profit



(c) Dynamics of average probability

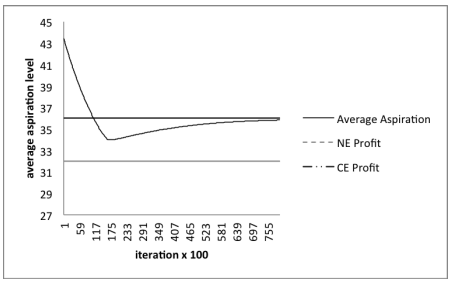
FIGURE 5.11: The outputs of not fully connected vs. low initial aspiration level in benchmark strategies

of average probability in Fig. 5.11(c). Typical behavior of this result is similar with fully connected vs. low initial aspiration level in subsection 5.2.1. Interaction scheme does not affect the market in low initial aspiration level.

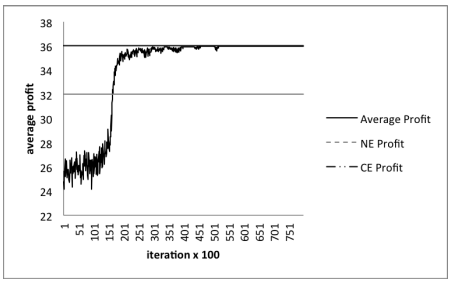
5.5.2.4 Not fully connected vs. high initial aspiration level in benchmark strategies

In this scenario we use scheme of interaction as described in Fig. 5.4. The initial aspiration level for each firm is $\{\rho_1(0) = 40, \rho_2(0) = 42, \rho_3(0) = 45, \rho_4(0) = 47\}$, all the initial aspiration level is higher than the profit of collusive equilibrium, i.e., $\rho_i(0) > \pi^{CE} = 36 \ \forall i \in \{1, 2, 3, 4\}$. We also set the habituation $h_1 = h_2 = h_3 = h_4 = 5 * 10^{-5}$. The initial probability is $pr_1(q, 0) = pr_2(q, 0) = pr_3(q, 0) = pr_4(q, 0) = \frac{1}{3}$, $\forall q \in QS3$, and the initial of the total average of stimulus of strategy q is $\omega_1(q, 0) = \omega_2(q, 0) = \omega_3(q, 0) = \omega_4(q, 0) = 0$, $\forall q \in QS3$.

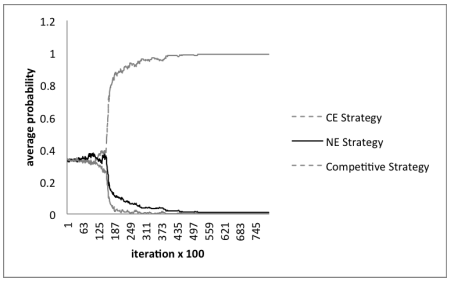
The results are similar to the fully connected scheme with high initial aspiration level. We can refer to Fig. 5.12(a), Fig. 5.12(b), and Fig. 5.12(c). High initial aspiration level is needed to reach the $\pi^{CE} = 36$ and interaction scheme does not effect the market.



(a) Dynamics of average aspiration level



(b) Dynamics of average profit



(c) Dynamics of average probability

FIGURE 5.12: The outputs of not fully connected vs. high initial aspiration level in benchmark strategies

5.6 Conclusion

We have shown that the proposed aspiration-based learning shaped by an information searching mechanism lead the market consisting of 4 firms to two equilibrium states, i.e., Nash equilibrium and collusive equilibrium. The Nash equilibrium is achieved in the competitive environment that is represented by having many strategies available to the firms. The convergence is not affected either by initial aspiration level (high or low) or the scheme of interaction (fully connected or not fully connected). However, the convergence time is more quickly to achieve in fully connected scheme with high initial aspiration level. The behavior of the firms is more competitive if they have many strategies.

This result is different from previous studies (see Altavilla et al., 2006; Dixon, 2000; Dixon et al., 2006) that assume all firms have the same aspiration level which is represented by overall average profit as a reference point. In these previous studies, the firms will try new strategies anytime their profits fall below the overall average profit (their aspiration level). Within this process they found that collusive behavior is likely to be observed in experimental duopolies (Altavilla et al., 2006; Dixon et al., 2006) and in the simulate duopolies (Dixon, 2000). The use of overall average profit as a reference point (aspiration level) to all firms tends to stimulate collusive behavior than the use of individual aspiration level (each firm has its own aspiration level). The reason for

this result is that a firm compares its goal (aspiration level) to all of its competitors and adjust its goal (aspiration level) based on this comparison. This process is quite similar to imitation process in the long term in which the firms imitate the aspiration level. The imitation behavior has been proved to increase the level of competition in the market (Huck et al., 1999; Vega-Redondo, 1997). Besides, the proposed learning model is built based on the reinforcement theory. Within this reinforcement process, the state space is more complex in terms of the probability over all available actions compare to satisfying theories. Therefore, the competitive behavior also increases if the firms have many strategies.

However, the Collusive equilibrium can be achieved if the firms have only three strategies. The initial aspiration level is crucial to reach CE. The initial aspiration level must be higher than $\pi^{CE} = 36$. With the higher initial aspiration level, at the beginning of each iteration, each firm experiments all strategies. If a firm chooses a strategy that give the firm a profit lower than $\pi^{CE} = 36$ (i.e., Nash strategy ($q^{NE} = 8$) with profit $\pi^{NE} = 32$ or Competitive strategy ($q^{CO} = 12$) with profit $\pi^{CO} = 0$), then the firm will not be satisfied. This strategy will not be favorable in the future (decrease the probability) and the firm will choose another strategy (i.e., Collusive strategy ($q^{CE} = 6$) with profit $\pi^{NE} = 36$). In the long-run the firm will settle at $q^{CE} = 6$ and get profit $\pi^{CE} = 36$.

On the other hand, if the firms start with lower aspiration level, i.e., lower than $\pi^{NE} = 32$, the experimentation can lead the firms to two states of convergence, i.e., NE or CE. If a firm chooses Nash strategy ($q^{NE} = 8$), then the firm will get a profit $\pi^{NE} = 32$ (higher than its aspiration level) and will be satisfied with this strategy (increase the probability). If this strategy prevails, then in the long-run the firm will settle at this strategy. Also, if a firm chooses collusive strategy ($q^{CE} = 6$), then the firm will get a profit $\pi^{CE} = 36$ (higher than its aspiration level) and will be satisfied with this strategy (increase the probability) as well. If this strategy prevails, then in the long-run the firm will settle at this strategy. The competitive strategy $q^{CO} = 12$ will have a small chance to be chosen because this strategy will give zero profit ($\pi^{CO} = 0$) for both firms. The chance to converge to either NE or CE is around 35% or around 65%, respectively.

The effect of interaction scheme is not crucial to reach the equilibrium. As long as there is a connectivity that involve all firms, the firms will get the information that coordinate their strategy. As we can see in the not fully connected scheme (Fig. 5.4), the first firm can get the information from the second firm, the second firm can get the information from the third firm, and the third firm can get the information from the fourth firm. Because the connection is in two directions, there is a chain of information in the long-run so that the firms can coordinate their strategy. This situation is also confirmed by Festinger (1954) in social comparison theory, i.e., pressure towards uniformity.

The proposed model gives us insight into the aspiration level, which determines the behavior of the firm in duopoly game. Besides, the number of strategies that are available to the firms would change the convergence of the market. However, in future research we would like to confirm our findings with experiment tests.

Chapter 6

Conclusion and Future Research

This thesis developed a new method of learning based on aspiration level and a new information searching mechanism to investigate individuals behavior in some problems, such as social dilemmas, learning organizational, and duopoly markets. These new models was motivated by the opportunities to improve the performance of aspiration-based learning model to investigate various problems, such as social dilemmas, organizational learning, and duopoly markets. General results of our model can be stated as follows:

Within the learning method shaped by sharing mechanism, the individuals tend to coordinate their action to equilibrium state. The performance of the system can be improved in some cases, e.g., social dilemmas and organizational learning and can proved two existing equilibriums, i.e., Nash equilibrium and collusive equilibrium in duopoly markets.

For the detail of the general results, we explained each implementation of the model and tried to describe some future research related for our model.

6.1 Summary of Model Implementation

We summarized the implementation of the model as follows:

1. In the social dilemmas' problem, the model can improve and maintain cooperative behavior in heterogeneous agents. Varying the learning rate can eliminate the heterogeneity of agents. The model also describes a characteristic of learning to reach an optimal outcome. The agents should slowly increase their learning rate after receiving negative stimulus and slowly decrease their learning rate after receiving positive stimulus. With interaction and sharing aspiration all players

can improve their cooperative behavior in a group. Players with different initial aspiration and learning model can adjust their aspiration through interaction and sharing aspiration process. The model has faster convergence time compared to Q-learning and Bush-Mosteller's model, overcomes heterogeneity in parameters setting, i.e., initial aspiration level, and habituation, overcome different model of learning, and can be used when there are more two players.

2. In organizational learning, balancing the process of exploration and exploitation in organizational learning can be achieved by individuals' experimentation and interaction to share their aspiration. Variability in beliefs perceived by individuals can be maintained via individuals' experimentation. Therefore, individuals can improve their knowledge. The process of mutual learning sharing aspiration level makes the heterogeneity in aspiration disappear. Individuals' experimentation is not a risk as long as the heterogeneity in individuals' aspiration level can be eliminated. The model improves the overall results compared to March's model and more robust in changing environment.

3. In duopoly markets, the model leads the market to two equilibrium states, i.e., Nash equilibrium and collusive equilibrium. The Nash equilibrium is achieved in the competitive environment that is represented by having many strategies available to the firms. The convergence is not affected either by initial aspiration level (high or low) or the scheme of interaction (fully connected or not fully connected). However, the convergence time is more quickly to achieve in fully connected scheme with high initial aspiration level. The behavior of the firms is more competitive if they have many strategies. The model can be used in heterogeneous agents, compared with Dixon's model that only uses homogenous agents, and more than 2 players can be involved, compared with 2 players in Q-learning.

Based on the above summary of the models, aspiration-based reinforcement learning shaped by sharing mechanism have the following features:

1. Solve the problem of heterogeneity
2. Promote coordination and cooperation.
3. Improve understanding about agents' behavior.
4. Can be applied in various domains.

6.2 Future Research

Our proposed model has proved the importance of aspiration-learning and sharing mechanism to investigation the behavior of agents by playing specific game. However, we only justify the model through simulation. In the future we would like to justify our model by using analytical approach and experimental approach. In analytical approach, we need to simplify our model, i.e., only consider two-players case to investigate precisely the dynamic of aspiration-based learning shaped by sharing mechanism. In experimental approach, we need to design experimental procedure that can be used to explore dependence of aspiration level and sharing mechanism among the subjects.

6.2.1 Analytical Approach

The first approach we want to use to justify the model is by using analytical approach, such as replicator dynamics. We want to consider the prisoner's dilemma problem and the structural change of the game by introducing the social learning dynamics (Deguchi, 2004). We want to introduce the concept of indirect control of an agent society by linking his/her aspiration level. Also, by introducing the information sharing mechanism to the agent society, we can analyze the structural change. The bifurcation analysis of the dynamical system can be used to analyze the structural change.

6.2.2 Experimental Approach

The second approach we want to use to justify the model is by using experimental approach. We want to consider the concept of social value orientation (SVO) (Murphy et al., 2011; Balliet et al., 2009). We can define the SVO as measuring the magnitude of the concern people have for others. By using experimental prisoner's dilemma played by human subject, we can analyze the effect of aspiration level in the learning process. For this purpose, Figure 6.1 shows the conceptual model of experimental approach. Every individual has his/her internal model, i.e., SVO and aspiration level, that make individual inclination to work together. There are two individual inclinations, i.e., general (Un) willingness to cooperate and generalized expectations of others. After that, they will play the prisoner's dilemma game or public goods game and this part depends on the environment setting. Based on the setting, we can analyze the results or the patterns to find factors that establish the patterns or to justify the learning model. With this analysis, we can bring the results to extend or to improve the simulation model. This model can also be used to involve the cultural or ethnical factors.

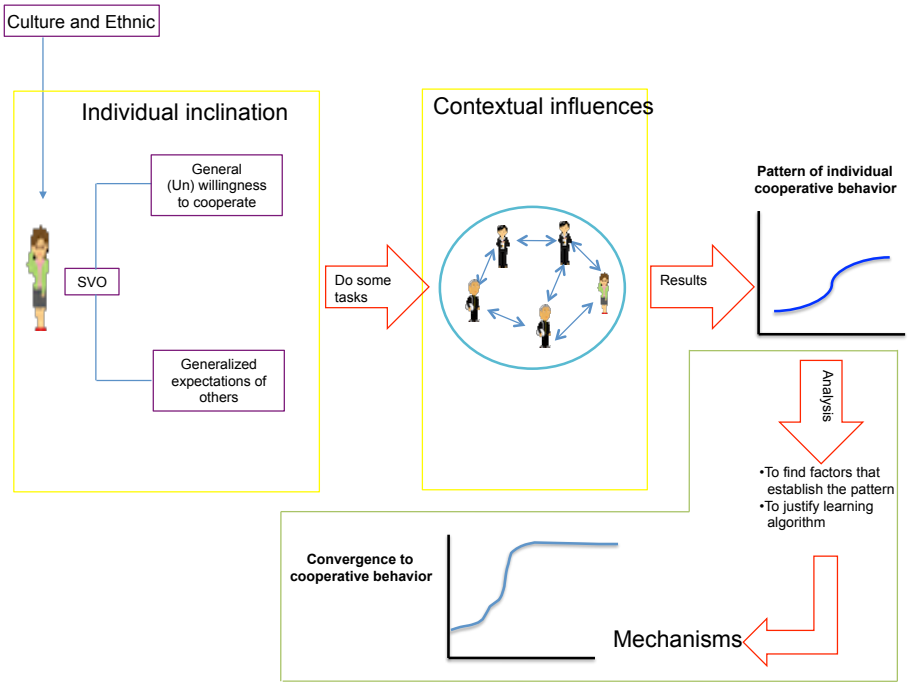


FIGURE 6.1: Conceptual Model of Experimental Approach

Bibliography

Altavilla, C., Luini, L. and Sbriglia, P. (2006) "Social Learning in Market Games," *Journal of Economic Behavior and Organization* 61. 632–652.

Anna, G. and Rapoport, A. (2006) "Embedding Social Dilemmas in Intergroup Competition Reduces Free-riding," *Organizational Behavior and Human Decision Making* 101. 184–199.

Balliet, D., Parks, C. and Joireman, J. (2009) "Social Value Orientation and Cooperation in Social Dilemmas: A Meta-Analysis," *Group Processes and Intergroup Relations* 12(4). 533–547.

Bendor, J., Diermeier, D. and Ting, M. (2004) *The Empirical Content of Behavioral Models of Adaptation,* Mimeo. Graduate School of Business Stanford University.

Bendor, J., Mookherjee, D. and Ray, D. (2001a) "Reinforcement Learning in Repeated Interaction Games," *Advances in Theoretical Economics.* 1(1). Article 3.

Bendor, J., Mookherjee, D. and Ray, D. (2001b) "Aspiration-Based Reinforcement Learning in Repeated Interaction Games: An Overview," *International Game Theory Review* 3. 159–174.

Bicchieri, C. and Lev-On, A. (2007) "Computer-Mediated Communication and Cooperation in Social Dilemmas: An Experimental Analysis," *Politics, Philosophy and Economics* 97. 170–176.

Borgers, T. (1996) "On the Relevance of Learning and Evolution to Economic Theory," *Econ. J.* 106. 1374–1385.

Borgers, T. and Sarin, R. (1997) "Dynamic Consistency and Non- expected Utility Models of Choice Under Uncertainty," *Journal of Economic Theory* 77. 1–14.

Bowling, M. and Veloso, M. (2002) "Multi-Agent Learning Using Variable Learning Rate," *Journal of Artificial Intelligence* pp. 136:215–250.

Brenner, T. (2004) *Agent Learning Representation Advice in Modeling Economic Learning,* Papers on Economics and Evolution. Jena: Max Planck Institute.

Bush, R. and Mosteller, F. (1955)*Stochastic Models of Learning,* New York: John Wiley and Sons.

Claus, C. and Boutilier, C. (1998)"The Dynamics of Reinforcement Learning in Cooperative Multi-Agent Systems," *In Proceedings of the 6th Conference On Artificial Intelligence (AAAI-98) and of the 11th Con- ference On Innovative Applications of Intelligence (IAAI-98)* pp. 764–752.

Cremer, D. D. and Vugt, M. V. (2002)"Intergroup and Intragroup Aspects of leadership in Social Dilemmas: A Relational Model of Cooperation," *Journal of Experimental Social Psychology* 38. 126–136.

Cyert, R. M. and March, J. G. (1963)"A Behavioral Theory of The firm," *Englewood Cliffs, N.J: Prentice-Hall.* .

Dawes, R. M. and Messick, D. M. (2000)"Social Dilemmas," *International Journal of Psychology* 35(2). 111–116.

Deguchi, H. (2004)*Economics as an Agent-Based Complex System: Toward Agent-Based Social Systems Science* Springer-Verlag Tokyo.

Dipyaman, B. and Sen, S. (2007)"Reaching Pareto-Optimality in Prisoners Dilemma Using Conditional joint Action Learning," *Auton Agent Multi-Agent Syst.* 15. 91–108.

Dixon, H. (2000)"Keeping Up with the Joneses: Competition and the Evolution of Collusion," *Journal of Economic Behavior and Organization* 43. 223–238.

Dixon, H., Sbriglia, P. and Somma, E. (2006)"Learning to Collude: An Experiment in Convergence and Equilibrium Selection in Oligopoly," *Research in Economic* 60. 155–167.

Dziubinski, M. and Roy, J. (2007)"Endogenous Selection of Aspiring and Rational rules in Coordination Games," *MPRA Paper No. 5941* .

Erev, I. and Rapoport, A. (1998)"Coordination, Magic, and Reinforcement Learning in a Market Entry Game ," *Games and Economic Behavior* 23. 146–175.

Erev, I. and Roth, A. (1998)"Predicting How People Play Games: Reinforcement Learning and Experimental Games with Unique Mixed Strategy Equilibrium," *American Economic Review* 88. 848–881.

Estes, W. (1954)"Individual Behavior in Uncertain Situations: An Interpretation in Terms of Statistical Association Theory," *In R. M. Thrall, C. H. Coombs, and R. L. Davis, editors, Decision Processes. Wiley, New York* .

Festinger, L. (1954)"A Theory of Social Comparison Processes," *Human Relations* 7. 117–140.

Flache, A. and Macy, M. (2002)"The Power Law of Learning," *Journal of Conflict Resolution* 46(5). 629–653.

Greve, H. R. (1998)"Performance, Aspirations, and Risky Organizational Change," *Administrative Science Quarterly* 43(1). 58–86.

HA, S. (1991)"Bounded Rationality and Organizational Learning," *Org Sci* 2(1). 125–134.

Holt, C. A. (1985)"An Experimental Test of the Consistent-Conjectures Hypothesis," *The American Economic Review* 75(3). 314–325.

Hopthrow, T. and Hulbert, L. G. (2005)"The Effect of Group Decision Making on Cooperation in Social Dilemmas," *Group Processes and Intergroup Relations* 8(1). 89–100.

Huck, S., Normann, H. and Oechssler, J. (1999)"Learning in Cournot OligopolyAn Experiment," *Economic Journal* 109. C80–C95.

Izquierdo, S. S., Izquierdo, L. R. and Gotts, N. M. (2008)"Reinforcement Learning Dynamics in Social Dilemmas," *J. Artif. Socie. Soc. Simul.* 11(2):1.

J, M. (1996)"Continuity and Change in Theories of Organizational Action," *Adm Sci Q* 41. 278–287.

Karandikar, R., Mookherjee, D., Ray, D. and Vega-Redondo, F. (1998)"Evolving Aspirations and Cooperation," *Journal of Economic Theory* 80(2). 292–331.

Kerr, N. and Kaufmann-Gilliland, C. (1994)"Communication, Commitment, and Cooperation in Social Dilemmas," *Journal of Personality and Social Psychology* 66. 513–529.

Kim, Y. (1999)"Satisficing and Optimality in 2x2 Common Interest Games," *Economic Theory* 13(2). 365–375.

Kimbrough, S. and Lu, M. (2003)*A Note on Q-Learning in the Cournot Game,* Proceedings of the Second Workshop on E-Business.

Kurzban, R. and Descioli, P. (2008)"Reciprocity in Groups: Information-seeking in a Public Goods Game," *European Journal of Social Psychology* 38. 138–158.

Kurzban, R. and Houser, D. (2001)"Individual Differences in Cooperation in Circular Public Goods Game," *European Journal of Social Psychology* 15. S37–S38.

L, A., P, I., JM, L. and RL, M. (2000)"Knowledge Transfer in Organizations: Learning from the Experience of Others," *Organ Behav Hum Decis Process* 82(1). 1–8.

L, B. and JG, M. (1988)"Organizational Learning," *Ann Rev Sociol* 14. 31–340.

Lant, T. K. (1992)"Aspiration Level Adaptation – An Empirical Exploration," *Management Science* 38(5). 623–644.

Levinthal, D. A. and March, J. (1981)"A Model of Adaptive Organizational Search," *J Econ Behav Org* 2. 307–333.

Levinthal, D. and March, J. (1993)"The Myopia of Learning," *Strateg Manag J* 14. 95–112.

Lupi, P. and Sbriglia, P. (2003)"Exploring Human Behavior and Learning in Experimental Cournot Settings," *Rivista Internazionale di Scienze Sociali* CXI. 373–395.

Macy, M. W. (1991)"Learning to Cooperate: Stochastic and Tacit Collusion in Social Exchange," *Am. J. Sociology* 97(3). 803–843.

Macy, M. W. and Flache, A. (2002)"Learning Dynamics in Social Dilemmas," *Proc. Natl. Acad. Sci. USA* 99. 7229–7236.

March, J. (1991)"Exploration and Exploitation in Organizational Learning," *Org Sci* 2. 71–87.

March, J. G. and Simon, H. A. (1958)"Organizations," *New York: Wiley.* .

Murphy, R., Ackermann, K. and Handgraaf, M. (2011)"Measuring Social Value Orientation," *Journal of Judgment and Decision Making* 6(8). 771–781.

Narendra, K. and Thathachar, M. (1989)"Learning Automata: An Introduction," *Englewood Cliffs: Prentice Hall* .

Oechssler, J. (2002)"Cooperation as a Result of Learning with Aspiration Levels," *Journal of Economic Behavior and Organization* 49(3). 405–409.

Palomino, F. and Vega-Redondo, F. (1999)"Convergence of Aspirations and (Partial) Cooperation in the Prisoners Dilemma," *International Journal of Game Theory* 28. 465–488.

Pazgal, A. (1997)"Satisficing Leads to Cooperation in Mutual Interest Games," *International Journal of Game Theory* 26. 439–453.

Quaglia, R. and Casey, C. D. (1996)"Toward a Theory of Student Aspirations," *Journal of Research in Rural Education* 12. 127–132.

Rapoport, A. and Amaldoss, W. (1999)"Social Dilemmas Embedded in Between-Group Competition: Effects of Contest and Distribution Rules," *M. Foddy, M. Smithson, S. Schneider, and M. Hogg (Eds.), Resolving Social Dilemmas: Dynamic, Structural, and Intergroup Aspects* pp. 66–99.

Roth, A. E. and Erev, I. (1995)"Learning in Extensive-Form Games: Experimental Data and Simple Dynamic Models in the Intermediate Term," *Games and Economic Behavior* 8. 164–212.

Selten, R. (1975)"Re-examination of the Perfectness Concept for Equilibrium Points in Extensive Games," *International Journal of Game Theory* 4(1). 25–55.

Simon, H. A. (1955)"A Behavioral Model of Rational Choice," *Quarterly Journal of Economics* 69. 99–118.

Stimpson, J. R. and Goodrich, M. A. (2003)"Learning to Cooperate in a Social Dilemma: A Satisficing Approach to Bargaining," *In Proceedings of the 20th International Conference on Machine Learning* pp. 728–735.

Sutton, R. and Barto, A. (1998)"Reinforcement Learning: An Introduction," *MIT Press* .

Thorndike, E. L. (1911)*Animal Intelligence: Experimental Studies,* New York: MacMillan.

TK, L. (1992)"Aspiration Level Adaption: An Empirical Exploration," *Manag Sci* 38. 623–644.

Tuomas, W. and Crites, R. H. (1995)"Multi-Agent Reinforcement Learning in The Iterated Prisoners Dilemma," *Biosystems* 37(1-2). 147–146.

Vega-Redondo, F. (1997)"The Evolution of Walrasian Behaviour," *Econometrica* 65. 375–384.

Vugt, M. V., Jepson, S. F., Hart, C. M. and Cremer, D. D. (2004)"Autocratic Leadership in Social Dilemmas: A Threat to Group Stability," *Journal of Experimental Social Psychology* 40. 1–13.

Xiao, L. and Boyd, S. (2004)"Fast Linear Iterations for Distributed Averaging," *Systems and Control Letters* 53. 65–78.