Department of Economics

Cooperation in Repeated Games, Bounded Rational Learning and the Adoption of Evolving Technologies

Javier Rivas

Thesis submitted for assessment with a view to obtaining the degree of Doctor of Economics of the European University Institute

Florence, December 2008

EUROPEAN UNIVERSITY INSTITUTE
Department of Economics

Cooperation in Repeated Games, Bounded Rational Learning and the Adoption of Evolving Technologies

Javier Rivas

Jury Members:

Karl Schlag, Supervisor, Universitat Pompeu Fabra
Pascal Courty, European University Institute
Edmund Hopkins, University of Edinburgh
Jordi Massó Carreras, University Autonoma Barcelona

A mi Madre y a mi Hermano.

Contents

Preface

Acknowledgments

This document contains the research I conducted from September 2004 to June 2008. The three chapters that follow this preface, plus the preface itself, constitutes my Ph.D. dissertation at the Department of Economics at the European University Institute (Florence, Italy).

There has been several people that, during these four years, have been of key importance to fulfilling the task of writing my PhD. thesis. I would like to start thanking my mother and brother, who, at first upset with my decision of moving to Italy, have always been available and ready to cheer me up in the hardest moments. This thesis is dedicated to them.

My supervisor, Professor Karl Schlag, has been of invaluable help for accomplishing the task of writing my Ph.D. thesis. His guidance during this four years have made me understand this profession in depth and, although I still have much to learn, has turned me into a great young academic with very high potential. There has been two other professors that I believe deserve special mention. Professor Pascal Courty has always been a great asset as a person with whom to discuss ideas from a different perspective. While Professor Schlag works on the theoretical side of our science, Professor Courty works on the applied side. This combination has been extremely useful to me as for any idea I came up with I always had two different yet equally valuable opinions.

Professor Larry Samuelson was my mentor during my visit to the University of Wisconsin-Madison from September 2006 to December 2006. During this period Professor Samuelson acted as my main advisor and for that I am indebted with him. Professor Samuelson was of great help for my development as a theoretical economist. Most of the second chapter of this dissertation was written while I was visiting the University of Wisconsin-Madison under the supervision of Professor Samuelson.

Other colleagues and friends have been of very valuable help and support for me during this four years. These include my friends here in Florence as well as my friends back in Alicante. You know who you are, thank you so much for everything.

In the remainder of this introduction I introduce, in an informal way, the three chapters that are to come. Hence, the next three subsections contain a non-technical summary of my Ph.D. dissertation that are aimed readers that are not familiar with the economic literature in general and with game theory in particular.

Summary

Chapter 1 - Friendship Selection[1]

In the first chapter of my dissertation I aim at understanding some of the most well known and recognized phenomena about friendship relations between human beings. Before getting deeper into these phenomena, however, let me start by saying few a words about the relationship between friendship relations and economics.

Economics is not the science of money. It is the science aimed at studying decision making in situations involving scarce resources. These types of decisions range from which university degree to study to where to go on our next holidays. It is natural to think about money when thinking about economics as money is a scarce resource that we constantly use in our everyday life.

Many economic theories exists on how to optimally allocate our time between work and leisure. Furthermore, many theories explain such things as how agents make job decisions, how many years of education to acquire, or whether to invest now in a pension fund or to wait for better fundamentals. However, very few theories exist that explain how the scarce resource of leisure is employed by agents. In particular, no economic theory exists on how people use their leisure time to choose with whom to have friendship relations. Given the relevance in society of the social contacts in general and friends in particular I study how friendship arises in a network of people. The first chapter of my dissertation is then aimed at understanding how friends are selected given the impossibility of being friends with everybody as friendship relations involve costs (effort, time, etc.).

I model friendship as repeated interaction between a group of agents. Interactions between the group of agents is modeled as a two-step decision. First, each agent decides with every other agent whether to join in a relationship or not. Joining a relationship involves a benefit that is not player specific and a cost that is player specific. This benefit is interpreted as the joy a player gets from being in a relationship while this cost is interpreted as the degree to which agents needs help in a relationship. Second, for every two agents that decided to be joined together in a relationship, they decide whether or not to help the other. Helping involves a cost for the person providing help but a benefit for the person receiving help.

If two agents do not agree on joining a relationship, they remain as strangers. If two agents decide on joining a relationship but are not helping each other, we say they are mates

[1]Many thanks to Karl Schlag, Pascal Courty, Anna Orlik, Ilan Eshel, Itai Agur and Sanne Zwart for very useful comments and discussions on this chapter. I also would like to thank the seminar participants at the European University Institute and University of Wisconsin-Madison.

because although having a relationship, this relationship is not strong enough to consider them as friends. Finally, if two agents decide on joining a relationship and are helping each other, we say they are friends.

As just mentioned in the paragraph above, each of the agents in the group is characterized by the degree to which they, if joined in a relationship with the other agent, need help. Hence, there are agents who constantly, and in each of its relationships, need help to a high degree and there are other agents who will barely need help at all. Players with a high degree of needing help prefer not to join a relationship over being mates with someone, as being in a relationship might involve a cost higher that the benefit from being in a relationship. Players with a low degree of needing help prefer to have a relationship of either of the two types (mates and friends) over not joining the relationship at all since the benefit from being in a relationship, independently on whether they receive and provide help, are greater than the cost of needing help. The degree to which a player needs help is what makes her different from the other players. We say that two agents are similar if they have a similar degree of needing help.

The decision of helping an agent with whom one is having a relationship is a type of Prisoners' Dilemma problem. Agents like to be helped as it involves benefits to them, however, since helping is costly, agents also prefer not to offer help. In this cooperative setting I find the following: for agents who need help to a relatively small degree, a friendship relationship is possible only if the differences in the degree to which each of this players need help is similar. That is, in this case similarity, also known as homophily, plays the key role in determining if two agents can become friends or not. This is because if an agent with a low degree of needing help joins in a relationship with an agent with a high degree of needing help, the low degree agent might have an incentive to stop offering help to the other agent and move their situation to a mate relationship.

However, when one of the players has a high degree of needing help, high enough so that a mate relationship is less beneficial than no relationship at all, then the difference between the degrees of needing help of each agent plays no role in determining whether the two agents can become friends or not. In this situation the incentives of breaking the friendship relation of the low degree agent might be smaller than if the other agent was also a low type. This is because if a friendship relation with the high type is broken, then these two agents will become strangers as the high type prefers not to be in a relationship over being in a mate relationship. Hence, the costs of breaking the relationship for the low degree type are high as breaking a relationship will mean not having any relationship at all. In this situation the key for determining if two player can become friends is the profit each player is getting from the relationship independently of the difference of their degrees of needing help.

Once the role of homophily in a friendship relation is understood, I move to study how friendships are selected when players have a time constraint by which they can only have a fixed maximum number of friends. Hence, the question now is not who can be friends with who, but rather the following: among those with whom an agent could be friends, which ones are actually going to become this agents friends? That is, if the possibility for two agents of becoming friends is not high enough, there is now a process for selecting friendships. I model the selection process as a random event by which at every moment only one agent is allowed to offer new relationships. This random selection process represents the fact that in the real world friendship relations arise as a result of a complex meeting-new-people process.

I find that unless all agents in the population are different enough it is impossible to predict which friendship relations will arise in equilibrium. Given that there exists a certain degree of substitutability between agents, if two or more agents are similar in that they have a similar degree of needing help, then the random process by which people make new friendship will determined the equilibrium outcome. I also find that the out of equilibrium length of a friendship relation is directly influenced by the degree to which agents discount the future. Impatient agents tend to have shorter friendship relations, an observation that is in line with empirical studies.

In the last part of this chapter I present robustness checks of the model and relate my results with the empirical literature on friendship. Among other findings, I find the model to be robust to different specifications of the cooperative strategies of the agents. I also illustrate how my results match empirical evidence reported on friendship.

Chapter 2 - Learning within a Markovian Environment[2]

In this chapter I explore human decision making in a situation where there are two alternatives and the outcome or benefits from each of these alternatives is unknown to the agents. Agents are faced with the situation of making a choice from a set of alternatives repeatedly over time. Although oblivious of the payoff players get from making the choices, they can learn from their past experiences or from observing the choices of other agents. This decision problem is faced by many of us in our everyday lives: such as whether to buy a PC or a Mac, whether to have fruit or a cake as a dessert in a restaurant, or whether to watch an action movie or a romantic movie at the theater.

The outcome of each of the two available choices depends on a random, non-stationary,

[2]Part of this chapter was written during my visit to the University of Wisconsin-Madison. I am grateful to the faculty at UW and the participants in the Theory Lunch. I would like to thank Karl Schlag and Larry Samuelson for useful discussions and comments. I would also like to thank Mark Le Quement for useful comments and the seminar audiences at the University of Alicante and at the European University Institute.

variable: the state of nature. This non stationarity is represented in the model as a Markov chain. That is, the probability of being at a given state of nature tomorrow depends on which state we are in today. To illustrate this case consider the simple situation in which a broker has to decide whether to buy or sell a given stock. The state of nature in this case is whether the stock goes up or down. I assume that the probability of having the stock going up is random and depends on whether the stock went up las period or not.

Agents are given a set of alternatives where to choose from but know nothing about the outcome of these alternatives. In particular they also ignore the fact that a Markov chain is governing the payoff of each of the available choices. Agents then learn which of the alternatives is better by repeatedly facing the same situation. They learn from their own payoff experiences and/or from the experiences of the others. The way agents learn is assumed to agree to the principle of reinforcement, whereby alternatives that where more successful in the past are more likely to be chosen. Moreover, I assume that there is a memory effect meaning that more recent payoff experiences weigh more in the present decision than less recent ones.

I study two different informational settings, one where after each choice, an agent knows the payoff he got and the payoff he would have gotten had he chosen the other alternative (foregone payoffs are observed). This setting can be interpreted as if there was some information transmission mechanism, like word-of-mouth, that makes agents aware of the performance of all the available options. Furthermore, I also study decision making in a another setting where after each choice each agent only knows the payoff he gets from his alternative (foregone payoffs are not observed).

Given these two informational settings just described, I study how the choices of a population evolve when agents learn according to the reinforcement principle. The results I find are intriguing and pose explanations to real life behavior that seemed puzzling before. In the informational setting where foregone payoffs are observed I find that the behavior of the population converges to a behavior very similar to the probability matching behavior. Probability matching behavior is better understood with an example. Imagine a pot with 100 balls, of which 60 are red and 40 are black. Suppose that I draw 5 balls with replacement and I ask you to guess the colors. Since at each draw there are more red balls in the urn, the probability of getting a red ball is higher than the probability of getting a black ball. Therefore, it is optimal to guess that the color will be red in all the 5 draws. However, in this type of decision problem it has been observed in experiments with human subjects that agents tend to guess red 3 times and black 2 times. Guessing red 3 out of 5 times means guessing red 60% of the time, which is exactly the proportion of red balls that he pot contains. Similarly, guessing black 2 out of 5 times means guessing black 40% of the time, again the fraction of black balls in the urn. The choice pattern whereby the frequency by which each choice is made equals the frequency by which each choice is optimal is known as probability matching.

As mentioned already, when foregone payoffs are observed, the behavior of the population converges to a behavior very similar to that in probability matching. The reason being the following: when foregone payoffs are observed, reinforcement is translated into being more likely to play tomorrow the action that was best today. Hence, each action will be reinforced a fraction of the time equal to the fraction of time that action is actually the best one. The result follows that in the long run each action is played a fraction of time that matches the fraction of times that action is actually the best one. As already pointed out, probability matching is not an optimal behavior. Reinforcement behavior, although sounding a plausible way of learning and being empirically relevant, may lead to suboptimal choices.

When foregone payoffs are not observed a very different behavior is observed. In this case choices of the population converge to a unique alternative, as opposite to the other informational setting, no mixing is observed. The alternative selected is the long run optimal one only if alternatives are different enough. That is, if there is not a significant difference in the long run payoff difference between alternatives then it is possible that the population ends up choosing always the suboptimal choice. This suboptimal lock-on could be observed if, for instance, the long run inferior alternative happens to be better for a long period of time. In this case, reinforcement will lead players to play the long run inferior alternative to the point where no one observes the performance of the other alternative. At this point, and once the long run optimal alternative happens to be better, no one notices it under the informational setting where foregone payoffs are not observed. Hence, no agent ever plays again the alternative that is best in the long run. This behavior explains why, for instance, we sometimes observe inferior products dominating the market as happened when the Betamax video format took over the VHS tapes.

I round off the analysis by showing, among all the possible ways of learning following the reinforcement principle, which ones are efficient in the sense that, when used by the agents, they end up choosing the long run best alternative. In this respect I show that efficient learning under reinforcement requires players to disregard the information from observing foregone payoffs, in case this information is available, and exhibit very cautious learning. Cautious learning implies that decisions are not very responsible to the feedback from the environment. Reinforcement learning, although a learning principle employed by real life subjects, does not use information optimally. Due to this, too much information can be harmful. This is why I get the striking result that using less information, by disregarding foregone payoffs, is the right thing to do if an agent learns according to reinforcement.

Chapter 3 - The Effects of the Market Structure in the Adoption of Evolving Technologies[3]

In the third chapter of my Ph.D. dissertation I study how does the market structure affects the speed at which new technologies are adopted. By the market structure I mean how the market power is shared between suppliers, firms that sell new technologies, and buyers, firms that buy new technologies.

In the model I present there is a set of firms, the suppliers, that have the rights of selling new technologies to another set of firms, the buyers. Technologies are constantly evolving and all suppliers have access to the same set of technologies. The role of the firms in the supply side is then to put a price to these technologies.

The different market structures differ in how the market power is shared. In the first market structure I consider, all the market power lies in the supply side. In this setting there is only one firm selling technologies and many firms in the demand side that compete for buying these technologies. In the second setting I consider, the opposite happens. The buyer holds all the market power. In this setting there are many firms selling technologies but only one firm interested in buying a new technology. Hence, in this setting firms in the supply side compete to selling the technology to the only firm on the demand side. The third setting considered is characterized by the fact that both supply and demand have a share of the market power. In this setting there is one firm on each side of the market and both firms compete to extract as much surplus as possible.

A point worth noticing is that in this chapter I aim at understanding how the different market structures can have an effect on the speed to which new technologies are adopted. It is usually argued that differences on speed of adoption are the result of the difference in the underlying evolution of the technologies themselves. That is, different paces of adoption of technologies are observed because not all the technologies evolve in the same way: the computational capacity of computers is greatly improved every year while it takes more than 5 years to see significant improvements in order technologies like cars. I propose an explanation of the differences in the timing of adoption that is based solely on the structure of the market where the technologies are sold. Hence, we prove that differences of speed can be explained in term of different market structures and independently on the nature of the technology at hands.

In my results I find some striking features about the speed of adoption of new technologies.

[3]I would like to thank Pascal Courty, Karl Schlag, Omar Licandro and Fernando Vega-Redondo as well as the seminar audience of the Micro Working Croup at the European University Institute for useful comments and discussions.

First, I show that when only one side of the market holds all the power and independently of which, the adoption of technologies is expected to occur at the same speed. That is, whether the suppliers or the buyers hold all the market power does not affect the pace of adoption. This is because as soon as the market power lies on only one side, the total surplus of the economy is maximized. The only actual difference between the two settings lies on which side of the market actually gets to keep all this surplus. I prove that actually the total surplus of the economy is maximized by comparing the speed of adoption when only one side of the economy holds all the market power with the Nash bargaining solution between supply and demand. In order words, if a social planner was to decide what is the optimal timing of adoption, the result will be the same as if we give all the market power, or bargaining power, to one side of the market.

When there is competition between suppliers and buyers in the sense that each side of the market competes for the total surplus of the economy, we find that adoption occurs at a slower pace. This surprising result implies that competition between suppliers and buyers is actually decreasing the total surplus of the economy as compared with the cases where only one side of the market holds all the power. Hence, from the social point of view, competition can be harmful in that the adoption of new technologies is delayed.

After this main result I then present some comparative statics results on how the speed of adoption is affected in the three market settings when the parameters determining the evolution of technologies change. I find that the timing of adoption is more sensible to the process determining the evolution of technologies when there is competition between supply and demand. That is, markets where only one side holds all the power should exhibit more similar behavior independently on the evolution of technologies than in markets where there is competition.

Another interesting fact revealed by the numerical exercise is that as the interest rate raises, the differences in timing of adoption between the setting where there is competition between supply and demand and the settings where only one side holds all the power converge to zero. This means that in high interest rate economies the market structure has less effect in the adoption of new technologies than in low interest rate economies.

Chapter 1 - Friendship Selection

1.0 Abstract

We model the formation of friendships as repeated cooperation within a set of heterogeneous players. The model builds around three of the most important facts about friendship: friends help each other, there is reciprocity in the relationship and people usually have few friends. In our results we explain how similarity between people affects the friendship selection. We also characterize when the friendship network does not depend on the random process by which people meet each other. Finally, we explore how players' patience influences the length of their friendship relations. Our results match and explain empirical evidence reported in social studies on friendship. For instance, our model explains why troublesome subjects have fewer friends.

1.1 Introduction

Social relationships represent one of the most basic needs of human beings. They arise quickly between subjects in any kind of environment and they condition the behavior of the subjects involved. Different degrees of social relationships can exist between individuals: family members, work mates, partners, friends, etc. Among all of them, friendship relations represent one of the most intriguing aspects of social relationships. While every person can identify his friends if asked, it is difficult to find a proper definition for what friendship means.

The most commonly mentioned characteristics of friendship relations are: helping, reciprocity and a limited number of friends (See, for example, Hruschka and Henrich (2004), Silk (2002), Hallinan (1979), de Vos and Zeggelink (1997), van de Bunt, van de Duijn and Snijders (1999) or Zeggelink (1995)). Mutual help in a friendship relation implies that friends help each other in case of necessity. The exchange or reciprocity means that people expect from their friends a similar attitude to the one that they take towards them. Finally, a limited number of friends simply means that subjects do not have as many friends as they would like since keeping up friendship relations takes time and effort.

The present paper presents a model that tries to reproduce these three facts: helping, reciprocity and a limited number of friends. The interactions between a group of players are modeled in the following repeated setting: each period every player has to decide whether to perform an activity with each of the other players in the population, one activity per pair of people. The activity might be going to the cinema, going on vacation, doing sports together, etc. Each player is characterized by an exogenous degree of needing help. This means that in each of the activities each player performs, she needs some help. This might be because she needs money, has had an accident, is sad, etc. After both players decide to do the activity together, they have to decide simultaneously whether to help the other or not. Helping involves a cost for the player who provides help but also a benefit for the player receiving help. In game-theoretical terms, we model the decision of helping as a cooperation game of a class of prisoners' dilemma game. We called this game the Helping Game. The degree to which a player needs help is exogenous, common knowledge and heterogeneous among the players. Finally, each player is able to provide help a limited number of times per time period. This represents the fact that helping is time-consuming. If two players are performing the activity and helping each other (playing the cooperative equilibrium) they are called friends. If they are performing the activity but not helping each other, they are called mates. If they are not performing the activity they will be called strangers.

As mentioned above, our aim is to construct a model that, based on helping, reciprocity and a limited number of friends, is able to explain some of the phenomena that we observe in the real world friendship relations. From the preceding paragraph it is clear how we make use of the helping and the limited number of friends. To implement reciprocity, the strategies that we use for supporting cooperation (providing help) will be Grim Trigger. According to Grim Trigger strategies, a player will keep on providing help to another player as long as this other player is also providing help to her. Because Grim Trigger strategies do not allow for forgiveness, in section 1.5.1 we check for the robustness of the results when instead players use Tit-for-Tat for supporting cooperation.

In our three most important results we explore the three following issues: role of similarity in friendship relations, uniqueness of equilibrium of the friendship network and length of the friendship relations. First, we manage to explain the role of similarity in friendship relations (similarity in the friendship context is often referred to as homophily). It has been reported in empirical studies that similar people (same hobbies, race, etc.) are more likely to have friendship relations. For example, Marmaros and Sacerdote (2004), using the number of emails exchanged between students from Dartmouth College, found that similarity in age, geographic closeness, race and interests increase the likelihood for two people to become friends. However, strong friendship relations between very different people can exist. Section 1.3 suggests a solution to this fact. According to the model, when two people are mates this means that they are having a relationship but their relationship is not strong enough to consider them friends. We find that similarity matters only if a "mate" relationship between two people is possible. Otherwise, similarity will play no role in determining if these two people can become friends and the only factor determining if a relationship is possible are the profits from the relationship each party gets.

Second, we show that it is in general impossible to predict the friendship relations that will prevail within a group of people in the long run. In particular we show that the equilibrium will depend on the order in which people meet each other. This order is modeled as a random process. In game-theoretical terms, the equilibrium is history dependent and the history follows a random process. What is interesting about the model we present is that we give two precise explanations for why the equilibrium may be history dependent. First, if people belonging to the group are not different enough in terms of their degree of needing help, then a certain degree of substitutability between people exists. In this case, the random process by which people meet each other will play a role in the final outcome of the process. Second, if no mechanism or social norm exists by which agents punish those agents who 'betray' their friends, then the random process by which people meet each other will again play a crucial role in determining the final outcome. In section 1.6 we relate this and other results with some empirical facts about friendship relations.

Many sociological, physiological and anthropological papers have modeled the process of friendship formation. For example, in a paper by Zeggelink (1995), friends have a fixed desired number of friends and each player is defined by a dichotomous variable (they are either type-1 player or type-2 player). Each player tries to have the desired number of friends and to maximize the similarity in his type with the type of his friends. The author performs simulations and finds that the players tend to group with the others of the same type. The taste for similarity is exogenously imposed whereas we make no assumption on this respect. In this respect, Hruschka and Henrich (2004) developed a model in which in each period players can choose with whom they want to play a prisoners' dilemma game. The model is focused on the evolutionary biological point of view of the cooperative relations. That is, they focus on the differences between the survival rates of cooperative players and selfish players.

The model presented is different also from the economic models of social networks pioneered by Jackson and Wolinsky (1995) and Bala and Goyal (2000). It differs from Jackson and Wolinsky (1995) in that in our model the payoff of the players is not determined uniquely by the state of the friendship network but also by the actions of the players against those with whom they do not share a friendship relation. The model presented, on the other hand, differs from Bala and Goyal (2000) in that when two players share a link, they then play a cooperative game and not a coordination game.

To our knowledge, only two papers examine the issue of social networks when players play a cooperative game. These are Lippert and Spagnolo (2005) and Vega-Redondo (2005). The first one focuses on the information transmission about the defectors in the network and on the different punishment mechanism for supporting cooperation. On the other hand, Vega-Redondo (2005) explores the amount of cooperation that will emerge in the network when the environment suffers from aggregate shocks to payoffs.

The rest of the paper is organized as follows. In Section 1.2 we develop the model. Section 1.3 explores the simplest case in which the population consists of only two players. Section 1.4 extends the model for more than two players. In Section 1.5 we discuss the robustness of the results and the assumptions as well as present some extensions. We relate our results with empirical findings on friendship relations in Section 1.6. Finally, Section 1.7 concludes.

1.2 The Model

1.2.1 Informal Discussion

Assume a population $\mathbb{N}$ of n players. Each player in the population is characterized by the degree to which she needs help $p \in (0, 1)$. Every time period $t = 1, 2, \ldots$ a player, say i, is selected by nature. This player can make 'phone calls' to the players with whom she intends to form a relationship. There are two types of relationships: friends and mates. When two players are not in a relationship, we say they are strangers. These three different states, friends, mates and strangers, are explained in more detail below. When player i calls player k, then players i and k decide simultaneously and non-cooperatively whether to enter into such a mutual relationship or not. Relationships carry a benefit to both players but also involve the the cost of additional additional cooperation. In any relationship, each party needs some help and the degree of help needed differs among players. Part of the relationship is an observable decision of whether or not to cooperate in the sense of providing help. So when two players have decided to join a mutual relationship they then non-cooperatively simultaneously decide whether or not to help the other. When both decide to help the other then we speak of friends; otherwise we speak of mates. If the relationship does not even arise because at least one of the two parties does not want to participate in the relationship then we speak of strangers.

The maximum number of times per period that a player can offer help is limited to $m \in \{1, \ldots, n-1\}$. This constraint reflects the fact that providing help is costly in terms of time. It implies that a player can only have at most m friends at every moment of time.

We limit the set of possible strategies of each player as follows. Only in a period in which a player makes or receives phone calls she can change her plan of action. Otherwise, she plays as she decided at the time of the last phone call. Only two types of plans of actions or strategies are considered: the cooperative and the defective.

In the cooperative plan the player acts as in Grim Trigger. A friendship is suggested, which means that, first, a relationship is suggested and then if the other agrees the first player suggests helping the other when in need of help. The Grim Trigger plan also specifies what to do if the other does not want to be friends or even to be mates: If a friendship does not arise then the player chooses whatever is best for herself in the one-shot game. Depending on the payoffs, this can be to not accept any relationship or to suggest forming a relationship. Finally, the Grim Trigger plan protects against later defections by proceeding, in case of a defection, as if a friendship did not even arise in the first place.

In the defective strategy the player rejects or breaks the friendship relation. If both players had a friendship relationship in the previous period then she breaks it. The player does so by not providing help to the other player but still receiving the benefits from the other player's helping her. The two players had not previously been in a friendship relationship with each other, she rejects a possible friendship relation and plays whatever is best for her in the one-shot game.

1.2.2 Formal Presentation

Assume a population $\mathbb{N}$ of N players that discount the future at a common rate $\delta \in (0,1)$. Each player $i \in \mathbb{N}$ is characterized by the degree to which she needs help $p_i \in (0,1)$. Every time period $t = 1, 2, \ldots$ every placer faces a one-shot game with every other player in the population. In this game, which we call the Relationship Game, the two players have to decide simultaneously whether to link l (suggest a relationship) to the other player or not n. If both players agree on having a relationship together, they then simultaneously decide whether to help H the other or not N. The subgame that starts after both players have decided to join a relationship (both chose l) is called the Helping Game.

The payoff scheme works as follows. First, at any given period the payoff of each player is the sum of the payoff she gets from playing the Relationship Game with all the other players in the population. Within each time the Relationship Game is played, if one of the players decides not to link with the other, then they both get 0 payoff. If both players play l but are not helping each other, then player i gets $A - p_i$ and player k gets $A - p_k$. Hence, they get a fixed amount $A \in (0,1)$ minus the degree to which they need help. If player i is helping player k but not the other way around, then player i gets $A - p_i - cp_k$ and player k gets $A - p_k + xp_k$. That is, player i has to pay the cost cp_k with $c \in (0,1)$ for helping player k and player k receives a benefit xp_k with $x \in (0,1)$ because of being helped by player i. The payoffs in the other situations follow the same logic. We assume $A \geq 1 - x$, so that being helped without providing help is always weakly preferred to not being linked. The Relationship Game is shown in extensive form in Figure 1.1.

For the reader's convenience, we present below the normal form of the Helping Game, which is the subgame that starts after both players have decided to join a relationship.

Rivas, Javier (2008), Cooperation in Repeated Games, Bounded Rational Learning and the Adoption of Evolving Technologies
European University Institute 10.2870/26189

Figure 1.1: Relationship Game

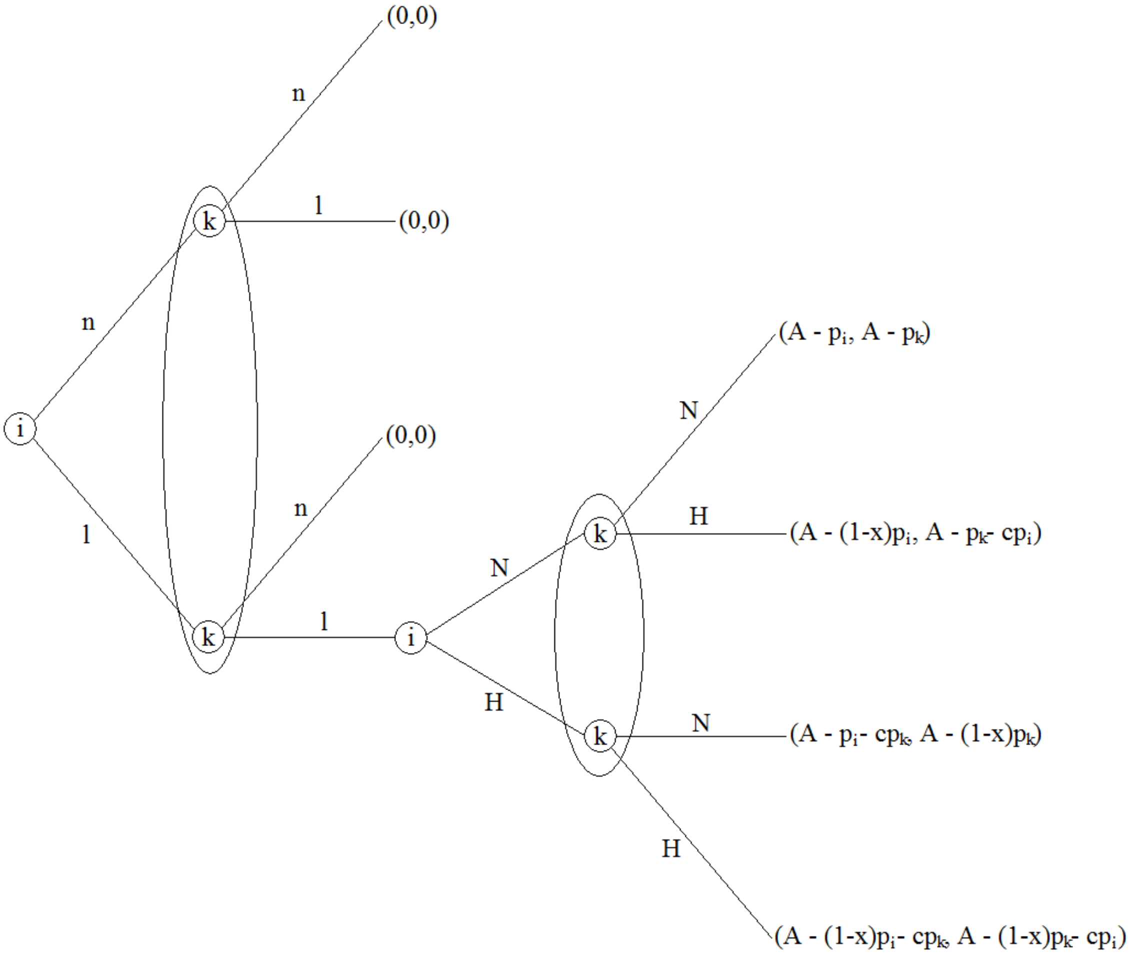


Table 1: Helping Game

	H	N
H	$A-(1-x)p_i-cp_k, A-(1-x)p_k-cp_i$	$A-p_i-cp_k, A-(1-x)p_k$
N	$A-(1-x)p_i, A-p_k-cp_i$	$A-p_i, A-p_k$

The following proposition characterizes the Nash equilibria of the Relationship Game between players i and k. Whenever we write $((x, y), (x', y'))$, this means that player i plays (x, y) and player k plays (x', y') with $x, x' \in \{l, n\}$ and $y, y' \in \{H, N\}$.

Proposition 1. *In the Relationship Game for any players $i, k \in N$:*

- *Nash equilibria: For each $p_i, p_k \in (0, 1)$, $((n, y), (n, y'))$ are Nash equilibria with $y, y' \in \{H, N\}$. If $p_i, p_k \leq A$ then $((l, N), (l, N))$ is also a Nash equilibrium.*
- *Sub-game perfect Nash equilibria: For each $p_i, p_j \in (0, 1)$, $((n, N), (n, N))$ is a sub-game*

perfect Nash equilibrium. If $p_i, p_k \leq A$, then $((l, N), (l, N))$ is also a sub-game perfect Nash equilibrium.

As mentioned above, if two players are playing (l, H) repeatedly against each other, we define them as friends. If they are playing (l, N) repeatedly against each other, they are mates. If two players play (n, N) repeatedly against each other, we say they are strangers. The words *betrays* and *betrayal* are used throughout the paper; below we write the formal definition of betray and betrayal.

Definition 1. *We say that a player i betrays another player k if they were friends in the last period, but in the current period i, still having a link with k, does not provide her with help. That is, they both played (l, H) against each other in the past round but i switches to play (l, N). We say there has been a betrayal between two players if they were friends in the last period but at least one of them betrays the other in the current period.*

As is well known from the Folk theorem in repeated games, infinitely many strategies can form Nash equilibria. Hence, we shall restrict the strategy space of the agents to make the model tractable. In our model, as already mentioned, players are only able to have two types of plans, the Cooperative Plan and the Defective Plan.

Cooperative Play according to Grim Trigger (defined below).

Defective Play (l, N) if you and the other player played (l, H) in the last round; play your weakly dominant strategy in the Relationship Game otherwise.

As can be infered from the Relationship Game and Proposition 1, whenever we write play your weakly dominant strategy it implies *play (n, N) if your degree of needing help is smaller than A; play (l, N) otherwise.*

Definition 2. *Define the Grim Trigger strategy for player $i \in \mathcal{N}$ played against any player $k \in \mathcal{N}$ as follows:*

- *If a play in any past period against k was either $((l, H), (l, N))$ or $((l, N), (l, H))$, then play your weakly dominant strategy.*
- *Otherwise, play (l, H).*

The Grim Trigger strategy prescribes helping unless there has been a betrayal in the past between the two players. Note that by the way we define the Grim Trigger strategy players are protected against possible deviations from the other player when both are playing (l, H).

Our choice of these two specific plans is motivated by two facts that we explain now and formally prove below. First, if players can play according to the Defective Plan, the

Cooperative Plan is the best (i.e. for more parameter values) way of supporting cooperation that does not involve the use of dominated actions. Second, the Defective Plan is the best possible deviation against the Cooperative Plan. Nevertheless, in section 1.5.1 we check for the robustness of the results when instead of Grim Trigger players are allowed to use the Tit-for-Tat, which allows players to forget deviations that occurred in the past.

Proposition 2. *Given the possibility of playing as in the Defective Plan, there is no strategy that does not involve the use of dominated actions and that can support the outcome* $((l, H), (l, H))$ *as a part of an equilibrium of the repeated Relationship Game for a bigger set of parameter values than the Cooperative Plan. Furthermore, the Defective Plan is the best possible deviation against the Cooperative Plan.*

Proof. See Appendix 1.A.1. □

We constrain the agents to provide help at most $m \in \{1, ..., n-1\}$ times per period and, hence, each player can have at most m friends in a given period. When each player is to decide with whom she can set up a friendship relation, she will do so in a pair-wise fashion. This means that, if i is to decide whether she can set up a friendship relation with k, i will take this decision as if there were no more players in the population. That is, as if $\mathcal{N} = \{i, k\}$. However, i will still take into account the upper-bound m. Hence, if i already has m friends, she will take into account that before setting up a friendship relation with k she must break one of her already existing friendship relations. On the other hand, if i has less than m friends, her decision of whether to set up a friendship relation with k will be taken as if $\mathcal{N} = \{i, k\}$. We make this assumption to make the model tractable. This assumption can be thought as a bound in the rationality of players by which they cannot fully take into account all the interactions that occur among all the players in the population when taking their decisions.

We refer to a friendship relation between i and k as pair-wise sustainable if the friendship relation is possible when $\mathcal{N} = \{i, k\}$. Thus, Proposition 3 in the next section, where we consider the two-player case, is telling us which friendship relations may exist in equilibrium.

Players are allowed to revise (or update) their strategies in the following way. Each period a player, say i, is selected by nature. This player can make "phone calls" to the players with whom she wants to play the Cooperative Plan. Player i then updates her strategies as follows. She plays the Cooperative Plan with whom she calls and plays the Defection Plan with the rest. The players who get a call from i can update only the plan or strategy they are playing against i. However, if a player gets a call but is already providing help to m players, then she can switch to play the Defective Plan against one of her friends in order to be able to play the Cooperative Plan with the player that called her. The rest of players do not update their strategies in any manner (for a more formal definition of the dynamics the reader is referred to Appendix 1.A.2). The fact that only one player per period is allow to make a phone call reflects the fact that real-world relationships don't happen instantaneously; rather, they are the result of a 'meeting people' process.

1.3 Two-Player Game

Assume that the population $\mathcal{N}$ consists of only two players, i and k.

Proposition 3. *A friendship relation between players i and k can be supported in the repeated Relationship Game when both players use Cooperative Plan if and only if the following holds:*

- *if* $p_i, p_k \leq A$ *then* $\frac{c}{\delta x} \leq \frac{p_i}{p_k} \leq \frac{\delta x}{c}$,
- *if* $p_i, p_k \leq \frac{A}{1-x}$ *but either* $A < p_i$ *or* $A < p_k$, *then* $A - p_i + xp_i - \frac{c}{\delta}p_k \geq 0$ *and* $A - p_k + xp_k - \frac{c}{\delta}p_i \geq 0$.

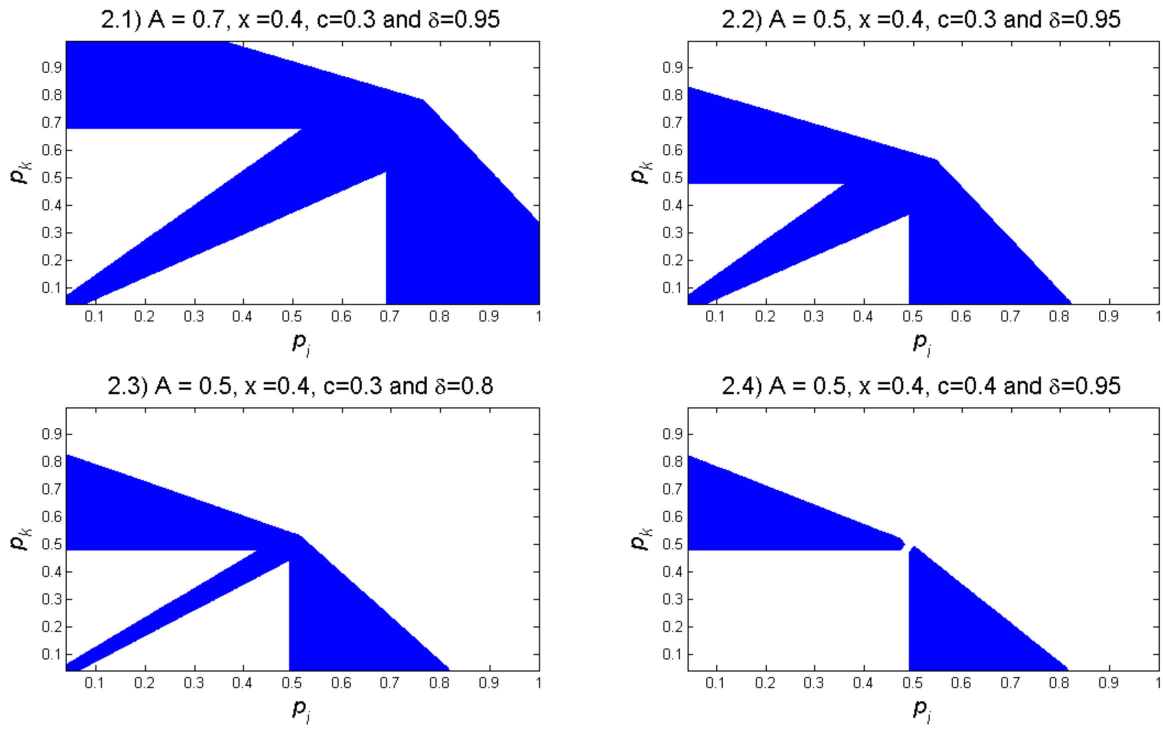
Proof. See Appendix 1.A.1. □

From Proposition 3 we conclude the following. First, if both p_i and p_k are smaller than a A, then the friendship relationship can only be supported if the relative difference between respective probabilities of needing help is sufficiently small. A player with low degree of needing help (less than A) will not accept a friendship relation with a player whose degree of needing help, although also smaller than A, is very different from his.

Second, if at least one of the players needs help with a degree higher than A and both players' need of help is below $\frac{A}{1-x}$, then players no longer care about the relative difference in probabilities of needing help but about their absolute values. In this case, as long as the inequalities $A - p_i + xp_i + \frac{c}{\delta}p_k \geq 0$ and $A - p_k + xp_k + \frac{c}{\delta}p_i > 0$ are satisfied, the friendship relationship can be supported. The relevant implication of this case is that player i cares now only about the balance between the costs and benefits of the relationship instead of, as in the previous case, being similar to the other player.

To have a better understanding of the implications of Proposition 1 we present Figure 1.2. It plots when, for a given value of the parameters A, x, c and δ a friendship relation is possible between players i and k. So if the coordinate (p_i, p_k) is shaded it is because for the given parameters a player whose degree of needing help is p_i can be a friend of a player whose degree of needing help is p_k and vice versa.

Figure 1.2: Friendship Relations



The common interesting feature of these graphs is the existence of non-convexity between the area when both p's are smaller than A and the area when one of the p's is higher than A. The intuition behind this result is that a player with a small degree of needing help will not want to be linked with a player with a much smaller degree of needing help. This may happen because she is afraid that this player may betray her (i.e. deviate from playing the Cooperative Plan) since this person may prefer her only has a mate. On the other hand, if their p's are close enough or the other player's degree of needing help is high, then he will be willing to have the friendship relation with that other person because she knows that: (1) she needs the other player as much as the other player needs her and since both are getting positive profits from the relationship no one will have incentives to terminate it, and, (2) if they are not having a relationship they will not even be mates as one of the player's degree of needing help is higher than A. In other words, the loss if one betrays the other is too high (they won't even be mates) in this second case.

Figure 2.4 (when $c > x$) merits special attention. It shows that no friendship relationship between players with a low degree of needing help will arise. Since having a friendship relationship is not very profitable in terms of x and c, the low-degree of needing help players will only like each other as mates and not as friends. This happens because the likelihood of betrayal is too high. For the same reason, the relationship between low degree players and moderate degree of needing help players will be possible. Players with a moderate degree of needing help won't want to have mates, only friends or strangers. Therefore, in this latter case, the cost of betrayal is very high. This makes the relationship more likely to be supported.

Proposition 3 states that the relevance of similarity for friendship selection differs accordingly to the type of players. For some pairs of players the similarity with their friends will matter and for some other players the similarity will be irrelevant. The thing that will matter in this latter case will be the balance between costs and benefits from the relationship.

One more thing is worth underlining. A player with a very low p may be "marginalized" among the players with low p because she needs "too little" help. Summing up the results of Proposition 3:

- If both p_i and p_k are small, the relevant thing for a friendship to be possible is the relative difference between p_i and p_k.
- If either p_i or p_k is not small, the relevant thing for a friendship to be possible is the absolute value of p_i and p_k.

1.4 n-Player Game

For a clearer understanding of the dynamics of the model when the population consists of more than two players, we present example 1. The example is drawn in Figure 1.3, where each node represents a player and a line between two players represents the fact that those two players are friends.

Example 1. *The simulation is conducted for* $N = \{1, 2, 3\}$, $p_1 = 0.4$, $p_2 = 0.45$, $p_3 = 0.55$, $A = 0.5$, $x = 0.6$, $c = 0.3$, $\delta = 0.7$ *and* $m = 1$. *In this setting all the possible friendship relations are pair-wise sustainable. To check this we can apply the result in Proposition 3 to the present example.*

In the first period, player 1 *is selected by nature to make calls. Since* $m = 1$*, player* 1 *can only provide help to at most one player. Hence, because* $p_2 < p_3$*, she will call player* 2 *and both of them will switch to play Grim Trigger against each other.*

In the second period, player 1 *is again the one allowed to make calls. Since this time she has one friend, she makes different considerations. Because of the fact that* $m = 1$*, she now wonders if betraying* 2 *and setting up a relationship with* 3 *is better than keeping the* *friendship relation with 2. Betraying player 2 is profitable for player 1 because providing help is costly. Hence, it may happen that the lower payoff associated with a friendship relation with player 3 is compensated by the one-period gains from betraying player 2. This is exactly the case in this particular example. Therefore, player 1 will call player 3 and they will switch to play Grim Trigger against each other. Moreover, player 1 will play (l, N) against player 2.*

In period 3, nature selects player 2. Since player 1 betrayed player 2, the friendship relation between them is no longer possible. This is due to the unforgivingness property of the Grim Trigger strategy. Hence, player 2 will call player 3. Player 3 is in a similar situation to the one faced by player 1 in the second period. In this particular example, player 3, as did player 1, finds the betrayal profitable. Hence, player 3 betrays player 1 to set up a friendship relation with player 2.

An equilibrium has been reached. Since player 1 betrayed player 2 and player 3 betrayed player 1, no new friendship relations can arise in the network.

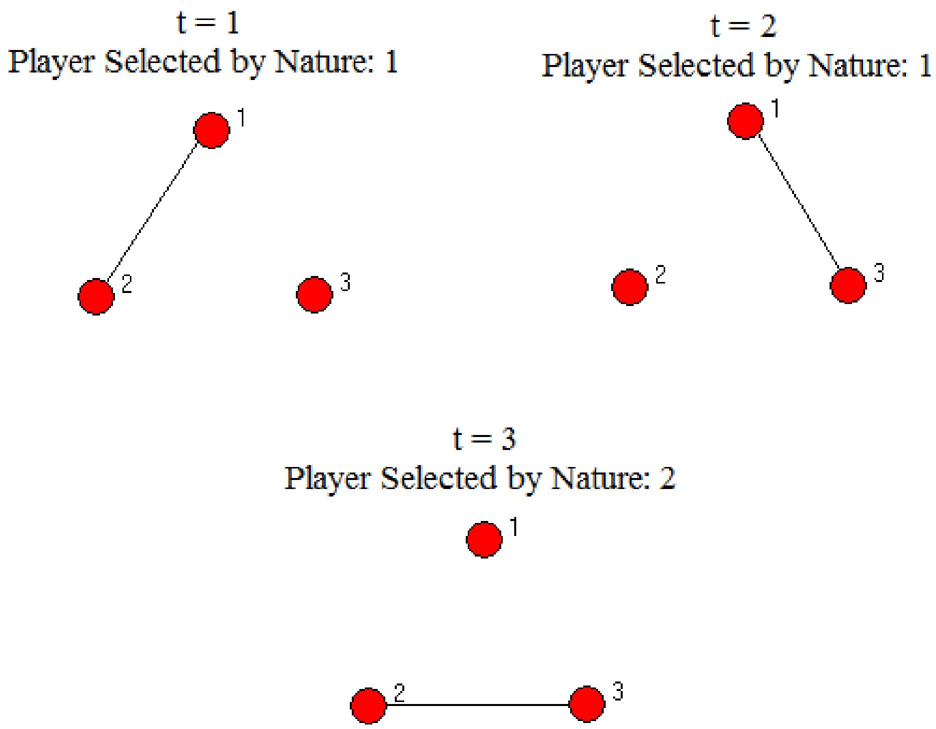
As we have seen, nature (or chance) plays an important role in determining which friendship relations can arise. If the players selected by nature were 2, 2 and 3 in this order, the equilibrium would have had players 1 and 3 as the only friends. This result together with other important ones is presented in the next subsection.

In the example above, the equilibrium in which players 2 and 3 are friends results only if 1 betrayed 2 (or vice versa) and 3 betrayed 1 (or vice versa). On the other hand, the equilibrium in which 1 and 3 are friends is possible only if 1 betrayed 2 (or vice versa) and 2 betrayed 3 (or vice versa). Therefore, the equilibrium in this case is history dependent.

The key to this result is that, if players 2 and 3 from the example above are similar enough, player 1 is not losing much by having a relationship with player 3 instead of with player 2. What loss there is can be compensated by the one-shot profits from betraying 2 today. Hence, the existence of a certain degree of substitutability between friends creates history dependence in the equilibria. This fact is exploited in Proposition 4.

Proposition 4 shows that convergence to equilibrium is always guaranteed, although the equilibrium can depend on the order in which players are selected by nature. This fact represents an important feature of friendship relations. The friendship relations that emerge in the real world are the result of a complex process of interactions between individuals in which unpredictable events may play a crucial role in the final outcome. The order of meeting people has important effects on one's long-term relationships. However, as we shall see below, there are some situations in which the convergence of the process to a uniquely determined equilibrium is guaranteed. These situations are: 1) when the players in the population are

Figure 1.3: Simulation



different enough from each other (Proposition 5), and, 2) when there exists some type of social rule by which betrayers are punished (Proposition 9).

Proposition 4. *The system converges with probability* 1 *to an equilibrium network architecture that can be history dependent. If* $A > c\delta$ *then there exists an* $\varepsilon > 0$ *such that if* $\min_{i,k\in\mathcal{N},i\neq k} |p_i - p_k| > \varepsilon$*, then the friendship network converges with probability* 1 *to a unique network architecture.*

Proof. See Appendix 1.A.1. □

Therefore, when players in the population are different enough, the process will converge to a unique equilibrium. In other words, the process has only one equilibrium that is not history dependent and the process will always converge to it. As mentioned before, there exists a certain degree of substitutability between friends. Hence, if players are different enough, no player will want to betray a friend to set up a relationship with a higher-degree-of-needing-help player. Once the substitutability between friends is eliminated, we can successfully predict the long-term friendship relations that will arise within the population. Note that conditions in Proposition 4 do not rule out the case where players with $p < A$ can have friendship relations between each other. Hence, even when players are different enough so that the friendship network converges to a unique equilibrium, similarity may play a role.

Another interesting feature of the model is that, when subjects are more patient, the duration of the friendship relations will tend to be longer. When a player is about to betray another one, she has to consider the fact that the betrayal will yield her a higher current payoff but possibly a lower future payoff (consider, for instance, the first betrayal in the case above). Hence, more patient subjects are less likely to betray another player to set up a friendship relation with a player who has a higher degree of needing help. Note that the decision of a player to betray one of her friends and to set up a friendship relation with a player who has a lower degree of needing help is independent of δ. This is formally stated in the following proposition.

Proposition 5. *Ceteris paribus, the length of the friendship relations depends positively on* δ.

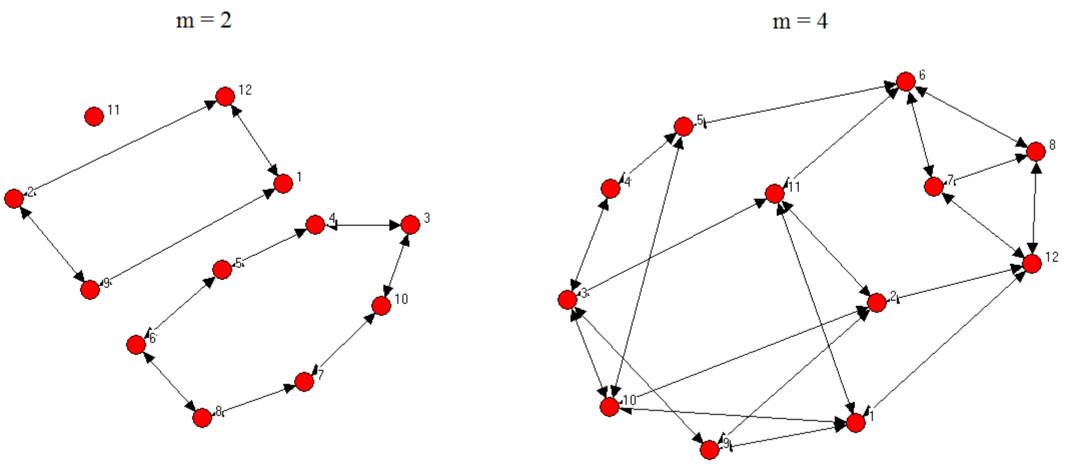
Proof. See Appendix 1.A.1. □

Now we present some more complex examples to get a better understanding of how the friendship network looks in equilibrium. Simulations are conducted for the same parameter values as in example 1. That is, $A = 0.5$, $x = 0.6$, $c = 0.3$, $\delta = 0.7$. We run simulations for two different population sizes, 12 and 19, and for two possible values for the maximum number of friends a player can have, $m = 2$ and $m = 4$.

Figure 1.4 show the result of the simulation for $m = 2$ (left-hand side) and $m = 4$ (right-hand side). The degrees to which each player needs help are given by $p_i = 0.05i$ with $i \in \{1, \ldots, 8\}$ and $p_i = 0.55 + 0.05i$ for $i \in \{9, \ldots, 12\}$. As we can see, when $m = 2$ two groups (components) are formed. These two components are not fully connected and exhibit the circle property. This is because each player is restricted to having at most two friends. However, if we allow them to have more than two friends, the two components merge and there is a slight increase in connectivity. Careful inspection of the graph at the right-hand side shows that we have two interconnected components, one with players $6, 7, 8$ and 12 and the other one with the rest.

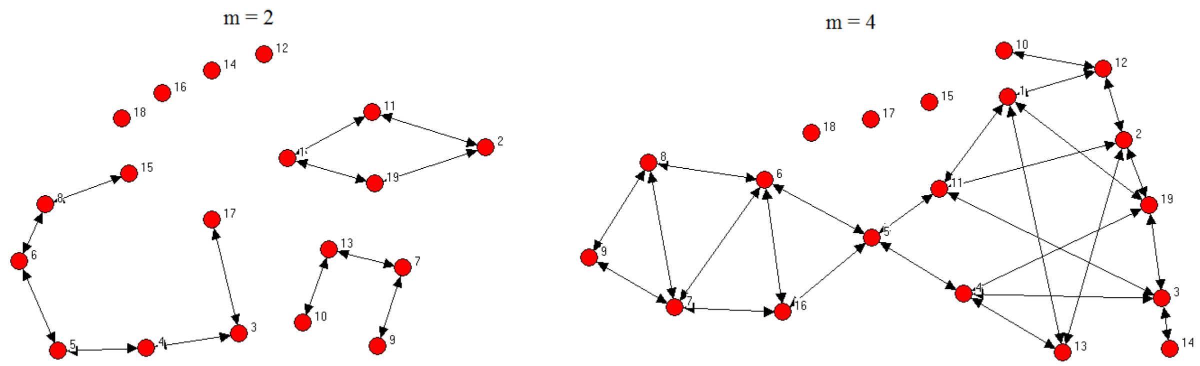
In Figure 1.5 we show a for the case where there are 19 players in the population and the degrees to which each player needs help are given by $p_i = 0.05i$ for $\{1, \ldots, 19\}$. When $m = 2$, we can see that there are three components plus four players that have no friendship relations at all. Again, once we allow the population to have more than 2 friends (right-hand side) the two components merge and we are left with a single component plus three players that are isolated. In the right-hand side graph, we can see the important role played by player 5, who

Figure 1.4: Simulation - 12 players



is linked to two subcomponents, the one that involves players $6, 7, 8, 9$ and 16 and the other that involves all the players except herself and players $15, 17$ and 18.

Figure 1.5: Simulation - 19 players



From Figures 1.4 and 1.5 we can see that, as one would expect from real-life friendship relations, the friendship networks arising in equilibrium are very complex and few or non generalities can be found. Apart from the qualitative results already presented, we have been unable to find any other generalization. Furthermore, for every seemingly general fact involving the existence of stars, components, circles, etc. and for every observation about the amount of connectivity, the characteristic of isolated players, the distance between players, etc. we were able to find an example such that, with a slight modification of the parameters, the fact or observation was no longer present..

1.5 Robustness Checks

1.5.1 Tit-for-Tat

In this section we check for the robustness of our results when players, instead of using Grim Trigger, use the Tit-for-Tat. Because of our way of modeling, we cannot use the standard definition of Tit-for-Tat. The problem arises due to the possibility of not being linked so we have to adapt the standard definition to our setting. We define Tit-for-Tat as follows. If player k betrays player i, player i will offer help to the other player again only if k offers help to i and at the same time i does not help k. Hence, after a betrayal, the friendship relation can be reestablished only if the betrayer 'pays back' to the betrayed for the harm done. Formally,

Definition 3. *Define the Tit-for-Tat strategy for player $i \in \mathcal{N}$ played against any player $k \in \mathcal{N}$ as follows:*

- *If i never betrayed k and k never betrayed i, play (l, H).*
- *Otherwise:*
  1. *If the play in the past period against k was $((l, N), (l, H))$ or $((l, H), (l, H))$ then play (l, H).*
  2. *If the play in the past period against k was $((n, \{H, N\}), (l, H))$ then play (l, N).*
  3. *Otherwise, play your weakly dominant strategy.*

The following result shows that the conditions for supporting friendship under Grim Trigger (Proposition 1) and Tit-for-Tat are the same apart from minor differences.

Proposition 6. *Under Tit-for-Tat strategies, a friendship relation between players i and k can be supported in the repeated game if and only if the following holds:*

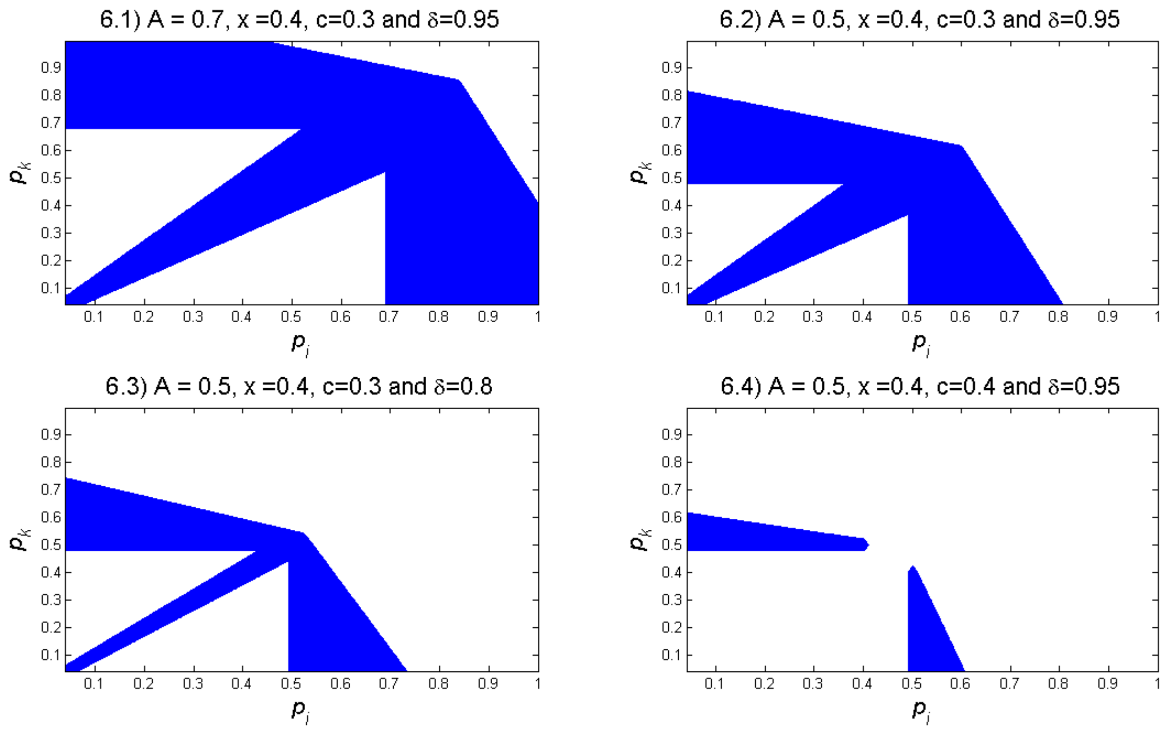
- *if $p_j \leq \frac{A}{1-x}$ and $p_{-j} \leq A$, then $\frac{p_j}{p_{-j}} \geq \frac{c}{\delta x}$*
- *if $p_j \leq \frac{A}{1-x}$ and $A < p_{-j} \leq \frac{A}{1-x}$, then $\frac{1}{1+\delta}(A - p_j) + xp_j - \frac{c}{\delta}p_{-j} \geq 0$*

for $j \in \{i, k\}$ and $-j \in \{i, k\} \setminus \{j\}$.

Proof. See Appendix 1.A.1. □

When both p_i and p_k are below A both Grim Trigger and Tit-for-Tat strategies yield the same conditions for supporting friendship. When either $p_i > A$ or $p_k > A$, the condition is

Figure 1.6: Tit-for-Tat



slightly different between the two settings. Figure 1.6 is a counterpart of Figure 1.2 for the case of Tit-for-Tat strategies formulation.

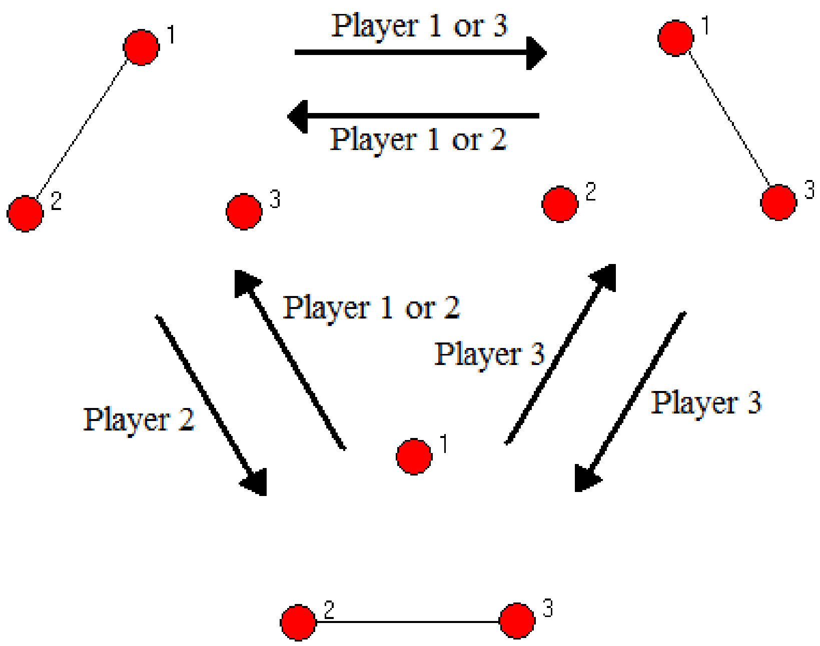
Under Tit-for-Tat strategy convergence to equilibrium is not guaranteed, that is, the friendship network may cycle between different configurations forever. The reason why this happens is that, because of the discount factor, it may be in some players' interest to continuously betray each other and became friends again. In Figure 5 we present an example of a situation in which the social network never reaches an equilibrium. The parameters used are the same as in example 1. Computations are presented in Appendix 1.A.4.

Remark 1. *Under Tit-for-Tat strategies, convergence to an equilibrium is not guaranteed.*

1.5.2 Social Punishment

Now we analyze a different issue. In the model presented in the main text, if a player betrays another, the rest of the players in the population do not react to the betrayal. We may think that if an agent is betraying her friends, it is less likely that new agents will want to set up a friendship relation with her. We explore a situation in which, if a player betrays another, all players will automatically switch to play (n, N) against the betrayer. Each player

Figure 1.7: Simulation - Tit-for-Tat



knows this fact when considering whether to betray one of her friends or not. We call this punishment mechanism the Social Punishment. This punishment mechanism may seem to be a bit too strong but we do not study here the effects of different punishment mechanisms (Kandori (1992) undertakes this issue). Here we restrict ourselves to the most basic and simple mechanism of social punishment.

In addition to this, we add one further plan to the two plans players already have at their disposal. The Friendly Ending Plan is now available for the players. By friendly ending we mean that the player who wants to break the friendship relation switches to play (n, N) instead of betraying the other player by playing (l, N).

Friendly Ending Play (n, N) if you and the other player played (l, H) in the last round; play your weakly dominant strategy in the Relationship Game otherwise.

Proposition 7. *Assume Social Punishment. For all A, x, c there exists a* $\hat{\delta} > 0$ *such that if* $\delta > \hat{\delta}$ *and* $p_i \neq p_j$ $\forall i, j \in \mathcal{N}$*, then the friendship network converges with probability* 1 *to a unique network architecture.*

Proof. See Appendix 1.A.1. □

Without the Social Punishment, we only need each player to be different enough ensure convergence to a unique equilibrium. On the other hand, with Social Punishment, we need each to be sufficiently patient and not to be equal to anybody else in the population.

The following result for the case of Social Punishment deserves attention. Define a *fully connected component* as the set of players who are all friends with each other and with no player outside the component.

Proposition 8. *Consider an equilibrium situation. If* $p_i \neq p_j$ $\forall i, j \in \mathcal{N}$, $\delta > \bar{\delta}$ *and Social Punishment exists, then there exists no fully connected component of* $m+1$ *or more players in which all of the players have a degree of needing help bigger than* A.

Proof. See Appendix 1.A.1. □

Proposition 8 implies that players with a high degree of needing help can not form big groups of friendship. Without Social Punishment a betrayal in an early period between people with high degree of needing help may make possible the existence in equilibrium of a component of more than $m+1$ players to exist in equilibrium.

1.6 Relating our Results with Some Empirical Facts About Friendship

We propose a repeated setting in which players are friends when they are helping each other. Many empirical studies show how important the exchange of help between friends is. For example, Walker (1995) interviewed 52 working- and middle-class subjects and found that one of the main functions of the friends was to provide help. She found that among the working-class this help was based on providing goods and services such as borrowing or lending small amounts of money or helping in finding a job. In turn, helping among the middle-class was based on emotional and intellectual support.

In their study on 185 Dutch students, Buunk and Prins (1998) found that in the relationship with their best friend, 73.6% of the subjects considered the friendship to be reciprocal. In our paper, players' reciprocity is translated into Grim Trigger: I help you as long as you help me. The Cooperative Plan is restrictive as it does not allow for forgiveness, which is a standard feature of friendship. We believe that using Grim Trigger is not so far from reality as betraying a friend is something very severe that is difficult to forgive. Betraying a friend causes direct and conscious harm, which is different to, for instance, having a small argument with a friend. Nevertheless, as a robustness check, section 1.5.1 presents the results for the case in which players instead use the Tit-for-Tat strategy, which allows for forgiveness.

According to Proposition 4, the friendship equilibrium can depend on the order in which players are selected by nature. We can think of the order in which players are selected by nature as being similar to the order in which players in the population meet each other. The player selected by nature is the one who can have the initiative to meet new people by making phone calls. The fact that the order in which people meet each other affects the long-term friendship relations was reported in an empirical study by Cloninger (1986). Cloninger found that meeting new people may result in breaking old and strong friendship relations because of the novelty of having new friends. In a forthcoming article, Whitmeyer and Yeingst (2008) refer to this characteristic of the friendship relations as *fickleness*.

Our result in Proposition 5 related the length of the friendship relations in a population to the patience of players, represented by the discount factor. In a sample with children from the fourth and sixth grades, Hallinan (1978) found that the length of the friendship relations was considerably longer among the sixth-grade children than among fourth grade children. Hence, considering that the discount factor decreases with age (see, for example, Read and Read (2004)), the result stated in the proposition matches the empirical result concerning friendship relations between children.

1.7 Conclusions

We have presented a model of friendship selection between a group of players. Each player can decide with whom of the other players in the group she wants to set up a friendship relation. The results of the paper state under which conditions friendship can arise between players. We find that when there are only two players, the decision to be friends between players whose degree of needing help is low depends on the relative difference between their degrees of needing help: the bigger the difference, the less likely they are to become friends. For players whose degree of needing help is high, we find that rather than caring about the relative difference in degrees of needing help, they look only at the absolute level of these values

When we move to analyze the case of a group of more than two people, we find that it is in general impossible to predict which friendship relations will be present in equilibrium. We present two explanations for why this happens. These are the existence of a certain degree of substitutability between friends and the non-existence of a social mechanism to punish the players betraying friends. We also find that the length of the friendship relations positively depends on the patience of the players.

The model presented here differs mainly from the existing models in psychology, anthropology and sociology in that it is solved analytically and in the fact no assumptions in the taste for friends are made. Moreover, it differs from the existing models of social networks in that: there exists heterogeneity between players, the cooperation game that players play in the network is micro-founded in friendship relations and the strategies of each player can be different in each one of the cooperative games that they play on each period.

The results found in the paper seem to match the findings reported in many empirical studies of friendship selection. In our opinion, the value of the paper lies in the fact that it gives a precise non-trivial explanation to some of the phenomena we find in the friendship relations among humans. Possible extensions of the model may include a more general setting in which the degrees of needing help are unknown but players can learn them or allowing for a more flexible dynamic setting with respect to how players change their strategies.

References

1. Bala, V. and Goyal, S (2000). "A Noncooperative Model of Network Formation", *Econometrica* 68 (5), 1181-1229.

2. Borgatti, S.P. (2002). *NetDraw: Graph Visualization Software*, Harvard: Analytic Technologies.

3. Buunk, B.P. and Prins, K. S. (1998). "Loneliness, Exchange Orientation, and Reciprocity in Friendships", *Personal Relationships* 5, 1-14.

4. Chwe, M. (2000), "Communication and Coordination in Social Networks", *Review of Economics Studies* 67, 1-16.

5. Cloninger, C. R. (1986), "A Unified Biosocial Theory of Personality and its Role in the Development of Anxiety States". *Psychiatric Developments* 3, 167–226.

6. V.J. Derlega and B.A. Winstead (1986). "Strategies in developing friendships", in *Friendship and Social Interaction.* New York: Springer Verlag.

7. de Vos H. and Zeggelink, E. (1997). "Reciprocal Altruism in Human Social Evolution: The viability of Reciprocal Altruism With Preference for "old-helping-partners". *Evolution and Human Behavior* 18, 261-278.

8. Franzen, A. (2000). "Does the Internet Makes us Lonely?", *European Sociological Review* 16, 427-438.

9. Fudenberg, D. and Tirole, J. (1991). *Game Theory*, The MIT Press.

10. Ghosh, P. and Ray, D. (1996). "Cooperation in Community Interaction without Information Flows". *The Review of Economic Studies* 63 (3), 491-519.

11. Goyal, S and Vega-Redondo, F. (2005). "Network Formation and Social Coordination", *Games and Economic Behavior* 50, 178-207.

12. Hallinan, M. T. (1978/79). "The process of Friendship Formation". *Social Networks* 1, 193-210.

13. Hruschka, D. J. and Henrich, J. (2004). "The Evolution of Friendship". mimeo.

14. Jackson, M. O. and Wolinsky, A. (1996). "A Strategic Model of Social and Economic Networks", *Journal of Economic Theory* 71, 44-74.

15. Jackson, M. O. and Watts, A. (2002). "On the Formation of Interaction Networks in Social Coordination Games", *Games and Economic Behavior* 41, 265-291.

16. Kandori, M. (1992). "Social Norms and Community Enforcement", *The Review of Economic Studies* 59 (1), 63-80.

17. Kreps, D. M., Milgrom P., Roberts, J and Wilson, R (1982). "Rational Cooperation in the Finitely Repeated Prisoners' Dilemma". *Journal of Economic Theory* 27, 245-252.

18. Lippert, S. and Spagnolo, G. (2005). "Networks of Relations and Social Capital". mimeo.

19. Liu, M. and Chen, X. (2003). "Friendship Networks and Social, School and Psychological Adjustment in Chinese Junior High School Students". *Psychology in the Schools* 40 (1), 5-17.

20. Marmaros, D. and Sacerdote, B. (2004). "How Do Friendships Form?" mimeo.

21. Silk, J. B. (2003). "Cooperation without Counting, the Puzzle of Friendship". *Genetic and Cultural Evolution of Cooperation.* MIT Press, 37-54.

22. Ueno, K. (2005). "The Effects of Friendship networks on Adolescent Depressive Symptoms". *Social Science Research* 34, forthcoming.

23. van de Bunt, G., G., van de Duijn, M., A., J. and Snijders, T. A. B. (1999). "Friendship Networks Through Time - An Actor-Oriented Dynamic Statistical Network Model". *Computational & Mathematical Organization Theory* 5 (2), 167-192.

24. Vega-Redondo (2005). "Building Up Social Capital in a Changing World", forthcoming in *Journal of Economic Dynamics and Control.*

25. Walker, K. (1995). ""Always There for Me": Friendship and Expectations among Middle- and Working Class Men and Women". *Sociological Forum* 10 (2), 273-296.

26. Whitmeyer, J. M. and Yeingst, C. N. (2008): "A dynamic model of friendly association networks". *Social Science Research*, forthcoming.

27. Zeggelink, E. P. H. (1995). "Evolving Friendship Networks: an individual-oriented approach implementing similarity". *Social Networks* 17, 83-110.

Appendix

1.A.1 Proofs

Proposition 2. *Given the possibility of playing as in the Defective Plan, there is no strategy that does not involve the use of dominated actions and that can support the outcome* $((l,H),(l,H))$ *as a part of an equilibrium of the repeated Relationship Game for a bigger set of parameter values than the Cooperative Plan. Furthermore, the Defective Plan is the best possible deviation against the Cooperative Plan.*

Proof. First, we prove the first part of the proposition. It is straightforward to notice that, given the possibility of playing the Defective Plan the only way of making the cooperative outcome easier to sustain, in the sense that it can be sustained for a bigger set of parameter values, is via the following strategy.

- If a play in any past period against k was either $((l,H),(l,N))$ or $((l,N),(l,H))$, then play (n,N).
- Otherwise, play (l,H).

That is, a strategy that threatens the opponent with playing (n,N) if a deviation occurs increases the parameter set for which a friendship relation is possible. Any other strategy different from the one above will imply less cooperation as the Defective Plan will make deviation profitable for a bigger parameter set. However, playing (n,N) is weakly dominated by playing (l,N) for all players $i \in \mathbb{N}$ with $p_i \geq A$, a contradiction.

The second part of the proposition follows easily. The best one-period gain can be achieved by playing (l,N) when the other player is playing (l,H), which is what the Defective Plan prescribes. Moreover, once deviation has occurred, the Cooperative Plan prescribes playing the weakly dominant action in the Relationship Game, and the best response to this is to also play the weakly dominant action in the Relationship Game, which is again what the Defective Plan prescribes. □

Proposition 3. *A friendship relation between players i and k can be supported in the repeated Relationship Game when both players use Cooperative Plan if and only if the following holds:*

- *if $p_i, p_k \leq A$, then $\frac{c}{\delta x} \leq \frac{p_i}{p_k} \leq \frac{\delta x}{c}$,*
- *if $p_i, p_k \leq \frac{A}{1-x}$ but either $A < p_i$ or $A < p_k$, then $A - p_i + xp_i - \frac{c}{\delta} p_k \geq 0$ and $A - p_k + xp_k - \frac{c}{\delta} p_i \geq 0$.*

Proof. The payoff to player i when both players play (l, H) equals $\frac{1}{1-\delta}(A - p_i + xp_i - cp_k)$. If player i deviates at time t from this strategy, then according to the definition of the Cooperative Plan, three things can happen:

Case 1. *If $p_i \leq A$ and $p_k \leq A$ then the most profitable deviation for player i is to play (l, N). This is weakly better for her than to play (n, N) if $p_i \leq A$. According to the Cooperative Plan, in the period after this deviation occurs, player k would switch to play (l, N) forever because $p_k < A$. Then, the payoff of the deviation for player i equals $(A - p_i - xp_i) + \frac{\delta}{1-\delta}(A - p_i)$. Hence, the increase of payoff for player i from the deviation is weakly negative if and only if:*

$$\frac{p_i}{p_k} \geq \frac{c}{\delta x} \tag{1.1}$$

Case 2. *If $p_i \leq \frac{A}{1-x}$ but $A < p_k$, then $A - p_k < 0$ and $A - p_i > 0$. In this case the best deviation for player i is to play (l, N) as shown before. But this time, however, because $p_k > A$ player k will switch to play (n, N) forever after player i's deviation occurs. The payoff of deviation from (l, H) for player i is given by $A - p_i - xp_i$. Hence, the increase of payoff for player i from the deviation is weakly negative if and only if:*

$$A - p_i + xp_i - \frac{c}{\delta} p_k \geq 0. \tag{1.2}$$

Case 3. *If $\frac{A}{1-x} \geq p_i > A$, then the best deviation for player i is to play (l, N) if the other player is playing (l, H) and to play (n, N) if the other player is not playing (l, H). So if they are both playing the Cooperative Plan and no deviation has occurred, if i deviates from (l, H), the action she will play is (l, N) (because $A - p_i + xp_i \geq 0$). The next period after the deviation the action played is (n, N). Hence, i's payoff is the same as in the previous case, so his incentives to deviate are the same as in the previous case. Therefore, equation (1.2) gives us the condition for player i to support the Friendship Equilibrium.*

Grouping the results of cases 1, 2 and 3 and their equivalent for player k gives us the result stated in the proposition. □

Proposition 4. *The system converges with probability 1 to an equilibrium network architecture that can be history dependent. If $A > c\delta$ then there exists a $\varepsilon > 0$ such that if*

$\min_{i,k\in\mathcal{N},i\neq k}|p_i - p_k| > \varepsilon$, *then the friendship network converges with probability* 1 *to a unique network architecture.*

Proof. First, we prove the first statement of the theorem. The fact that the equilibrium network architecture may be history dependent was already shown in example 1. Moreover, the system will always converge to an equilibrium because of the following. Given that players are using Grim Trigger as the strategy for supporting cooperation, if one player betrayed one of her friends then they won't become friends ever again. Hence, the process will eventually get to a point in which no player will want to betray her friends nor to change the strategy she is currently playing against the other players. Once this happens, the process has reached an equilibrium.

To prove the second statement of the theorem we proceed as follows. As mentioned earlier, the process is not ergodic because of the substitutability between players. That is, given the order in which players are selected by nature, it may happen that a player betrays one of her friends to set up a friendship relation with a third one whose degree of needing help is higher. As we show in the next paragraph, this will never happen if players are different enough. If players are different enough, then the unique equilibrium can be constructed in a fashion that we will specify below.

The increase in the profit for player i from betraying a friend, say player k, for setting up a friendship relation with another player j with $p_k < p_j$ is at most $c\left(p_k - \frac{\delta}{1-\delta}(p_j - p_k)\right) + \frac{1}{1-\delta}(A - xp_i + p_i)$. Hence, if $c(p_k - \delta p_j) + A - (1-x)p_i) < 0$ for all $i, j, k \in \mathcal{N}$, that is, if $-c\delta + A < 0$ and p_j and p_k are different enough, then the profits from betraying will be negative and the only betrayal that will occur will be those in which one player betrays another for setting up a friendship relation with a third one that has a smaller degree of needing help.

To construct the equilibrium network when players are different enough we proceed as follows. Take the player with the lowest degree of needing help in the population, say i. Define the combination of relationships between i and the rest of the players that maximize i's payoff for a given m. This combination of friendship relations, call it f_i, is uniquely determined if all the players are different from each other. Now take the player that has the second lowest degree of needing help in the group, say k. Define the combination of relationships between k and the rest of the players that maximize k's payoff for a given m and considering that the friendship relation prescribed by f_i has to hold. Continue in this fashion until the player with the highest degree of needing help. This results in a friendship network $F = f_1 \cup f_2 \cup \cdots \cup f_N$.

It is clear that if players are different enough F is an equilibrium network as no player can improve her situation by betraying a friend in order to set up a friendship relation with a different player. F is indeed the only equilibrium network; this follows from the fact that in any network configuration different to F, there exists at least one player that can improve her situation by changing her current strategy. To see this, consider a network configuration different to F. Take the player with the lowest degree of needing help that has her friendship relations different that what F prescribes. If she breaks her links and offers links to the players with whom she should be linked according to F, these links will be accepted and she will improve her payoff (by construction of the network F).

Now we show that the process converges with probability one to network F. To do so we only have to consider the fact that (1) for any network different from F there is a positive probability from moving to a different network, (2) once the network F is reached, the process remains there forever, and (3), for any network there is a positive probability of reaching the network configuration F in a finite number of steps. Statements (1) and (2) were proven in the preceding paragraphs (by showing that F is the unique equilibrium network). To show (3) it is enough to notice that at any point in time there is a positive probability that the players allowed to revise their strategy in each period are ordered from the one with the lowest degree of needing help to the one with the highest degree of needing help. However, if this is the case, the network that the process reaches is exactly F, as we wanted to show. □

Proposition 5. *Ceteris paribus, the length of the friendship relations depends positively on δ.*

Proof. For any friendship relation between two players, the chances that one of the players will betray the other negatively depends on the discount factor. This is so because when a player betrays another player, she is increasing her present payoff for a possible decrease of her future payoff. Hence, the higher the discount factor, the less likely a player will betray one of her friends. Therefore, given a set of parameters and a population, the speed at which the friendship relations change depends negatively on the discount factor. □

Proposition 6. *Under Tit-for-Tat strategies, a friendship relation between players i and k can be supported in the repeated game if and only if the following holds:*

- *if* $p_j \leq \frac{A}{1-x}$ *and* $p_{-j} \leq A$, *then* $\frac{p_j}{p_{-j}} \geq \frac{c}{\delta x}$
- *if* $p_j \leq \frac{A}{1-x}$ *and* $A < p_{-j} \leq \frac{A}{1-x}$, *then* $\frac{1}{1+\delta}(A - p_j) + xp_j - \frac{c}{\delta}p_{-j} \geq 0$

for $j \in \{i, k\}$ *and* $-j \in \{i, k\} \setminus \{j\}$.

Proof. We structure the proof of this result similarly to the proof of Proposition 1 but with the only difference that once a player betrays the other, it may be in her interest to play (l, H)

in the next round after betrayal so as to bring the helping situation back. The payoff of player i that the situation in which both players play (l, H) forever equals $\frac{1}{1-\delta}(A - p_i + xp_i - cp_k)$. If player i deviates at time t from this strategy, according to the definition of the Tit-for-Tat strategy two things can happen:

Case 1. *If $p_i \leq \frac{A}{1-x}$ and $p_k \leq A$, then the best deviation for player i is to play (l, N). In the period after this deviation occurred, player k will switch to play (l, N) because $p_k \leq A$. If player i then plays (l, N) or (n, N) forever, we are in the same case as Grim Trigger, i.e. the case in which condition 1.1 has to hold. On the other hand, if player i plays (l, H) from after the period she deviated on, player k will also come back to playing (l, H). Note that for player i there is no difference between trying to restore the friendship relationship immediately after he betrayed player k T periods after the betrayal has taken place. The payoff of the deviation for player i equals $(A - p_i + xp_i) + \delta(A - p_i - cp_k) + \frac{\delta^2}{1-\delta}(A - p_i - xp_i - cp_k)$. Hence, the increase of payoff for player i from the deviation is weakly negative if and only if:*

$$\frac{p_i}{p_k} \geq \frac{c}{\delta x} \tag{1.3}$$

Note that in this case the condition for friendship to be possible is the same as in the case with Grim Trigger.

Case 2. *If $p_i \leq \frac{A}{1-x}$ but $A < p_k$, then $A - p_k < 0$ and $A - p_i > 0$. In this case the best deviation for player i is to play (l, N) as shown before. But this time, however, because $p_k > A$ player k will switch to play (n, N) after player i's deviation occurs. If player i then plays (l, N) forever, we are in the same case as Grim Trigger, i.e. the case in which the condition 1.2 has to hold. On the other hand, imagine that player i plays (l, H) from after the period she deviated on. Then, according to the Tit-for-Tat strategy, player k will play first (l, N) and then (l, H) forever. The payoff of deviation from (l, H) for player i is given by $A - p_i + xp_i + \frac{\delta^2}{1-\delta}(A - p_i - cp_k) + \frac{\delta^3}{1-\delta}(A - p_i + xp_i - cp_k)$. Hence, the increase of payoff for player i from the deviation is weakly negative if and only if:*

$$\frac{1}{1+\delta}(A - p_i) + xp_i - \frac{c}{\delta}p_k > 0 \tag{1.4}$$

Note that condition 1.4 is stronger than condition 1.2.

Grouping the results of cases 1, 2 and 3 and their equivalent for player k gives us the result stated in the proposition. □

Proposition 7. *Assume Social Punishment. For all A, x, c there exists a $\hat{\delta} > 0$ such that if $\delta > \hat{\delta}$ and $p_i \neq p_j$ $\forall i, j \in \mathcal{N}$, then the friendship network converges with probability 1 to a unique network architecture.*

Proof. If a player i is to betray another player under the Social Punishment setting, she knows that from the moment of her betrayal she will get a payoff of 0 forever. Hence, when deciding whether to betray or not, the player takes into account the present period increase in her profits with the future decrease in her payoff. Hence, if player i is patient enough she won't be interested in betraying any of her friends ever. This result, combined with the fact that all players are different, shows that, using the same arguments as in Proposition 4, a unique network exists and the system converges to it with certainty. □

Proposition 8. *Consider an equilibrium situation. If δ is high enough, $p_i \neq p_j$ $\forall i, j \in \mathcal{N}$ and there exists Social Punishment, then there exists no component of $m+1$ or more players in which all of the players have degree of needing help bigger than A.*

Proof. The proof goes by contradiction. Take a group of $k > m+1$ players among which all have their degree of needing help bigger than A. Take the $m+1$ players of the component with the lowest degree of needing help. Because $k > m+1$, at least one of them won't be linked with the other m (if not these $m+1$ players will form a closed component which by assumption is not the case). Take a player among the $m+1$ with the lowest degree of needing help in the component who is not linked with the other m with the lowest degree of needing help in the component. If she makes calls to the players with whom she is not linked and have the lowest degree of needing help in the component, the calls will result in new friendships. Note that this won't happen if some players have a degree of needing help smaller than A. Hence, the initial situation was not an equilibrium. □

1.A.2 Dynamics

The dynamics of the model work as follows.

1. At $t = 0$ each player is playing the strategy "play (n, N) against all players in all the rounds".

2. In period t for $t = 1, 2, \ldots$ the following sequence of events takes place:

   (a) A player $i \in \mathcal{N}$ is selected by nature. This player can make calls to the other players.

   (b) Every player $k \in \mathcal{N}_{-i}$ plays, if she gets a call from i, according to one of the possible schemes:

      i. If cooperation between k and i is not pair-wisely sustainable. Then k plays the same strategy she played last period against all the players in the population.

ii. If cooperation between k and i is pair-wisely sustainable and player k is providing help less than m times. Then player k plays Grim Trigger with i and plays the same strategy she played last period against the rest of the players.

iii. If cooperation between k and i is pair-wisely sustainable and k is providing help exactly m times. Let j be the player with the highest degree of needing help among those who k is currently helping. If, moreover, the discounted present value of the profits from playing the Cooperative strategy with i plus playing the Defective strategy j are higher than the profits of player k from playing the Cooperative strategy against j, then player k switches to play the Cooperative strategy with i, the Defective strategy with j and plays the same strategy she played last period against the rest of players. Otherwise, k plays the same strategy she played last period against all the players in the population.

(c) Player i, the one selected by nature in the current period, makes calls to the other players and changes her current strategy against them. She does so knowing that the players who get a call will react as stated in step b. She makes the calls and changes her strategy in such a way as to maximize her present value payoff myopically, i.e.

i. she will decide whom to call and play the Cooperative strategy with

ii. she will play the Defective strategy with the players she does not call.

(d) Players who get a call from i play according to step b.

(e) All other players don't change strategy.

1.A.3 Example: A Simple Case

The simulation in example 1 is conducted for $N = \{1, 2, 3\}$, $p_1 = 0.4$, $p_2 = 0.45$, $p_3 = 0.55$, $A = 0.5$, $x = 0.6$, $c = 0.3$, $\delta = 0.7$ and $m = 1$. First we check that all friendship relations are possible. To do so we only have to apply Proposition 3 to the present example. Players 1 and 2 can be friends because $\frac{0.3}{0.6\times0.7} \leq \frac{0.4}{0.45} \leq \frac{0.6\times0.7}{0.3}$. Players 1 and 3 can be friends because $0.5-0.4+0.6\times0.4-\frac{0.3}{0.7}\times0.55 = 0.104 \geq 0$ and $0.5-0.55+0.6\times0.55-\frac{0.3}{0.7}\times0.4 = 0.108 \geq 0$. Finally, players 2 and 3 can be friends because $0.5-0.45+0.6\times0.45-\frac{0.3}{0.7}\times0.55 = 0.084 \geq 0$ and $0.5 - 0.55 + 0.6 \times 0.55 - \frac{0.3}{0.7} \times 0.45 = 0.087 \geq 0$.

In period 1, player 1 is selected by nature. Since she can set up friendship relations with the other two players but she is constrained to have at most one friendship relation player 1 will choose to call player 2. This is true simply because $p_2 < p_3$ and hence, the stream of payoffs for player 1 is higher if she sets up a friendship relation with player 2.

In particular, the stream of payoffs if player 1 sets up a relationship with player 2 equals $\frac{1}{1-0.7}(0.5 - 0.4 + 0.6 \times 0.4 - 0.3 \times 0.45) = 0.683$. On the other hand, if player 1 sets up a friendship relation with player 3, her stream of payoffs equals $\frac{1}{1-0.7}(0.5 - 0.4 + 0.6 \times 0.4 - 0.3 \times 0.55) = 0.583$. Player 2 will respond to the call of player 1 by switching to play Grim Trigger with her since $\frac{0.3}{0.6 \times 0.7} \leq \frac{0.4}{0.45} \leq \frac{0.6 \times 0.7}{0.3}$ holds.

In period 2, player 1 is again selected by nature. Now her decision is whether or not to betray player 2. In this example, player 1 will switch to the strategy "play (l, N) if you and the other player played (l, H) in the last round; play your weakly dominant strategy in the Relationship Game otherwise" against player 2 and will call player 3 and play Grim Trigger against her. That is, player 1 will betray player 2. The next period after this deviation occurs, both player 1 and player 2 will switch to play (l, N) against each other (because $p_1 < p_2 < A$). To see that player 1 will betray player 2 and call player 3 and play Grim Trigger against her, we consider her payoff with this change of her strategies. If player 1 betrays player 3 and sets up a relationship with player 3, her payoff equals $(0.5 - 0.4 + 0.6 \times 0.4) + \frac{0.7}{1-0.7}(0.5 - 0.4) + \frac{1}{1-0.7}(0.5 - 0.4 + 0.6 \times 0.4 - 0.3 \times 0.55) = 1.156$. On the other hand, if player 1 keeps her friendship relation with player 2, she will get a payoff equal to: $\frac{1}{1-0.7}(0.5 - 0.4 + 0.6 \times 0.4 - 0.3 \times 0.45) = 0.683$. Hence, player 1 will betray player 2 and set up a friendship relation with player 3.

In period 3, player 2 is selected by nature. She will call player 3 instead of player 1 because the betrayal that happened in period 2 now makes the friendship between player 1 and 2 impossible forever. In this example, we have that, in response to player 2's call, player 3 will betray player 1 to set up a relationship with player 2 even though the profit of player 3 is higher if she has a friendship relation with player 1. The stream of payoffs of player 3 from betraying player 1 by setting up a relationship with player 2 equals: $(0.5 - 0.55 + 0.6 \times 0.55) + \frac{1}{1-0.7}(0.5 - 0.55 + 0.6 \times 0.55 - 0.3 \times 0.45) = 0.763$. On the other hand, the stream of payoffs of player 3 if she keeps her friendship relation with player 1 equals: $\frac{1}{1-0.7}(0.5 - 0.55 + 0.6 \times 0.55 - 0.3 \times 0.4) = 0.533$. Hence, player 3 will betray player 1.

After period 3, the network is in equilibrium. No player can increase her profit by changing the strategy as it can be easily verified.

1.A.4 Example: Nonexistence of Equilibrium under Tit-for-Tat

We now show with an example that under Tit-for-tat there may not exist equilibrium. We use the same set of parameters as in example 1. That is, $N = \{1, 2, 3\}, p_1 = 0.4, p_2 = 0.45, p_3 = 0.55, A = 0.5, x = 0.6, c = 0.3, \delta = 0.7$ and $m = 1$.

First, we check that a friendship relation is possible between any two players in the group.

To do so, we only have to apply Proposition 6 to the present example in the same fashion as we applied Proposition 3 in Appendix 1.A.3.

We show now that, in this particular case, the process will never converge no matter how nature selects the players. As we showed in the proceeding paragraph, all friendship relations are possible. Now we show that for any given friendship relation in this group, there is always a profitable betrayal independent of the history of past play. Imagine that players 1 and 2 are friends and both players betrayed player 3 recently. If one of these two players wants to set up a friendship relation with player 3, they will have to first 'pay back' and offer help to player 3. Imagine that player 2 is selected by nature: she will betray player 1 and switch to play (l, H) against player 3 if and only if $(0.5-0.45+0.6\times 0.45)+0.7(0.5-0.45-0.3\times 0.55)+\frac{0.7}{1-0.7}(0.5-0.45)\cdots+\frac{0.7^2}{1-0.7}(0.5-0.45+0.6\times 0.45-0.3\times 0.55) > \frac{1}{1-0.7}(0.5-0.45+0.6\times 0.45-0.3\times 0.4)$. This inequality holds true. Note that because $p_1 < p_2$ and $p_3 < A$, if player 2 finds it profitable to retake her friendship with player 3 so will player 1 . Also note that if player 1 (2) finds it profitable to betray 2 (1) to retake her friendship relation with 3, it is straightforward to show that since $p_3 > p_2 > p_1$, if player 1 (2) is having a friendship relation with 3, she will find it profitable to betray player 3 and to retake (or start) a friendship relation with 2 (1). Also, in this example player 3 finds it profitable to betray player 1 (2) to retake her friendship relation with player 2 (1). However, this is not needed for the result we want to show.

So we have that for any history of past play (or, more intuitively, history of past betrayals), there always exists at least one player that can increase her profit by changing strategy. Hence, the process never converges to an equilibrium.

1.A.5 Formal Definitions of the Sets of Strategies

Let i and k stand for the two typical elements of $\mathcal{N}$ and let $\mathcal{N}_{-i} = \mathcal{N} \setminus \{i\}$. Define the set of actions in the Relationship Game as $A = \{l, n\} \times \{H, N\}$ and let A_i be the set of actions of each player i against every other player in the Relationship Game, $A_i = (A_{ij})_{j\ in \mathcal{N}_{-i}}$ with $A_{ij} \in A$. Let H_{ik}^t be the set of all possible histories between players i and k till the beginning of time $t \geq 0$. Hence, we have that $\left(h_{ik}^1, \ldots, h_{ik}^{t-1}\right) \in H_{ik}^t$ for $t \geq 1$ and $H_{ik}^0 = \varnothing$ with $h_{ik}^s \in \{A_{ik} \times A_{ki}\}$ for $s \in \{1, \ldots, t-1\}$. Define $H_i^t = (H_{ij}^t)_{j \in \mathcal{N}_{-i}}$.

Let L^t be the sequence of players selected by nature till time t, hence $L^t = (l^\tau)_{\tau=1}^t$ with $l^\tau \in \mathcal{N}$. Define the set of strategies of each player i against player k given the players selected by nature each period and set of all possible histories *between i and all the other players* as Σ_{ik}. Hence, if $\sigma_{ik} \in \Sigma_{ik}$ then:

$$\sigma_{ik} : \cup_{t=0}^{\infty} \left\{L^t \times H_i^t\right\} \rightarrow A.$$

Therefore, a strategy σ_{ik} is a plan that maps all the possible histories and all the possible combinations of players selected by nature into the set of actions. We make use of pair-wise strategies. That is, if player i is to decide which action to take against player k, i will only consider the past history between i and k as if $\mathcal{N} = \{i, k\}$. Formally, denote the pair-wise set of strategies for each player i against player k given the players selected by nature in every period and set of all possible histories *between players* i *and* k by Σ_{ik}^p. Hence, if $\sigma_{ik}^p \in \Sigma_{ik}^p$, then:

$$\sigma_{ik}^p : \cup_{t=0}^{\infty} \left\{ L^t \times H_{ik}^t \right\} \rightarrow A.$$

We are using the superscript p to refer to the fact that the strategy is pair-wise. For each i define $\Sigma_i = (\Sigma_{ij})_{j \in \mathcal{N}_{-i}}$ and $\Sigma_i^p = \left(\Sigma_{ij}^p \right)_{j \in \mathcal{N}_{-i}}$.

We write $\pi_{ik}(\sigma_i, \sigma_{-i})$ as the discounted present value payoff for player i when he plays the Relationship game against player k when i's strategy is σ_i and the rest of players are playing a strategy σ_{-i}. Define the best response of each player i as $\sigma_i^{BR} = (\sigma_{ij}^{BR})_{j \in \mathcal{N}_{-i}}$ where:

$$\begin{aligned} \sigma_{ij}^{BR} \in & \arg \max_{\sigma_{ij}^p \in \Sigma_{ij}^p} \pi_{ij}(\sigma_{ij}^p, \sigma_{-i}) \\ & st : \# \left\{ \sigma_i^p : \sigma_i^p \in (l, H) \right\} \leq m \end{aligned}$$

Put in words, each player maximizes her payoff taking each relationship pair-wisely subject to the constraint of not offering help more than m times.

Finally, we reduce the strategy space to the case in which, for each player i, her strategy against every player k consists in either the Cooperative strategy or the Defective strategy. For every i, k let $\hat{\Sigma}_i^p$ be this strategy space.

Chapter 2 - Learning within a Markovian Environment

2.0 Abstract

We investigate learning in a setting where each period a population has to choose between two actions and the payoff of each action is unknown by the players. The population learns according to reinforcement and the environment is non-stationary, meaning that there is correlation between the payoff of each action today and the payoff of each action in the past. We show that when players observe realized and foregone payoffs, a suboptimal mixed strategy is selected. On the other hand, when players only observe realized payoffs, a unique action, which is optimal if actions perform different enough, is selected in the long run. When looking for efficient reinforcement learning rules, we find that it is optimal to disregard the information from foregone payoffs and to learn as if only realized payoffs were observed.

2.1 Introduction

Imagine the simple decision problem in which every period individuals in a population have to choose between two alternatives. The payoff of these two alternatives is not know by the players. What is more, the payoff of the alternatives could vary over time according to some distribution also unknown for the players.

This decision problem is faced by many of us in our everyday lives: whether to buy a PC or a Mac, whether to have fruit or a cake as a dessert in a restaurant, or whether to watch an action movie or a romantic movie at the theater. Although oblivious of the payoff we will get from making these choices, we might have some information that can help in choosing the better alternative. This information could have been obtained, for instance, from our own experiences in the past or via word-of-mouth communication.

In this paper we study how the choices made by a population evolve in the setting just described. The model we present has two major features about how players learn and about how the payoffs change. First, players learn according to reinforcement, whereby actions that where successful in the past are more likely to be chosen. Second, the underlying distribution determining the payoff of each action is non-stationary. This means that the payoff today of a given action depends on the payoff it yielded in the past. In particular, we consider the case in which payoffs depend deterministically on the state of nature. The state of nature changes following a Markov chain. Hence, the probability of being at a given state tomorrow depends on which state we are in today. Players are ignorant of this fact; they simply observe that the payoff of available actions changes over time.

In the learning literature, as well as in the economic literature in general, randomness determining the outcome of certain events or actions is almost always assumed to follow a stationary i.i.d. process. This assumption is clearly made for the sake of technical simplicity, as real life phenomena, such as financial markets, gambling, population biology, statistical mechanics, etc., quite often follow non-stationary processes. To our knowledge, only Ben-Porath et al. (1993) and Rustichini (1999) deal with the evolutionary properties of models where nature follows a non-stationary process.

Ben-Porath et al. (1993) present an evolutionary model that is framed within a changing environment. They study two types of environments: one in which the change is deterministic and another in which the changes in environment follow a Markov chain. In their model, players' actions are subject to random mutations. They characterize the mutation rate that maximizes population growth in the long run.

Rustichini (1999) presents a paper that focuses on the optimality of two different population dynamics within a Markovian environment. In his model, the environment changes according to a Markov chain, and for any state in the chain there is a unique action that maximizes payoff. Rustichini (1999) studies the optimality properties of linear and exponential (logit) adjustment process when players have infinite memory. An adjustment process or learning rule is simply a map between information and strategies. Rustichini (1999) considers two different informational settings about payoffs of actions. In one of these settings players observe the performance of all the actions (realized and foregone payoffs are observed), while in the other they only observe the performance of the action chosen (only realized payoffs are observed).

As in Rustichini (1999), we consider two informational settings: one in which both realized and foregone payoffs are observed and another in which only realized payoffs are observed. There are two main differences between Rustichini's work and ours. First, we consider a very general set of learning rules instead of only two specific rules. Second, and most importantly, in our setting players don't use the whole history of past payoff realizations. Instead, as prescribed by reinforcement, players learn using the information they have from their most recent payoff experiences. The reason why we are interested in a setting where players have limited memory is that empirical and theoretical literature in psychology and economics agrees that limited memory is a better assumption for modeling human behavior than infinite memory (see for example, Rubinstein (1998), Hirshleifer and Welch (2002) and Conlisk (1996)).

As already mentioned, the learning rules considered in this paper have the property of being reinforcing. According to reinforcement learning, actions that were more successful today are more likely to be adapted for tomorrow. Reinforcement has been found to be one of the main driving forces of human behavior in repeated decision problems. For some detailed expositions on reinforcement learning and its relationship with real life behavior the reader is referred to Roth and Erev (1995), Erev and Roth (1998) and Camerer and Ho

(1999).

When both realized and foregone payoffs are observed, reinforcement is translated into being more likely to play tomorrow the action that was better today. For this setting, we use a generalization of the best response behavior that we call the Stochastic Better Response. Under the Stochastic Better Response, the probability of playing tomorrow a given action increases if and only if today that action was better than the other one. The magnitude of the change in probabilities of playing either action depends on the specific functional form used. The Stochastic Better Response is a very general learning rule that allows players to respond to the magnitude and not just the ordering of payoffs of each action. Note that the Stochastic Better Response is a different concept from the Stochastic Better Reply Dynamics (Josephson (2007)). The Stochastic Better Reply Dynamics are the dynamics for the evolution of strategies resulting when players use the better response, which is a particular case of the Stochastic Better Response.

When foregone payoffs are not observed, players can not directly compare the performance of both actions within the same time period. In this case, players reinforce (possibly negatively) the action they played. How much they reinforce this action will depend on the payoff achieved. We use a general case of the Cross (1973) learning rule that also generalizes the rules in Börgers, Morales and Sarin (2004) (BMS, henceforth). We call this rule the General Reinforcement Rule. Note that players could use the General Reinforcement Rule even if they observe foregone payoffs. While this implies that players are disregarding information, we will show that it may be optimal to do so.

Under the Cross Learning Rule, players increase the probability of playing the action just played by the payoff yielded by that action. An interesting result shown by Börgers and Sarin (1997) is that a population that plays according the Cross Learning Rule exhibits a behavior that converges to replicator dynamics.

The rules in BMS can incorporate aspiration levels (exogenous or endogenous): in other words, if the payoff of the action chosen is higher than the aspiration level, then the probability of playing that action increases for the next period. On the other hand, if the payoff achieved by the action chosen is smaller than the aspiration level, then the probability of playing that action decreases for next period. The rules in BMS are linear on payoffs. We relax this by allowing for any increasing function on realized payoff.

In the case where foregone payoffs are observed, we show that the continuous time limit of the evolution of strategies converges to a situation where every period every action is played with a constant probability bounded away from 1. The specific value of the probability by which each action is played at every period will depend on two things: first, the difference in payoffs between the two actions and the specific form of Stochastic Better Response used, and, second, on the probabilities that the limiting distribution of the Markov chain for states puts on each state. The behavior found in this setting is a generalization to what is know as probability matching. Under probability matching, if an action is best a fraction x of the time, then in any given period it is played with probability x. The best reply matching behavior is clearly suboptimal. While some experimental papers report that this behavior is observed in real life (see, for example, Rubinstein (2002), Siegel and Goldstein (1959)), there does not seem to be consensus as to whether probability matching is in fact present in the behavior of real life agents (see, for instance, Vulkan (2000) and Shanks et al. (2002)).

The results found in this informational setting are also closely related to the findings by Kosfeld et al. (2002). They study a setting where a finite set of players repeatedly play a normal-form game. Players adapt their strategies by increasing the probability of playing a certain action only if this action is a best reply to the actions played by the other agents. Hence, the rule they use is a particular case of the Stochastic Better Response in which the magnitude of payoffs is irrelevant for the updating of strategies. Our setting is also different from theirs in that players do not play against other players but against nature and in that we consider a general class of rules instead of only one. Kosfeld et al. (2002) find that the continuous time limit of the system converges to a best-reply matching equilibrium. In a best-reply matching equilibrium each player plays an action with a probability that is equal to the probability that this action is a best response to the actions of the other players. The probability matching behavior found in this paper for games against nature is the equivalent to the best-reply matching equilibrium found in Kosfeld et al. (2002). In Section 2.5.1 this issue is discussed in more depth.

In our second informational setting, when foregone payoffs are not observed, we show that the population may end up playing a suboptimal action. The population surely selects the action that has higher average payoff only if the difference between the average payoff of the two actions is high enough. Hence, the system may lock-on to a suboptimal action. In this respect, our work extends Ellison and Fudenberg (1995) results to a general set of learning rules and an environment that may not be stationary.

Our results are rounded off by characterizing the efficient rules for both informational settings. A striking result is that when foregone payoffs are observed, it is optimal to ignore the extra information conveyed by the payoff of the action not chosen. That is, players are better off by learning using the General Reinforcement Rule, which only uses the information of the realized payoff. This is due to the fact that observing foregone payoffs leads players to adopt the action that is best today but may be not the best in the long run. That is, players are "distracted" by observing the performance of all the actions. When foregone payoffs are not observed, we show that if players use learning rules that diminish the magnitude of payoffs, that is, that have very cautious and show slow learning, then the population learns the optimal action. These results from are in contrast to those of Rustichini's (1999). In Rustichini (1999), when the population uses the exponential rule (fast learning) the best action is selected only in situations where foregone payoffs are observed, whereas if populations uses the linear rule (slow learning) best action is selected only in situations where foregone payoffs are not observed. Here, instead, we find that under reinforcement learning it is optimal to disregard foregone payoffs and to exhibit slow learning in both informational settings.

This paper's contribution to the literature is twofold. Our first contribution to the literature is the introduction new techniques for dealing with correlated states of nature. As mentioned, very few papers have studied the situation in which the future realization of the state of nature depends on its past realizations. Most papers on learning consider either that the environment does not change or that it changes independently of past realizations. This is due to the technical difficulties involved in dealing with correlated realizations of states. In this paper we show how these difficulties can, at least partially, be overcome. The proofs for the result for the Stochastic Better Response demonstrate how dependent randomness can be dealt with by showing that for any possible realization of states of nature, the position of the system in the future can be approximated by the differences in speed of convergence towards each action.

The proof of the result for the case where foregone payoffs are not observed extends Ellison and Fudenberg's (1995) result to the case where the distribution of payoffs is not stationary. We show that the behavior of a system that evolves according to a Markov Chain can be approximated by the behavior of a system in which the probability of each state occurring is independent and equal to the limiting distribution of the Markov Chain.

Our second contribution is the extension of the knowledge about stimulus response learning models and evolutionary models. The differences in the behavior of the population under the two informational settings are very intriguing and of interesting application for real life situations. For instance, why can inferior technologies come to dominate the market? A well known example is that when the video format VHS took over from the superior format Betamax. The model can explain that if the two technologies are not too different in terms of performance, the stochastic evolution of nature can lead the population to lock on the suboptimal choice forever. In the example with video formats, during the first months after the release of both technologies, Betamax tapes could not hold an entire movie. This caused the population to slowly adopt the VHS format. Once the true potential of Betamax was revealed, it was too late, consumers had already locked on the inferior technology.

The rest of the paper is organized as follows. Section 2.2 presents the model. The two informational settings are introduced in Section 2.3. Results are developed in Section 2.4. Section 2.5 presents a discussion and a deeper comparison of this work with the existing literature. Finally, Section 2.6 concludes.

2.2 The Model

Consider a continuum of identical players of measure 1. Every period $t = 0, 1, \ldots$ players in the population have to choose between action 1 or action 2. The payoff of each player at time t depends on her action and on the current state of nature $s^t \in \{1, \ldots, m\}$. If a player chooses action i and the state equals j then she gets a payoff $\pi_{ij} \in [0, 1]$ with $i \in \{1, 2\}$ and $j \in \{1, \ldots, m\}$. Note that the payoff of each player does not depend on the actions played by others but only on her own action and the state of nature. We assume there is no weakly dominant action. That is, there exists no $i \in \{1, 2\}$ such that $\pi_{ij} \geq \pi_{-ij}$ for all $j \in \{1, \ldots, m\}$. Without loss of generality we assume that for some $h < m$, $\pi_{1j} \geq \pi_{2j}$ for $j \leq h$ and $\pi_{2j} > \pi_{1j}$ for $j > h$. That is, in the first h states action 1 yields at least the same payoff as action 2. In the remaining states, action 2 yields more payoff than action 1. Finally, we define π_j as the vector of payoffs of action 1 and action 2 in state j, $\pi_j = (\pi_{1j}, \pi_{2j})$.

The sequence of states of nature $\{s^t\}_{t=0}^{\infty}$ follows a discrete Markov process P with $m \geq 2$ states. The probability of transiting from state i to state j is given by $\theta_{ij} \in [0, 1]$. We assume the Markov chain to be irreducible and aperiodic. Hence, if $\theta_{ij} = 0$ for some $i, j \in \{1, \ldots, m\}$ then there exists a sequences of states $k_1, k_2, \ldots, k_n \in \{1, \ldots, m\}$ with $n \leq m$ such that $\theta_{ik_1}, \theta_{k_1,k_2}, \ldots, \theta_{k_n,j} \neq 0$. We define $\lambda \in [0, 1]^m$ as the limiting distribution of the Markov chain P where λ_i is the weight the limit distribution puts in state i. An environment is defined then by the payoff vectors together with a transition matrix, $\{(\pi_1, \ldots, \pi_m), P\}$.

A strategy is the probability of playing each action at a given period. We denote by $\sigma_i^t \in [0, 1]$ with $i \in \{1, 2\}$ and $t \in \{0, 1, \ldots\}$ the probability of playing action i at time t. Define $\sigma^* = (\sigma_1^*, \sigma_2^*) \in [0, 1]^2$ as the strategy that maximizes payoff in the long run. Formally, for any $(\bar{\sigma}_1, \bar{\sigma}_2) \in [0, 1]^2$ we have that

$$\sum_{j=1}^{m} \lambda_j \left(\sigma_1^* \pi_{1j} + \sigma_2^* \pi_{2j} \right) \geq \sum_{j=1}^{m} \lambda_j \left(\bar{\sigma}_1 \pi_{1j} + \bar{\sigma}_2 \pi_{2j} \right).$$

Since we are dealing with a continuum of population, Law of Large Numbers applies and we have that σ_i^t is also the fraction of players playing action i at time t. In an abuse of notation, throughout the paper we will refer to σ_i^t as both the probability for a single player of playing action i at time t and the fraction of the population playing action i at time t.

Note that given our setting, the sequence $\sigma_i = \{\sigma_i^t\}_{t=0}^{\infty}$ is an irreducible and aperiodic Markov process on $[0,1]$ for $i \in \{1,2\}$. The aim of the paper is to characterize, if it exists, the invariant distribution of such process.

The timing within each time period works as follows. First, players choose actions according to their strategies. Then, nature decides the state. Third, payoffs are realized and players observe their payoff and possibly forgone payoffs. The possibility of observing foregone payoffs depends on the informational setting being considered. Finally, players update their strategies.

When updating their strategies, players use the following information: their strategy at the beginning of the period, the action they played and the payoff they got and possibly the payoff the other action would have yielded (foregone payoffs). Formally, a learning rule is a function $b : [0,1]^2 \times \{1,2\}^2 \times [0,1]^2 \rightarrow [0,1]^2$. That is, a function that maps three arguments, strategies for the present period, action played and payoff gotten and action not played and foregone payoff, into the strategies for the following period. The functional form of b will depend on the specific learning rule under consideration.

2.3 Informational Settings

2.3.1 Forgone Payoffs are Observed

When both realized and foregone payoffs are observed, players best respond to the environment by increasing the probability of playing at the next period the action that was most successful at the present period. We use a generalization of the best response behavior that we call the Stochastic Better Response.

We write $\sigma_i^{t+1}|_j$ to denote the value of σ_i^{t+1} given that at period t the state of nature, s^t, was j. The Stochastic Better Response is defined by

$$\sigma_1^{t+1}|_j = \begin{cases} \sigma_1^t + \sigma_2^t \mu f(\pi_j) & \text{if } \pi_{1j} \geq \pi_{2j} \\ \sigma_1^t - \sigma_1^t \mu f(\pi_j) & \text{otherwise,} \end{cases}$$

where $\mu > 0$ is a learning speed parameter. The function $f : [0,1]^2 \rightarrow [0,1]$ maps the payoff of the action that yielded higher payoff and the payoff of the other action into a number between 0 and 1. This function is interpreted as the probability of adopting or learning the action that was best given today's state of nature. The only requirement on f is that it must be weakly increasing in the payoff of the action that yielded higher payoff and weakly decreasing in the payoff of the other action. That is, f is weakly increasing (decreasing) in π_{ij} only if $\pi_{ij} > (<) \pi_{-ij}$. We set $f(\pi_j) = 0$ if and only if $\pi_{1j} = \pi_{2j}$. In other words, we assume that the population does not change strategies if and only if both actions yielded the same payoff. The function f could also be a constant. In the case where the function f is constant and equals 1, the learning rule is equivalent to the standard best response in which players show inertia with probability $1-\mu$ (as in Samuelson (1994) and Kosfeld et al. (2002)).

The intuition behind the Stochastic Better Response is the following. In each period, all players observe the payoff of the action chosen and the payoff of the other action. Then every player updates her strategy in the following way. The probability of playing action i in the next period is increased if and only if action i yielded higher payoff than the other action in the current period. The increase in the probability of playing action i will depend on the difference in payoffs between the two actions.

A different interpretation of this same rule uses the fact that σ_i can be considered as the fraction of population playing action i deterministically. Under this interpretation, at every period, players that did not play the best action will change their actions (best response to the environment) with some probability. The probability of changing action depends on the difference in payoff between the two actions. The Stochastic Better Response is an individual learning rule because actions played by other players have no effect on the updating of the one's own strategy.

As an example, we can look at two possible ways of writing the Stochastic Better Response. In the first one below, payoffs enter exponentially in the function f.

$$\sigma_1^{t+1}|_j = \begin{cases} \sigma_1^t + \sigma_2^t \mu \frac{e^{\pi_{1j}} - e^{\pi_{2j}}}{e^{\pi_{1j}} + e^{\pi_{2j}}} & \text{if } \pi_{1j} \geq \pi_{2j} \\ \sigma_1^t - \sigma_1^t \mu \frac{e^{\pi_{2j}} - e^{\pi_{1j}}}{e^{\pi_{1j}} + e^{\pi_{2j}}} & \text{otherwise} \end{cases} \tag{2.1}$$

A second example could be the following, where only the payoff of the best action at the current period enters in f and f is linear.

$$\sigma_1^{t+1}|_j = \begin{cases} \sigma_1^t + \sigma_2^t \mu \pi_{1j} & \text{if } \pi_{1j} \geq \pi_{2j} \\ \sigma_1^t - \sigma_1^t \mu \pi_{2j} & \text{otherwise} \end{cases}$$

2.3.2 Foregone Payoffs are not Observed

When foregone payoffs are not observed, players have no means of directly comparing the performance of both actions within the same time period. In this case, players reinforce (possibly negatively) the action they played. How much they reinforce this action will depend on the payoff achieved. We use a general case of the Cross (1973) learning rule that also generalizes the rules in BMS. We call this rule the General Reinforcement Rule.

Let $\sigma_i^{t+1}|_{kj}$ be the probability by which a player plays action i at time $t+1$ given that action k was played at time t and state at time t, s^t, was j. The General Reinforcement Rule is defined by

$$\begin{aligned}\sigma_1^{t+1}|_{1j} &= \sigma_1^t + \sigma_2^t g(\pi_{1j}),\\ \sigma_1^{t+1}|_{2j} &= \sigma_1^t - \sigma_1^t g(\pi_{2j}),\end{aligned}$$

and similarly for $\sigma_2^{t+1}|_{1j}$ and $\sigma_2^{t+1}|_{2j}$. The only assumption we make in $g : [0,1] \to [-1,1]$ is that it must be weakly increasing in its argument. If $g(\pi_{ij}) = \pi_{ij}$ then we have the Cross Learning Rule. For the rules in BMS we have that $g(\pi_{ij}) = A_{ij} + B_{ij}\pi_{ij}$ for given $A_{ij} \in \mathbb{R}$ and $B_{ij} \in \mathbb{R}$ for $i \in \{1,2\}$ and $j \in \{1,\ldots,m\}$. BMS show that setting $A_{ij} = -\min\{1-\sigma_1^0, \sigma_1^0\}/\max\{1-\sigma_1^0, \sigma_1^0\}$ and $B_{ij} = 1/\max\{1-\sigma_1^0, \sigma_1^0\}$ for all i, j results in the best monotone rule. A rule is defined to be monotone if the expected probability of playing the action that is best given today's state increases. A rule is said to be the best monotone rule if the expected increase in playing the best action from one period to another is highest among all monotone rules. Since BMS study a setting in which the evolution of nature follows a stationary distribution, the action that is best today is the action that is best at every period. In our setting the action that is best today may not be the best action tomorrow due to the Markovian evolution of the states of nature. This particular difference will have important consequences in the optimality properties of the rules in BMS.

2.4 Results

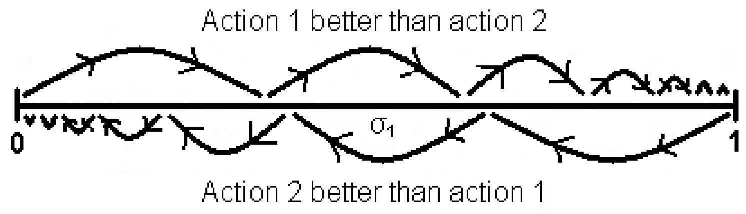
2.4.1 Results - Foregone payoffs are Observed

Before going to the formal results, we present a small discussion on the behavior of the system under the Stochastic Better Response. First, note that the biggest difference in the behavior of the two rules that we consider lies in the way they behave when σ_i is close to the corners (0 and 1). In particular, under the Stochastic Better Response the corners are not absorbing while the opposite occurs under the General Reinforcement Rule.

Assume for this short discussion that there are only two states of nature. Under the Stochastic Better Response, the speed at which a player adopts an action slows down as the probability of playing that action increases. That is, consider that action 1 is played with a high probability and that today action 1 yielded a higher payoff than action 2. Then the increase in the probability of playing action 1 will be small. On the other hand, consider that action 1 is played with a small probability and today action 1 yielded higher payoff than action 2. In this case the probability of playing action 1 next period increases sharply.

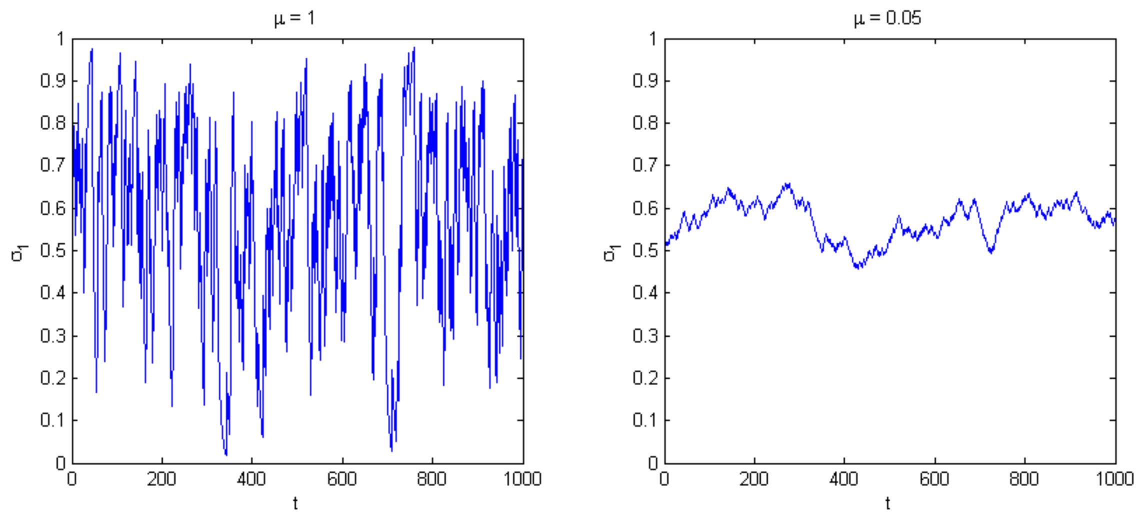
Figure 2.1 shows the movements of the probability of playing action i (σ_i) as a response to an action being better than the other in the current period. As above, assume that an action is played with a high probability. Then the increase in playing that action in case it yielded a higher payoff than the other action at the present period is low.

Figure 2.1: Stochastic Better Response



As one could possibly guess already, the Stochastic Better Response will not converge to any of the corners. To study convergency, we consider the limit case when μ, which can be viewed as the size of the changes in σ_i, gets arbitrarily small. Once such a limit is taken, the Stochastic Better Response converges to a single point. This issue can be seen much more clear by looking at Figure 2.2, where a simulation is conducted. The specific learning rule used is given by equation 2.1. The value of the parameters is set to $m = 2$, $\pi_{11} = 0.5, \pi_{12} = 0.3, \pi_{21} = 0.1, \pi_{22} = 0.6$ and $\theta_{12} = \theta_{21} = 0.3$. The initial value σ_1 was set to $\sigma_1^0 = 0.5$. The figure depicts the same simulation, the same random seed, for two situations: one in which $\mu = 1$ and another in which $\mu = 0.05$.

Figure 2.2: Simulation - Stochastic Better Response



By studying the behavior of the system when μ is made arbitrarily small we are characterizing the continuous time limit of σ_i. When μ is taken to zero the adjustment in the strategies is made arbitrarily small while keeping constant the speed at which the environment changes. For other papers that use this continuous time limit approximation in settings somewhat different from ours see, for example, Börgers and Sarin (1997) and Benaïm and Weibull (2003).

The following proposition characterizes the convergence of (σ_1, σ_2) under the Stochastic Better Response when μ is arbitrarily small. Later in this section we present a sketch of the proof. The formal proof is contained in the Appendix.

Proposition 1. *Define*

$$\tilde{\sigma} = \frac{\sum_{j:\pi_{1j} \geq \pi_{2j}} \lambda_j f(\pi_j)}{\sum_{j=1}^{m} \lambda_j f(\pi_j)}.$$

For any $\varepsilon > 0$ there exists a $\bar{\mu} > 0$ such that if $\mu < \bar{\mu}$ then

$$P\left(\lim_{t \to \infty} \left|\sigma_1^t - \tilde{\sigma}\right| > \varepsilon\right) = 0.$$

The interpretation of the result is the following. For simplicity of the exposition let us focus on the evolution of the variable σ_1 and assume again that there are only two states of nature. The point $\tilde{\sigma}$ corresponds to the situation where an increase in σ_1^t due to action 1 yielding higher payoff at time t than action 2 would be equivalent to the decrease in σ_1^t from action 2 yielding more payoff than action 1. That is, with $m = 2$, $\tilde{\sigma}$ is the σ_1^t is such that $\left|\sigma_1^{t+1}|_1 - \sigma_1^t\right| = \left|\sigma_1^{t+1}|_2 - \sigma_1^t\right|$. In Figure 1, the point $\tilde{\sigma}$ would be such that the size of the arrows (or jumps) towards the left from a given point σ_1^t is the same as the size of the arrows towards the right from this same point σ_1^t. Hence, $\tilde{\sigma}$ is the point where the marginal movements towards action 1 and towards action 2 are equalized.

One can easily check that $\tilde{\sigma} < 1$, so it will never be the case that the best action in the long run is played with probability 1. For the general case where the Markov chain has m states, action 1 is strictly better than action 2 if and only if $\sum_{j=1}^{m} \lambda_j \pi_{1j} > \sum_{j=1}^{m} \lambda_j \pi_{2j}$; this inequality holds in the simulation in Figure 2.2. However, for that simulation we have that $\tilde{\sigma} = 0.57$. That is, in the long run at any given period action 1 is played with probability of 0.57. This behavior is clearly suboptimal as if $\sum_{j=1}^{m} \lambda_j \pi_{1j} > \sum_{j=1}^{m} \lambda_j \pi_{2j}$ then the σ_1^t that maximizes payoff in the long run is $\sigma^* = 1$.

Let us now look at a sketch of the proof. To studying the convergence of the sequence σ_1 we first show that it suffices to study the convergence of a sequence $y = \{y^t\}_{t=\hat{t}}^{\infty}$, for $\hat{t}$ large enough, which evolves in a world with just 2 states of nature and symmetric transition matrix.

First, define the sequence $\hat{\sigma}_1 = \{\hat{\sigma}_1^t\}_{t=\hat{t}}^{\infty}$ as $\hat{\sigma}_1^{\hat{t}} = \sigma_1^{\hat{t}}$ and recursively for $t > \hat{t}$

$$\hat{\sigma}_1^{t+1} = \begin{cases} \hat{\sigma}_1^t + \hat{\sigma}_2^t \mu f(\pi_1) & \text{with probability } \lambda_1 \\ & \vdots \\ \hat{\sigma}_1^t + \hat{\sigma}_2^t \mu f(\pi_h) & \text{with probability } \lambda_h \\ \hat{\sigma}_1^t - \hat{\sigma}_1^t \mu f(\pi_{h+1}) & \text{with probability } \lambda_{h+1} \\ & \vdots \\ \hat{\sigma}_1^t - \hat{\sigma}_1^t \mu f(\pi_m) & \text{with probability } \lambda_m \end{cases}.$$

Then we have that for any given $t > \hat{t}$,

$$P\left(|E_0(\sigma_1^t) - E_0(\hat{\sigma}_1^t)| > \varepsilon\right) = 0. \tag{2.2}$$

Hence, the expected value of both σ_1 and $\hat{\sigma}_1$ converge in probability to the same value. This is because the transition matrix P is irreducible and aperiodic. Now define the sequence $y = \{y^t\}_{t=\hat{t}}^{\infty}$ as $y^{\hat{t}} = \hat{\sigma}_1^{\hat{t}}$ and define recursively

$$y^{t+1} = \begin{cases} y^t + 2(1-y^t)\mu \sum_{j:\pi_{1j} \geq \pi_{2j}} \lambda_j f(\Pi_j) & \text{with probability } 1/2 \\ y^t - 2y^t \mu \sum_{j:\pi_{1j} < \pi_{2j}} \lambda_j f(\Pi_j) & \text{with probability } 1/2 \end{cases}.$$

Note that the variable y evolves according to the expected movement in the long run of the variable $\hat{\sigma}_1$. It can be easily seen that $y^t = \hat{\sigma}_1^t$ implies $E_0(y^{t+1}) = E_0(\hat{\sigma}_1^{t+1})$. Hence, since $y^{\hat{t}} = \hat{\sigma}_1^{\hat{t}}$, the distribution of both y^t and $\hat{\sigma}_1^t$ is aperiodic and both $E_0(y^{\hat{t}+1})$ and $E_0(\hat{\sigma}_1^{\hat{t}+1})$ are linear in their arguments, we can state that $E_0(y^{\hat{t}+k}) = E_0(\hat{\sigma}_1^{\hat{t}+k})$ for any $k \in \mathbb{N}$. Moreover, we have that for any $t > \hat{t}$, equation 2.2 must hold. Hence, we have that for any $\varepsilon > 0$ and $t > \hat{t}$,

$$P\left(|E_0(\sigma_1^t) - E_0(y^t)| > \varepsilon\right) = 0.$$

Furthermore, by making μ arbitrarily small we make the variance of both random variables y^t and σ_1^t to shrink to zero. Thus, their limiting distribution puts weight on a single point. In other words, y and σ_1 must converge in probability to a fixed value $\bar{y}$ and $\bar{\sigma}$ respectively. Since $E_0(y^{\hat{t}+k})$ converges to $E_0(\sigma_1^{\hat{t}+k})$ for all $k \in \mathbb{N}$, we must have that $\bar{y} = \bar{\sigma}$. Hence, instead of studying the convergence of the variable σ_1 we focus on the convergence of the variable y. This is more formally stated in Lemma 2 in the Appendix.

Note now that the point $y^t = \tilde{\sigma}$, with $\tilde{\sigma}$ as defined in Proposition 1, solves the equation

$$y^t + 2(1-y^t)\mu \sum_{j:\pi_{1j} \geq \pi_{2j}} \lambda_j f(\pi_j) = y^t - 2y^t \mu \sum_{j:\pi_{1j} < \pi_{2j}} \lambda_j f(\pi_j).$$

Define the sequence $y_1 = \{y_1^t\}_{t=\hat{t}}^{\infty}$ as follows

$$y_1^t = \begin{cases} y^t & \text{if } y^t \geq \tilde{\sigma} \\ \tilde{\sigma} & \text{otherwise} \end{cases}.$$

Hence, we have that $E_0(y^t) \leq E_0(y_1^t)$ for all $t > \hat{t}$. Note that $E_0(y_1^{t+1}) \leq E_0(y_1^t)$. Therefore, y_1 is a super-martingale with lower bound $\tilde{\sigma}$. Thus, by the martingale convergence theorem, y_1 converges in probability to $\tilde{\sigma}$. This implies that for t large enough, $E_0(y^t) \leq \tilde{\sigma}$.

Define now the sequence $y_2 = \{y_2^t\}_{t=\hat{t}}^{\infty}$ as follows

$$y_2^t = \begin{cases} y^t & \text{if } y^t \leq \tilde{\sigma} \\ \tilde{\sigma} & \text{otherwise} \end{cases}.$$

Therefore, we have that $E_0(y^t) \geq E_0(y_2^t)$ for all $t > \hat{t}$. Note that $E_0(y_1^{t+1}) \geq E_0(y_2^t)$. Hence, y_2 is a sub-martingale with upper bound $\tilde{\sigma}$. Thus, by the martingale convergence theorem, y_2 converges in probability to $\tilde{\sigma}$. This implies that for t large enough, $E_0(y^t) \geq \tilde{\sigma}$.

Hence, we know that for t large enough, $E_0(y^t) \leq \tilde{\sigma}$ and $E_0(y^t) \geq \tilde{\sigma}$. This implies that for all $t > \hat{t}$, $E_0(y^t) = \tilde{\sigma}$. Since the variance of y shrinks to zero as μ is made arbitrarily small, we have that y converges in probability to $\tilde{\sigma}$ as μ is made arbitrarily small. Combined with the fact that y converges in probability to σ_1, this implies that σ_1 converges in probability to $\tilde{\sigma}$.

2.4.2 Results - Foregone Payoffs are not Observed

We recall that the probability by which a player plays action i at time $t+1$ given that action k was played at time t and state at time t was j is denoted by $\sigma_i^{t+1}|_{kj}$ and given by

$$\begin{aligned} \sigma_1^{t+1}|_{1j} &= \sigma_1^t + \sigma_2^t g(\pi_{1j}), \\ \sigma_1^{t+1}|_{2j} &= \sigma_1^t - \sigma_1^t g(\pi_{2j}). \end{aligned}$$

Hence, $\sigma_1^{t+1}|_j$, which is the probability of playing action 1 at time $t+1$ given that state was j, equals $\sigma_1^t + \sigma_2^t g(\pi_{1j})$ if action 1 was played at time t and $\sigma_1^t - \sigma_1^t g(\pi_{2j})$ if action 2 was played at time t. Action i with $i \in \{1, 2\}$ is played at time t with probability σ_i^t. Hence, since we are dealing with a continuum of players, we can use Law of Large Numbers to state that

$$\sigma_1^{t+1}|_j = \sigma_1^t \sigma_1^{t+1}|_{1j} + \sigma_2^t \sigma_1^{t+1}|_{2j}.$$

This can be rewritten as

$$\sigma_1^{t+1}|_j = \sigma_1^t \left(\sigma_1^t + (1 - \sigma_1^t) g(\pi_{1j})\right) + (1 - \sigma_1^t) \left(\sigma_1^t - \sigma_1^t g(\pi_{2j})\right).$$

Thus, it follows that

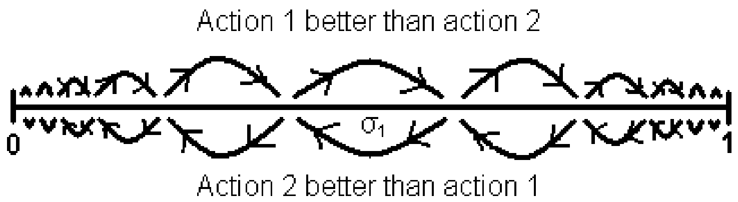
$$\sigma_i^{t+1}|_j = \sigma_i^t \left(1 + (1 - \sigma_i^t)\left[g(\pi_{ij}) - g(\pi_{-ij})\right]\right). \tag{2.3}$$

Note that if we set $g(\pi_{ij}) = \pi_{ij}$, as in the Cross Learning Rule, the resulting law of motion for σ_i is the discreet time version of the Replicator Dynamics. That is, if $g(\pi_{ij}) = \pi_{ij}$ then we have that

$$\sigma_i^{t+1}|_j \quad = \quad \sigma_i^t + \sigma_i^t \left(\pi_{ij} - [\sigma_i^t \pi_{ij} + \sigma_{-i}^t \pi_{-ij}]\right).$$

The General Reinforcement Rule behaves completely differently to the Stochastic Better Response. Under the General Reinforcement Rule, the changes in the variable σ_i^t become smaller as σ_i^t gets closer to either bound. For example, consider that action 1 is played with a high probability. Then the change in σ_i will be small independently of whether action 1 yielded higher payoff than action 2 or the other way around. Figure 2.3 shows the movements of σ_1 under the General Reinforcement Rule as a response to the environment.

Figure 2.3: General Reinforcement Rule

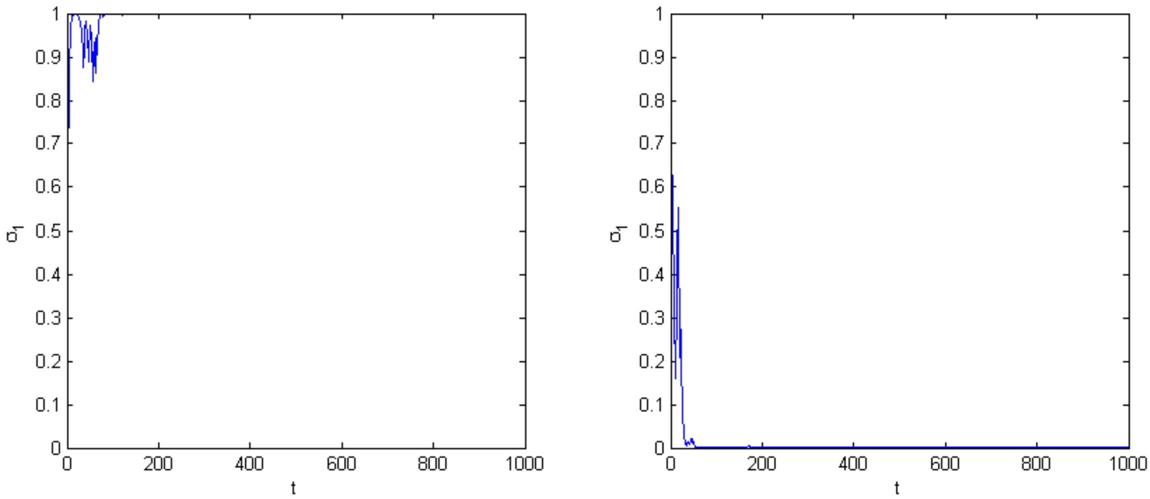


As we see, the process will spend almost no time in intermediate values of σ_i. This will allow us to draw our conclusions from analyzing only the behavior of σ_i in the neighborhoods of its bounds. In this respect, our analysis will partially rely on the approach by Ellison and Fudenberg (1995).

Figure 2.4 shows a simulation for the General Reinforcement Rule for the case where $g(\pi_{ij}) = \pi_{ij}$ and with the same parameters as the ones used in Figure 2.2. The figure plots the result of the same simulation performed with two different random seeds.

It can be seen that the General Reinforcement Rule quickly converges to a situation in which all the population plays the same action a fraction 1 of the time. An interesting thing to note is that the action selected by the General Reinforcement Rule does not coincide necessarily with the action that is best in the long run. The simulation on the right-hand side shows a situation in which the General Reinforcement Rule converges to a situation where all players in the population are playing the suboptimal action. As we will see, this is the result of the two actions performing not too differently in terms of payoffs in the long run.

Figure 2.4: Simulation - General Reinforcement Rule



The following proposition, whose proof is presented in the Appendix, characterizes the convergence of the sequence σ_1.

Proposition 2. *Define* $\gamma_j = 1 + g(\pi_{1j}) - g(\pi_{2j})$ *and* $\hat{\gamma}_j = 1 + g(\pi_{2j}) - g(\pi_{1j})$ *and consider the two inequalities:*

$$\sum_{j=1}^{m} \lambda_j \log \gamma_j > 0, \tag{2.4}$$

$$\sum_{j=1}^{m} \lambda_j \log \hat{\gamma}_j > 0. \tag{2.5}$$

1. *If both (2.4) and (2.5) hold then* $\lim_{t\to\infty} \sigma_1^t$ *does not exist.*
2. *If (2.4) holds but (2.5) does not then* $\lim_{t\to\infty} \sigma_1^t = 1$.
3. *If (2.5) holds but (2.4) does not then* $\lim_{t\to\infty} \sigma_1^t = 0$.
4. *If neither (2.4) nor (2.5) hold then* $\lim_{t\to\infty} \sigma_1^t$ *has full support over* $\{0, 1\}$.

Since $\sigma_2 = 1 - \sigma_1$ the convergence of the sequence σ_2 follows for the proposition above. An important fact revealed by proposition above is that the process may fail to converge to the best action. Consider for simplicity the Cross Learning Rule, where $g(\pi_{ij}) = \pi_{ij}$. Action 1 is weakly better than action 2 in the long run if and only if $\sum_{j=1}^{m} \lambda_j \pi_{1j} \geq \sum_{j=1}^{m} \lambda_j \pi_{2j}$. This condition can be rewritten as $\sum_{j=1}^{m} \lambda_j \gamma_j \geq 1$. However, even if $\sum_{j=1}^{m} \lambda_j \gamma_j \geq 1$ holds, it may still happen that $\sum_{j=1}^{m} \lambda_j \log \gamma_j < 0$ holds and hence σ_1 may not converge to 1. To make this point more clear consider the case in which $m = 2$ and $\lambda_1 = \lambda_2 = 0.5$. That is, there are only two states of nature and both states are equally likely in the long run. The following corollary characterizes the convergence of σ_1 in this case when action 1 is better in the long run than action 2.

Corollary 1. *Assume* $g(\pi_{ij}) = \pi_{ij}$, $m = 2$, $\lambda_1 = \lambda_2 = 0.5$ *and* $\pi_{11} + \pi_{12} > \pi_{21} + \pi_{22}$.

- *If* $\pi_{11} + \pi_{12} - \pi_{21} - \pi_{22} - (\pi_{11} - \pi_{21})(\pi_{22} - \pi_{12}) > 0$ *then* $\lim_{t\to\infty} \sigma_1 = 1$.

- *Otherwise,* $\lim_{t\to\infty} \sigma_1$ *has full support over* $\{0, 1\}$.

Proof. We can rewrite inequalities (2.4) and (2.5) from Proposition 2 for the case with $m = 2$ and $\lambda_1 = \lambda_2 = 0.5$ as follows:

$$\log \gamma_1 + \log \gamma_2 > 0 \quad (2.6)$$

$$\log \hat{\gamma}_1 + \log \hat{\gamma}_2 > 0. \quad (2.7)$$

The conditions (2.6) and (2.7) can be rewritten as $\gamma_1\gamma_2 > 1$ and $\hat{\gamma}_1\hat{\gamma}_2 > 1$. These in turn can be rewritten as

$$\pi_{11} + \pi_{12} - \pi_{21} - \pi_{22} - (\pi_{11} - \pi_{21})(\pi_{22} - \pi_{12}) > 0, \quad (2.8)$$

$$\pi_{21} + \pi_{22} - \pi_{11} - \pi_{12} - (\pi_{11} - \pi_{21})(\pi_{22} - \pi_{12}) > 0. \quad (2.9)$$

It can be easily seen that equation (2.9) is never holding. Hence, by Proposition 2, if the inequality (2.8) holds then we have that $\lim_{t\to\infty} \sigma_1 = 1$, whereas if (2.8) does not hold we have that $\lim_{t\to\infty} \sigma_1$ has full support over $\{1, 2\}$. □

For the process to select the best action, the two actions need to perform significantly differently. That is, having action 1 better than action 2, $\pi_{11} + \pi_{12} - \pi_{21} - \pi_{22} > 0$, is not enough for the process to select the best action.

Now we present the intuition for the proof of Proposition 2 for the case where $g(\pi_{ij}) = \pi_{ij}$. The proof of Proposition 2 relies partially on the analysis by Ellison and Fudenberg (1995).

In Ellison and Fudenberg (1995), the realization of states of nature is independent of past values of states. In order to be able to apply Ellison and Fudenberg's analysis to our setting, we proceed as follows. Given that the transition matrix P is irreducible and aperiodic, the state of nature many periods ahead is independent of the state of nature today. This means that by the law of large numbers, we can take the probability of each state being realized many periods ahead as the limiting probability placed on it by the Markov chain. Therefore, for the rest of the exposition we consider that the realization of states is independent of past values. For a formal proof the reader is referred to Lemma 4 in the Appendix.

Assume, for the simplicity of the exposition, that there are only two states of nature. Let $1-p$ be the probability by which state 1 occurs. Since the process spends almost no time at its intermediate values, it suffices to examine the convergence of the variable σ_i when it is close to its boundary values (0 and 1). To make the exposition clearer, we focus on the sequence $\sigma_2 = 1 - \sigma_1$. Imagine that σ_2 is arbitrarily close to 0. Then we can rewrite (2.3) as follows:

$$\sigma_2^{t+1}|_j = \gamma_j \sigma_2^t + o(\sigma_2^t) \tag{2.10}$$

where $\gamma_j = 1 + g(\pi_{2j}) - g(\pi_{1j})$ for $j \in \{1, 2\}$ and $o(\sigma_2^t)$ is a term of order higher than σ_2 and hence is negligible when σ_2 is arbitrarily small. Without loss of generality we can assume that $\pi_{11} > \pi_{21}$, which implies $\pi_{12} < \pi_{22}$. Then we can rewrite (2.10) as

$$\sigma_2^{t+1}|_j = \begin{cases} \gamma_1 \sigma_2^t + o(\sigma_2^t) & \text{if } \pi_{1j} \geq \pi_{2j} \\ \gamma_2 \sigma_2^t + o(\sigma_2^t) & \text{otherwise} \end{cases}.$$

Since $\pi_{11} > \pi_{21}$ and $\pi_{12} < \pi_{22}$ we have that $\gamma_2 > 1 > \gamma_1 > 0$. Finally, note that $\pi_{1j} \geq \pi_{2j}$ with probability $1-p$ and $\pi_{1j} < \pi_{2j}$ with probability p.

The sequence σ_2 converges to 0, or σ_1 converges to 1, if and only if the sequence $x = \{x^t\}_{t=0}^{\infty}$ with $x^t = \log \sigma_2^t$ converges to $-\infty$. The process for x when σ_2^t is close to 0 can be approximated by

$$x^{t+1} = \begin{cases} \log \gamma_1 + x^t & \text{with probability } 1-p \\ \log \gamma_2 + x^t & \text{with probability } p \end{cases}.$$

Therefore, $E_t(x^{t+1}) = (1-p)\log \gamma_1 + p \log \gamma_2 + x^t$. Hence, if $(1-p)\log \gamma_1 + p\log \gamma_2 > 0$ then $E_t(x^{t+1}) > x^t$, which implies that x is a sub-martingale. Thus, by the Martingale Convergence Theorem, if $(1-p)\log \gamma_1 + p \log \gamma_2 > 0$ then x cannot converge to $-\infty$ and hence σ_2 cannot converge to 0. Which implies that σ_1 does not converge to 1.

Ellison and Fudenberg's (1995) result is presented here for the readers' convenience.

Lemma 1 (Ellison and Fudenberg (1995)). *Let z^t be a Markov Process on (0,1) with*

$$z^{t+1} = \begin{cases} \gamma_1 z^t + o(z_t) & \textit{with probability } 1-p \\ \gamma_2 z^t + o(z_t) & \textit{with probability } p \end{cases}.$$

Suppose that $\gamma_1 < 1 < \gamma_2$.

(a) If

$$\frac{p}{1-p} > -\frac{\log(\gamma_1)}{\log(\gamma_2)},$$

then z^t cannot converge to 0 with positive probability.

(b) *If*

$$\frac{p}{1-p} < -\frac{\log(\gamma_1)}{\log(\gamma_2)},$$

then there are $\delta > 0$ *and* $\varepsilon > 0$ *such that if* $z^0 < \delta$ *then* $P\left(\lim_{t\to\infty} z^t = 0\right) \geq \varepsilon$.

(c) *If*

$$\frac{p}{1-p} > -\frac{\log(\gamma_1)}{\log(\gamma_1)},$$

there is a $\bar{z} > 0$ *such that for all* $z^0 > 0$ *and all* $t \in \{0, 1, \dots\}$, $P\left(z^t < \bar{z}\right) = 0$.

2.4.3 Efficient Learning Rules

We say that a learning rule is efficient if it is able to select to optimal action in the long run. An interesting result is that if foregone payoffs are observed, then it is optimal to disregard this information and to act as if only realized payoffs were observed.

When players observe the performance of both actions they can be "distracted" towards the suboptimal action by the Markov chain. This is because even if the population plays the optimal action with a high probability they can still observe the performance of the suboptimal action. Hence, since the suboptimal action is the best action for some states of nature, randomness can constantly lead some players in the population to adopt the suboptimal action for many periods in time. Thus, the continuous time limit of the process converges to a situation in which the suboptimal action is played with a positive probability. This is formally proven in the next proposition.

Proposition 3. *Under the Stochastic Better Response, for some* $\varepsilon > 0$ *there exists no* $f : [0,1]^2 \to [0,1]$ *such that for all the environments* $(\{\pi_1, \dots, \pi_m\}, P)$ *we have that* $|\tilde{\sigma}_1 - \sigma_1^*| < \varepsilon$.

Proof. Assume, without loss of generality, that $\sum_{j=1}^m \lambda_j \pi_{1j} > \sum_{j=1}^m \lambda_j \pi_{2j}$. Hence, we have that $\sigma_1^* = 1$.

The proof goes by contradiction. Assume that for all $\varepsilon > 0$ there exists a function $f : [0,1]^2 \to [0,1]$ such that for all the environments $(\{\pi_1, \dots, \pi_m\}, P)$, $|\tilde{\sigma}_1 - \sigma_1^*| < \varepsilon$. This can be rewritten as follows: there exists a sequence of functions $f = \{f_n\}_{n=0}^{\infty}$ with $f_n : [0,1]^2 \to [0,1]$ for all $n \geq 0$ such that for all the environments we have that

$$\lim_{\varepsilon\to 0} \lim_{n\to\infty} \tilde{\sigma}_1(f_n) = \sigma_1^* = 1,$$

where $\tilde{\sigma}_1(f_n)$ is the value of $\tilde{\sigma}_1$ associated with the function f_n.

The limit above holds if and only if

$$\lim_{\varepsilon\to 0} \lim_{n\to\infty} \frac{\sum_{j:\pi_{1j}\geq\pi_{2j}} \lambda_j f_n(\pi_j)}{\sum_{j:\pi_{1j}<\pi_{2j}} \lambda_j f_n(\pi_j)} = \infty \tag{2.11}$$

holds.

Take now an environment $E = (\{\pi_1, \pi_2\}, P)$ where $0 < \pi_{11} < \pi_{22}$ and $\pi_{ij} = 0$ for all $i \neq j$. We could consider more general environments but that will only complicate the exposition leaving the logic of the proof unchanged. P is such that action 1 is the optimal one in the long run. That is, given $\pi_{11} < \pi_{22}$ and $\pi_{ij} = 0$ for all $i \neq j$, P is such that $\sum_{j=1}^{2} \lambda_j \pi_{1j} > \sum_{j=1}^{2} \lambda_j \pi_{2j}$. In this situation, equation (2.11) implies that

$$\lim_{\varepsilon \to 0} \lim_{n \to \infty} \frac{\lambda_1 f_n(\pi_1)}{1 - \lambda_1 f_n(\pi_2)} = \infty. \tag{2.12}$$

Given that the transition matrix P is irreducible we have that $\lambda_1 \in (0, 1)$. Thus, we must have that (2.12) holds if and only if the following limit holds.

$$\lim_{\varepsilon \to 0} \lim_{n \to \infty} \frac{f_n(\pi_1)}{f_n(\pi_2)} = \infty \tag{2.13}$$

However, given that $\pi_{11} < \pi_{22}$ and $\pi_{ij} = 0$ for all $i \neq j$, we have that $f_n(\pi_1) < f_n(\pi_2)$ for all $n > 0$. Hence, the sequence f is such that equation (2.13) cannot hold for the environment E, a contradiction. □

The logic behind the proof is that if a learning rule makes the population to select the optimal action in a given environment E', then the rule must magnify the payoffs of each action. This can be seen in equation (2.11), where, according to the learning rule, payoffs are magnify to infinity. However, if this is the case, an environment E can be found such that there is a very rare state for which the payoff of the suboptimal action is much bigger than the payoff of the optimal action for that state. In this situation, the learning rule that make the population to select the best action for environment E' will fail to do so in environment E.

When only realized payoffs are observed, a different force operates. Once the population is almost always playing the optimal action, it is very difficult for players to take notice of the periods in which the suboptimal action is giving more payoff than the optimal action. A drawback for the population under this informational setting is that if both actions perform not too differently in terms of payoffs, the population may lock on the suboptimal action forever. However, a learning rule can be designed such that this inefficiency is avoided.

The next result states two important features about efficiency rules under the General Reinforcement Rule. The first one is that if learning is sufficiently cautious in that the magnitude of payoffs is diminished then the population will select the optimal action. The second important feature is that how cautious the learning has to be depends on how big the difference in the long run average payoff of both actions is. The more both actions differ in terms of long run performance, the more cautious the learning has to be. This implies that while a learning rule that is very cautious may not be able to make the population to select the best action, this will only happen in environments where the two actions perform very similarly in the long run. Hence, when cautious learning is exhibited, the possible loss in payoff from not selecting the best action is small.

Proposition 4. *Under the General Reinforcement Rule, assume* $g : [0,1] \to [-1,1]$ *is given by*

$$g(\pi_{ij}) = x\pi_{ij}$$

where

$$x = \frac{1 + 4\varepsilon - \sqrt{1 + 8\varepsilon}}{4\varepsilon}$$

for some $\varepsilon > 0$*. If* $|\sum_{j=1}^{m} \lambda_j \pi_{1j} - \sum_{j=1}^{m} \lambda_j \pi_{2j}| > \varepsilon$*, then we have that* $\lim_{t\to\infty} \sigma_1^t = \sigma_1^*$.

Proof. Assume, without loss of generality, that $\sum_{j=1}^{m} \lambda_j \pi_{1j} > \sum_{j=1}^{m} \lambda_j \pi_{2j}$. Hence, we have that $\sigma_1^* = 1$. Moreover, given the inequality $|\sum_{j=1}^{m} \lambda_j \pi_{1j} - \sum_{j=1}^{m} \lambda_j \pi_{2j}| > \varepsilon$, we must have that $\sum_{j=1}^{m} \lambda_j (x\pi_{1j} - x\pi_{2j}) > x\varepsilon$ for all $x > 0$.

Using the first order Taylor series for the logarithmic function around 1 we get that

$$\log(1 + x\pi_{1j} - x\pi_{2j}) = x\pi_{1j} - x\pi_{2j} + R_1(1 + x\pi_{1j} - x\pi_{2j}),$$

where $R_1(1 + x\pi_{1j} - x\pi_{2j})$ is the remainder term and $x > 0$. Using the Lagrange form we can rewrite the remainder term as

$$R_1(1 + x\pi_{1j} - x\pi_{2j}) = \frac{-1/y^2}{2}(1 + x\pi_{1j} - x\pi_{2j} - 1)^2,$$

where y lies between 1 and $1+x\pi_{1j} - x\pi_{2j}$. We can bound the absolute value of the remainder term in the following way:

$$\begin{aligned} |R_1(1 + x\pi_{1j} - x\pi_{2j})| &\leq \frac{1/(1-x)^2}{2}(x\pi_{1j} - x\pi_{2j})^2 \\ &\leq \frac{x^2}{2(1-x)^2}. \end{aligned}$$

Moreover, we have that

$$\begin{aligned} \log(1 + x\pi_{1j} - x\pi_{2j}) &= x\pi_{1j} - x\pi_{2j} + R_1(1 + x\pi_{1j} - x\pi_{2j}) \\ &\geq x\pi_{1j} - x\pi_{2j} - |R_1(1 + x\pi_{1j} - x\pi_{2j})|. \end{aligned}$$

This can be rewritten as

$$\begin{aligned} \sum_{j=1}^{m} \lambda_j \log(1 + x\pi_{1j} - x\pi_{2j}) &\geq \sum_{j=1}^{m} \lambda_j \left(2x\pi_{1j} - x\pi_{2j} - |R_1(1 + x\pi_{1j} - x\pi_{2j})|\right) \\ &> x\varepsilon - \frac{x^2}{2(1-x)^2}. \end{aligned}$$

Rivas, Javier (2008), Cooperation in Repeated Games, Bounded Rational Learning and the Adoption of Evolving Technologies
European University Institute 10.2870/26189

If we take $x > 0$ to be the minimum solution to the equation

$$x\varepsilon - \frac{x^2}{2(1-x)^2} = 0,$$

we get that

$$x = \frac{1 + 4\varepsilon - \sqrt{1+8\varepsilon}}{4\varepsilon}. \tag{2.14}$$

Thus, setting $x > 0$ as in equation (2.14) yields

$$\sum_{j=1}^{m} \lambda_j \log \gamma_j \quad > \quad 0. \tag{2.15}$$

Similar arguments show that

$$\begin{aligned} \sum_{j=1}^{m} \lambda_j \log(1 - x\pi_{1j} + x\pi_{2j}) \quad &\leq \quad -\sum_{j=1}^{m} \lambda_j x\pi_{1j} - x\pi_{2j} + |R_1(1 - x\pi_{1j} + x\pi_{2j})| \\ &< \quad -x\varepsilon + \frac{x^2}{2(1-x)^2}. \end{aligned}$$

Hence, setting again $x > 0$ as in equation (2.14) yields

$$\sum_{j=1}^{m} \lambda_j \log \hat{\gamma}_j \quad < \quad 0. \tag{2.16}$$

Finally, combining inequalities (2.15) and (2.16) with Proposition 2 we get that if $g(\pi_{ij}) = x\pi_{ij}$, where we set $x > 0$ as in equation (2.14), and if $|\sum_{j=1}^{m} \lambda_j \pi_{1j} - \sum_{j=1}^{m} \lambda_j \pi_{2j}| > \varepsilon$, then we have that $\lim_{t\to\infty} \sigma_1^t = \sigma_1^*$. $\square$

Note that if we set $g(\pi_{ij})$ as in Proposition 4, then $\lim_{\varepsilon\to 0} g(\pi_{ij}) = 0$. That is, a rule that makes the population able to select the best action in all the environments must exhibit arbitrarily slow learning.

2.5 Discussion

A way of enriching the model could be by adding idiosyncratic perturbations to payoffs. This could be done by adding ε_{ht} to each payoff π_{ij}. ε_{ht} are normally distributed zero mean random variables that are independent across players h and time t. Since the rules we consider under both scenarios can treat payoffs in a non-linear way, it is not true that the process will converge to the same values as compared to the case without noise. The reason is the same as why, for instance, $E(x^2) \neq E\left((x+\varepsilon)^2\right)$ with $E(\varepsilon) = 0$. However, it can easily be

Rivas, Javier (2008), Cooperation in Repeated Games, Bounded Rational Learning and the Adoption of Evolving Technologies
European University Institute
10.2870/26189

verified that adding noise makes no difference to our results for all the learning rules that treat payoffs linearly. Rules that treat payoffs linearly include the standard best response and the bernoulli best response, for the case where foregone payoffs are observed, and the Cross Learning Rule and the rules in BMS, for the case where foregone payoffs are not observed.

One might argue that if players had means of comparing the payoff of the same action across different time periods, they could recall different payoff realizations over time and have significantly more information about the world they are living in. However, as showed by Rustichini (1999) in a setting very similar to ours, even if players had infinite memory and could make this comparison, it is not true that they will learn the best action for sure.

2.5.1 Relating our results for the Stochastic Better Response with Kosfeld et al. (2002)

Kosfeld et al. (2002) present a setting where a finite set of players play a normal-form game. Each period players update their strategies myopically in the following way. They increase the probability of playing an action if and only if that action is a best response to the action played by the other players. If there are many actions that are a best response, the increase in probability is shared equally among the actions that are a best response. Formally, let $\sigma_i^t(j)$ be the probability by which player j plays action i at time t. Define s_{-j} as the actions played by all the players but j. Finally, let $B_j(s_{-j})$ be the set of actions that are a best response for player j to s_{-j} and let $|B_j(s_{-j})|$ be the cardinality of $B_j(s_{-j})$. The evolution in the strategies of every player j is governed by

$$\sigma_i^{t+1}(j) = \begin{cases} (1-\mu)\sigma_i^t(j) + \mu/|B_j(s_{-j})| & \text{if } s_j \in B_j(s_{-j}) \\ (1-\mu)\sigma_i^t(j) & \text{otherwise,} \end{cases} \tag{2.17}$$

where $\mu \in (0,1)$ is exogenously given.

Comparing this rule with the Stochastic Better Response there are two points worth noting. First, the rule in Kosfeld et al. (2002) is a particular case of the Stochastic Better Response. Second, and most importantly, in our model players play against nature and not against themselves. Hence, in Kosfeld et al.ś (2002) setting, players best respond to the actions of other players while in our setting players best respond to the actions of nature.

Kosfeld et al. (2002) show that the continuous time limit of their process, when μ is made arbitrarily small, converges to a so-called Best-Reply Matching Equilibrium. In a Best-Reply Matching Equilibrium, for every player, the probability of playing a given action is equal to the probability by which that action is a best response given the strategies of the other players.

Their result and our result for the Stochastic Better Response have the same intuition behind them and in some situations are equivalent. Given that in our setting there are only two action we can rewrite (2.17) as follows.

$$\sigma_1^t|_j = \begin{cases} \sigma_1^t + \sigma_2^t \mu & \text{if } \pi_{1j} \geq \pi_{2j} \\ \sigma_1^t - \sigma_1^t \mu & \text{otherwise} \end{cases}$$

In Proposition 1 we proved that the sequence σ_1 defined above converges in probability to

$$\begin{aligned} \hat{\sigma} &= \frac{\sum_{j:\pi_{1j} \geq \pi_{2j}} \lambda_j}{\sum_{j=1}^{m} \lambda_j} \\ &= \sum_{j:\pi_{1j} \geq \pi_{2j}} \lambda_j. \end{aligned}$$

That is, σ_i^t, which is the probability of playing action i, converges to the limiting probability that action i is a best response to the environment. Hence, the population strategies match the nature's strategies, exactly as predicted by the Best-Reply Matching Equilibrium.

In our results for the Stochastic Better Response we consider a much bigger set of rules than do Kosfeld et al. (2002). In particular, Kosfeld et al. (2002) only consider one rule. However, for the specific rule used by Kosfeld et al. (2002), their results and ours come from two different settings, as in their setting players play against each other while in our setting players play against nature.

2.6 Conclusions

In this paper we investigated learning within an environment that changes according to a Markov chain and where players learn according to reinforcement. The payoff of each possible action depends on the state of nature. Since transition between states follows a Markov Chain, there is correlation between today's state and tomorrow's state of nature. We studied two different scenarios, one in which realized and foregone payoffs are observed and another in which only realized payoffs are observed. Our contribution to the literature relies on the fact that we studied reinforcement learning in a setting where the realization of the state of nature is correlated with the past.

The literature has focused on the study of learning only in a setting where the realization of states (or the shocks to payoffs) is independent of its past values. The reason for this is the technical complexities involved in dealing with the correlated realization of states.

There are several questions left for further research. For the case where foregone payoffs are observed, we only characterized the asymptotic distribution when the learning step goes to zero. For the case where foregone payoffs are not observed we are unable to quantify the probabilities of reaching each endpoint where the process does not converge deterministically to a single point.

The present piece of work explores learning in two very general scenarios but there are other settings that could be of interest. For instance, how does local interaction affect learning when the environment changes according to a Markov chain? What if there are non-stochastic idiosyncratic payoff differences among players? Our paper also tried to shed some light on the techniques that could be used for dealing with such environments. We expect that in the future more papers dealing with non stationary environments will appear.

References

1. Ben-Porath, E., Dekel, E. and Rustichini, A. (1993): "On the Relationship between Mutation Rates and Growth Rates in Changing Environment". *Games and Economic Behavior* 5 (4), 576-603.

2. Benaïm, M. and Weibull, J. (2003): "Deterministic Approximation of Stochastic Evolution in Games". *Econometrica* 71 (3), 873-903.

3. Börgers, T., Morales, A. and Sarin, R. (2004): "Expedient and Monotone Learning Rules". *Econometrica* 71 (2), 383-405.

4. Börgers, and Sarin, R. (1997): "Learning Through Reinforcement and Replicator Dynamics". *Journal of Economic Theory* 77, 1-14.

5. Camerer, C. and Ho, T. H. (1999): "Experienced-Weighted Attraction Learning in Normal Form Games". *Econometrica* 67 (4), 827-874.

6. Conlisk, J. (1996): "Why Bounded Rationality?". *The Journal of Economic Literature* 34, 669-700.

7. Cross, J. (1973): "A Stochastic Learning Model of Economic Behavior". *The Quarterly Journal of Economics* 87, 239-266.

8. Ellison, G. and Fudenberg, D. (1995): "Word-of-Mouth Communication and Social Learning". *The Quarterly Journal of Economics* 110 (1), 93-125.

9. Erev, I. and Roth, A. (1998): "Predicting How People Play Games: Reinforcement Learning in Experimental Games with Unique, Mixed Strategy Equilibria". *The American Economic Review* 88 (4), 848-881.

10. Fudenberg, D. and Harris, C. (1992): "Evolutionary Dynamics with Aggregate Shocks". *Journal of Economic Theory* 57, 420-441.

11. Hirshleifer, D. and Welch, I. (2002): "An Economic Approach to the Psychology of Change: Amnesia, Inertia, and Impulsiveness". *Journal of Economics & Management Strategy* 11 (3), 379-421.

12. Kosfeld, M., Droste, E. and Voorneveld, M. (2002): "A Myopic Adjustment Leading to Best Reply Matching". *Games and Economic Behavior* 40, 270-298.

13. Rubinstein, A. (1998): "Modeling Bounded Rationality". The MIT press, Cambridge, Massachusetts.

14. Rubinstein, A. (2002): "Irrational Diversification in Multiple Decision Problems". *European Economic Review* 46, 1369-1378.

15. Rustichini, A. (1999): "Optimal Properties of Stimulus Response Learning Models". *Games and Economic Behavior* 29, 244-273.

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

17. Samuelson, L. (1994): "Stochastic Stability in Games with Alternative Best Replies". *Journal of Economic Theory* 64 (1), 35-65.

18. Shanks, D., Tunney, R. and McCarthy, J. (2002): "A Re-examination of Probability Matching and Rational Choice". *Journal of Behavioral Decision Making* 15, 233-250.

19. Siegel, S. and Goldstein, D. A. (1959): "Decision Making Behavior in a Two-Choice Uncertain Outcome Situation". *Journal of Experimental Psychology* 57 (1), 37-42.

20. Vulkan, N. (2000): "An Economist's Perspective on Probability Matching". *Journal of Economic Surveys* 14 (1), 101118.

Appendix

2.A.1 Proof of Proposition 1

We begin by proving the following lemma.

Lemma 2. *For any* $\varepsilon > 0$ *there exists a* $\hat{\mu} > 0$, $\hat{t}(\varepsilon) > 0$ *and a sequence* $y = \{y^t\}_{t=\hat{t}}^{\infty}$ *given by* $y^{\hat{t}} = \sigma_1^{\hat{t}}$ *and recursively for* $t > \hat{t}$

$$y^{t+1} = \begin{cases} y^t + 2(1-y^t)\mu \sum_{j:\pi_{1j} \geq \pi_{2j}} \lambda_j f(\pi_j) & \textit{with probability } 1/2 \\ y^t + 2y^t \mu \sum_{j:\pi_{1j} < \pi_{2j}} \lambda_j f(\pi_j) & \textit{with probability } 1/2 \end{cases},$$

such that for any $\mu < \hat{\mu}$ *we have that*

$$P\left(\lim_{t\to\infty} |\sigma_1^t - y^t| > \varepsilon\right) = 0.$$

Proof. In the main text we defined $h < m$ as the minimum natural number such that $\pi_{1j} \geq \pi_{2j}$ for $j \leq h$ and $\pi_{2j} > \pi_{1j}$ for $j > h$. For any given $\varepsilon > 0$ define now the sequence $\hat{\sigma}_1 = \{\hat{\sigma}_1^t\}_{t=\hat{t}(\varepsilon)}^{\infty}$ as $\hat{\sigma}_1^{\hat{t}(\varepsilon)} = \sigma_1^{\hat{t}(\varepsilon)}$ and recursively for $t > \hat{t}(\varepsilon)$

$$\hat{\sigma}_1^{t+1} = \begin{cases} \hat{\sigma}_1^t + \hat{\sigma}_2^t \mu f(\pi_1) & \text{with probability } \lambda_1 \\ & \vdots \\ \hat{\sigma}_1^t + \hat{\sigma}_2^t \mu f(\pi_h) & \text{with probability } \lambda_h \\ \hat{\sigma}_1^t - \hat{\sigma}_1^t \mu f(\pi_{h+1}) & \text{with probability } \lambda_{h+1} \\ & \vdots \\ \hat{\sigma}_1^t - \hat{\sigma}_1^t \mu f(\pi_m) & \text{with probability } \lambda_m \end{cases}.$$

Fix $\hat{t}(\varepsilon)$ to be the minimum natural number such that for any given $t > \hat{t}(\varepsilon)$,

$$P\left(|E_0(\sigma_1^t) - E_0(\hat{\sigma}_1^t)| > \varepsilon\right) = 0. \tag{2.18}$$

The existence of such $\hat{t}(\varepsilon)$ is guaranteed by the fact that the transition matrix P is irreducible and aperiodic and by the Perron-Frobenius theorem applied to P. In an abuse of notation, from now on we will simply write $\hat{t}$ to denote $\hat{t}(\varepsilon)$.

Since E_0 is linear in both $\hat{\sigma}_1^t$ and y^t, we have that for all $t > \hat{t}$, $\hat{\sigma}_1^t = y^t$ if and only if $E_0(\hat{\sigma}_1^{t+1}) = E_0(y^{t+1})$. Thus, given that $y^{\hat{t}} = \hat{\sigma}_1^{\hat{t}}$, that E_0 is linear in both $\hat{\sigma}_1^t$ and y^t and that the distribution of both y and $\hat{\sigma}_1$ is aperiodic, we have that

$$E_0(y^{\hat{t}+k}) = E_0(\hat{\sigma}_1^{\hat{t}+k}) \tag{2.19}$$

for all $k \in \mathbb{N}$.

Given the definition of y and equations (2.18) and (2.19) we must have that for any $\varepsilon > 0$ and any $t > \hat{t}$,

$$P\left(|E_0(\sigma_1^t) - E_0(y^{t+1})| > \varepsilon\right) = 0. \tag{2.20}$$

Given the specification of σ_1 and the definitions of $\hat{\sigma}_1$ and y, as μ gets arbitrarily small, the variance of σ_1, $\hat{\sigma}_1$ and y gets arbitrarily small as well. Formally, for any $\varepsilon > 0$ there exists a $\hat{\mu} > 0$ and a $t > \hat{t}$ such that for any $\mu < \hat{\mu}$ and $k \in \mathbb{N}$ we have that $Var_t(\sigma_1^{t+k}) < \varepsilon$, $Var_t(\hat{\sigma}_1^{t+k}) < \varepsilon$ and $Var_t(y^{t+k}) < \varepsilon$.

Assume that σ_1 does not converge in probability to y. As μ goes to zero the variance of both σ_1 and y goes to zero. Hence, both variables will converge in probability to a single point. That is, for all $\delta > 0$ there exists $\bar{\sigma}_1$, $\bar{y}$, $\bar{\mu} > 0$ and $\bar{t} \in \mathbb{N}$ such that for all $\mu < \bar{\mu}$ and $t > \bar{t}$, $P\left(|\sigma_1^t - \bar{\sigma}_1| > \delta\right) = 0$ and $P\left(|y_1^t - \bar{y}| > \delta\right) = 0$. This can also be rewritten as $P\left(|E_0(\sigma_1^t) - \bar{\sigma}_1| > \delta\right) = 0$ and $P\left(|E_0(y_1^t) - \bar{y}| > \delta\right) = 0$.

If $\bar{\sigma}_1 \neq \bar{y}$, them we must have that exists a $\gamma > 0$ and a $t \in \mathbb{N}$ such that

$$P\left(|E_0(\sigma_1^{t+k}) - E_0(y^{t+k})| > \gamma\right) > 0$$

for all $k \in \mathbb{N}$, which contradicts equation (2.20). Hence, given that $P\left(|\sigma_1^t - \bar{\sigma}_1| > \delta\right) = 0$, $P\left(|y_1^t - \bar{y}| > \delta\right) = 0$ and $\bar{\sigma}_1 = \bar{y}$, we must have that for any $\varepsilon > 0$ there exists a $\hat{\mu}$ such that for all $\mu < \hat{\mu}$,

$$P\left(\lim_{t\to\infty} |\sigma_1^t - y^t| > \varepsilon\right) = 0.$$

□

In the next lemma we establish that y converges in probability to $\tilde{\sigma}$.

Lemma 3. *For any $\varepsilon > 0$ there exists a $\hat{\mu} > 0$ such that for any $\mu < \hat{\mu}$ we have that*

$$P\left(\lim_{t\to\infty} |y^t - \tilde{\sigma}| > \varepsilon\right) = 0.$$

Proof. First, note that the point $y^t = \tilde{\sigma}$, with $\tilde{\sigma}$ as defined in Proposition 1, solves the equation

$$y^t + 2(1-y^t)\mu \sum_{j:\pi_{1j}\geq\pi_{2j}} \lambda_j f(\pi_j) = y^t - 2y^t\mu \sum_{j:\pi_{1j}<\pi_{2j}} \lambda_j f(\pi_j).$$

Define now the sequence $y_1 = \{y_1^t\}_{t=\hat{t}}^{\infty}$ as follows

$$y_1^t = \begin{cases} y^t & \text{if } y^t \geq \tilde{\sigma} \\ \tilde{\sigma} & \text{otherwise} \end{cases}.$$

Hence, we have that $E_0(y^t) \leq E_0(y_1^t)$ for all $t > \hat{t}$. Note that if $y^t > \tilde{\sigma}$ then we have that $E_0(y^{t+1}) < E_0(y^t)$. This implies that $E_0(y_1^{t+1}) < E_0(y_1^t)$ for all $y_1^t > \tilde{\sigma}$ and $E_0(y_1^{t+1}) = E_0(y_1^t)$ for $y_1^t = \tilde{\sigma}$. Therefore, y_1 is a super-martingale with lower-bound $\tilde{\sigma}$. Thus, by the Martingale convergence theorem, $\lim_{t\to\infty} y_1^t$ exists. Given that $E_0(y_1^{t+1}) < E_0(y_1^t)$ for all $y_1^t > \tilde{\sigma}$ and $E_0(y_1^{t+1}) = E_0(y_1^t)$ for $y_1^t = \tilde{\sigma}$, we must have that $\lim_{t\to\infty} y_1^t = \tilde{\sigma}$. This implies that y_1 converges in probability to $\tilde{\sigma}$.

Define now the sequence $y_2 = \{y_2^t\}_{t=\hat{t}}^{\infty}$ as follows:

$$y_1^t = \begin{cases} y^t & \text{if } y^t \leq \tilde{\sigma} \\ \tilde{\sigma} & \text{otherwise} \end{cases} .$$

Hence, we have that $E_0(y^t) \geq E_0(y_1^t)$ for all $t > \hat{t}$. Note that if $y < \tilde{\sigma}$ then we have that $E_0(y^{t+1}) > E_0(y^t)$. This implies that $E_0(y_2^{t+1}) > E_0(y_2^t)$ for all $y_2^t < \tilde{\sigma}$ and $E_0(y_2^{t+1}) = E_0(y_2^t)$ for $y_2^t = \tilde{\sigma}$. Therefore, y_2 is a sub-martingale with upper-bound $\tilde{\sigma}$. Thus, by the Martingale convergence theorem, $\lim_{t\to\infty} y_2^t$ exists. Given that $E_0(y_2^{t+1}) > E_0(y_2^t)$ for all $y_2^t < \tilde{\sigma}$ and $E_0(y_2^{t+1}) = E_0(y_2^t)$ for $y_1^t = \tilde{\sigma}$, we must have that $\lim_{t\to\infty} y_2^t = \tilde{\sigma}$. This implies that y_2 converges in probability to $\tilde{\sigma}$.

Hence, we have that for any $\varepsilon > 0$ exists a $\hat{\mu}$ such that for all $\mu < \hat{\mu}$,

$$\begin{aligned} P\left(\lim_{t\to\infty} |y_1^t - \tilde{\sigma}| > \varepsilon\right) &= 0 \\ P\left(\lim_{t\to\infty} |y_2^t - \tilde{\sigma}| > \varepsilon\right) &= 0. \end{aligned}$$

We know, given the definition of y, that for any $\varepsilon > 0$ there exists a $\hat{\mu} > 0$ and a $t > \bar{t}$ such that for any $\mu < \hat{\mu}$ and $h > t$ we have that $Var_t(y^{t+h}) < \varepsilon$. This, together with the fact that $E_0(y^t) \leq E_0(y_1^t)$ and $E_0(y^t) \geq E_0(y_1^t)$ for all $t > \hat{t}$ implies that for all $t > \max\{\bar{t}, \hat{t}\}$ we must have that $\lim_{t\to\infty} y^t = \tilde{\sigma}$. This implies that y converges in probability to $\tilde{\sigma}$. □

Now we are able to prove the result in Proposition 1.

Proof of Proposition 1. We know from Lemma 2 that σ_1 converges in probability to y. From Lemma 3 we also know that y converges in probability to $\tilde{\sigma}$. Hence, we must have that σ_1 converges in probability to $\tilde{\sigma}$. This is the result of the Proposition. □

2.A.2 Proof of Proposition 2

Whenever σ_1^t is arbitrarily close to 0 we have that

$$\sigma_1^{t+1}|_j = \sigma_1^t(1 + g(\pi_{2j}) - g(\pi_{1j})) + o(\sigma_1^t).$$

Define $\gamma_j = 1 + g(\pi_{2j}) - g(\pi_{1j})$ for all $j \in \{1, \ldots, m\}$. Hence, given that g is increasing, we have that $\gamma_i \leq 1 < \gamma_j$ if and only if $\pi_{1i} \geq \pi_{2i}$ and $\pi_{1j} < \pi_{2j}$. We can approximate the equation for the evolution of the sequence σ_1 when σ_1^t is arbitrarily close to 0 as follows:

$$\sigma_1^{t+1}|_j = \gamma_j \sigma_1^t.$$

Lemma 4. *For any $\bar{\sigma}_1^t \in (0,1)$ and any $\varepsilon > 0$ there exists a $\sigma_1^t < \bar{\sigma}_1^t$ and a $\bar{k} \in \mathbb{N}$ such that for $k > \bar{k}$*

$$P\left(\left|\sigma_1^{t+k} - \hat{\sigma}_1^{t+k}\right| > \varepsilon\right) = 0,$$

where $\hat{\sigma}_1^{t+\bar{k}} = \sigma_1^{t+\bar{k}}$ and

$$\hat{\sigma}_1^{t+k+1} = \begin{cases} \gamma_1 \sigma_1^{t+k} & \text{with probability } \lambda_1 \\ & \vdots \\ \gamma_m \sigma_1^{t+k} & \text{with probability } \lambda_m \end{cases}$$

for $k > \bar{k}$.

Proof. Given that the transition matrix P is irreducible and aperiodic and that the number of states is finite, we have the standard result that the empirical distribution of states converges to the limiting distribution of states. This can be rewritten as: for any $\delta > 0$ there exists a $\bar{k}(\delta) \in \mathbb{N}$ such that for $k > \bar{k}(\delta)$,

$$P\left(\left|\frac{\sum_{t=0}^{k} \mathbb{1}_{\{s^t=j\}}}{k+1} - \lambda_j\right| > \delta\right) = 0 \tag{2.21}$$

for all $j \in \{1, \ldots, m\}$.

We have seen before that if σ_1^t is arbitrarily close to 0 we can write $\sigma_1^{t+1}|j = \gamma_j \sigma_1^t$. In other words, for any $\kappa > 0$ there exists a $\bar{\sigma}_1(\kappa) \in (0,1)$ such that if $\sigma_1^t < \bar{\sigma}_1(\kappa)$ then

$$P\left(\left|\sigma_1^{t+1}|_j - \gamma_j \sigma_1^t\right| > \kappa\right) = 0$$

for all $j \in \{1, \ldots, m\}$. This result can also be expressed as follows. For any $\kappa > 0$ and any $k \in \mathbb{N}$ there exists a $\bar{\sigma}_1(\kappa) \in (0,1)$ such that if $\sigma_1^t < \bar{\sigma}_1(\kappa)$ then

$$P\left(\left|\sigma_1^{t+k+1}|_j - \gamma_j \sigma_1^{t+k}\right| > \kappa\right) = 0. \tag{2.22}$$

Hence, we have the following two facts. First, the probability of a state being realized a sufficiently far way number of periods converges to the limiting distribution of the Markov chain. Second, that $\sigma_1^{t+1}|j$ behaves as $\gamma_j \sigma_1^t$ if σ_1^t is sufficiently small. Then, for k sufficiently large and σ_1^t sufficiently close to 0 we have that for all $j \in \{1, \ldots, m\}$, $\sigma_1^{t+k+1} = \gamma_j \sigma_1^{t+k}$ with probability λ_j. In other words, combining the results in equations (2.21) and (2.22) we can write that for all $\varepsilon > 0$ there exists a $\bar{k}(\varepsilon) \in \mathbb{N}$ and $\bar{\sigma_1}(\varepsilon) \in (0,1)$, such that for all $k > \bar{k}(\varepsilon)$ and $\sigma_1^t < \bar{\sigma_1}(\varepsilon)$ we have that

$$P\left(\left|\sigma_1^{t+k} - \hat{\sigma}_1^{t+k}\right| > \varepsilon\right) = 0,$$

where $\hat{\sigma}_1^{t+\bar{k}} = \sigma_1^{t+\bar{k}}$ and

$$\hat{\sigma}_1^{t+k+1} = \begin{cases} \gamma_1 \sigma_1^{t+k} & \text{with probability } \lambda_1 \\ & \vdots \\ \gamma_m \sigma_1^{t+k} & \text{with probability } \lambda_m \end{cases}$$

for $k > \bar{k}$. □

Lemma 5. *The sequence σ_1 cannot converge to 0 if*

$$\sum_{j=1}^{m} \lambda_j \log \gamma_j > 0.$$

There is a positive probability that the sequence σ_1^t converges to 0 if

$$\sum_{j=1}^{m} \lambda_j \log \gamma_j < 0.$$

Proof. Reasoning as in the proof of Lemma 1 in Ellison and Fudenberg (1995), the sequence σ_1 can converge to zero if and only if the sequence $y = \log \sigma_1$ can converge to $-\infty$. Using again the proof from Lemma 1 in Ellison and Fudenberg (1995) and Lemma 4 in this Appendix, the sequence y can converge to $-\infty$ only if $\sum_{j=1}^{m} \lambda_j \log \gamma_j < 0$. The result follows. □

To study the situation in which the process is arbitrarily close to 1, we proceed as follows. First, we define $w^t = 1 - \sigma_1^t$. Then we apply the analysis above to the variable w^t. Define $\hat{\gamma}_j = 1 + g(\pi_{2j}) - g(\pi_{1j})$. Then we have that for all $\varepsilon > 0$ there exists a $\bar{k} \in \mathbb{N}$ and $\bar{w} \in (0,1)$ such that for all $k > \bar{k}$ and $w^t < \bar{w}$ we have that

$$P\left(\left|w^{t+k} - \hat{w}^{t+k}\right| > \varepsilon\right) = 0,$$

where $\hat{w}^{t+\bar{k}} = w^{t+\bar{k}}$ and

$$\hat{w}^{t+k+1} = \begin{cases} \hat{\gamma}_1 w^{t+k} & \text{with probability } \lambda_1 \\ & \vdots \\ \hat{\gamma}_m w^{t+k} & \text{with probability } \lambda_m \end{cases}$$

for $k > \bar{k}$.

An analogous to Lemma 5 when σ_1^t is close to 1 is the following:

Lemma 6. *The sequence σ_1 cannot converge to 1 if*

$$\sum_{j=1}^{m} \lambda_j \log \hat{\gamma}_j > 0.$$

There is a positive probability that the sequence σ_1 converges to 1 if

$$\sum_{j=1}^{m} \lambda_j \log \hat{\gamma}_j < 0.$$

Summing up the results from lemmas 5 and 6 the result in Proposition 2 follows.

Chapter 3 - The Effects of the Market Structure on the Adoption of Evolving Technologies

3.0 Abstract

We study the speed at which new technologies are adopted depending on how the market power is shared between suppliers and buyers. The suppliers consists of firms that own technologies and sell then to the firms that demand the technologies, which then produce output using these technologies. Three different market structures are considered: one where the suppliers have all the market power; one where the buyers have all the market power and a another where market power is shared and there is competition between suppliers and buyers. Our results suggest, among other things, that competition reduces the pace of adoption of new technologies.

3.1 Introduction

The adoption of new technologies is regarded as one of the main contributors to economic growth (see, for instance, Lucas (1993), Barro and Sala-i-Martin (1995)). The differences in timing of the establishment of new technologies in countries or firms can lead to very different growth rates. Adopting new technologies too quickly may be disadvantageous given the sunk cost the establishment of a new technology carries. On the other hand, delaying the adoption of a new technology can lead to high opportunity costs or to a disadvantageous position with respect to competitors. This tradeoff has been widely studied in the literature.

The literature so far has focused on the pace of the adoption of new technologies from the perspective of a firm, which by adopting a new technology incurs a fixed cost that will be compensated over time by the benefits from having a better technology. In this respect the problem of adopting new technologies was reduced to two basic settings. In the first one, the problem of the firm was an optimal stopping problem (see for example Farzin et al. (1998) and Jovanovic and Nyarko (1996)) where the firm has to decide, given a fixed price, at which point in time to adopt a new technology. In the second setting, firms adopting new technologies play a game in which earlier adoption leads to high costs but to a temporary advantageous position against competitors (see for example Götz (1999) and Chamley and Gale (1994)). We change these two approaches and consider instead the game played between the firms adopting new technologies and the firms that create and price the new technologies. Hence, in our model as opposed to the existing literature, the price of the different available technologies is endogenous. This allows us to study how the timing of the adoption of new technologies is affected by the competition resulting from the interaction between supply and demand.

In our model there is an exogenous process that determines the evolution of a technology parameter. Technologies differ in how productive they are, so different values of the technology parameter means different technologies. The firms selling technologies own the technologies and have no influence in their evolution. Their only role is to price the different available technologies. Firms buying technologies have to decide at which point in time to adopt a new technology. The game is then the following: sellers have the tradeoff between price and time, with higher price means higher income but at a later date. On the other hand, the buyers have the tradeoff between early adoption, which implies an earlier increase in productivity, and late adoption, which implies a greater increase in productivity since a more advanced technology is adopted. Our model explains then how these trade offs are solved when we consider three different market structures that are distinct in how the market power is shared among the buyers and the suppliers.

In the first market structure we consider there is only one firm selling technologies and many firms willing to buy technologies. Hence, in this setting the supply side holds all the market power and buyers act as a price takers. In the second market setting there are many firms supplying technologies and only one firm interested in buying it. In this setting the demand side holds all the market power and, therefore, sellers compete in prices and make profits equal to their outside option of not participating in the market. In the last market setting we consider there is one firm on each side of the market. In this last setting suppliers and buyers compete for the surplus in the economy.

With our model we explore how the different market structures affect the adoption of new technologies. This helps us in understanding why in some industries there is a huge gap between the release of a new technology and its adoption while in some others new technologies are adopted instantly as soon as they are released.

The present paper tries to shed light on the issue mentioned above, speed of adoption, with respect to different market structures. Rather than focusing on the nature of the technology itself or other factors, we chose to study how market power can explain these two issues. We do not claim market power is the only reason why we observe different types of behavior. However, as we shall show, it is a factor that can explain these differences by itself and should be taken into account. Furthermore, this is the first paper that to our knowledge deals with the interaction between sellers of technologies and buyers of technologies. Hence, as a first step to understanding the interaction between supply and demand in technology markets, some simplifying assumptions are required. We assume that each buyer can only buy a new technology once. Similarly, each seller is only allowed to sell a technology once. Doraszelski (2004) and Dixit and Pindyck (1994) presented a model of technology adoption where there is only demand and the pricing of technologies is exogenous. They assume, as we do, that firms only buy a new technology once. Doraszelski (2001) shows that the firm's decision problem, if we allow the firm to buy new technologies more than once, is of the same form after each adoption of a new technology. Unfortunately, the assumption that each seller can only sell technologies once is harder to justify and made only for analytical convenience.

In our results we find that if there is competition between suppliers and buyers, then the adoption of new technologies occurs at a slower pace than when either suppliers or buyers hold all the market power. This suggests that competition between both sides of the market, instead of competition within each side, can delay the adoption of new technologies. A striking result is that when only one side of the market holds all the power adoption occurs at the same pace independently of which one that is. When only one side of the market holds all the power the total surplus in the economy is maximized. Hence, from the point of view of the speed of adoption of new technologies the only actual difference between the situation where only one side holds all the market power is how the total surplus in the economy is divided between the sides of the market.

From the theoretical point of view, many models study the optimal timing of technology adoption. Jovanovic and Nyarko (1996) present a model where the decision maker increases productivity by either learning by doing or by switching to a better technology. The effect of learning by doing for any given technology is bounded and hence there comes a point in which the only way to improve productivity is by upgrading to a better technology. Adopting a new technology is not costly in monetary terms but in productivity terms. When a new technology is adopted, it takes time to learn how to use it. Hence, a new technology brings more possibilities of growth for the long run but decreases productivity in the short run. Karp and Lee (2001) extend this model by introducing discount factors.

Farzin et al. (1998) present a model where the increase in productivity caused by the adoption of the newest technology is know only in expected terms. Adopting a new technology has a sunk cost that is independent of the productivity level of the new technology. In a recent work by Doraszelski (2004) a distinction between technological breakthroughs and engineering refinements is introduced.

Götz (1999) introduces a model of monopolistic competition where the technological improvement happens just once and the cost of adopting such technological improvement decreases over time. In Götz's model delaying the adoption of a new technology gives a comparative advantage in the long run but it is disadvantageous in the short run.

In a paper by Chamley and Gale (1994) a population faces the decision of whether to adopt a new technology or not. The performance of the new technology is unknown but there are information externalities arising when other players adopt the technology. Delaying the adoption of a new technology has benefits given the extra information about that technology that is accumulated over time. However, an early adoption to the new technology can be beneficial given the monopoly power that it provides.

From the empirical perspective, there is no doubt that the timing of technology adoption has been a concern. Hoppe (2002) presents a literature review on this topic. To cite some, Karshenas and Stoneman (1993) present a study on the diffusion of CNC (computer numerically controlled machine tools) in the UK engineering industry. Factors determining the delaying in the adoption of the new technology were found to be, among others, the learning effects and the cost of the new technology. Weiss (1994) studied the adoption of a new process technology called the surface-mount technology by printed circuit board manufacturers.

In the remainder of this section we present a survey on the relevant literature. In section 3.2 the model is presented. Section 3.3 presents our findings for the three different market structures considered. In section 3.4 we present a comparative statics analysis. Finally, section 3.5 concludes.

3.2 The Model

Consider a continuous time model where the two sides of the market, suppliers and buyers, play a repeated game. On the supply side of the market there are firms selling technologies, sellers, while on the demand side of the market there are firms buying technologies, buyers. An exogenous process determines the evolution of new technologies, which are sold by the sellers to the buyers. The seller firms have to put a price to these technologies while buyer firms produce output given an initial level of technology and decide when to buy a better technology. We define $n_s \geq 1$ to be the number of sellers and $n_d \geq 1$ to be the number of buyers. All firms are assumed to be risk neutral.

Three different market settings are considered. These three market settings are explained in detail in their respective subsections, but here we briefly introduce them to the reader. In the first market setting we considered, the supply side holds all the market power and firms in the demand side act as price taker. Consequently, for this setting we assume there is one firm supplying technologies and many firm interested in buying them. Alternatively to one firm selling technologies one can assume that many sellers collude to gain all the market power. In the second market setting, the demand holds all the market power and, hence, it extracts all the surplus in the economy. Hence, in this setting there is one firm interested in buying technologies and many firms producing technologies. The third setting we consider has the supply, that consists of a single firm, and demand, that also consists of a single firm, competing for the surplus in the economy.

As an illustrative example, consider the case of a research lab developing new patents. Sellers then can embed these new patents into products that they then release onto the market. To be a bit more precise consider the case of micro processors for computers. Intel develops new micro processors that are then sold by computer manufactures to consumers. In this example, Intel is represented in the model by the exogenous process determining the evolution of technologies, computer manufactures are represented by the firms selling technologies and consumers deriving utility from buying computers is represented in the model by the firms in the demand side that produce output given a level of technology.

Technologies are denoted by a parameter θ where higher θ means better, more efficient, technology. Seller firms own these technologies. Buyer firms are all endowed with the same initial level of technology, denoted by θ_0. As a simplification, we assume that only two different levels of technology coexists in the economy at any given point in time t. These are the initial level of technology, θ_0, and the newest technology $\theta(t)$. Sellers have no control over the evolution or the level of technology available in the economy. It owns the technologies and its only role is to put a price to the newest technology. If a firm from the supply side sells a given technology θ then it has to pay a fix cost of $C > 0$ for the transaction to take place. All sellers have access to the same set of technologies and pay the same amount C when selling a technology to a firm in the demand side.

From t to $t + \mathrm{d}t$ the state of the newest technology, $\theta(t)$, evolves according to a Geometric Brownian Motion:

$$\mathrm{d}\theta(t) \quad = \quad \alpha\theta(t)\mathrm{d}t + \sigma\theta(t)\mathrm{d}z(t),$$

where $\alpha, \sigma > 0$ and z is a Brownian Motion (defined in the Appendix 3.A.1). The fact that the technologies evolve following a Geometric Brownian Motion and the values of θ_0, α, σ and z are all common knowledge.

The assumption that the evolution of technologies follows a Geometric Brownian Motion is made simply for analytical convenience. Closed form solutions can be found for this dynamics but not for other dynamics that could be used, like Poisson processes or standard Brownian Motion. As it will be clear later when the analysis is presented, the specific motion assumed for the technology has no qualitative implications for our results.

Note that at any point in time there may be technological regress, $\theta(t + \mathrm{d}t) < \theta(t)$. This fact has no implications for the model and our results as if it is optimal not to buy the technology θ at a given price then it is not optimal to buy technology $\tilde{\theta} < \theta$ at this same price. To interpret the technological regress we can assume that when there is technological regress these new technologies are inventions that simply did not work out and were never made public.

The problem for the buyers at each point in time is whether to adopt the newest technology or to stick with the technology currently in use. Following Doraszelski (2004) and Dixit and Pindyck (1994) we assume that the adoption of a new technology is a one-time irreversible decision. Hence, once a buyer has adopted a new technology, it is stuck forever with that technology. Doraszelski (2001) shows that the firm's decision problem, if we allow the firm to buy new technologies more than once, is of the same form after each adoption of a new technology. From a mathematical point of view, the problem of the firms on the demand side is an optimal stopping problem. Similarly, we assume that each seller only sells technologies once. Hence, in the market settings where there is only one firm supplying technologies, once this firm has sold a technology no further adoption of technologies occurs. This latter assumption is made for analytical convenience as it reduces the strategy space in a more tractable way.

The timing of the game played between sellers and buyers is as follows: At time $t = 0$ each seller $i \in \{1, \ldots, n_s\}$ decides on a price to charge for the newest technology $I_i \geq 0$. Then, at every period $t > 0$ and given the current level of technology $\theta(t)$, each buyer decides whether to buy technology $\theta(t)$ at price $I = \min_{i \in \{1,\ldots,n_s\}} I_i$ or to wait.

Let $I_i \geq 0$ be the price charged by seller $i \in \{1, \ldots, n_s\}$. Let $\mathcal{H}(t)$ be the history of prices, technologies, and decisions of the buyers up to period t. Hence, the element $h(k) \in \mathcal{H}(t)$ for $k \leq t$ consists of the level of technology at time k, $\theta(k)$, the prices of all sellers $\{I_i\}_{i=1}^{n_s}$ and the decision of all buyers at time k of whether to buy the technology $\theta(k)$ at given prices $\{I_i\}_{i=1}^{n_s}$ or not.

A strategy for a seller $i \in \{1, \ldots, n_s\}$ consists of a price I_i that depends on previous history $\mathcal{H}(t)$: $I_i : \mathbb{N}^2 \to \mathbb{R}$. Note that we are using I_i for both the action and the strategy. This should not give rise to any confusion in our context. The price in the economy of the newest technology is given by $I = \min_{i \in \{1,\ldots,n_s\}} I_i$. A strategy for a buyer consists of the function $S : \mathcal{H} \times I \times \theta \to d$ where $d = \{\text{buy}, \text{ wait}\}$ is the decision of the firm of whether to buy the technology θ at a price I or to wait. The equilibrium concept we use throughout the paper is the standard Nash equilibrium (henceforth NE).

After adopting a new technology θ, the discounted stream of profits of a buyer is given by

$$\int_0^\infty \pi(\theta) e^{-rs} \mathrm{d}s - I \quad = \quad \frac{\pi(\theta)}{r} - I$$

where $\pi(\theta)$ is the instantaneous profit of a firm on the demand side from using a technology θ and $r > 0$ is the interest rate. We assume π to be strictly increasing. Whenever necessary we use a specific functional form for π. We chose to focus on the more natural example, the one that can be derived from the Cobb-Douglas production function (see Farzin et. al. (1998)). In this case we have that $\pi(\theta) = \phi\theta^b$ where $\phi > 0$ and $b > 1$.

When a seller sells a technology θ it has to pay a fixed cost $C > 0$ for the transaction to take place. In order to make the selling of technologies possible we assume $C < \pi(\theta_0)/r$. If technology θ is sold at price I at time t then the present value of the profits of a buyer are given by

$$(I - C)\, E(e^{-rt}).$$

In case a set of firms want to buy a given level of technology at the same price, then each of these buyers has equal probability of making the purchase. Similarly in case a set of firms wants to sell a given technology at the same price, then each of these sellers has equal probability of selling the technology.

We define θ^* as the value of θ at which the first purchase of a new technology takes place. Hence, θ^* is a function of the price I. In Farzin et. al. (1998) Doraszelski (2004) and Dixit and Pindyck (1994) θ^* represents the level at which it is optimal to switch technology. In our paper the interpretation of θ^* is the same as in theirs. Moreover, the value of θ^* is our measure of the speed at which new technologies are adopted. Higher θ means that more time has to pass before a new technology is adopted and hence we say that the adoption occurs at a slower pace.

3.3 Speed of Adoption of Technologies

3.3.1 The Supply Side Holds All the Market Power

In this setting there is one seller while there are at least 2 firms interested in buying the technology. That is, $n_s = 1$ and $n_d > 1$. As mentioned before, if two or more buyers want to buy the technology θ at a given price I then each of these firms has equal probability of being the one that actually buys the new technology. Given that the seller only sells a technology once and that the instant initial profits of the buyers equal $\pi(\theta_0)$, the seller sets up a price such that in the NE the buyer that buys the technology does not increase its profits. Otherwise there will be another buyer willing to pay more for the technology and still make positive profits. Moreover, the profits of the buyer must be at least the same as its profits from not buying the technology as otherwise this firm is better off by not buying a new technology and sticking to the original one, θ_0. Hence, in the NE of the game at hand we must have that,

$$\frac{\pi(\theta^*)}{r} - I = \frac{\pi(\theta_0)}{r}. \qquad (3.1)$$

In the NE of the game the strategy of the buyers is such that (3.1) must hold. Another condition for a NE is that the seller must play a best response to the other firms strategies. Since in equilibrium (3.1) must hold, if we solve the maximization problem of the seller subject to (3.1) the trigger level θ^* obtained will be part of the NE of the game.

The problem of the seller is to maximize its expected profits given the trigger level θ^* from equation (3.1). Let τ denote the hitting time of θ on θ^*. That is, τ is the infimum point in time where $\theta \geq \theta^*$. Then $E(e^{-r\tau})$ gives the expected discount factor at which the seller values selling the technology. For a given θ^* we can compute the value of $E(e^{-r\tau})$ using the analysis found in Dixit and Pindyck (1999). For the readers convenience we reproduce this analysis in Appendix 3.A.2. When θ follows a Geometric Brownian Motion the expected discount factor is given by

$$E(e^{-r\tau}) = \left(\frac{\theta}{\theta^*}\right)^{\beta},$$

where $\beta > 1$ is given by

$$\beta = \frac{1}{2} - \frac{\alpha}{\sigma^2} + \sqrt{\left(\frac{\alpha}{\sigma^2} - \frac{1}{2}\right)^2 + \frac{2r}{\sigma^2}}. \tag{3.2}$$

In this paper, we use $E(e^{-r\tau})$ as the measure for the speed at which new technologies are adopted. A higher expected discount factor means a higher delay in the adoption of a new technology. Since $E(e^{-r\tau})$ depends ultimately on θ^*, higher θ^* means slower adoption.

The problem of the seller is to choose I, to maximize profits given the effect of I on θ^* and, hence, on $E(e^{-r\tau})$. The problem of the firm in the supply side is then

$$\max_I \quad (I - C)E(e^{-r\tau}).$$

Which, after plugging in the value of $E(e^{-r\tau})$ leads to

$$\max_I \quad \theta^{*-\beta}(I)(I - C)$$

where $\theta^*(I)$ is given implicitly in equation (3.1).

The first order condition yields

$$\frac{1}{\beta(I - C)} = \frac{\pi^{-1\prime}(\pi(\theta_0)) + rI)}{\pi^{-1}(\pi(\theta_0) + rI)} r. \tag{3.3}$$

In order for the optimal price I not to explode we need to impose the following condition on β.

Assumption 1.

$$\beta \quad > \quad \lim_{x\to\infty} \frac{\pi^{-1}(x)}{x\pi^{-1'}(x)}$$

Assumption 1 is satisfied by the Cobb-Douglas profit function if and only if $\beta > b$. From equation 3.3 we can get the equilibrium value of I. From I, we can then compute the value of θ^* in equilibrium using equation (3.1).

Using the Cobb-Douglas production function and from (3.3), after some rearrangement we get that the equilibrium levels of θ^* and I are given by

$$\begin{aligned} \theta^* &= \left[\frac{\beta}{\beta - b}\left(\theta_0^b + \frac{r}{\phi}C\right)\right]^{1/b}, \\ I &= \frac{b}{\beta - b}\frac{\phi\theta_0^b}{r} + \frac{\beta}{\beta - b}C. \end{aligned} \tag{3.4}$$

Note that although all agents are risk neutral, changing σ, which is the variance of the technology parameter, changes the optimal levels. This is because increasing σ decreases $E(e^{-r\tau})$ and hence allows for a higher price for the new technologies while keeping θ constant. We explore this issue and perform other comparative statics in Section 3.4.

3.3.2 The Demand Side Holds All the Market Power

In this subsection we consider the case where the demand side has all the market power. In this setting there is only one buyer and many, $n_s \geq 2$, sellers competing in price. As mentioned earlier, if two or more sellers offer the technology at the same price and a buyer decides to buy a new technology, then each seller has equal probability of being the one that actually sells the technology to the buyer. Therefore, given that there is only one firm buying technologies and that the firm buying technologies can only buy a new technology once, sellers compete to the point where in equilibrium their profits from selling the technology must be zero. Otherwise, standard competition a la Bertrand arguments apply and a seller could decide on lowering the price of the technology to attract the buyer and still make positive profits. Hence, an equilibrium condition is then

$$I \quad = \quad C. \tag{3.5}$$

In this setting, and similarly to what occurred in the previous setting, in the NE the zero profit condition (3.5) must hold. Hence, for computing the optimal level of θ^* we must compute the best response of the buyer given equation (3.5). The problem of the buyer is then

$$\max_{\theta^*} \quad E\left(\int_0^\tau \pi(\theta_0)e^{-rs}\mathrm{d}s + \int_\tau^\infty \pi(\theta^*)e^{-rs}\mathrm{d}s - Ie^{-r\tau}\right).$$

where $I = C$. The expected value is taken over all possible values of τ. The first term of the objective function represents the profits of the firm before it buys the technology at time τ. During this first periods it produces using its initial level of technology θ_0. The second term in the objective function is the profits once the firm bought a new technology at time τ. From this point in time onwards the firm produces using the technology purchased at time τ. Finally, the third term discounts the cost of buying the technology at time τ.

After some algebra and plugging in equation (3.5) we can rewrite the maximization problem of the buyer as (derivation found in the Appendix 3.A.3).

$$\max_{\theta^*} \quad \theta^{*-\beta} \left(\pi(\theta^*) - \pi(\theta_0) - rC\right).$$

Taking the first order condition we get that

$$\pi(\theta^*)\beta - \pi'(\theta^*)\theta^* \quad = \quad \beta(\pi(\theta_0) + rC). \tag{3.6}$$

The equilibrium level of θ^* is then given implicitly by equation (3.6). If we use the Cobb-Douglas production function we get that the optimal levels are given by

$$\begin{aligned} \theta^* &= \left[\frac{\beta}{\beta - b}\left(\theta_0^b + \frac{rC}{\phi}\right)\right]^{1/b}, \\ I &= C. \end{aligned} \tag{3.7}$$

Note that the equilibrium level of θ^* when the demand holds all the market power is the same as the one obtained when the supply holds all the market power. This is due to the fact that when only one side of the market holds all the power, the total surplus of the economy is maximized, the Pareto optimal allocation is achieved. The only actual difference between the two settings lies on which side extracts all this surplus. We prove this statement below by showing that the trigger level θ^* under the two market settings above (given by equations (3.4) and (3.7)) coincides with the Nash bargaining solution.

Proposition 1. *The value of θ^* when only one side of the market holds all the power (sections 3.3.1 and 3.3.2) is Pareto Optimal in that it maximizes total surplus in the economy.*

Proof. We solve for the Pareto optimal value of θ^* by solving for the Nash bargaining solution. For this purpose we need to solve the the maximization of the Nash bargaining function (Nash (1950)):

$$\max_{(\theta^*, I)} \quad \left[(I - C)\left(\tfrac{\theta}{\theta^*}\right)^{\beta}\right]\left[\left(\tfrac{\pi(\theta^*)}{r} - I - \tfrac{\pi(\theta_0)}{r}\right)\left(\tfrac{\theta}{\theta^*}\right)^{\beta}\right].$$

The first order conditions of the problem above lead to

$$\begin{aligned}\pi(\theta^*) - \pi(\theta_0) - 2rI + rC &= 0,\\ -2\beta(\pi(\theta^*) - \pi(\theta_0) - rI) + \theta\pi'(\theta^*) &= 0.\end{aligned}$$

Using the Cobb-Douglas production function and after some algebra we get that

$$\theta^* = \left[\frac{\beta}{\beta - b}\left(\theta_0^b + \frac{rC}{\phi}\right)\right]^{1/b}, \qquad (3.8)$$

$$I = \frac{\theta_0^b(\beta - \phi(\beta - b)) + \frac{rC}{\phi}(\beta + r(\beta - b))}{2r(\beta - b)}.$$

Comparing the values of $\theta*$ in equations (3.4) and (3.8) gives the desired result. □

In the next result we formalize the fact that the speed of adoption is the same independently on which side of the economy holds all the market power.

Proposition 2. *Independently on whether $n_d > 1 = n_s$ or $n_s > 1 = n_d$ (sections 3.3.1 and 3.3.2 respectively) we have that*

$$\theta^* = \left[\frac{\beta}{\beta - b}\left(\theta_0^b + \frac{rC}{\phi}\right)\right]^{1/b}.$$

Proof. Follows from (3.4) and (3.7). □

3.3.3 Competition between Supply and Demand

For this setting we assume that there is only one seller and only one buyer. We solve for the NE of the game by backwards induction. First, we compute the best response of the buyer for any given pair I, θ. This will give us a value of θ^* given any price I and level of technology θ. Then, given this trigger level θ^* as a function of I, θ, we compute the optimal price I. This will give us the unique levels of θ^* and I in the NE of the game.

The problem of the buyer is similar to the problem it faced in the setting where the buyer holds all the market power. The only difference is that now I is taken as exogenous:

$$\max_{\theta^*} \quad E\left(\int_0^\tau \pi(\theta_0)e^{-rs}\mathrm{d}s + \int_\tau^\infty \pi(\theta^*)e^{-rs}\mathrm{d}s - Ie^{-r\tau}\right).$$

After some algebra, the derivations of which are found in Appendix 3.A.4, the maximization problem becomes

$$\max_{\theta^*} \quad \theta^{*-\beta}\left[\frac{\pi(\theta^*)}{r} - \frac{\pi(\theta_0)}{r} - I\right].$$

The first order condition of the problem leads to

$$\pi(\theta^*)\beta - \pi'(\theta^*)\theta^* \quad = \quad \beta(rI + \pi(\theta_0)). \tag{3.9}$$

Equation (3.9) is the implicit function for the optimal trigger level θ^* as a function of the price I. We now turn to the problem of the seller. The objective function is, in this case,

$$\max_{I} \quad \theta^{*-\beta}(I)(I - C).$$

From equation (3.9) we get the value of I as a function of θ^*. In order to simplify computations, we maximize over θ^* instead of over I and consider the function I given implicity in equation (3.9). Therefore, the problem of the seller then becomes

$$\max_{\theta^*} \quad \theta^{*-\beta}\left(\pi(\theta^*)\beta - \pi'(\theta^*)\theta^* - \pi(\theta_0)\beta - \beta rC\right).$$

The first order condition leads to

$$-\pi(\theta^*)\beta^2 + \pi'(\theta^*)\theta^*(2\beta - 1) - \pi''(\theta^*)\theta^{*2} + \pi(\theta_0)\beta^2 + \beta^2 rC \quad = \quad 0. \tag{3.10}$$

The system of equations (3.9) and (3.10) implicitly determine then the optimal levels of θ^* and I. Again we use the specific functional form for the profit function to get more information about the process of adoption of new technologies. If we assume a Cobb-Douglas production function we get that

$$\theta^* \quad = \quad \left[\left(\frac{\beta}{\beta - b}\right)^2 \left(\theta_0^b + \frac{rC}{\phi}\right)\right]^{1/b}, \tag{3.11}$$

$$I \quad = \quad \frac{b}{\beta - b}\frac{\phi\theta_0^b}{r} + \frac{\beta}{\beta - b}C.$$

Hence, as we can infer from (3.11), under competition between supply and demand, the adoption of technologies is expected to occur at a slower rate.

Proposition 3. *The value of θ^* under competition between supply and demand (section 3.3.3) is higher than when one side of the market holds all the power (sections 3.3.1 and 3.3.2).*

Proof. Follows from (3.4), (3.7) and (3.11). □

An important observation is that while the adoption of new technologies occurs at a slower pace under competition between supply and demand, the level of technology adopted in that market setting is higher. This might lead one to think that the long run productivity of the firm on the demand side is higher under competition between supply and demand. This is due to our simplifying assumption that firms can only buy a new technology once. As we mentioned already, Doraszelski (2001) shows that the firm's decision problem, if we allow the firm to change technologies more than once, is of the same form after each adoption of a new technology. Hence, allowing the firms to buy new technologies more than once will still mean that the switch to better technologies occurs later under competition between supply and demand. Thus, long run productivity is still lower under competition between supply and demand.

3.4 Comparative Statics

In Figure 3.1 some comparative static results are presented. The figure depicts the level of technology adopted, θ^*, and the expected discount factor, which is our measure of the speed at which new technologies are adopted. We explore the changes in these two variables when the variance of the process governing the evolution of technologies σ, the trend or the expected evolution of technology α, and the interest rate r, change. The value of the parameters ϕ, b, r and θ_0 are set to the same values as in Farzin et. al. (1998). The value of θ, α and σ are set such that the expected discount factor implies a delay in the adoption of new technologies of around 16 periods, which is about the value Farzin et al. (1998) use (17.79 periods in their paper). The value the parameters we use are then $\phi = 151.32, b = 1.25, r = 0.1, \theta_0 = 1, \theta = \theta_0, \alpha = 0.05$ and $\sigma = 0.01$.

From Figure 3.1 there are three facts that are worth noting. First, an increase in either α or σ has bigger effect in the level of the technology adopted when there is competition between supply and demand. Hence, when only one side of the market holds all the power the level of technology adopted is less sensitive to the process governing the evolution of technologies. Markets for different technologies should exhibit more diverse behavior under competition between supply and demand than when only one side holds all the market power.

A second feature that deserves attention is that as the interest rate rises, the level of technology that is adopted tends to converge to the same value under both competition between supply and demand and the case when only one side holds all the market power. This means that in economies with high interest rates, the structure of the market for technologies has less effect than when compared with low interest rate economies.

A third fact which the numerical analysis reveals is that the effects of the variance of the process for technology, σ, has a small effect if any on the speed of adoption. That is, industries where there the evolution of technologies is very random with big improvements in short timespace and periods with almost no improvements should not influence the speed at which the adoption occurs.

3.5 Conclusions

In this paper we investigated how different market structures affect the speed at which new technologies are adopted. A game between the demand side, firms buying technology, and the supply side, firms selling technology, was presented. Three different market scenarios were considered, one in which the supply holds all the market power, another in which the demand holds all the market power, and a third setting where there is competition between both supply and demand.

In our results, we explained how these three different market structures affect the adoption of technologies. The speed of adoption when one side of the market holds all the market power is the same independently of which side holds the power. However, when no side of the economy has all the market power, the competition between supply and demand case, then the adoption occurs at a slower pace. This suggests that competition between the two sides of the market might decrease the speed of adoption and that competition within each side might increase the speed of adoption.

The literature so far has only focused on the optimal timing of adoption of new technologies from the perspective of a firm that faces an exogenous process of technological change where the price of new technologies is also exogenous. To our knowledge, this is the first paper that incorporates the pricing of new technologies as something endogenous that results from the interaction between firms buying the technology and firms selling the technology. Our results show interesting insights about the effects of the market structure on the adoption of new technologies. The results found explain empirical phenomena often attributed to the differences in the technologies themselves. We show that market structure itself can account for, at least some, of this observed phenomena.

References

1. Barro, R. & Sala-i-Martin, X. (1995): "Economic Growth". McGraw-Hill, New York.

2. Berndt, E. R. & Rappaport, N. J. (2001): "Price and Quality of Desktop and Mobile Personal Computers: A Quarter- Century". *The American Economic Review Papers and Proceedings* 91 (2), 268-273.

3. Chamley, C. & Gale, D. (1994): "Information Revelation and Strategic Delay in a Model of Investment". *Econometrica* 62 (5), 1065-1085.

4. Doraszelski, U. (2004): "Innovations, Improvements, and the Optimal Adoption of New Technologies". *Journal of Economic Dynamics and Control* 28, 1461-1480.

5. Jovanovic, B. & Nyarko, Y. (1996): "Learning by Doing and the Choice of Technology". *Econometrica* 64 (6), 1299-1310.

6. Farzin, Y., Huisman, K. & Kort, P. (1998): "Optimal Timing of Technology Adoption". *Journal of Economic Dynamics and Control* 22, 779-799.

7. Götz, G. (1999): "Monopolistic Competition and the Diffusion of New Technology". *The RAND Journal of Economics* 30 (4), 679-693.

8. Hoppe, H. (2002): "The Timing of New Technology Adoption: Theoretical Models and Empirical Evidence". *The Manchester School* 70 (1), 56-76.

9. Karp, L. & Lee, I. (2001): "Learning-by-Doing and the Choice of Technology: The Role of Patience". *Journal of Economic Theory* 100, 73-92.

10. Karshenas, M. & Stoneman, P. (1993): "Rank, Stock, Order, and Epidemic Effects in the Diffusion of New Process Technologies: An Empirical Model". *The RAND Journal of Economics* 24 (4), 503-528.

11. Lucas, R. (1993): "Making a Miracle". *Econometrica* 61, 251-272.

12. Nash, J. (1950): "The Bargaining Problem". *Econometrica* 18, 155-162.

13. Prentice, D. & Xin, X. (2004): "Constructing for a Quality-Adjusted Price Index for a Heterogeneous Monopoly". *The Manchester School* 72 (4), 423-442.

14. Stoneman, P. & Kwon, M. (1996): "Technology Adoption and Firm Profitability". *The Economic Journal* 106 (437), 952-962.

15. Weiss, A. (1994): "The Effects of Expectations on Technology Adoption: Some Empirical Evidence". *Journal of Industrial Economics* 42, 341-360.

Appendix

3.A.1 Brownian Motion

Let z be a continuous random variable. Denote by $z(t)$ the value of z at time t. We say that z is a Brownian Motion if

1. $z(0) = 0$,

2. $z(t)$ is almost surely continuous for all t and

3. $z(t) - z(t-s) \sim \mathcal{N}(0, t-s)$.

3.A.2 Derivation of the Expected Discount Factor

Note that, for a given θ^*, the only variable the expected discount factor depend on is the current level of technology θ. Then we can define the function f as

$$f(\theta) = E(e^{-r\tau}).$$

If $\theta = \theta^*$ then it is obvious that $\tau = 0$ and hence $f(\theta^*) = 1$. Assume then that $\theta < \theta^*$. Choose dt small enough so that θ won't surpass θ^* in the next time interval. Then we have that the problem of computing the hitting time restarts at the point $\theta + \mathrm{d}\theta$. That is,

$$\begin{aligned} f(\theta) &= e^{-r\mathrm{d}t} E(f(\theta + \mathrm{d}\theta)) \\ &= e^{-r\mathrm{d}t} \left[f(\theta) + E(\mathrm{d}f(\theta)) \right]. \end{aligned} \tag{3.12}$$

Given that θ follows a Geometric Brownian Motion we can expand $\mathrm{d}f(\theta)$ using Itô's Lemma. In this case we get

$$\mathrm{d}f(\theta) = (\alpha\theta\mathrm{d}t + \sigma\theta\mathrm{t}z)f'(\theta) + \frac{1}{2}\sigma^2\theta^2 f''(\theta)\mathrm{d}t.$$

Note that $E(\mathrm{d}z) = 0$. Furthermore, using the Taylor expansion and ignoring the terms of order $\mathrm{d}t^2$ and higher we can state that $e^{-r\mathrm{d}t} = 1 - \rho\mathrm{d}t$. Hence,

$$e^{-r\mathrm{d}t} E(\mathrm{d}f(\theta)) = (1 - \rho\mathrm{d}t)\alpha\theta f'(\theta)\mathrm{d}t + \frac{1}{2}\sigma^2\theta^2 f''(\theta)\mathrm{d}t.$$

Therefore, ignoring once more the terms of order equal or higher than $\mathrm{d}t^2$, we get that

$$0 = -\rho f(\theta) + \alpha\theta f'(\theta) + \frac{1}{2}\sigma^2\theta^2 f''(\theta). \tag{3.13}$$

Equation (3.13) is a second order linear differential equation in f with the boundary conditions $f(\theta^*) = 1$ and $f(\theta) \to 0$ as the difference $\theta^* - \theta$ becomes large. The general solution to the second order linear differential equation is given by

$$f(\theta) = C_1\theta^{\beta} + C_2\theta^{\beta'}$$

where β and β' are the roots to the characteristic equation in x

$$0 = -\rho + \alpha x + \frac{1}{2}\sigma^2 x(x-1).$$

Let β be the positive root exceeding unity to the above equation, that is

$$\beta = \frac{1}{2} - \frac{\alpha}{\sigma^2} + \sqrt{\left(\frac{\alpha}{\sigma^2} - \frac{1}{2}\right)^2 + \frac{2r}{\sigma^2}}. \tag{3.14}$$

Using the terminal conditions we can set $C_1 = \theta^{*-\beta}$ and $C_2 = 0$. This leads to

$$f(\theta) = \left(\frac{\theta}{\theta^*}\right)^\beta$$

where β is given in equation (3.14). Recalling that $f(\theta) = E(e^{-r\tau})$ gives the desired result.

3.A.3 Problem of the Firm from the Demand Side when the Demand Holds All the Market Power

The problem of the buyer is given by

$$\max_{\theta^*} \quad E\left(\int_0^\tau \pi(\theta_0)e^{-rs}\mathrm{d}s + \int_\tau^\infty \pi(\theta^*)e^{-rs}\mathrm{d}s - Ie^{-r\tau}\right).$$

After computing the integrals above the maximization problem becomes

$$\max_{\theta^*} \quad \left\{\pi(\theta_0)E\left(\tfrac{1}{-r}e^{-r\tau}\right) + \pi(\theta^*)E\left(\lim_{x\to\infty} \tfrac{1}{-r}e^{-rx} - \tfrac{1}{-r}e^{-r\tau}\right) - IE(e^{-r\tau})\right\}.$$

Which in turn can be rewritten as

$$\max_{\theta^*} \quad \left\{\tfrac{\pi(\theta_0)}{r}E\left(1 - e^{-r\tau}\right) + \tfrac{\pi(\theta^*)}{r}E(e^{-r\tau}) - IE(e^{-r\tau})\right\}. \tag{3.15}$$

Plugging in the value of $E(e^{-r\tau})$, using the fact that $I = C$ and dropping the constants from the maximization problem we get that

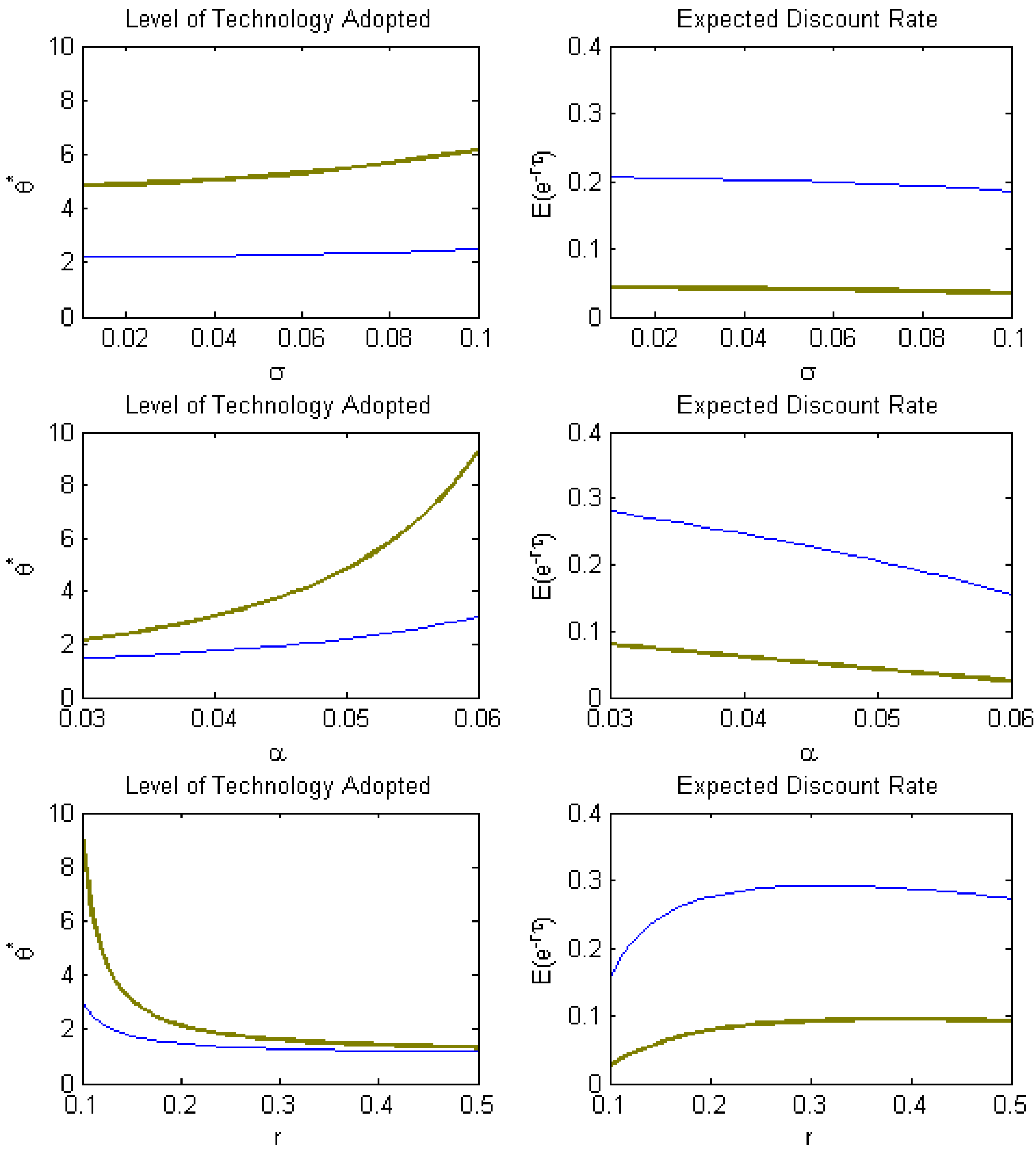
$$\max_{\theta^*} \quad \theta^{*-\beta}\left(\pi(\theta^*) - \pi(\theta_0) - rC\right).$$

3.A.4 Problem of the Firm from the Demand Side Under Competition between Supply and Demand

The same analysis applies here as in Appendix 3.A.3 to get equation (3.15). From this equation, by pugging the value of $E(e^{-r\tau})$ and dropping the constants from the maximization problem we get that

$$\max_{\theta^*} \quad \theta^{*-\beta}\left(\tfrac{\pi(\theta^*)}{r} - \tfrac{\pi(\theta_0)}{r} - I\right).$$

Figure 3.1: Comparative Statics



The thick green line corresponds to the case where there is competition between supply and demand while the thin blue line corresponds to the case where the market power lies with only one side of the economy.