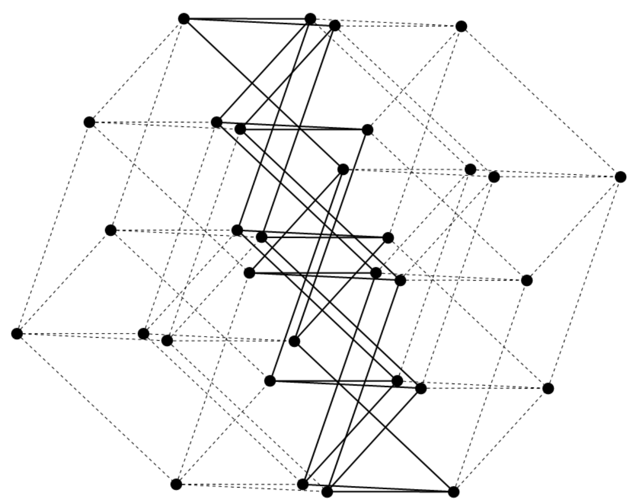
Utrecht University

MSc Mathematics

The Minority Game: Individual and Social Learning



Author:
Stathis Grigoropoulos

Supervisor:
Dr. Matthijs Ruijgrok

Second Reader:
Dr. Paul A. Zegeling

April 29, 2014

Contents

1 Introduction **1**

2 Learning in Game Theory **5**
2.1 Rationality and Game Theory 5
2.2 Bounded Rationality 6
2.3 What is Learning 8
2.4 Learning Theories 8
2.4.1 Behaviorism 8
2.4.2 Cognitivism 9
2.4.3 Social Cognitive Learning 10
2.5 Multi-Agent Systems Learning 10
2.5.1 Belief Learning 10
2.5.2 Reinforcement Learning 11
2.5.3 Rule Based Learning 13
2.5.4 Evolutionary Learning and Imitation 14

3 Minority Game **17**
3.1 The Minority Game 17
3.2 Alternative Learning Methods in the Minority Game 24
3.3 Minority Stage Game and Nash Equilibria 26
3.4 Minority Game as a Congestion Game 28
3.5 Imitation in the Minority game 30
3.5.1 Multi-Population Replicator Dynamics 31
3.6 Three-Player Minority Game 34
3.6.1 Replicator Dynamics 35
3.7 Minority Game and Human Behavior 38

4 Experiments **41**
4.1 Multi-Population Learning 42
4.2 Single-Population Learning 49

5 Conclusions **61**

Abstract

Learning has been given much attention in Artificial Intelligence (AI) and Game Theory (GT) disciplines, as it is the key to intelligent and rational behavior. However in a multiagent setting, as in Multi-agent Systems (MAS), where the environment changes according to the actions of the players, the participants cannot afford to be fully rational and resort to heuristics. In such cases classic Game Theory fails to provide convergence results of the adjustment process, thus losing predictive power. Evolutionary Game Theory (EGT), motivated from biology, has been proven suitable for analyzing bounded rationality and heuristic learning using the robust replicator dynamics. In this thesis we use a famous congestion game with many (odd) participants called the Minority Game (MG) as a learning paradigm. The most critical learning methods used in the MG are reviewed, motivated from both economics and machine learning perspective along with their results. Continuing, individual-reinforcement learning through replicator dynamics is analyzed and the asymptotic properties of the learning procedure in the MG are provided. Moreover, we compare individual learning with social learning through imitation using agent-based simulations. The two types of learning do share common convergence characteristics, but differ in the resource allocation schemes and in terms of robustness. Individual-reinforcement learning is a more utilitarian process maximizing system efficiency with disregard to single-agent performance. On the other hand, social imitation can provide a more egalitarian setting where individual scores are almost equal.

Chapter 1

Introduction

“All men by nature desire to know”. This is how Aristorle began his Metaphysics. However, what is learning and what type of learning processes can we identify in a context with many interacting agents? Can we analyze mathematically the expected behaviour of learning players for the benefit of Artificial Intelligence (AI), Social and Economic Sciences? Although the motivation of Evolutionary Game Theory (EGT) stems from biology, since the work of Maynard Smith, it has been realized that individual or social learning can be modeled through EGT and the replicator equations [49, Preface]. This report revisits the not-so-simple question of how agents learn to play and what do they learn to play? [80]. Not surprisingly, different adaptive methods lead to different learning predictions [53].

We narrow our focus on a simple congestion game played by many bounded rational agents, called the Minority Game (MG). The Minority Game, a paradigm of Complex Adaptive Systems (CAS), attracted much attention in the Statistical Physics and Multi-Agent Systems (MAS) literature [92],[93, p. 175]. In this report, we provide a survey of the numerous learning methods used in MG. Subsequently, we study the Minority Game through EGT and apply the pairwise imitation protocol that leads to the well-studied replicator equations [90]. We derive the learning outcomes of agents through analytical treatment and agent based simulations. Interestingly, individual and social imitation, in the MG do share similarities in terms of learning outcomes. However, as we will present in this thesis, there are subtle differences that can benefit a learning process within an optimization setting or an economic and social one.

Game Theory provides the necessary mathematical formalism to model interacting agents. In a nutshell, non-cooperative Game Theory analyses a strategic situation (game) played by fully rational players. Specifically, the players know all the details of the game, including each other’s preferences over all possible outcomes [101, ch. 1]. The richness and depth of the analysis is tremendous, with the downside that it can handle only a small number of heterogeneous agents [22, ch. 1]. Moreover, classic Game Theory, outside a certain strategic environment, fails to explain human behavior and decision making in many cases [39, ch. 1]. Concretely, human experiments showed that humans do not always play in fully rational terms, but rather deviate from this behavior [49, Preface]. This discrepancy between theory and experiments, impacted the way game theoretical tools were used to analyze learning.

The primary solution concept used in strategic interactions, is the Nash Equilibrium (NE) [71]. In words, it is the situation where no agent can improve his payoff by unilaterally deviating from his behavior [35]. Classically, in terms of learning, NE took an eductive justification, which relies on the players ability to reach equilibrium through careful thinking. Since agents are fully rational, they can always predict and optimally respond to their opponents actions [80]. However, in the case when the players do not behave rationally, initially the research community assumed that NE loses its predictive power [35]. In the next paragraphs, we will witness that is not true at all.

Following the aforementioned events, bounded rationality was introduced as the basic concept to model the cognitive limitations, in behavior and decision making of humans [94]. Although, the definition of bounded rationality is still debatable, limiting the cognitive capabilities of agents in a game context, provided the necessary framework for evolutionary learning and adaption to appear [49, ch. 1]. One way to describe bounded rationality is by the use of simple "rules of thumb" for every day decision making, i.e. heuristics [41]. Heuristic rules, can be considered as the outcome of an ongoing learning process among agents.

Learning models can be classified in individual learning, social imitation and belief learning [80]. In individual learning, success and failure directly influence agent choices and behavior. Learning theories that describe such a procedure are mainly Behaviorism and Cognitive theories [75]. From a MAS learning perspective, individual learning is interpreted as various types of reinforcement learning. Social learning occurs in the cases where success and failure of other players influence choice probabilities. Social Cognitivism is the equivalent psychological theory representing this human phenomenon. Lastly, Belief learning is a learning model originating from economic theories, where experiences affect players beliefs. The main difference between Belief-based and reinforcement individual learning is that in the former one should hold explicit beliefs for the rest of the players and play a strategy that yields the highest payoff according to these beliefs. On the contrary, in reinforcement learning the player adjusts his/her strategy taking into account only his/her payoff, i.e the agent might not even know the number of the existence of other players [32]. Moreover, we should divide theories talking about learning into descriptive and prescriptive. That is, theories analyzing learning and teaching respectively [93, p 194]. In this report we will deal with individual and social learning, following descriptive paradigms. We will not impose a specific goal of learning throughout our analysis, but rather observe the convergence points of evolution (if any).

An evident mathematical inquiry arises, namely how can we model individual and social learning using game theory as our basis? Evolutionary Game Theory (EGT) helps to answer the aforementioned question. EGT imagines that a game is played in a repeated fashion, by socially or biologically bounded agents who are drawn from large populations in a random fashion [101]. At each point in time, each agent only plays a pure strategy. This specific strategy can represent the heuristic rule this particular player follows to play the game. In the original biological inspiration of EGT, the agents "reproduce" in a manner proportional to their payoff (fitness) [101]. This same process can be translated into imitation learning between players, in the case when the population has a constant size [101]. Clearly, one weak assumption is of that a game can be repeated with the same settings over time. However, when interaction is anonymous, i.e from a large population without having interest of the player identity, EGT framework appears to be more justifiable [80] . Moreover, the possibility that boundedly rational agents reach a NE by means of some adaptive procedure, justifies the importance of this solution concept [80].

One simple model that serves as a paradigm of adaptation is the El Farol Bar problem (EFBP) [7]. The El Farol the problem goes as follows. We have a popular bar in Santa Fe with limited capacity that organizes each Thursday a Jazz-music night. Given that the number of potential customers (players) is fixed. Suppose the bar is crowded, then no-one will have a good time and each customer prefers staying at home. So each week players have to choose one out of two possible actions: stay at home or go to the bar. The players that are in the minority win the game.

As stated by Brian Arthur, in the Foreward of the book [22]."Legend is indeed correct: in 1988 on Thursday nights Galway musician Gerry Carty played Irish music at the El Farol, and the bar was not pleasant if crowed. Each week I mulled whether it was worth showing up, and I mulled that others also mulled". Arthur's agents (players) are equipped with "predictors" of the bar attendance. An example of a "predictor" is, assume the attendance to be as last Thursday or two weeks ago. The "predictors" in turn define a behavior for each agent. The major contribution of Arthur's seminal paper was that it showed in a clear manner the limitations of the game theoretic perspective of pure strategic reasoning for a complex game with many participants. Arthur proposed a model of agents equipped with bounded rationality and inductive reasoning to tackle complex problems of everyday life.

Following the success of the El Farol model, the Minority Game (MG) was introduced by Damien Challet and Yi-Cheng Zhang as a concise mathematical formulation of the original model [21]. The MG, although a simplification of the original problem, managed to preserve the dynamics and characteristics of the El Farol model. In the MG, an odd number of agents must choose one of two choices independently at each turn (stay of go to the bar). The players who end up on the minority side win [22].

The MG is a simple congestion game with many anonymous participants [53]. In a congestion game (Rosenthal,1973), players use several facilities from a common pool. The costs or benefits that a player derives from a facility depends on the number of users of that facility. A congestion game is therefore a natural game to model scarcity of common resources [38]. Particularly, MG has been used primarily to model financial market time series [44],[33]. Moreover, it has been used to optimize cognitive and normal wireless networks resource allocation [60], [64], to increase distributed systems efficiency [2] and to exploit the computational potential of multi-core clusters [25]. In addition the MG has uses in the energy management for Smart Buildings [104]and an intuitive use in road traffic optimization [37]. The MG sparked a lot of interest in the Statistical Physics literature, as originally game theoretical analysis was avoided. There are over two hundred relevant articles and the MG has its own web-page with a list of most of them [19]. Although Statistical mechanics investigations led to great insight on the MG behavior, it did so in an aggregate level [22]. On this report we will employ EGT to answer how the agents in each one setting of the MG behave, at least in the simplest memoryless case. Furthermore, it is interesting to investigate what agents learn through the MG platform along with the dynamics of their behavior.

This article is organized as follows. In the following chapter, we review the role of rationality in Game Theory and the emergence of bounded rationality. Continuing, we present the definition of learning and its different cognitive and social interpretation and theories 2.4. Concretely, we connect critical adaptation methods used in Multi-Agent Systems and Game Theory literature with their psychological and behavioral counterpart theories (chapter 2, section 5). In chapter 3 we describe formally the original MG and analyze its stage game. Moreover, we discuss relevant literature containing results in terms of outcomes and Nash Equilibria of the MG where alternative methods of learning have been applied. Furthermore, we analyze the MG as congestion game and refine the Nash Equilibrium of the MG using imitation learning and replicator dynamics. Chapter 4 consists of the validation of the MG analytical results through computer simulations. Moreover, we investigate social imitation in the MG, a novel scenario, using agent-based simulations. Lastly, chapter 5 provides a short conclusion of our work on learning in the MG.

Chapter 2

Learning in Game Theory

2.1 Rationality and Game Theory

Game theory studies, through mathematical models, the conflict or cooperation of intelligent and rational decision-makers [69, p. 1]. Intelligence and rationality, in terms of players, are assumptions based on Von Neumann's principal theories about economic behavior [56]. Namely, early models of economics and Game Theory assumed that players in a game setting are capable, as utility maximizers [56, p. 5], to search the best solution or action looking through all the possible moves of their opponent(s). To analyze and learn the outcome of the game when played by rational agents, John Forbes Nash, Jr. in his seminal work [71] introduced the Nash equilibrium as a solution concept. Furthermore, when the interaction in a certain strategic game of decision-makers gets repeated, the players can use the knowledge of the previous outcomes of each stage game and the Nash equilibrium to deduct long run play. For instance, in the iterated prisoners dilemma, by backward induction we can conclude that defect is the game theoretical best strategy. However, the predicted results fall short on explaining the experimental findings and the emergence of cooperation when real players play iterated prisoners dilemma[5]. In addition, Axelrod devised tournaments in order to put to the test several strategies and find the best one. It was proved that the most successful strategy in the repeated prisoners dilemma was Anatol Rapoport's simple tit-for-tat (TFT) [9]. This was one of the first cues to show that simplistic rules, such as imitate the move of your opponent, could outperform complex strategies in game settings.

On one hand, perfect rationality is a very plausible assumption when the stakes are high and the players are fully informed of the payoffs and rules of the game [70]. Moreover, the assumption of perfect rationality greatly simplifies the design of the interaction mechanism among the agents and economists can prove many theorems for the economies that are inhabited by rational players [87]. One the other hand, when situations get more complex (for instance when we have more than two players and multiple possible actions), the burden imposed by rationality to the decision-makers is heavy. An illustrative example is the game of chess, where we still have only two players and we assume that the agents have all the information needed at their disposal, that is know all the possible moves at each step of the game. A perfectly rational agent should be equipped with an arbitrary large memory to remember all the states of the game, the moves of their opponents and in addition the cognitive ability to perform multiple induction arguments to deduct an optimal policy of play in the long run [8]. Through introspection and laboratory experiments, one is lead to the conclusion that even in quite simple decision problems, people often fail to conform to some of the basic assumptions of rational decision theory [8]. In parallel, this pursuit of rationality forced the emerging field of artificial intelligence to place extensive computational capabilities in agents, for the sole purpose to compute and behave using optimal - rational actions [39, p. 1-4]. Consequently, learning in game and decision theory for a period of time was confined in studying Nash Equilibrium as learning outcome and goal [34, p. 1].

2.2 Bounded Rationality

Question arise however, since humans and their decisions cannot be modeled as fully logical processes. Which are the ways of human thinking and resolving in a world of finite cognitive abilities? Simon in his seminal work, introduced bounded rationality as the basic concept to model the cognitive limitations, in behavior and decision making, of humans [94]. A precise definition of bounded rationality is not still available in the literature. However, we can start clarification by stating what is not bounded rationality [39, p. 13 - 19]. Certainly, bounded rationality is not irrationality and does not explain behavior stemming from abnormality and mental illness. Moreover, an agent, in order to complement his limited cognitive capabilities, applies (or is subjected to) adaptive and learning mechanisms after he/she has simplified a complex situation. Thus, a decision-maker can be viewed as a satisficer, one seeking a satisfactory solution instead of the optimal one [10]. Finally, a person in an decision process may not only relax its expectations of optimal solution in terms of one utility, but also have desire or goal for more than one type or number of utilities. An example could be that an agent is aiming in emotional reward in addition to some monetary utility (i.e a decision maker can possess a utility vector)[94]. Furthermore, one of the most simple and fundamental ideas in bounded rationality is that no rationality at all is required to arrive at a Nash equilibrium [8]. Thus, contrary to the belief that uncertainty, and not fully rationality, would eliminate the notion of Nash equilibria, the NE solution concept can be regarded as the long-run outcome of bounded-rational play where agents strive for optimality [34, p. 1 - 10].

In order to understand bounded rationality as a resolving framework, we must turn our attention to the evaluation process inside the decision making cycle. A decision is realized in many costly steps such as gathering and organizing information. However, the evaluation step is crucial and it is dependant on the learning procedures of the agent [94]. Gigerenzer suggests economic agents use simple rules to come to a decision, using minimum computational resources [41]. These "fast and frugal" decision methods are called heuristics, and can be seen as a "shortcuts" of mind when the space of choices is arbitrary large or unknown. The human brain can generalize knowledge from a few specific examples, discover reusable patterns in everyday life and exhibit near optimal behavior by following a few simple rules of thumb [81]. Moreover, it has been observed that heuristic searches are used by experts in decision making and are a result of accumulated learning and experience in a given situation [41]. Heuristics are used heavily in the computer science field as techniques designed for solving a problem more quickly when classic methods are too slow, or for finding an approximate solution when classic methods fail to find any exact solution. A heuristic can quickly produce a solution that is acceptable in contrast to an exhaustive search for an exact optimal solution in a prohibitively long time [86, chap. 4]. The main difference with the physiological heuristics is that in most cases, the solution space is static and the environment is not dynamic. For instance, in a graph search problem, algorithms such as the A* search algorithm, have the ability to explore partially the solution space and take the optimum solution in each step [46]. In contrast, psychological heuristics, are somewhat "hard-coded" mental rules derived from personal experience or social knowledge that are applied directly bypassing most of the searching in the possible solutions. In addition any heuristic proposed for real life problems, should account the involvement of uncertain and changing environment that is also affected by each agents decision [95, chap. 1]. Thus, in AI science field, heuristics were already considered as a sign of intelligence and a way to quickly get satisfactory results [77]. On the contrary, psychologists, viewed heuristics as cognitive biases and source of error in decision process of the human mind [40]. Heuristics can be considered as the building blocks of mental frameworks of behavior, i.e as methods of learning and adaptation [42]. The mental frameworks or models are called schemata [85] and it is often the case that schemata represent mental stereotypes that ignore stimuli contrariant to the fixed belief. A more elaborate explanation of the mind schema, will be presented along with cognitive learning theories in a following subsection.

Gigerenzer in his ground breaking research, showed that heuristics do have attractive properties and their usage in every day decision problems is not unjustifiable. Moreover, Gigerenzer and collaborator's experiments showed that in many cases, simple heuristics outperform more sophisticated decision algorithms [43, 42, 41]. As heuristic methods in decision making can be considered efficient mappings of bounded rationality [42], the question still remains on what type of behavior we learn following heuristic rules in given situations. In the same spirit, WB Arthur's paper gave a clear representation of the limitations of rationality and proposed a model of agents equipped with bounded cognition following inductive reasoning to tackle complex problems of everyday life [7]. Arthur's model, and the simplification by Challet and Young was a concise way of mathematically presenting behavior rules(i.e schemas) and heuristics in agents decision process. Studying the original model and the various extensions we can use the tools of game theory to derive insightful results towards the learning outcomes of the population as a whole and the individual players.

2.3 What is Learning

Before we analyze learning as a concept and procedure inside the decision making cycle, we must first define what learning is and review learning theories that connect behavioral models used in economics and artificial intelligence. Formally speaking learning can be viewed as (the observable or not) process or product of information acquisition or modification and the transformation of it to reusable knowledge as result of experience. Learning is extended to alteration in potential behavior, skill and emotions by the synthesis of different stimuli and information of the subjects environment [15],[58],[47],[75]. This technical definition of learning, although it describes what learning is from a broad point of view, is far from a complete explanation of how learning works, motivated and accomplished in various different settings. Hence, numerous learning theories have been developed to address key questions, such as how does learning happen and which environments enhance it [1]. Furthermore, questions arise on a interdisciplinary science level regarding the importance of memory and the inner workings of the brain, when learning takes place [3],[75]. The scientific field of human learning is vast and undergoes intense research as a prominent field, a full review can be found in [47],[75] and their references therein. In this chapter we narrow our focus to the main learning theories that have contributed their methods to adaptive multi agent systems and repetitive game theoretic situations. Specifically, we will present the specific learning algorithms used in the literature for the MG.

Learning theories can be divided into descriptive and prescriptive theories [93, p 194], [98]. Descriptive learning theories are concerned about how learning occurs and construct models that can be used to explain learning results. On the other hand prescriptive learning theories strive to find optimal methods and techniques of learning when the learning goal is fixed. The research of prescriptive learning methods greatly support instructional design sciences [83]. We continue to analyze descriptive theories of learning, as we mostly care to model the learning process where subjects are not in an explicit teaching environment.

2.4 Learning Theories

2.4.1 Behaviorism

Descriptive learning theories can be viewed through three general "perspectives", namely behaviorism, social cognitive theory and cognitive theory [75]. Behaviorism is a view that assumes learning to be a passive process responding to environmental stimuli(Stimuli - Response, S-R). Classical behaviorists assume that learning will lead to an observable behavior change. The learner starts off as a clean slate (tabula rasa) and behavior is shaped through positive reinforcement or negative. Both directions of reinforcement increase the probability that the preexisting behavior will reoccur. In contrast, punishment (both positive and negative) is meant to decrease the possibility that the preexisting behavior will happen again [75]. Behaviorist theory reigned around 1950, succeeding connectionism and Thorndike's "Law of Effect" [47, p. 24-48]. From behaviorism theory stems the reinforcement learning method applied widely in machine learning and multi-agent systems [63, p. 293-314],[16],[93]. Another variant algorithm with the same concept of reinforcement, also widely used, is Q-learning [explanation and citation needed].

2.4.2 Cognitivism

The cognitivist response replaced behaviorism in 1960 as the dominant learning theory. Cognitivism argues that thinking is not only a behavior response as behaviorism assumes. Thinking is a process on its own, therefore requiring researching the responses to various stimulus conditions, in order to infer about the cognitive procedures in the human brain [75, p. 157]. Cognitivism contains three main perspectives, namely constructivism, information processing and contextual theories. The latter two views represent models of how a subject evaluates information and the educational setting when the acquiring is taking place respectively. Most notably, in the case of information processing theories, human brains are modeled as a computer equipped with bounded memory capacity. Information Processing theory along with the cognitive restrictions of humans proposed by Simon, leads to insightful organizational design models of businesses [36]. Constructivism, and its main theorist Jean Piaget (1980) assume that learning, is the progressive reorganization of knowledge as result of experience [79]. Piaget introduces the aforementioned mental schemata as the general framework where information and knowledge are processed and connected [78]. However, Piaget failed to formalize the mechanism underlying the creation of schema and the progression stages of knowledge development he had observed. Drescher in his seminal work, formalizes Piagetian schemata inner workings with the introduction of the schema mechanism [29]. Schema mechanism can be viewed as the generic model of learning in beings capable of learning, regardless of its origin. By origin we refer to biological, electronic, emotional or other abstract initiations of a learning sequence [30, p. 8 - 11]. A schema mechanism is comprised by schemas, actions and items. Formally a schema is a tripartite structure consisting of a context, action and result. A schema poses that when some context conditions are met and a schema action is taken, then the result conditions will be obtained. Moreover, each action is an event that can change the state of the world. Finally, an item is a state component and each item is a proposition of the state of the world [30, p. 9]. In summary, congitivism from the point of view of artificial intelligence can be seen as the general framework of connecting representations of information to knowledge structures. AI is a direct descendant of cognitivism and employs methods from the vast arsenal of machine learning heuristics to increase its cognitive capabilities. However, modeling artificial intelligence through cognitivism is just one approach, other exists such as enaction [99, p. 53 - 62].

2.4.3 Social Cognitive Learning

Social cognitive theory is concerned with the ways in which people learn from observation [75, p. 140]. Social cognitivism is a generalization of social learning. Social learning theory can be considered as the intermediate step between behaviorism and cognitivism. Through social stimulus-response and imitation, a subject can be reinforced to learn from a third person inside the social environment [75, p. 144]. The conditioning of the imitator can lead to delayed imitation, where the change in behavior is not visible from the beginning. When we talk about imitation, we actually include various types of information that can be transfered through the social network. Three types of information have been distinguished to be transfered with the imitation process. Namely those are, actions, goals and results and depending on which type of information the imitator perceives, the result of social learning is different. Thus, we can have imitation of actions or behavior with or without the knowledge of the goals of the imitated person that can reproduce the desired result [26, chap. 9]. Artificial Intelligence has benefited by the social cognitivism paradigms, towards studying and creating machines that can respond, learn and acquire common sense through social interactions [66, chapter 6].

2.5 Multi-Agent Systems Learning

In this section we review the most important learning procedures used in game theoretic contexts, suited for the problem analysis of this thesis. The reader is prompted to the references in the end of paragraph for a more thorough look in algorithmic learning. Surprisingly, many early learning algorithms used in artificial intelligence do not have explicit connection to psychology and learning theories. Moreover, apart from the individual learning schemes, multi-agent systems can learn through evolution as a population. The idea of evolutionary learning was developed by biologists and later incorporated in the artificial intelligence domain. [93, chap. 7], [16, chap. 6], [11]. In this section we review the most important learning procedures used in game theoretic contexts, suited for the problem analysis of this report.

2.5.1 Belief Learning

Methods of belief learning where initially used for calculating iteratively Nash equilibria on zero-sum games. However, belief based learning was reinterpreted as adaptive models of behavior in a repeated setting [93, p. 196].

Fictitious Play

One of the earliest learning rules used in strategic situation is Fictitious play. Initially, fictitious play was proposed as an iterative method to compute Nash-equilibria of zero-sum games. The rule is straightforward always choose a best reply to the actions of the other players in the previous periods. That is, if A is the set of the opponent's actions, and $\forall a \in A$ we have

$$P(a) = \frac{w(a)}{\sum_{a' \in A} w(a')},$$

the probability of a in opponent's mixed strategy and $w(a)$ the number of times opponent has played action a. Fictitious play is an interesting heuristic method as it is very simple and gives rise to strong results concerning behavior in equilibrium. There are variations of the updating rule, such as smooth or weighted fictitious play that can refine the learning process even further. However, as discussed in this chapter, fictitious play is not a plausible model of human learning and the belief update setting is mathematically constraining [93, p. 200].

Bayesian Learning

Bayesian or Rational learning follows the trail of fictitious play and generalizes it. Namely, it allows players to have a richer set of beliefs about the strategies of their opponents. We begin by a set of initial beliefs, as in fictitious play, and we use Bayesian updating to update these beliefs. Given the set of possible histories $h \in H$, we have

$$P_i(s|h) = \frac{P_i(h|s)P_i(s)}{\sum_{s' \in S} P_i(h|s')P_i(s')},$$

the probability that an opponent is playing strategy $s \in S$.
Rational learning is a very intuitive learning model and provides strong guarantees of convergence to the true strategies of the players. However, the complicated analysis required due to the vast space of possible opponent strategies, makes the Bayesian model an unsuitable contestant to explain human thinking. On the contrary, in many cases in everyday decision making, people tend to prefer handling specific information rather than general beliefs.

2.5.2 Reinforcement Learning

Reinforcement models stem directly from the behaviorist learning theories. Reinforcement methods, were originally designed for one agent problems and afterwards generalized in economic decisions and games [95]. In short, an agent associates values with states of the environment, by observing the rewards received when visiting those different states. In the case of repeated games, however, all changes in the expected reward are due to alteration in strategy by the players. There is no changing environment state that agents can explore and assign a certain value. Therefore, repeated games are sometimes also called stateless games [73]. Moreover, reinforcement learning does not explicitly model the opponent's strategy, in contrast, it assigns a value in each next available action or strategy, that characterizes the expected reward following the respected action/strategy. We distinguish two main methods to assign action/strategy-value pairs. Namely, the Roth-Erev rule and the stateless Q-learning method.

Roth-Erev Reinforcement Learning

Based on the work of psychologists Bush and Mosteller, E. Roth and I. Erev, associate propensities with each action. Those propensities in turn, translate to probabilities in taking that action [31].

In the basic model, each player is assumed to start with equal propensities for each available strategy. That is, for all players n

$$q_{nk} = q_{nj} \forall k, j \in S \text{ set of strategies.}$$

If player n plays his $k-th$ pure strategy at time t and receives a reinforcement $R(x)$, then the propensities are updated according to the following formula.

$$q_{nj} = \begin{cases} q_{nj} + R(x) & \text{if } j = k \\ q_{nj} & \text{otherwise,} \end{cases}$$

Where $R(x)$ is a suitable reinforcement function when the payoff is $x \in \mathbb{R}$. The probability that player n will play his $k-th$ strategy at time t is defined by:

$$p_{nk} = \frac{q_{nk}(t)}{\sum_{j \in S} q_{nj}(t)}.$$

Q-Learning

Q-learning is another important variation of reinforcement learning. In Q-learning, the agent creates direct action-Q-value pairs rather than mapping strategies. In a stateless environment for actions $a \in A$, the Q-update function is:

$$Q(a) \leftarrow Q(a)\lambda(r - Q(a)) \ \forall a \in A,$$

where r is the reward received, and λ is the learning rate $(0 \leq \lambda \leq 1)$.

What comes next is the question of which action an agent should follow. The agent action selection is governed by the fundamental trade-off between exploitation and exploration in each turn. Namely, the agent intuitively could follow the greedy rule of taking the action with the highest score. However, in that case the room for exploration of maybe better alternatives would be zero [51, 95]. Therefore, a simple alternative is to follow the greedy heuristic most of the time, but every once in a while with small probability ε, select an action at random. This method is called ε-greedy method. Although near-greedy methods are simple and popular way of balancing exploitation and exploration, they have a drawback on effective exploration. When a ε-greedy heuristic, choose an alternative of the current best action, it assumes the same weight of all other alternatives. Thus, it might be possible to choose the worst alternative action, leading to suboptimal results. For that reason, the softmax action selection method was introduced. The softmax method still gives the highest selection probability to the greedy action, however the rest of the alternatives are ranked according to their value estimates. Commonly, the softmax method uses the Boltzmann distribution to assign probabilities to action $a \in A$ at time t play, as follows.

$$p(a) = \frac{e^{Q_t(a)/\tau}}{\sum\limits_{b \in A} e^{Q_t(b)/\tau}}, \tag{2.1}$$

where τ is a positive parameter called the temperature. High temperatures sets actions with nearly same probabilities. On the other hand, low temperatures create a greater difference in selection probability for actions that differ in their value estimates. When $\tau \to 0$, softmax selection becomes the same as taking greedy action. The above choice method is also called logit choice rule in a game theoretic context. Reinforcement learning is a very active in multi-agent systems learning and provides convergence to Nash equilibrium for zero-sum stochastic games. However, there is no such guarantees for general sum stochastic games.

2.5.3 Rule Based Learning

In Rule learning, players are equipped with decision rules that map histories of play into strategy choices. In contrast to learning which specific strategies to choose, agents learn which rules to use for optimal play. One example in pure game theoretical setting, is a player assigning weights in a set of trigger strategies he has in his disposal according to performance in play (for instance tit-for-tat). Afterwards, using a decision heuristic the agent can use the action dictated by the proper rule. Intuitively, the set of rules can be expressed as the way an expert weights different cues of evidence (for instance a doctor viewing symptoms of a patient). Rule based learning can be considered as a representation of Piaget's schemata of the mind and inductive learning procedure. Machine learning decision trees and inductive rules are closely connected, however numerous results suggest that simple rules, i.e simplistic decision trees outperform complex ones [43, p. 97 - 140]. Therefore, a closer match to inductive rules learning is the pruned decision trees. There is an intensive research on the best pruning sequence that simplify a decision tree to resemble a human expert. Although we cannot offer a full review on the subject of pruning and decision trees, it suffices to say that finding the simplest inductive schemes from a decision tree is an intractable problem [65]. In the field of economics and learning, the most known and well-studied game that introduces inductive rules is the El Farol Bar problem and subsequently the Minority game. WB. Arthur, using the EFBP, successfully presented the learning process of the human mind to economists and game theorists believing in perfect rationality. He did so by following the psychological rules described in this section, strictly remaining inside bounded rationality limits. In the following chapter, we will provide a formal representation of the game, as it is our main research environment.

2.5.4 Evolutionary Learning and Imitation

The application of game theory in evolving populations, gave rise to evolutionary game theory. Agent-based simulations can benefit greatly from the use of evolutionary learning as it can provide robust results and lead to emergent conventions. In evolutionary settings, agents have a policy hard-coded in their "gene" and follow it until they are deleted from the evolutionary process. It should be noted that there are two main processes followed by evolutionary models. The one highlights the workings of the mutation process and population growth respectively. From the fist class of evolution, the machine learning genetic algorithms are derived. On the other hand, population growth and dynamics are best described by replicator dynamics.In the latter case we shift our focus from the individual agent behavior changes to the population, over time. Intuitively, the most successful gene-strategy has the ability to reproduce more children in the population. Moreover, social imitation can be modeled as an evolutionary process and it is a very interesting field to investigate, as it is mostly unexplored [101, p. 152]. The Replicator dynamics model is a paradigm that provides strong convergence to Nash equilibria for a wide class of games. Therefore, in this thesis we will follow the replicator dynamics learning model, in order to clearly analyze the solution concepts of the MG and derive useful results and conclusions concerning the behavior of the agents in the aggregate and individual level. In this subsection we will present how the imitation dynamics, a stochastic process, can be well approximated through a mean-dynamic, an ordinary differential equation set by the expected direction of the evolutionary process. Subsequently, we seemingly connect replicator dynamics with social imitation and reinforcement learning.

We consider a game played by a single population, where agents play equivalent roles. Let there be N players, each of whom takes a pure strategy from the set $S = \{1 \dots n\}$. We call population state x the element of the simplex $X = \{x \in \mathbb{R}^n_+ \,:\, \sum_{j \in S} x_j = 1\}$, with x_j the fraction of agents playing strategy j. A population game is identified by a continuous vector-valued function that maps population states to payoffs, i.e $F \,:\, X \mapsto \mathbb{R}^n$. The payoff of strategy i when population state is x, is described by the scalar $F_i(x)$.

Population state x^* is a Nash Equilibrium of F, when no agent can benefit and improve his profit by switching unilaterally from strategy i to strategy j. Specifically, x^* is a NE if and only if:

$$F_i(x^*) \geq F_j(x^*) \; \forall j \in S \tag{2.2}$$

$$\forall \; i \in S \text{ s.t.} \tag{2.3}$$

$$x^* > 0. \tag{2.4}$$

In population games modeled as above, agents are matched randomly and play their strategies, producing their respected payoffs. However, population games can also embody congestion games, where all the players take part in the game. Since MG is a congestion game we will deal with their formulation in detail in a subsequent chapter. Continuing, we present the basic elements of imitation models. The foundations of population model dynamics are built upon a notion called revision protocol.

Definition 1. *A revision protocol is a map $\rho : \mathbb{R}^n \times X \mapsto \mathbb{R}$ that takes as input payoff vectors π and population states x and returns non-negative matrices as output.*

Specifically, agents in a population are equipped with a time rate R, at which they review their strategy choice. Thus, the expected number of revision opportunities of N agents playing strategy i in state x, over the next dt time units, is approximately Nx_iRdt. Player i who receives a revision opportunity, switches to strategy j with probability ρ_{ij}/R, where $p_{ij}(\pi, x)$ scalar is called the conditional switch rate from strategy i to j. Hence, the expected number of switches in dt is $Nx_i\rho_{ij}dt$. Therefore, the expected change in the number of agents playing strategy i in time Δt units is

$$N\Delta x_i = N\left(\sum_{j\in S} x_j\rho_{ji}(F(x), x) - x_i \sum_{j\in S} \rho_{ij}(F(x), x)\right)\Delta t. \tag{2.5}$$

Let $\Delta t \to 0$ and $N \to 0$ such that $N \cdot \Delta t$. If we divide the above equation by N, we get the differential equation for the change rate in the portion of agents using strategy i.

$$\dot{x_i} = \sum_{j\in S} x_j\rho_{ji}(F(x), x) - x_i \sum_{j\in S} \rho_{ij}(F(x), x). \tag{2.6}$$

Equation (2.6) is the mean dynamic with revision protocol ρ in population game F. The first term is the inflow, whereas the last term captures the outflow of agents switching from strategy i to other strategies. Commonly, social imitation is modeled through the revision protocol called proportional imitation protocol $\rho_{ij}(\pi, x) = x_j[\pi_j - \pi_i]_+$ [90]. This protocol generates the mean dynamic

$$\dot{x}_i = x_i\hat{F}_i(x), \tag{2.7}$$

with $\hat{F}_i(x) = F_i(x) - \bar{F}(x)$ and $\bar{F}(x) = \sum_{i\in S} x_iF_i(x)$. Equation (2.7) is the well-studied replicator dynamic of evolutionary game theory [90]. Therefore, all the results derived for the replicator dynamic can be used to analyze social imitation. The heuristic rule described in the pairwise imitation protocol, is imitate a random agent from the street only if his payoff is better (with probability proportional to the difference). Other revision protocols are also studied in the literature of evolutionary game theory that generate replicator dynamics as their mean field, we direct the reader to Weibull (1995) [101].

In a slightly different direction, it has been established that evolutionary dynamics and reinforcement learning are equivalent in the continuous time limit [11]. This crucial result holds for the multi-population models of imitation. Suppose we have n populations playing an n-player game. Specifically, a random agent from each population is selected and they play the game at each round. Following the steps of the single population model described above, we have the pairwise proportional imitation revision protocol, $\rho_{hk}(\pi, x^i) = x^i_k[\pi_k - \pi_h]_+$, for each population $i \in N$ with population state x^i. Concretely, an agent with a revision opportunity, selects an agent to possibly imitate inside his own population. By the law of large numbers, the flow of agents switching from strategy x_h to x_k yields the following differential equations:

$$\dot{x}^i_i = \sum_{k \in S} x^i_k \rho_{kh}(F(x^i), x^i) - x^i_h \sum_{k \in S} \rho_{hk}(F(x^i), x^i) dt \; \forall i \in N, \forall h, k \in S. \quad (2.8)$$

Substituting the revision protocol and making the calculations we end up with the multi-population replicator dynamics [90] (multiplied by a factor of two).

$$\dot{x^i}_h = x^i_h \hat{F}_h(x^i), \; \forall h \in S, \forall i \in N. \quad (2.9)$$

Establishing the connection between replicator dynamics and imitation, we are able to justify the relation between learning and the well-studied replicator dynamics. In the case of the multi-population model, the imitation taking place inside each population can be seen as a reinforcement process. Specifically, one can think that each agent in a population represents a different voice or opinion and as the revision opportunities arrive the current best performing opinion gets reinforced [11]. The single population case represents the social imitation of agents, where players can observe a better strategy of another agent and learn. In the next chapter we will analyze the outcomes of the MG, using imitation as learning in both single and multi population models. It is interesting to see, whether those two learning processes lead to different results.

Chapter 3

Minority Game

3.1 The Minority Game

In the seminal paper of the MG model [21], we have an odd population of N agents competing in a repeated one-shot game ($N = 2k + 1, k \geq 1$) where communication is not allowed. At each time step (round) t of the game, every agent has to choose between one of two possible actions, either "A" or "B" ("buy" and "sell" bid respectively in a market context). These two choices are represented by integers "1" and "-1" respectively. We denote the action of agent i at time t as $\alpha_i(t) \in \{-1, 1\}$. The minority choice wins the round at each time step and all the winning agents are rewarded, following a predefined reward function $u_i(t)$. In the paper of Challet and Zhang [21], the step and the linear payoff schemes were proposed as ways of awarding points to the successful agents. In the step payoff scheme, one point is awarded to every successful prediction and none otherwise. In the linear payoff scheme, the awarded points have a linear dependence with the number of players that choose the minority side. That is, the payoff increases linearly as less people select the minority side [21]. When all the players have performed an action, the winning side is made available through a public signal and is maintained as an evolving history sequence denoted as $\mu(t)$. When the winning side, i.e the minority, is the agents taking action "-1" or "1", the public signal transmitted is the number "-1" or "1" respectively. By construction, the MG is a negative sum game, as the winners are always less than the losers. Each agent has a memory size M and is equipped with a set of fixed inductive rules S_i drawn from a rule pool. These rules are the equivalent of the "predictors" in the EL Farol Bar problem and help the agents decide which action to take at each time step. Moreover, the structure of the "predictors", follows closely the psychological heuristics rules. Concretely, each agent possess a number of decision trees, each composed by one node and each leaf is a mapping of each possible history pattern to an action. Effectively, these "frugal" decision trees can be represented as lookup tables. We will refer to each of the aforementioned history - actions mappings as strategies

History	Action
-1 -1 -1	1
-1 -1 1	-1
-1 1 -1	-1
-1 1 1	1
1 -1 -1	-1
1 -1 1	1
1 1 -1	1
1 1 1	-1

Table 3.1: A strategy example with memory size $M = 3$.

for the remainder of this thesis. The number of all possible patterns of history with size M is 2^M. Therefore, the number of the possible strategies is 2^{2^M}, a number that gets extremely large even for a modest M value. Moreover, since each agent holds different strategies, the MG becomes asymmetric. An instance of an agent's strategy with $M = 3$ is depicted in table 3.1.

Formally, let N agents be equipped with memory of length $M \in \mathbb{N}$ and draw actions from the set $A = \{-1, 1\}$. Therefore the possible histories are the set $H = \times A^M$, with $|H| = 2^M$. Therefore, the set of possible strategies is $S = \{f : H \mapsto A\}$. I.e the set of mappings from each history to an action with $|S| = |A|^{|H|}$. Let agent $i \in N$, own $s_{ij} \in S_i$ strategies, with $S_i \subset S$ and $|S_i| = n \geq 2 \; \forall i \in N$.
All the $s_{ij} \in S_i$ strategies of an agent i have to predict at every round of the game, and points are given to those strategies (no matter whether they are being selected to perform the action) that give correct predictions. The scores of all the strategies are summed and called as the *virtual* points of the strategies. These scores start at zero in the basic Minority Game, following the "tabula rasa" of the behaviorism learning theories. At each round of the game, agents make their decisions according to the strategy with the highest virtual score in that particular moment. Suppose there are many strategies with the highest score, then one of these strategies is randomly employed. Therefore, the agents use a greedy heuristic rule to pick the strategy and subsequently the proper action, i.e they always pick the best response at any given moment. The learning method used by the minority game, can be described as learning with fixed behavioral rules, inherited by the rule based learning models [54]. Moreover, players themselves who make the winning decisions are also rewarded with points, and these are called the real points of the agents.

We define the *attendance* $Att(t)$ as the sum of actions from all agents at time t. We denote the prediction of strategy s_{ij} of agent i under the information $\mu(t)$

to be $\alpha_{s_{ij}}^{\mu(t)}$ at time t. Thus, the attendance is expressed as

$$Att(t) = \sum_{i=1}^{N} \alpha_{s_{is}}^{\mu(t)} = \sum_{i=1}^{N} \alpha_i(t) \tag{3.1}$$

where s_{is}, $s \in n$, characterizes the best strategy of agent i at time t. That is,

$$s_{is} = \arg \max_{s=1..n} U_{s_{is}}(t), \ \forall i \in N \tag{3.2}$$

with $U_{s_{is}}(t)$ the virtual score of the strategy $s_{is} \in S_i$ of agent i at time t. Consequently,

$$U_{s_{ij}}(t+1) = U_{s_{ij}}(t) + \text{sgn}(1 - \text{sgn}[a_{i,s_{ij}}^{\mu(t)}(t) Att(t)]) \ \ \forall i \in N, j \in n. \tag{3.3}$$

The history $\mu(t)$ is updated as accordingly with the last minority outcome. We mention, that the history $\mu(t)$ can also conveniently represented as a bitstring, by applying the convention that the winning side has label 0, 1 when the minority actions are "-1" and 1 respectively. This convention does not change the behavior of the model and helps the computer implementation. However, it does change the mathematical formulation and in the current analysis, we will preserve the winning labels to be $\{-1, 1\}$. Notably equations 3.2, 3.3, denote the adaptation process for each agent, as the ranking of a strategy, thus the action, changes through time with respect to its virtual scores and the evolving history string $\mu(t)$. Moreover, the payoff function for each agent i at time t is as follows:

$$u_i(t) = \text{sgn}(1 - \text{sgn}[a_{i,s_{is}}^{\mu(t)}(t) Att(t)]) \ \ \forall i \in N. \tag{3.4}$$

The above reward scheme is the mathematical representation of the step pay-off as a function of the actions of the agents for each time t and signifies the real points gained by players of the MG. Although other pay-off schemes can be implemented, it has been found that most of the MG properties are independent of the gain, as long as the minority rule is kept intact [59]. The minority rule is kept by setting the comfort level, in accordance with the EFBP, as $Att(t) = 0$. In the MG, contrary to the EFBP, the agents cannot attain the comfort level but rather fluctuate around it. The MG model was initially studied through extensive simulations [21, 23, 92, 61] and it was observed that the time average of attendance $Att(t)$ always has a value of 0 ($\langle Att(t) \rangle = 0$), regardless of the parameters of the game. Therefore, the MG exhibited the same behaviour as the main conclusion of Arthur's EFBP [7]. Next quantity of interest was to measure the fluctuations of the attendance around the comfort level. In [92], the variance of the attendance was introduced to measure the efficiency of resource allocation in the game, that is if the minority side is optimally occupied. The variance is given by

$$\sigma^2 = \langle Att(t)^2 \rangle - \langle Att(t) \rangle^2. \tag{3.5}$$

A high value of variance corresponds to large fluctuations in attendance and hence an inefficient game. On the other hand, low variance corresponds to an efficient game. The behavior of variance, also called volatility in a market context, presented non-trivial behavior. Extensive simulations showed that σ^2/N is a function of the control parameter

$$P = \frac{2^M}{N}, \tag{3.6}$$

for each value of S [92]. In [92], extensive experiments established that the behavior of the system is similar for each S. Thus, in this report we will continue presenting findings for the case when $S = 2$, as it is a well explored case. As shown in Figure 3.1, when P is large, the value of σ^2/N approaches the value associated with random choice play. That is, when all agents choose an action α_i by coin toss probability. Specifically, let X be a stochastic variable which has a binomial distribution with parameters N and p. Let Y be a stochastic variable defined by $Y = lX + b$, then the expected value of Y is $lNp+b$ and the variance of Y is $(l^2)Np(1-p)$. Therefore, in our case we have $l = |A| = 2$, $b = -N$ and assuming a binomial distribution of agent's actions with probability $p = 0.5$, the variance of the attendance can be obtained as $\sigma^2/N = 0.5 \times (1 - 0.5) \times 4 = 1$. At low values of P the average value of the variance of the game is very large. Specifically, it scales as $\sigma^2/N \approx P^{-1}$, hence the losing side is greater than $N/2$ and agents behave as a herd, i.e switching sides approximately as a single unit [92]. In the case of intermediate values of P, the volatility σ^2/N is less than the random regime, and experiments showed that the minimum is $P_c \approx 0.5$ [92]. Later analytical treatment showed that the minimum value is slightly lower and $P_c \approx 0.337$ [20]. In the region where the value P_c resides, the size of the losing group is close to the minimum value of $N/2$ and the structure of the divisions allow the identification of two distinct phases around the critical value P_c. It should be noted however that in the MG, coordination is not complete and a best possible solution is not reached. Specifically, the case where agents alternate in groups of $(N-1)/2$ and $(N+1)/2$, which results in $\sigma^2/N = 1/N$.

In addition to volatility, other macroscopic quantities presented interesting behavior between the two phases. R. Savit et al. [92], examined the distributions of winning probabilities for a certain action after various history strings and discovered that these distributions are quite different in the two distinct phases. We define $P(1|\mu_k)$ to be the conditional probability of action with label "1" to be the minority after history string μ_k of length k. In Figures 3.2a, 3.2b, are shown the histograms of the conditional probability $P(1|\mu_k)$ of experiments with $N = 101$, $S = 2$ and $M = 4, 6$ respectively.

As shown in Figure 3.2a, we observe that the $P(1|\mu_k)$ histogram is flat at the value 0.5 when $P < P_c$. Whereas, in the region $P > P_c$ the histogram is not flat. Concretely, as concluded in [92], there is no extractable information for any history string of length $k = M$, when $P < P_c$. On the other hand, in the cases when $P > P_c$, the difference in probabilities for history strings, signify a predictability of the next minority side of the sequence. Thus, the phase with $P < P_c$ is called symmetric or unpredictable and the phase with $P > P_c$ is

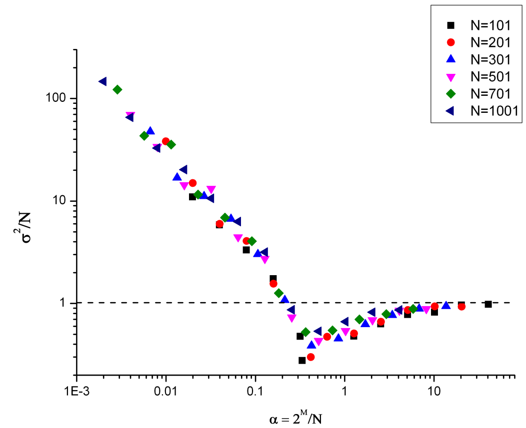


Figure 3.1: Volatility with respect to the control parameter $P = 2^M/N$ for $S = 2$ and various number of agents. The dashed line is the volatility in the random play.

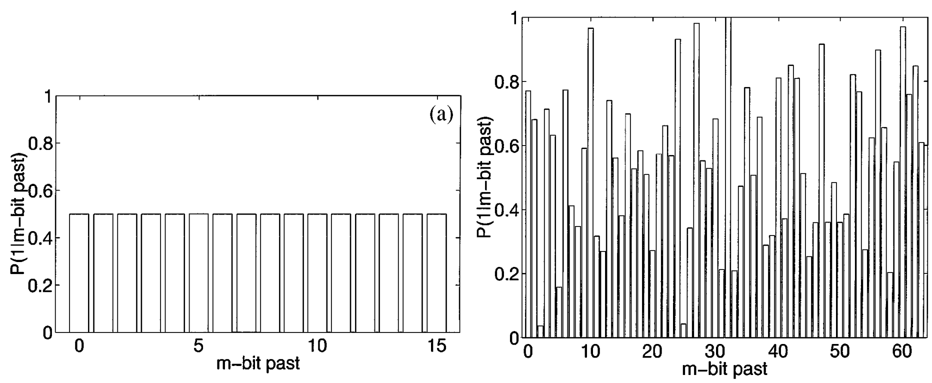


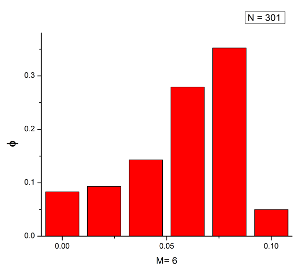
Figure 3.2: (a) A histogram of the conditional probability $P(1|\mu_k)$ with $k = 4$ for the game played with $M = 4$. The bin numbers, when transformed in binary form, yield the corresponding strings μ. (b) A histogram of the conditional probability $P(1|\mu_k)$ with $k = 6$ and $M = 6$ [92].

called asymmetric or predictable. These observations where sharpened in [27] along with the confirmation of a phase transition at critical value $P_c = 0.337$ where the minimum σ^2 takes place (for $S = 2$). As a measure of non-uniformity of the winning probabilities given certain history-information, predictability H

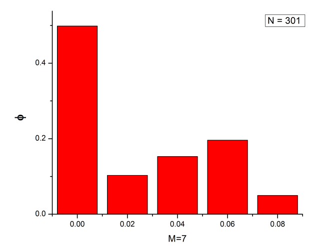
was defined as follows.

$$H = \frac{1}{2^M} \sum_{\mu=1}^{2^M} \langle \mathrm{sgn}(Att)|\mu\rangle^2. \tag{3.7}$$

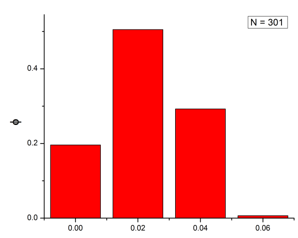
Hence, in the symmetric phase we have $\langle \mathrm{sgn}(Att)|\mu\rangle = 0 \; \forall \mu$. On the contrary in the asymmetric phase we get $\langle \mathrm{sgn}(Att)|\mu\rangle \neq 0$ for at least one μ. Finally in [27], it was found that for a fixed M, H is a decreasing function of the number of agents N. Furthermore, important phenomena of MG arise from the microscopic state change, when agents switch strategy. In article [27], the concept of frozen agents was introduced, as the fraction ϕ of agents that play the same strategy all the time. In Figures 3.3a, 3.3b, 3.3c are plotted the average distributions of the frozen agents ϕ as histograms for $N = 301$ and $M = 6, 7, 11$ respectively.



(a) $M = 6$



(b) $M = 7$



(c) $M = 11$

Figure 3.3: Normalized histograms of average distributions of the frozen agents ϕ for $N = 301$ and $M = 6, 7, 11$.

We note the increase of frozen agents in the instance of $M = 7$ in Figure 3.3b close to the critical point. Moreover, in Figure 3.4 we present the information predictability and the fraction of frozen agents w.r.t to the control parameter P, for various memory sizes. We remark that, $H = 0$ for $P \lesssim 0.3$ and $H \neq 0$ for $P > 0.3$.

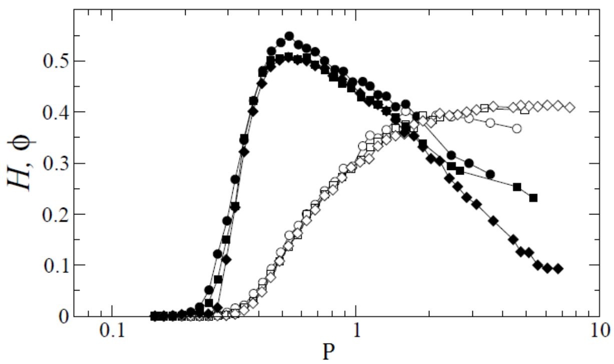


Figure 3.4: Information H (open symbols) and fraction of frozen agents ϕ (full symbols) as a function of the control parameter $P = 2^M/N$ for $s = 2$ and $M = 5, 6, 7$ (circles, squares and diamonds in respective order)[68].

The discontinuity on predictability and frozen agents shown in the Figures above, certified the existence of phase transition at some critical value P_c and paved the way to solve analytically the MG for $S = 2$ using replica symmetry [27]. The statistical mechanics approach of the MG, gave accurate answers for σ^2 in the region of $P > P_c$. However in the case of $P \leq P_c$, due to the degeneracy of the predictability H, the theoretical results did not provide predictions in accordance to the experimental ones [22, p. 49 - 57]. At this point the literature concerning the MG started to grow, following three main research streams. The first one extended the MG (or simplified it), in order to study in detail the analytical interpretation of the MG, using statistical tools or a functional approach [22, p. 57 - 61]. The second research path focused on the application of the MG in modeling financial markets. The Minority Game and a trading market share crucial common features. Namely, that it is intractable to calculate an optimal strategy and agents should differentiate among themselves in both cases. However, financial markets are much more complex involving different capitals, time horizons and trader needs. Therefore, the second stream of the MG literature involved variations of the model aimed to reproduce realistic market fluctuations. The interested reader in the above schemes is referred to the book [22, ch. 3] and the references therein, for a detailed description of the simplifications and the analytical methods used on the MG. A detailed review of the market models based on the MG and their corresponding results can be found in [22, ch. 4]. An extensive list of the variations of the MG can be found in the official website of the MG maintained by one of the authors of the original model [19]. Finally, extensions on the MG were proposed to answer a research question posed in some degree in the original paper, that is, how efficiently a group of adaptive agents allocating limited resources can coordinate. The fact that agents, given bounded rationality and following a simple learning regime, are able to coordinate better than random, was seen as non-trivial result. In the early research of the MG, Darwinism was proposed, where at time step τ the worst-performing player is replaced with a clone of the best, expect that one strategy is redrawn with a small probability to allow searching the whole span of strategies [23]. Darwinism increased the efficiency of the system in the critical phase, however maintaining the fluctuations of attendance. Thus, a "quest for better coordination" was initiated, followed by a series of proposed alternative methods of learning for the MG. Specifically, a question overlapping the coordination scheme and the modeling of financial markets is the fixed rules learning method, used in the standard MG, able to model the behavior of agents in real life cases and specifically in a market context [22, p. 87]. When the problem faced by a player is simple and does not have a significant impact, for instance choosing between two roads to get back home, it might be the case that the MG fixed rules can model successfully the situation. On the other hand, this not true when players are faced with important decisions in a trading scheme [62]. What is more, the agents of the MG do not play strategically. Agents, discussed so far do not play the game against $N-1$ other players, but rather against the signal μ for each t, which is insufficient to lead to an optimal outcome [22, p.93]. However, this fact along with the limited adaptation capability of the agents (only a few unchanged strategies given to each agent), raises the main question of this report, i.e is fixed rules really learning and in extend, how can we model a learning process within the framework of bounded rationality. In game theoretic terms, the research question can be translated to a refinement of the Nash Equilibria through various learning techniques in MG. Thus, in the next section we will review the most important learning methods applied in the MG, along with their results, particularly following cases where agents have limited capabilities. In latter sections, we analyze MG as a congestion game and refine the NE using evolutionary game theory and Imitation learning.

3.2 Alternative Learning Methods in the Minority Game

The vast analysis of the MG in the literature using econophysics methods, led to insightful results concerning real market price fluctuations. By dramatically simplifying the individual agent's sophistication, although maintaining the diversity of beliefs and opinions, the MG manages to model the collective behavior of the financial markets [22, ch. 4]. Nonetheless, the statistical analysis of the macroscopic quantities used for the MG, does not discuss the changes in behavior of the agents in the individual level. Moreover, The Minority Game, is a challenging platform for adaptive agents that pursue their own selfish goals. As a consequence, various learning schemes are tested in the literature, using the MG environment as testbed. In general, the different models proposed for the MG increase sophistication and intelligence of the agents, striving however to remain inside the bounded rationality limits. In [62], as a first step towards modeling markets with the MG, the learning method proposes that each agent should take into account his own impact of the market. Namely, the learning dynamics are updated as follows.

$$U_{i,j}(t+1) = U_{s_{ij}}(t) + \text{sgn}(1 - \text{sgn}[a_{i,s_{ij}}^{\mu(t)}(t)A(t)]) + \eta\delta_{is}(t) \;\; \forall i \in N, j \in S. \tag{3.8}$$

Where, the additional term rewards the strategy actually played by the agent. It is clear that for $\eta = 0$, we resort to the original case. Therefore, the aforementioned term, can be viewed as a reinforcement procedure towards an optimal picking of strategies. However, to correctly set the parameter η, each agents requires to know what payoff she would have got if she had played any strategy s, including the not used strategies $s_{ij} \neq s_{is}$. Since the agents in the modified MG model must account for their own action, they are called "sophisticated" agents. In turn, the agents using the learning method of the standard MG, are called "naive". Additionally as noted in the previous section,the macroscopic quantities of the MG remain unaffected by the modified MG. Furthermore, in [62] the softmax decision heuristic, coupled with the Boltzmann distribution, is used to assign probabilities to strategies at each time step, in order to ease mathematical calculations. The paper concludes that with exponential learning and full information, agents coordinate on a Nash equilibrium. The convergence depends on the initial conditions $U_{ij}(0)$, different initial beliefs select different Nash equilibria. The Nash equilibria of the aforementioned model are also discussed in [27] , where is shown that the number of Nash equilibria grows exponentially with N.

In a slight different direction, in [50] agents undergo the process of evolution and are equipped with one strategy S. The version of the MG described in [50], is also referred as Evolutionary Minority Game (EMG) in the literature [48]. To make decisions, all agents are assigned a different probability p_i at the beginning, with $0 \leq p_i \leq 1$, which is defined as the probability that agent i acts according to the strategy S, i.e. follow the recent winning action or the last outcome for that M-bit history. With a probability $1 - p_i$, agent i chooses the choice opposite to the past winning action for that history. This probability p_i takes the role of the "gene" of agents as in the evolutionary game theory models. Thus, the scores are rewarded or penalized subject to p_i. Moreover, agents can change the value p_i within a certain range, if the failures drop below a specific threshold. The results of this game showed that agents tend to self-segregate in to opposing groups of $p = 0$ and $p = 1$.

Reinforcement learning algorithms are also used in the El Farol Bar Problem and naturally extend to the Minority Game. Notably in [102], the Roth-Erev reinforcement learning algorithm is proved to converge to a Nash equilibrium of the stage game of the El Farol Bar Problem for any comfort level. Therefore, this critical result can be naturally extended to the Minority Game. Moreover, in [4] it is shown numerically that Q learning yields a stationary state close to a Nash equilibrium. In [17], through simulations, the convergence of Roth-Erev reinforcement learning to NE is confirmed. Notably in [17], the experiments using reinforcement learning methods in the MG converge to a certain type of Nash equilibrium with agents evenly divided into the two actions and one playing a mixed strategy. As we will describe in detail in the next section, this is an efficient state of the game.

In [55, 100, 45] interacting neural networks are trained in the history of the MG and develop a good strategy towards competition. Moreover, it is shown that a system of neural networks possesses several advantages compared to the original learning algorithm of decision tables. Finally in the spirit of machine learning algorithms, in [96] the agents use one-point genetic crossover mechanism to mutate their strategies at hand. The genetic model, not surprisingly, reaches rapidly an efficient state with minimal fluctuations.

It is clear that many learning algorithms stemming from different motivations, have been tested in the MG benchmark. The vast majority of experiments has optimal agent coordination as a goal. As presented in this chapter, with few exceptions, that is accomplished. However, as we will explain in detail in the next subsection, the points of coordination of the agents are many in the Minority Game. Therefore, where do players converge or how do agents manage to coordinate are still questions to be researched. Moreover, the classification of individual agent behaviors in the MG has been seldomly addressed. In order to answer the aforementioned queries, one should first analyze the hierarchical structure of the Nash Equilibria of the MG. An initial effort was performed in [62, 27], where the existence of a complex hierarchical organization of NE of the MG is signaled. However as concluded in [62], the efforts to study Nash equilibria in the MG is far from complete. In addition to the different learning dynamics between the "naive" and "sophisticated" agents in the MG, as we will see clearly in the next section, the Nash equilibrium points are many. Thus, how do "sophisticated" agents manage to converge into the same equilibrium is an interesting question. Moreover, the time it takes for agents to learn a certain equilibrium in MG is another captivating research query [22, p. 95 - 97].

In this report we will contribute towards the classification of different behaviors in the MG and the types of coordination achieved. To this end, we will first discuss the Nash equilibria of the MG stage game. Continuing, following evolutionary game theory, we will describe the NE where the replicator dynamics learning regime converge. As mentioned in the previous chapter, replicator dynamics lead to Nash equilibria of the underlying game; which in turn makes it a primary tool for analyzing types of NE, especially when many exist.

3.3 Minority Stage Game and Nash Equilibria

The game theoretic interpretation of the Minority Game is a single MG stage game with a state of the world as μ. We define the Minority Stage Game as follows.

Definition 2. *Define the Minority stage game as the one shot strategic game* $\Gamma =< N, \Delta, u_i >$ *consisting of:*

- N *players indexed by* $i \in \{1, \dots, 2k+1\}, k \in \mathbb{N}, N = \{1, \dots, 2k+1\}$,

- *a finite set of strategies $A_i = \{-1, 1\}$ indexed by α , where α_i denotes the strategy of player i and*
- *a payoff function $u_i : \alpha_i \times \alpha_{-i} \mapsto \mathbb{R} = \{0, 1\}$, where $\alpha_{-i} = \prod_{i \neq j} \alpha_j$. More formally,*

$$u_i = \begin{cases} 1 & \textit{if } -\alpha_i \sum_{j=1}^{N} \alpha_j \geq 0 \\ 0 & \textit{otherwise.} \end{cases} \tag{3.9}$$

Note that the "inversion" symmetry $u_i(-\alpha_i, -\alpha_{-i}) = u_i(\alpha_i, \alpha_{-i})$ implies that the two actions are a priori equivalent: there cannot be any best actions, because otherwise everybody would do that and lose [62]. Furthermore, the set of mixed strategies of player i is denoted by $\Delta(A_i)$. We describe a mixed strategy profile by $\alpha \in \times_{i \in N} \Delta(A_i)$. Let us now characterize the equilibria of the Minority stage game. We denote the set of Nash equilibria of the stage game as $\bar{Y}$. In [102] it is shown that $\bar{Y}$ contains a finite number of elements. We have three general types of Nash equilibria, namely:

- Pure Strategy Nash Equilibria. I.e., when all players play a pure strategy.
- Symmetric Mixed Strategy Nash Equilibria. That is, the agents choose the same mixed strategy to play.
- Asymmetric Mixed Strategy Nash Equilibria. Specifically, the NE when some players choose a pure strategy and the rest a mixed strategy.

The elements of a pure strategy NE can be easily defined.

Proposition 1. *A pure strategy profile is a Nash equilibrium if and only if one of the actions $\mathbb{A} = \{-1, 1\}$ is chosen by exactly k of the $2k + 1$ players [97].*

Proposition 2. *The number of pure strategy Nash Equilibria in the stage game of the original MG is $2\binom{N}{\frac{N-1}{2}}$.*

Proof. The number of pure strategy Nash equilibria in the Minority stage Game is the sum of two parts. The first part is the number of ways $\frac{N}{2} - 1$ different players can be chosen out of the set of N players at a time with minority side "0". Similarly, the second part is the same number as the first part with the difference that the chosen set of players are labeled with winning side "1". $\Diamond$

We continue to characterize the asymmetric Nash equilibria of the underlying game, where some players follow a pure strategy and the remaining ones a mixed strategy. The agents playing a mixed strategy are called mixers.

Lemma 1. *Let be $\alpha \in \times_{i \in N} \Delta(A_i)$ a Nash equilibrium with a non-empty set of mixers. Then all mixers use the same mixed strategy [102, 53].*

In addition, the asymmetric Nash equilibria of the MG stage game can be divided into subtypes [54]. Namely, we define the type (l, r, λ) of an asymmetric Nash equilibrium, where $l, r \in 0, 1..2k + 1$ denote the number of players choosing pure strategy "-1" or "+1", $\lambda \in (0, 1)$ the probability with which the rest of players (mixers) $z(l, r, \lambda) = (2k + 1) - (l + r) > 0$ play pure strategy "-1".

Let $u_{-1}(l, r, \lambda), u_{+1}(l, r, \lambda)$ denote the expected payoff of an agent choosing -1 and +1 respectively. A Nash Equilibrium is defined if and only if

$$u_{-1}(l + 1, r, \lambda) = u_{+1}(l, r + 1, \lambda). \tag{3.10}$$

These equilibria are of type $z(k, k, \lambda)$ for any $\lambda \in (0, 1)$. In this case, the mixer uses an arbitrary mixed strategy, whereas the remaining $2k$ players are spread evenly over the two pure strategies.

Moreover, equilibria with more than one mixer exist. With, $l + r \leq 2k - 1$, there is a Nash equilibrium of type (l, r, λ) if and only if $\max\{l, r\} < k$. The analogous probability $\lambda \in (0, 1)$, solves equation 3.10 and it can be shown to be unique [53].

Following the above, there exist a unique symmetric mixed strategy Nash Equilibrium where all players choose one the two actions with probability $p = 1/2$. It is clear from the stage game equilibrium analysis that the solution points are many, creating the difficulty of n-players to coordinate to a specific solution. Finally, for the sake of completeness we define:

Definition 3. *The Minority Game is the infinite repeated Minority Stage Game.*

3.4 Minority Game as a Congestion Game

A large part of the current literature on the inefficiency of equilibria concerns congestion games [72, ch. 18, p. 461]. Congestion games are an active research area, as they can model situations where many agents strategically interact in order to utilize common resources. Introduced by W. Rosenthal [84], congestion games model instances when the payoff of each player depends on the choice of resources along with the number of players choosing the same resource [84]. Examples that congestion games can model successfully are the route choice in a road network or selfish packet routing in complex structures, such as the Internet [72, ch. 18]. Congestion games are equipped with many attractive properties that may provide further refinement of the Nash equilibria of the MG. Most importantly, congestion games are potential games [84]. Therefore, there exist a single scalar-valued function that characterizes the game [88, p. 53].

Minority Game can be naturally modeled as a congestion game associated with a congestion model, as the two available choices to the agents can represent two distinct resources.

A congestion model $(N, M, (A_i)_{i \in N}, (c_j)_{j \in M})$ is described as follows:

- N the number of players.

- $M = \{1..m\}$ the number of resources.
- A_i the set of strategies of player i, where each $a_i \in A_i$ is a non empty set of resources.
- For $j \in M$, $c_j \in \mathbb{R}^n$ denotes the vector of benefits, where c_{jk} is the cost (e.g cost or payoff) related to each user of resource j, if there are exactly k players using that resource.

The congestion game associated with a congestion model is a game with the set of N players, with sets of strategies $(A_i)_{i\in N}$ and with cost function defined as:
Let $A = \times_{i\in N} A_i$ be the set of all possible players pure strategy vectors. For any $\vec{a} \in A$ and for any $j \in M$, let $\sigma_j(\vec{a}) = \#\{i \in N : \ j \in \vec{a}\}$ be the number of players using resource j, with $\vec{a}$ the current profile. We have the overall cost function for player i [93, p. 174]:

$$C_i = \sum_{j\in a_i} c_j(\sigma_j(\vec{a})) = -u_i(\vec{a}). \tag{3.11}$$

Moreover, it holds that

$$\sum_{i\in\{\sigma_j(\vec{a})\}} u_i(\vec{a}) = -\sum c_j(\sigma_j(\vec{a})), \ \forall j \in M, \vec{a} \in A. \tag{3.12}$$

That is the total cost of using resource j is the opposite of the total benefit of agents using this resource. We remark that congestion games have an anonymity property. Specifically, players care about how many others use a given resource, rather than which do so [93, p. 175]. Concretely, The MG is a congestion game with $M = 2$ resources labeled as $\{-1, +1\}$, A_i the set of strategies of each player i and a cost function

$$c_{jk} = \begin{cases} -1 & \text{if } k < \frac{N}{2} \\ 0 & \text{otherwise,} \end{cases} \tag{3.13}$$

for each resource $j \in M$ and $C_i = -u_i \ \forall i \in N$. The translation of the MG as a congestion game helps us to use critical properties of the latter to analyze the Nash Equilibria of the game.

Theorem 1. *Every Congestion Game is a potential game and admits the exact potential function (or just potential) of the form [67]:*

$$P(\vec{a}) = \sum_{j\in M} \sum_{k=1}^{\sigma_j(\vec{a})} c_j(k) \tag{3.14}$$

The potential function captures the equality of the players with respect to the available resources. Specifically, it shows the change in the cost of using a facility is due to the number of agents using a resource, rather than who is using the facility. We define a potential of a game $G = \langle N, (A_i), u_i \rangle$ in strategic form, with A the collection of all deterministic strategy vectors in G.

Definition 4. *A function $P : A \mapsto R$ is a potential of the game G if $\forall \vec{a} \in A, \forall a_i, b_i \in A_i$ $u_i(b_i, \vec{a}_{-i}) - u_i(a_i, \vec{a}_{-i}) = P(b_i, \vec{a}_{-i}) - P(a_i, \vec{a}_{-i})$ [67].*

Thus, the Minority Game is a potential game [97] and based on the above, we can derive the potential function of the Minority Game. Equation 3.14, can be written as

$$\sum_{k=1}^{\sigma_{-1}(\vec{a})} c_{-1,k} + \sum_{k=1}^{\sigma_{+1}(\vec{a})} c_{+1,k}, \tag{3.15}$$

where $j \in \{-1, +1\}$ labels the two available resources, actions of the Minority Game, which posses equal congestion cost. Let μ, λ the number of players to choose side $-1, +1$, respectively, in a strategy profile $\vec{a} \in A$, with $\mu + \lambda = 2h + 1 = N, h \in \mathbb{R}$. Without loss of generality, we assume $\mu > \lambda$, that is λ is the minority. Therefore 3.15 becomes

$$P = \sum_{k=1}^{\mu} c_k + \sum_{k=1}^{\lambda} c_k = \sum_{k=1}^{(h+1)+c} c_k + \sum_{k=1}^{(h-c)} c_k =$$

$$-h + c(h+1) - (h-c) = -h + 0 - \lambda = -h + \sum_{i=1}^{N} u_i(\vec{a}) = \tag{3.16}$$

$$-h + \sum_{i=1}^{N} u_i(\vec{a}). \tag{3.17}$$

Where $\mu = (k+1) + c,\ \lambda = k - c$ with $c \in \mathbb{R}$. Moreover, we have the following Lemma [67]:

Lemma 2. *Let P_1 and P_2 be potentials of a finite game G. There exist a constant c such that*

$$P_1(\vec{a}) - P_2(\vec{a}) = c\ \forall \vec{a} \in A.$$

Thus, we state the proposition concerning the potential function of the MG.

Proposition 3. *An exact potential of the MG is the sum of the payoffs of all the $N = 2h + 1 \in \mathbb{R}$ players. Therefore,*

$$P(\vec{a}) = \sum_{i=1}^{N} u_i(\vec{a}) \text{ with } \forall \vec{a} \in A \tag{3.18}$$

Proof. If follows directly from Lemma 2 and equation 3.16, with constant $c = -h$. ◊

3.5 Imitation in the Minority game

It is interesting to analyze the Minority Game through imitation. There are a lot of ways with which imitation can be modeled in a population of strategically interacting agents. That is, if we take into account the different heuristic rules agents apply to imitate, the mutation or noise in imitation process and the various (social) networks that players can be a part of.

As introduced in section 2.5, imitation can be interpreted in two ways, depending on the type of the population game played. In single-population games, imitation is social learning and each agent in the population represent a potential player using a mixed strategy. On the other hand, in the n-population case, each agent of a population represent a different opinion inside the mind of the (human)-player. Therefore, each population is the player and the imitation process reinforces the best performing ideas of each player. In a nutshell, single population imitation represents social learning and multi-population imitation models individual learning.

3.5.1 Multi-Population Replicator Dynamics

For a model of pure imitation, we assume that agents imitate without noise. Moreover, all reviewing players follow the heuristic of adopting the strategy of "the first man they meet of the street" with a probability proportional to their score difference. Suppose that the review rate is linearly decreasing in the average payoff, then the process of imitation can be modeled through the standard replicator dynamics [101, ch. 4].
The Minority Game is a n-person game, therefore we use the standard n-population replicator dynamics [101]. Let $N = \{1, ..., 2k+1\}$ be a set of populations, with each population representing an agents role i in the MG. Thus, each population can be divided into two subpopulations, one for each of the pure strategies in the minority game. A population state is a vector $\vec{a} = \alpha = (\alpha_1, ..., \alpha_{2k+1})$ or point in the polyhedron $\Delta(A)$ of the mixed strategy profiles. Moreover, each component α_i is a point in the simplex $\Delta(A_i)$, denoting the proportion of agents programmed to play the pure strategy $a_i \in A_i$. Time is continuous and indexed by t and agents – one from each population – are continuously drawn uniformly at random from these populations to play the minority game. The imitation dynamics modeled through replicator dynamics, are expressed as follows:

$$\forall i \in N,\ \forall \alpha_i \in A_i\ :\quad \dot{\alpha_i}(a_i) = \alpha_i(a_i)(u_i(a_i, \alpha_{-i}) - u_i(\alpha_i, \alpha_{-i})). \tag{3.19}$$

This system of differential equations defines the continuous time multi-population replicator dynamics [101]. Concretely, the growth rate $\frac{\dot{\alpha}(a_i)}{\alpha(a_i)}$ of a pure strategy $a_i \in A_i$ in population $i \in N$ of constant size, is equal to the difference in payoffs of the pure strategy and the current average payoffs for the population. Therefore, the population shares of strategies that do better than average will be imitated more often, while the shares of the other strategies will decline. Moreover, we can see that the subpopulations associated with the pure best replies to the current population state have the highest growth rates.

Equations 3.19, define a continuous solution mapping $\psi : \mathbb{R} \times (\times_{i \in N} \Delta(A_i)) \mapsto \times_{i \in N} \Delta(A_i)$ $\forall$ time t, $\forall$ initial conditions $\alpha^0 \in \times_{i \in N} \Delta(A_i)$. The solution mapping ψ assigns the population state $\alpha^{\psi} = \psi(t, \alpha^0)\ \forall t, \alpha^0$ and the solution trajectory of an initial condition α^0 is the graph of the solution mapping $\psi(\cdot, \alpha^0)$

A population state $\alpha \in \times_{i \in N} \Delta(A_i)$ is a stationary state of the replicator dynamics 3.19, if and only if for each population $i \in N$ every pure strategy $a_i \in A_i$ used by some players in the population gives the same rewards. Thus, $\dot{\alpha_i}(a_i) = 0\ \forall i \in N, \forall \alpha_i \in A_i$. Let T be the set of stationary states of equations 3.19, with $T = \{\alpha \in \times_{j \in N} \Delta(A_j) | \forall i \in N, \forall \alpha_i \in A_i\ :\ \dot{\alpha_i}(a_i) = 0\}$. Suppose $\alpha \in T$, then, by definition, α is a pure or a mixed strategy. In the case of the latter, following Lemma 1, all mixers must use the same strategy. If there is more than one mixer, the common mixed strategy, determined by equation 3.10, is unique and defined by the number of players choosing the two pure strategies. We conclude that the set T of stationary states can be partitioned into three subsets [53]. Namely,

T1 : The connected set of Nash equilibria with at most one mixer,

T2 : Nash equilibria with more than one mixer and

T3 : non-equilibrium profiles of the type (l, r, λ).

$$\begin{cases} l, r \in 2k+1, \\ l + r \leq 2k+1, \\ \text{if } l + r < 2k+1 \text{ then } \lambda \in (0,1) \text{ is uniquely defined by solving equation 3.10.} \end{cases}$$

Furthermore, we analyze the stationary states of the MG under replicator dynamics. We consider two types of stability, namely Lyapunov stability and asymptotic stability. Concretely, a population state $\alpha \in \times_{i \in N} \Delta(A_i)$ is Lyapunov stable if every neighborhood B of α contains a neighborhood B^0 of α such that $\psi(t, a^0) \in B$ for every $x^0 \in B \cap \times_{i \in N} \Delta(A_i)$ and $t \geq 0$. Moreover, a stationary state is asymptotically stable if it is Lyapunov stable, and, in addition, there exists a neighborhood B^*, with $\lim_{t \to \infty} \psi(t, a^0) = \alpha\ \forall \alpha^0 \in B^* \cap \times_{i \in N} \Delta(A_i)$. In words, if all solutions of the population system that start out near an equilibrium profile α stay near α through time, then α is Lyapunov stable. Consequently, a population state α is asymptotically stable if it is Lyapunov stable and a small perturbation in the population shares, results in the movement of the system to the original state α.

The Minority Game analysis as a congestion game in the previous section, proves fruitful as we can rewrite the replicator dynamics equation using the potential function U of the MG. Specifically, using definition 4 of the potential function of the pure strategies and extending it to mixed strategy space, using expectations, in equations 3.19, we get

$$\forall i \in N,\ \forall \alpha_i \in A_i\ :\quad \dot{\alpha_i}(a_i) = \alpha_i(a_i)(U(a_i, \alpha_{-i}) - U(\alpha_i, \alpha_{-i})). \tag{3.20}$$

Thus, the following proposition holds.

Proposition 4. *The potential function U of the minority game is a Lyapunov function for the replicator dynamic: for each trajectory $(\alpha(t))_{t \in [0, \infty]}$, we have $\frac{dU}{dt} \geq 0$. Equality holds at the stationary states [53].*

Proof. We have

$$\frac{dU(\alpha)}{dt} = \sum_{i=1}^{N} \sum_{a_i \in A} \frac{\partial U(\alpha)}{\partial \alpha_i(a_i)} \dot{\alpha_i(a_i)}. \tag{3.21}$$

Since $U(\alpha) = \alpha_i(a_i)U(1, \alpha_{-i})\ \forall i \in N, a_i \in A$ equation 3.21 becomes

$$\frac{dU(\alpha)}{dt} = \sum_{i=1}^{N} \sum_{a_i \in A} U(a_i, \alpha_i) \dot{\alpha_i(a_i)}$$

$$= \sum_{i=1}^{N} \sum_{a_i \in A} U(a_i, \alpha_i)(\alpha_i(a_i)(U(a_i, \alpha_{-i}) - U(\alpha_i, \alpha_{-i}))$$

$$= \sum_{i=1}^{N} \sum_{a_i \in A} (\alpha_i(a_i)U(a_i, \alpha_{-i})^2 - U(\alpha_i, \alpha_{-i})^2)$$

$$= \sum_{i=1}^{N} (\mathbb{E}_{\alpha_i}[U(a_i, \alpha_{-i})^2] - (\mathbb{E}_{\alpha_i}[U(a_i, \alpha_{-i})])^2)$$

$$= \sum_{i=1}^{N} \mathrm{Var}_{\alpha_i} U(a_i, \alpha_{-i}) \geq 0.$$

Where equality holds only when variances are zero, i.e α is a stationary point of the replicator dynamics. ♢

Finally, we can conclude the following proposition

Proposition 5. *The collection of Nash equilibria with at most one mixer in $T1$ is asymptotically stable under the replicator dynamics. Moreover, stationary states in $T2$ and $T3$ are not Lyapunov stable [53].*

Proof. The Nash Equilibria of the set $S1$, are global maxima of the potential function U. That is, all pure strategy profiles along with the profiles with one mixer maximize the function U. By Theorem 6.4 of Weibull [101], this connected set of global maxima of the Lyapunov function U is asymptotically stable.

The elements of $S2$ are not Lyapunov stable, as is the case for points in $S3$. Suppose $\alpha^* \in S2$, be a NE with more than one mixer. Furthermore, suppose it is Lyapunov stable. Since it is an isolated point of the collection of stationary states, there exist a neighborhood D of α^* whose closure contains only the stationary state α^*, i.e $cl(D) \cap S2 = \{\alpha^*\}$. Lyapunov stability states that, as long as the initial state $\alpha(0)$ is within a sufficiently small neighborhood D' of α^*, the solution trajectory $\alpha(t)$ $t \in [0, \infty)$ remains in D. Let $i \in N$ be one the mixers in the NE α^*. Player i is indifferent between the two pure strategies, therefore:

$$U(\alpha^*) = U(-1, \alpha^*_{-i}) = U(1, \alpha^*_{-i})$$

Therefore, $U(\gamma_i, \alpha^*_{-i}) = U(\alpha^*)\ \forall \gamma_i \in \Delta(A_i)$. Moreover, for $\gamma_i \neq \alpha^*_i$ sufficiently close to α^*_i, we have that $(\gamma_i, \alpha^*_{-i}) \in D'$. Thus, the solution trajectory $s_\gamma = \gamma(t)\ t \in [0, \infty)$ with initial condition $\gamma(0) = (\gamma_i, \alpha^*_{-i}) \in D$. From Proposition 4, since the potential function U is maximized to reach a stationary state, we have that the Lyapunov function U strictly increases along the trajectory, until it may reach a stationary state. Let $\gamma^* \in \times_{j \in N} \Delta(A_j)$ be the limit point of s_γ. Then, there is a strictly increasing sequence of time points $t_m \to \infty$ with $\lim_{m\to\infty} \gamma(t_m) \to \gamma^*$. Lemma A.1 of [89, p. 104] certifies that such a limit point exists and has to be a stationary point. Since $cl(D) \cap S2 = \{\alpha^*\}$ and the trajectory lies in D, therefore $\gamma^* = \alpha^*$. Thus, $\lim_{m\to\infty} U(\gamma(t_m)) = U(\alpha^*) = U(\gamma(0))$. But this contradicts that the Lyapunov function is increasing along the trajectory. Hence, α^* is not Lyapunov stable.

The same reasoning applies for $\alpha^* \in S3$. As α^* is not NE, a mixer i will deviate slightly to profit (remaining in D'), maximizing the potential function, however still having α^* as a limit point. $\diamondsuit$

Thus, as a corollary of the above result, we note that the symmetric NE of the MG is not Lyapunov stable, following imitation learning regime.

3.6 Three-Player Minority Game

We analyze the case of the three player MG, in order to derive and visualize useful results concerning the Nash equilibria of the stage game and the convergence of the replicator dynamics. We first derive the payoff function for each player i with $i \in \{1, 2, 3\}$, allowing mixed strategies. The payoff matrix of the three-player MG can be viewed in Table 3.2.

Table 3.2: Payoff matrix of the three-player MG. A1,A2 and A3 denote agents 1,2,3 respectively with actions $\{-1, 1\}$. The utility matrix is split into two submatrices using agent A3 actions as a divider. The payoffs for each agent are presented w.r.t to their number, i.e. payoff 010 means payoff for A1=0, A2=1 and A3=0.

	A3	-1		A3	1
A1/A2	-1	1	A1/A2	-1	1
-1	000	010	-1	100	100
1	100	001	1	010	000

Since player i has two pure strategies from the set $A_i = \{-1, 1\}$, he can play a strategy $a_i \in A_i$ at each stage of the game. The mixed-strategy simplex is defined as $\Delta(A_i) = \{\alpha_i \in \mathbb{R}^2_+ : \sum_{a_i \in A_i} \alpha_{ia_i} = 1\}$. Since all probabilities are non-negative and sum up to one, we can express the mixed-strategy simplex of player i as the line segment Δ^1_i, without loss of information. We denote $x, y, z \in \mathbb{R}[0, 1]$ the probabilities for each player to play strategy $a_i = 1$ respectively.

Therefore, a complete mixed strategy profile $\alpha \in \times_{i \in N} \Delta_i$ of the players can be expressed as a point $\alpha(x, y, z) \in \mathbb{R}^3[0, 1]$.
Following the fact that the expected payoff of a mixed strategy is exactly the proportion of the payoff of the pure strategy, we have the utility function of player 1 defined as

$$\begin{aligned} u_1(x, y, z) = x \cdot u_1(1, y, z) + (1 - x) \cdot u_1(0, y, z) = \\ x \cdot (y \cdot u_1(0, 0, z) + (1 - y) \cdot u_1(0, 1, z)) + \\ (1 - x) \cdot (y \cdot u_1(1, 0, z) + (1 - y) \cdot u_1(1, 1, z)) \Rightarrow \\ u_1(x, y, z) = x - xy - xz + yz. \end{aligned} \tag{3.22}$$

In a similar manner we derive the payoff functions of player 2 and 3 with respect to the mixed strategy profile. We have,

$$u_2(x, y, z) = y - yx - yz + xz \tag{3.23}$$

$$u_3(x, y, z) = z - zx - zy + yx \tag{3.24}$$

Therefore, using equation 3.16 we have the potential function for the three player game as follows:

$$U(x, y, z) = u_1 + u_2 + u_3 = x + y + z - xy - yz - zx \tag{3.25}$$

It is easy to see that $\max U(x, y, z) = 1$ and the points where the potential function is maximized, represent the “utilitarian” Nash equilibria of the game. Furthermore, if we constrain the maximization of the potential function, posing equality in utilities, we end up with the unique symmetric mixed Nash equilibrium $(1/2, 1/2, 1/2)$. The “egalitarian” symmetric Nash equilibrium, preserves equality, at the expense of inefficiency of utilities in the aggregate level $U(1/2, 1/2, 1/2) = 3/4$. In Figure 3.5, we present the contour of the maximized potential function, with respect to the mixed strategies of the players. The highlighted edges of the mixed strategy space (cube) connect the pure Nash equilibria of the game that are the “utilitarian” solutions of the game. Namely, it maximizes the sum of the utilities. Moreover, the line of Nash equilibria in Figure 3.5, is exactly the connected set of equilibria with one mixer. Moreover, to gain an insight on the curvature of the potential function, we plot in Figure 3.6 the surface of the parametrized function $U(x, y, z) = 1$ with respect to an extended space $\mathbb{R}^3[-2, 2]$. Due to the boundary limits of the strategy space, Nash equilibria with one mixer are found in a single connected line in $\mathbb{R}^3[0, 1]$.

3.6.1 Replicator Dynamics

Following equations 3.25 and 3.20 , the replicator dynamics of the three player MG can be defined as below, using the potential function $U(x, y, z)$:

$$\dot{x} = x(U(1, y, z) - U(x, y, z)) \tag{3.26}$$

$$\dot{y} = y(U(x, 1, z) - U(x, y, z)) \tag{3.27}$$

$$\dot{z} = z(U(x, y, 1) - U(x, y, z)). \tag{3.28}$$

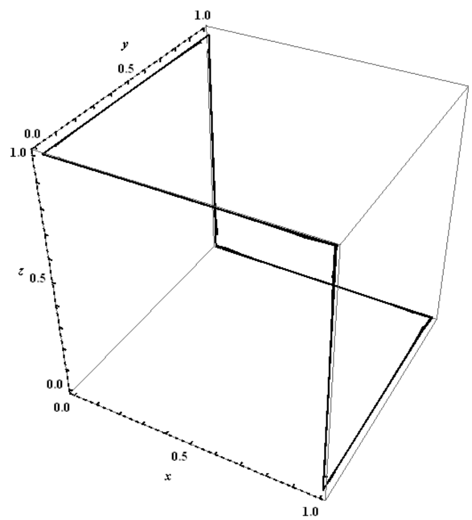


Figure 3.5: The bold line represents the points of the mixed strategy space where the potential function $U(x, y, z)$ is maximized ($U(x, y, z) = 1$). The line is the set of Nash equilibria with one mixer.

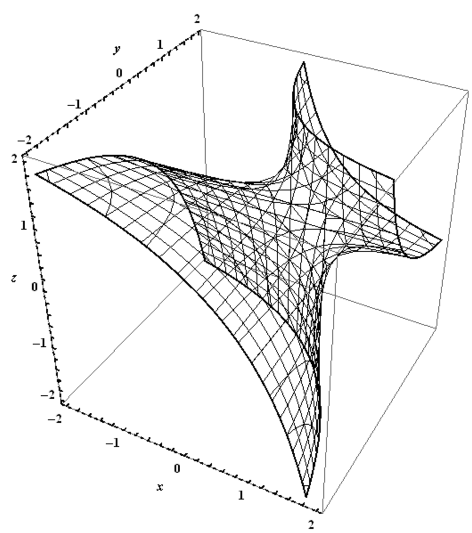


Figure 3.6: The potential function surface $U(x, y, z) = 1$ w.r.t. an extended space $\mathbb{R}^3[-2, 2]$.

From equation 3.18, we have

$$\dot{x} = (1 - y - z)(1 - x)x \tag{3.29}$$

$$\dot{y} = (1 - z - x)(1 - y)y \tag{3.30}$$

$$\dot{z} = (1 - x - y)(1 - z)z. \tag{3.31}$$

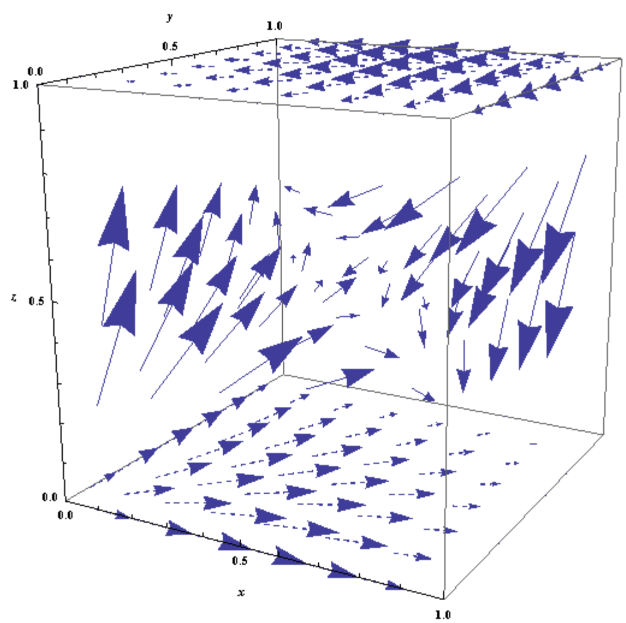


Figure 3.7: Three player MG replicator dynamics, vector field. The field grows until it reaches Nash equilibrium. The planes with $z = 0, 1$ are depicted with dashed arrows.

Notably, non-equilibrium pure strategy profiles are rest points of the replicator dynamics, as well as the unique mixed Nash equilibrium of $x = y = z = 1/2$. None of the aforementioned profiles ,however, are stable, in contrast with the Nash equilibrium combinations of two agents playing opposite pure strategies and one player mixing between them with a probability $l \in [0, 1]$. In Figure 3.7, we have the replicator dynamics vector field of the three player MG derived from equations 3.29 . The potential function is eager to grow inside the strategy space, until it reaches the set of the MG Nash equilibria on the boundary. We remark that when a player uses a pure strategy A_i (e.g. $z = 0$), the replicator dynamics grow in the A_i plane. Figure 3.7 shows the plane for the case of $z = 0, 1$ in dashed arrows. By symmetry, graphs representing the rest of the vector planes are similar with Figure 3.7. Furthermore, to illustrate the coordinative character of the MG, in Figure 3.8 the $(2D)$ vector field is plotted on the plane with $x = y$. Since $x = y$ the solution converges to a NE with pure strategies. Namely, the strategy triples $(0, 0, 1)$ or $(1, 1, 0)$. Moreover, the instability of the mixed NE $(1/2, 1/2, 1/2)$ becomes clear.

In Figure 3.9, we present the trajectories of MG replicator dynamics using various initial conditions. Each strategy triplet is represented as a point in the cubed mixed strategy space. Each trajectory remarks the change of strategies until convergence. The initial strategies of each game stand in the point denoted by the beginning of each curve and the solution profile at each end vertex (ball). Clearly, regardless of the initial conditions, excluding non-equilibrium saddle

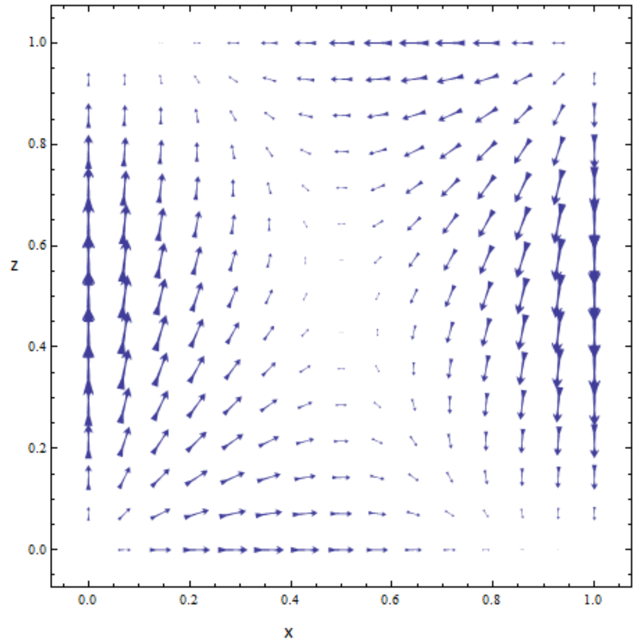


Figure 3.8: Three player MG replicator dynamics vector field in the $x = y$ plane. The field points to pure strategy NE.

points, all plays quickly converge towards the black line highlighting the Nash equilibria of the MG and stabilize.

3.7 Minority Game and Human Behavior

Several experimental studies have been conducted to explore which strategies people use in the MG. Specifically, in the simulation studies discussed, the strategies used by the agents are arbitrary and deterministic. I.e, they do not allow for randomization in most cases. Results from laboratory experiments may help understand which strategies actual humans play in the MG. Studies [12, 13, 76, 91, 14] agree that people always play better than the random case, i.e. better than the symmetric Nash equilibrium. However, the dynamics of these games in the laboratory are not trivial. Particularly in [91], it is shown that oscillations of the fluctuations occur. Moreover, in [76] it was found that the length M of the public information seems to have no influence on the summary statistics and the dynamics of volatility. This fact comes to agreement with the theoretical literature accounting for the irrelevancy of the agents memory in the MG results when they posses the same amount of memory [18]. We remark that when agents posses different memory sizes, the theoretical results

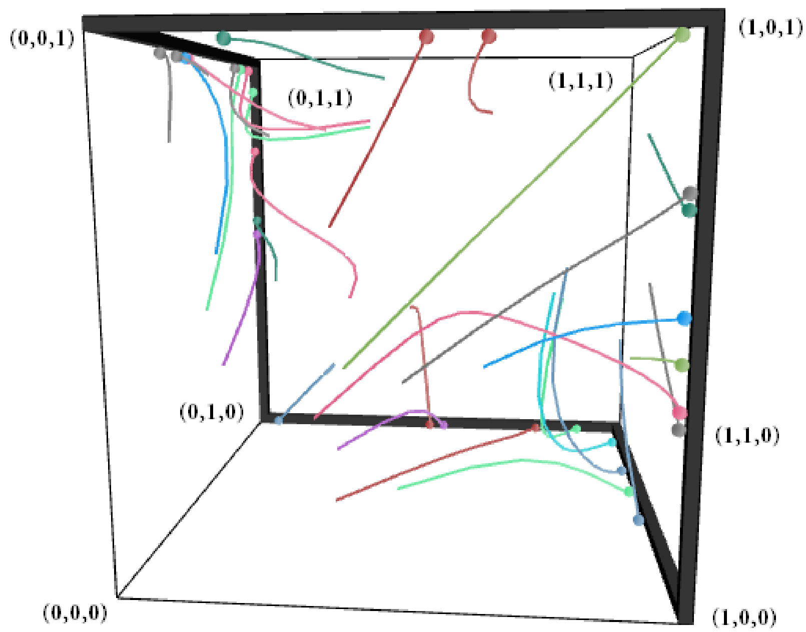


Figure 3.9: Multiple ($J = 30$), three player MG replicator dynamics, with random initial strategies. All games converge to a MG NE (thick black line).

differ from the homogeneous case [6]. Furthermore in [13], it was noted that there was no significant improvement in the individual play, when players where given the information of their opponents play. Moreover according to [91, 14], participants tend to repeat the same action at the end of the experiment when they have enough information and win consistently or they are bored of the game.

In the experiments on the MG discussed above, effort was put to motivate humans to play simultaneously the MG. However, such experiments are notoriously tricky in social sciences [52]. To overcome the laboratory difficulties in [57], one human player was placed against artificial agents through a web interface. As the players progress through the game, a record of all the playing data was kept. The most surprising result is that humans definitely outperform the computer agents when $M \geq 4$. The $M = 3$ barrier marks a transition where human player tend to disregard history and revert to oversimplified strategies. Remarkably, [91] also points out the criticality of $M = 3$, associating it with the well know psychology 7 ± 2 rule, the number of things a human can remember simultaneously. When $M = 3$ the number of possible histories is $2^3 = 8$, thus the connection with the aforementioned psychological rule.

Another difficulty presented in the MG experiments with real players, is the inability to definitely determine if the participants randomize or not. To that extent in [24], a three-player minority game experiment is studied, where participants can explicitly use mixed strategies. Moreover, there is random rematching of groups after each round, in order to highlight the mixed strategy Nash equilibrium as more obvious candidate for individual behavior. The results indicate heterogeneity in decision rules, and approximately a quarter of the participants is best described by the symmetric mixed strategy Nash equilibrium. In the same spirit in [28], a three-player minority game experiment is considered, where each player is represented by a team of three participants. Teams are video monitored and their discussion is analyzed to discover strategies used for playing the MG. The video recordings reveal that teams rarely use a randomization strategy and that they tend to focus more on their own past actions than on opposing teams over time, specifically in the case when they have been successful. It is however unclear whether these strategies are affected by being taken by a group and not by a sole individual.

It is clear that more experiments must be conducted in order to acquire a better understanding of the strategies used in the Minority Game. The web based experiments open new directions, although recording and analyzing players can be proven equally fruitful towards unlocking their strategic choices.

Chapter 4

Experiments

In this chapter simulations are performed for the individual, reinforcement learning and social imitation, represented by multi-population and single-population replicator dynamics, respectively. Simulations were prototyped initially using the well-known agent-based platform NetLogo [103],[82]. Moreover, custom implementations using Java programming language were created to gain speed in the computer experiments and retain cross-platform compatibility [74]. Individual learning comes first as the analytical solution has been presented and the simulations follow as a validation of the theory in Chapter 3. In the case of social learning, a mathematical model was not attempted. Therefore, experiments are provided to investigate the rich phenomena of the social imitation algorithm. Experiments include reporting attendance and volatility over time to measure the efficiency of the system. Moreover, individual imitative agents behavior is investigated, by monitoring their payoffs in the MG. In all experiments one unit of time t is the equivalent of one round (step).

4.1 Multi-Population Learning

A brief algorithmic analysis on the multi-population system reveals that volatility can reach the lower bound of $LB(\sigma^2/N) = 0$. This is the case when the mixer chooses a pure strategy, resulting in a constant minority side. However, we will see through experimentation that this instance cannot be socially optimal. We present below the algorithm of the individual learning model following the pairwise imitation protocol (Algorithm 1).

Input: N odd Populations (Pop_i) each with n agents (ag_{ij}).
Each Pop_i has an initial strategy p_i, i.e the probability of playing action $\{1\}$.
Assign agent (ag_{ij}) a deterministic action by the probability p_i.
A revision rate R and a number of Imitators Im.
A probability change factor $f \in (0, 1]$.

Set $score(ag_{ij}) = 0 \; \forall i, j$
Start time t.
loop
  Pick N random agents, one from each Pop_i.
  Play the MG with prescribed actions and update $score$.
  if $t \mod R = 0$ **then**
    for Each Pop_i **do**
      Pick Im random agents (the Imitators)
      Each Imitator Im_{ij}, chooses a Reference Ref_{ik}
      if $score(Ref_{ik}) > score(Im_{ij})$ **then**
        Im_{ij} imitates Ref_{ik} action with probability
        $x = f \times (score(Ref_{ik}) - score(Im_{ij}))/R$
      end if
      Set $score(ag_{ij}) = 0 \; \forall i, j$
    end for
  end if
  increment t step 1.
end loop

Algorithm 1: The Multi-Population Pairwise Imitation protocol algorithm

All indexes i, j of the algorithm start counting from zero. The first experiment is performed using the following parameters as shown in Table 4.1.

In Figure 4.1, we present the attendance of the system with eleven players w.r.t time.

As we can view in Figure 4.1, the attendance fluctuations constantly decrease until convergence of the algorithm to one, out of the many, Nash Equilibria with one mixer. In the NE state, we have $|Att| = 1$, which appears around step $t = 3500$. Specifically for experiment of Table 4.1, the system converged to the NE with the Population strategies as show in Figure 4.2. Here, p_i denotes the probability of Population i to play action 1 (or $1 - p$ to play action 0).

Parameter	Value
Populations	11
Agents in Pop.	35
Imitators	18
Review Rounds	1
Prob. Change Factor	0.7

Table 4.1: Parameters for the first experiment using Multiple Populations as Players.

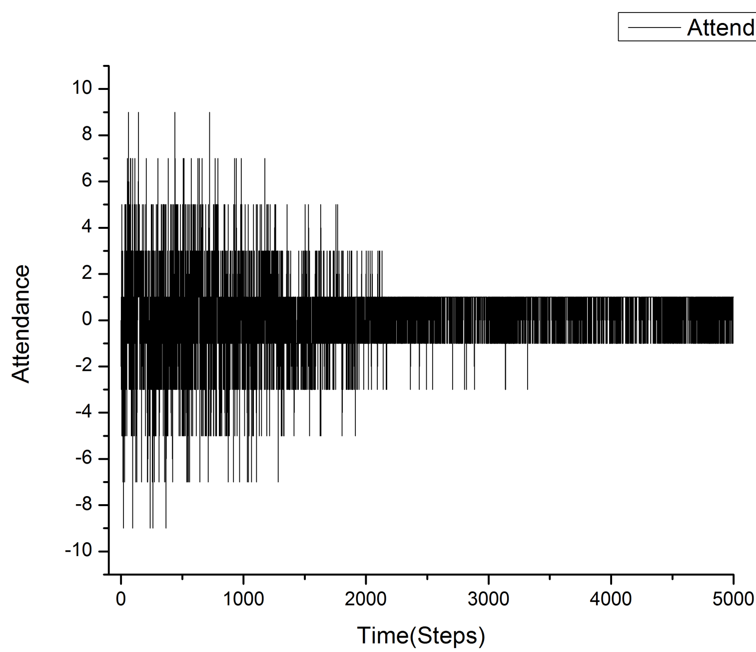


Figure 4.1: Attendance w.r.t time of the Multi-Population MG with parameters of Table 4.1.

Specifically, p_i is calculated as the fraction of agents in population i who play action 1. Therefore, we certify the segregation of $(N-1)/2$ players in each pure strategy and the one mixer (Player 2, counting from zero).

In order to analyze the efficiency of the solution in terms of resource allocation, the volatility of the system is measured w.r.t to time of the experiment of Table 4.1. The volatility is presented in Figure 4.3 where, after some initial fluctuations it drops, converging to the minimum value (dashed red line). Since the actions of the rest of the populations is deterministic in NE, the minimum

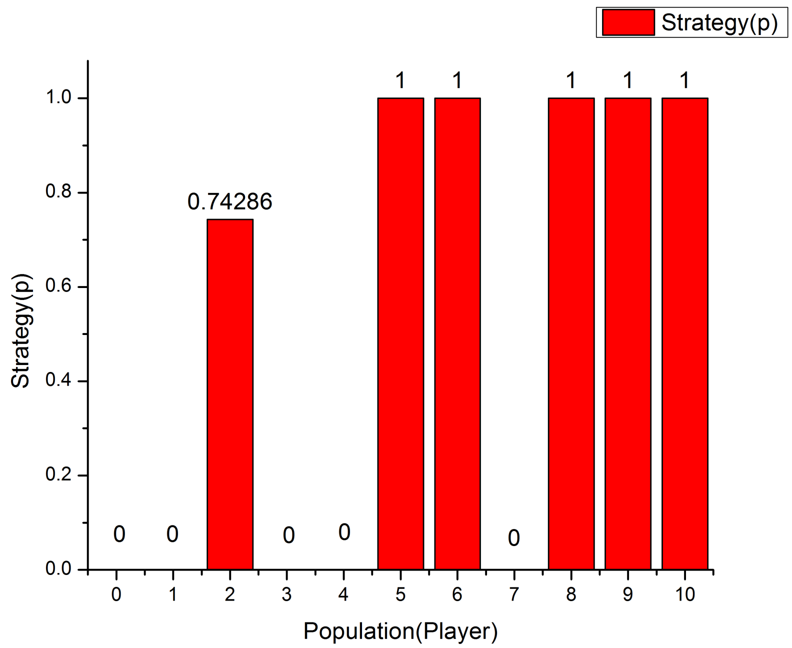


Figure 4.2: Strategies of the MG Players with parameters of Table 4.1 ($t = 3500$).

value of volatility and mean attendance is a function of the mixer's strategy. In the current experiment Population 2 has strategy $p \approx 0.74$, therefore the mean attendance becomes $\langle Att \rangle = -1 \cdot (1 - 0.74) + 0.74 = 0.48$. Consequently, the minimum variance is:

$$\sigma^2 = (0.26 \cdot (-1 - 0.48)^2 + 0.74 \cdot (1 - 0.48)^2) \Rightarrow$$
$$\sigma^2/N \approx 0.07, \text{ with } N = 11.$$

We remark that in the experiments we calculate the volatility over the whole time series, resulting in the slow but steady convergence to the minimum after time step $t = 20000$.

As it has been described through analysis and experiments, the individual learning regime is optimal in terms of attendance levels. Specifically, the maximum number of players gains a point in each turn, however without nourishing individual performance. In Figure 4.4 it is shown the scores of three representative populations over time. The score of each population is measured by averaging over the points of the agents consisting the population. Population(player) 0 gains the highest score compared to the other two and along with him the players that use the same strategy. What is more, the strategy of the unique

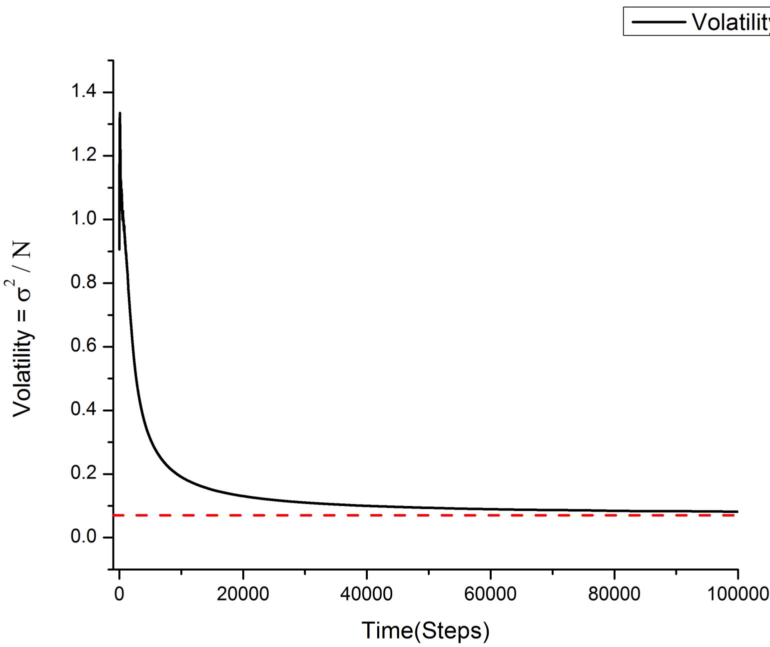


Figure 4.3: Volatility w.r.t time of the Multi-Population MG with parameters of Table 4.1.

mixer is what creates the difference in average payoffs, denoting the minority side by its strategy. In experiment of Table 4.1, the mixer strategy is favoring the populations playing action $\{-1\}$. Furthermore, the mixer population is always on the losing side after the algorithm convergences to NE. The difference in performance between the populations can be viewed in a clear manner in Figure 4.5, where we have a snapshot of the scores at time $t = 150000$. Players 0,1,3,4,7 form a group of equal high scores compared to the rest of the players minus the mixer (Player 2).

Continuing the experiments, the robustness of the system is tested, in terms of learning speed. Therefore, the number of populations is increased, the number of agents in each population is decreased and the probability of change factor is set to 1. Thus, the rule becomes to imitate surely in the case when a better agent is met. Moreover, the number of imitators is set the same as number of agents in a population. Therefore, in each round every agent has a chance to imitate another agent. The parameters of the second experiment are placed in Table 4.2.

The attendance and volatility of the experiment are presented in Figure 4.6. We observe the fast convergence of the algorithm in time $t = 1000$. Thus, since the procedure is robust, it is in the hands of the designer to set the desirable

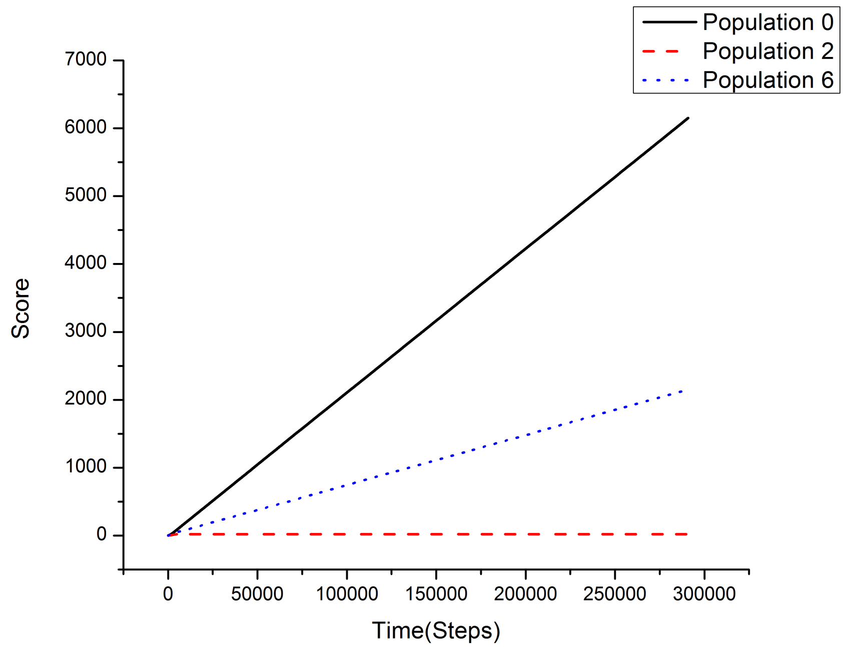


Figure 4.4: Average Scores w.r.t. time of the experiment of Table 4.1.

Parameter	Value
Populations	101
Agents in Pop.	18
Imitators	18
Review Rounds	1
Prob. Change Factor	0.7

Table 4.2: Parameters for the second experiment using Multiple Populations as Players.

learning parameters according to the purpose of the algorithm.

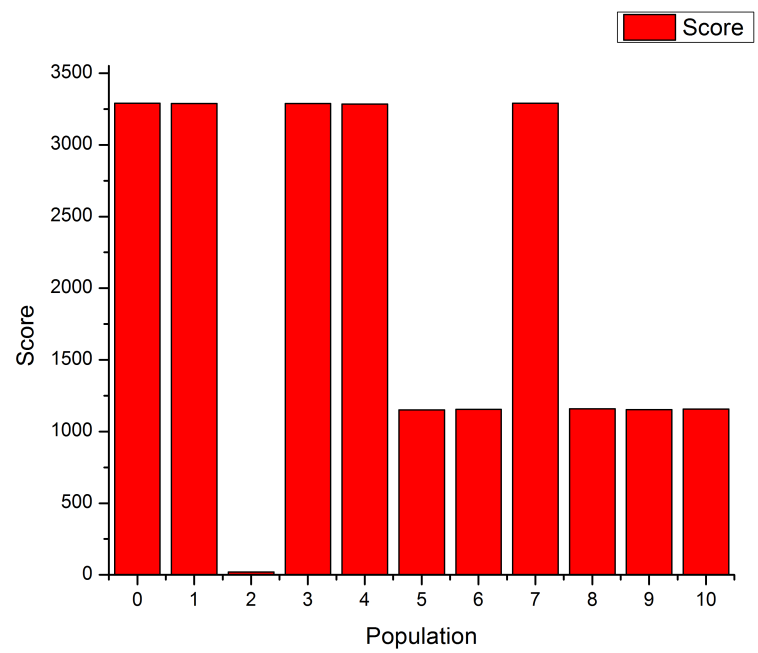


Figure 4.5: Scores in time $t = 150000$ of players of the experiment with parameters of Table 4.1.

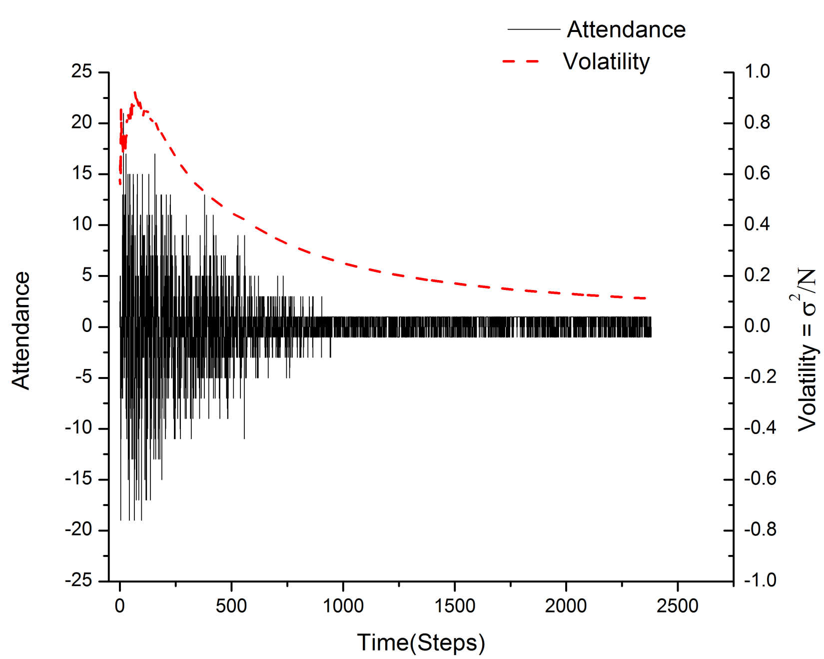


Figure 4.6: Attendance and Volatility w.r.t. time of the experiment of Table 4.2. Mixer has strategy $p = 0.60$.

The mixer in the second experiment converged to strategy $p = 0.60$, i.e only sightly favoring the action 0 as the minority. Thus, the populations following pure strategy should have almost equivalent scores. In Figure 4.7, a snapshot of the player scores is presented in time $t = 2000$, in sorted order. The mixer has the lowest payoff and he will continue to lose in position 0 (red color bar in color mode). The populations choosing action 1 are visible with the light blue color, followed by the players with pure strategy $p = 0$ in the dark blue color and the slightly higher payoffs. We observe that the difference in scores among the populations playing pure strategies is lower than the previous experiment.

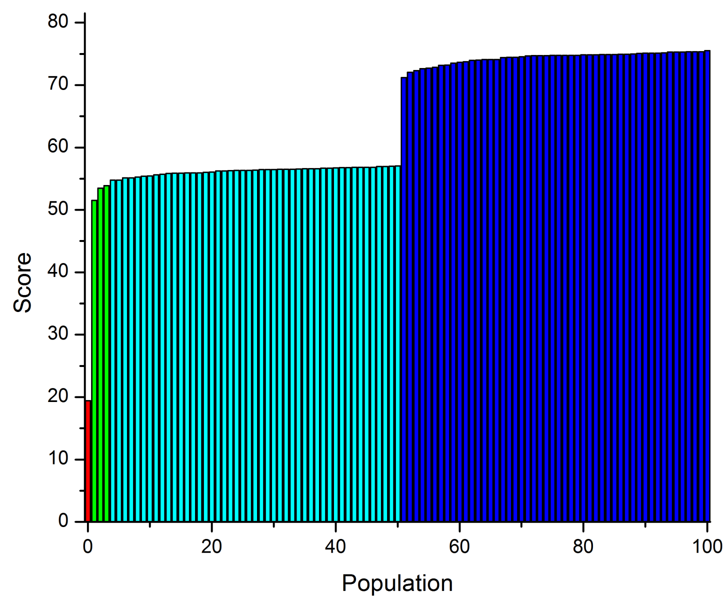


Figure 4.7: Average scores in time $t = 2000$ of players of the experiment with parameters of Table 4.2.

Individual learning, is efficient terms of acquiring maximum sum of utilities from the system. However, individual performance is affected by the strategy of the mixer. Even in the case where the mixer plays symmetrically, he is the one that will always lose when the algorithm has converged. Therefore, multi-population model in MG is a "utilitarian" solution, not keeping payoff equality among players. Continuing, the single-population model is investigated, where we will find similarities and differences with the multi-population model.

4.2 Single-Population Learning

Social imitation can be modeled through a single population of agents playing the MG using a specific mixed strategy. Following the pairwise imitation protocol, agents revise their play when given the opportunity. Since there is no relevant literature we will perform simulations to provide insight of the model's phenomena. The algorithm of social imitation is presented below (Algorithm 2).

Input: N odd Agents (Ag_i).
  Each Ag_i has a strategy p_i, i.e the probability of playing action $\{1\}$.
  Assign agent (Ag_i) a strategy p_i.
  A number of agents that play the MG $NumA \leq N$.
  A revision rate R and a number of Imitators Im.
  A probability change factor $f \in (0, 1]$.

  Set $score(ag_{ij}) = 0 \; \forall i, j$
  Start time t.
  loop
    Pick $NumA$ random agents, one from each Pop_i.
    Play the MG with actions dictated by agent strategies and update *score*.
    if $t \mod R = 0$ **then**
      for Each Pop_i **do**
        Pick Im random agents (the Imitators)
        Each Imitator Im_{ij}, chooses a Reference Ref_{ik}
        if $score(Ref_{ik}) > score(Im_{ij})$ **then**
          Im_{ij} imitates Ref_{ik} strategy with probability
          $x = f \times (score(Ref_{ik}) - score(Im_{ij}))/R$
        end if
        Set $score(ag_{ij}) = 0 \; \forall i, j$
      end for
    end if
    increment t step 1.
  end loop

Algorithm 2: The Multi-Population Pairwise Imitation protocol algorithm

Procedure 2 is very similar to Algorithm 1, however in this case the agents play as individuals. In order to preserve the microfoundations of the model, a parameter controlling how many agents of the available ones are playing the MG [80], is introduced. Moreover, with respect to producing quality simulations it is beneficial to have a population with $N > 51$. Below this number, the algorithm becomes really sensitive to initial distribution of strategies.

Preliminary experiments show that segregation can arise in a population of imitative agents. In this thesis, in order to be consistent with the reviewed literature [90], all the available players are set to play the game ($NumA = N$).

In Table 4.3, the parameters of the fist experiment are presented.

Parameter	Value
Agents	99, 143, 195
Imitators	3
Review Rounds	3
Prob. Change Factor	0.7

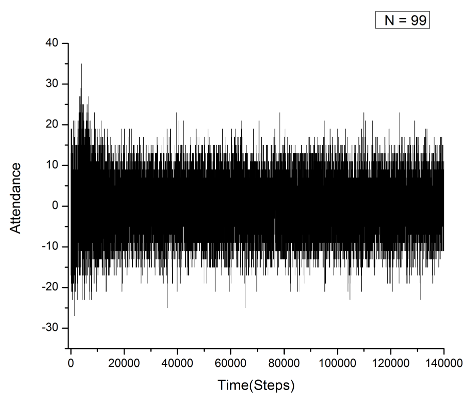
Table 4.3: Parameters for the first experiment using a Single Population of Players.

In Figure 4.8 we show the attendances w.r.t time for number of agents N of Table 4.3 respectively.

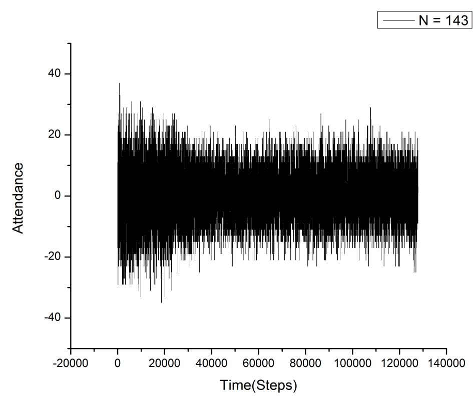
It is evident from Figure 4.8 above that the system does not converge to a NE of the MG game. However, the attendance values fluctuate around the comfort level $Att = 0$ with mean $\langle Att \rangle = 0$, throughout all realizations. Moreover, the fluctuations of attendance do resemble the behavior of the original learning algorithm of MG with histories. A closer look at the attendance graphs, reveals that for all experiment cases of N, there is a significant reduction around step $t = 30000$ of the attendance fluctuations. The decrease of fluctuations, although a noisy process, seems not to depend on the number of agents in the population. Therefore, it is of real interest to study the mixed strategies used by the agents in the stationary state. In Figure 4.9 typical strategies used by the agents are presented, along with the fraction of the population using them respectively for the parameters of Table 4.3, in stationary state. The results of Figure 4.9, are normalized over the total population number. Segregation is visible, however the strategies used are not always the pure ones of $p = 0$ and $p = 1$. This effect is due to the initial distributions of the runs, where the pure strategies can get lost in the early stages of imitation.

As is shown in Figure 4.9 in the $N = 99$ case(green triangle), the population is split in almost equal sized groups when playing the extreme strategies. However, as we see in $N = 143, 195$ cases, when the converged strategies are not the extreme ones, the two respective fractions playing each strategy are not equal. Moreover, as shown by the variance point bars, the strategy distributions are not constant, but rather oscillate with a rate equal to the number of the imitators. Naturally, the next inquiry regards which strategies do appear in the stationary state. In Figure 4.10, the histogram of the strategies played in the stationary state over a five thousand realizations for $N = 99$, are showed. The system preference to converge to the pure strategies is clear, followed by the complete absence of the near symmetric strategies. What is more, experiments show that the results are independent from the number of agents and in each case the surviving strategies are always two. In Figure 4.11, we plot the pairs of strategies in stationary state of the multiple realizations for $N = 99$. The clustering in the extreme strategies region is clear with few exceptions.

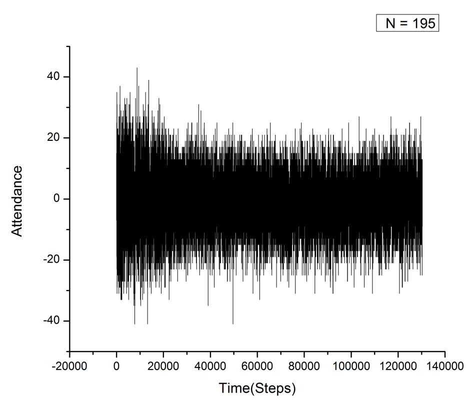
Concerning the efficiency of the system and the maximization of utility in a



(a) $N = 99$



(b) $N = 143$



(c) $N = 195$

Figure 4.8: Attendance w.r.t time for $N = 99, 143, 195$ of Table 4.3 respectively.

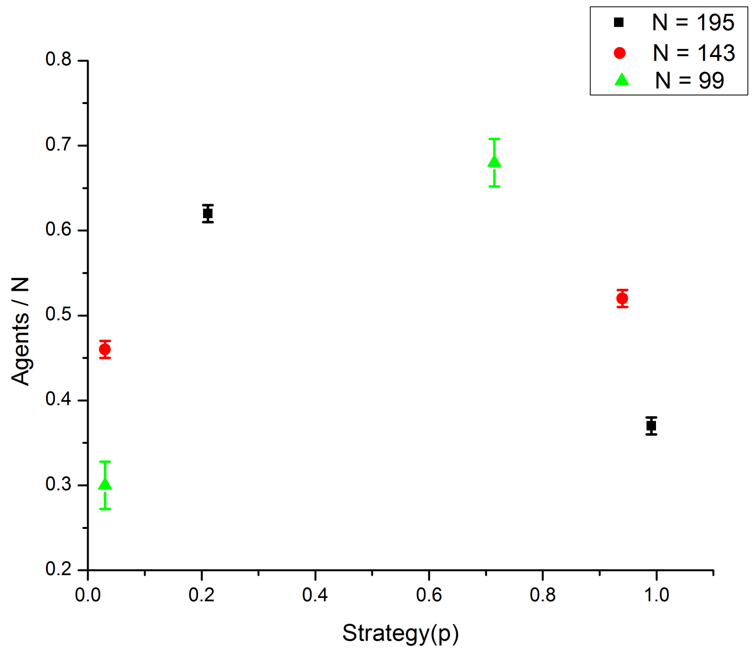


Figure 4.9: Strategy snapshot in stationary state w.r.t agent fraction for population parameters of Table 4.3, $t = 120000$. The point bars represent the variance of the agents in each case.

social setting, the volatility of the system with the parameters of Table 4.3 is plotted, in the stationary state. Experiments show that volatility is a function of the strategies in converged state rather than the number of agents in the population. In Figure 4.12, the normalized variance of all population sizes of Table 4.3 is presented. In this specific case all three experiments converged approximately to the same pair of strategies $(0.1, 0.9)$.

Moreover, all volatilities drop below the random play threshold of value 1, converging to $\sigma^2/N \approx 0.35$. The convergence goes smoothly for $N = 99, 195$, however not for $N = 143$ where the imitation process was not directed immediately to the extreme strategies. Thus, resulting in a higher volatility until convergence to the stationary state. In order to visualize the dependence of volatility with respect to the strategies in the stationary state, we plot Figure 4.13. Volatility in Figure 4.13 is averaged over many realizations for each strategy pair (Strategy 1, Strategy 2) for $N = 99$. We remark the linear decrease of volatility as Strategy 2 approaches the $p = 1$ for values of Strategy 1. This behavior is consistent except for the case when Strategy 1 has value $p = 0.4$, where volatility remains constant for all values of Strategy 2. Moreover as Strategy 1 reaches the value $p = 0.4$ the minimum volatility at Strategy 2 value $p = 1$

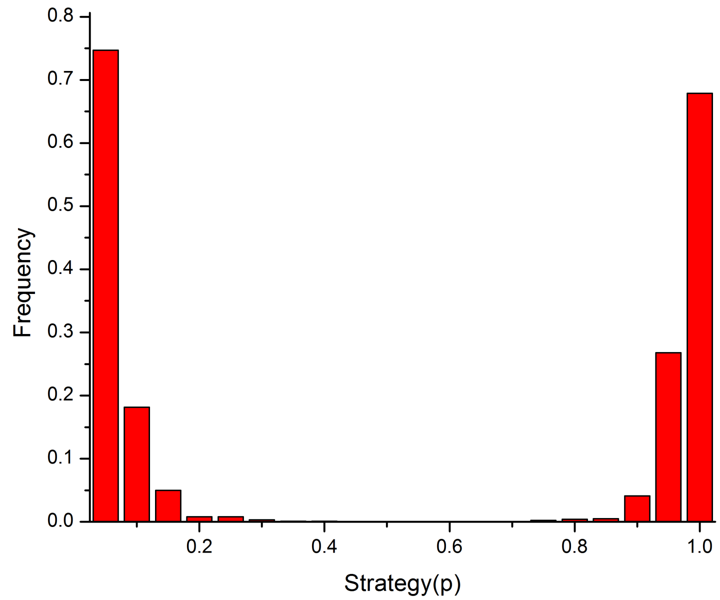


Figure 4.10: Histogram of strategies at stationary state, with parameters of Table 4.3 and $N = 99$.

increases linearly. As expected the minimum value of volatility for the system is found at the strategy pair $(Strategy\,1 = 0, Strategy\,2 = 1)$.

In a broad view, the results concerning efficiency surely favor individual learning, as in that case we can optimally utilize the system resources. Nevertheless, the efficiency of reinforcement learning has a price paid by the inequality of scores among the players. On the contrary, social learning provides a more "egalitarian" solution, where agent payoff do not have great differences. In Figure 4.14, we present the minimum, maximum and average payoff of the agent in a population of $N = 99$ over time. We observe that the minimum and maximum score stay close to the average and grow with the same rate. Moreover in Figure 4.15, typical individual performance of three random agents is plotted w.r.t time. Notably, there are changes in the position of the agents according to the maximum score, i.e there is no specific agent performing better than the rest for the whole duration of the experiment. Intuitively, we understand that the best performance in score will be imitated, thus losing its advantageous position in time.

To provide further insight of the imitation process, we present Table 4.4, consisting of five typical results of our experiments with population of $N = 99$. In the first column we have the index number, second is the pair of strategies at

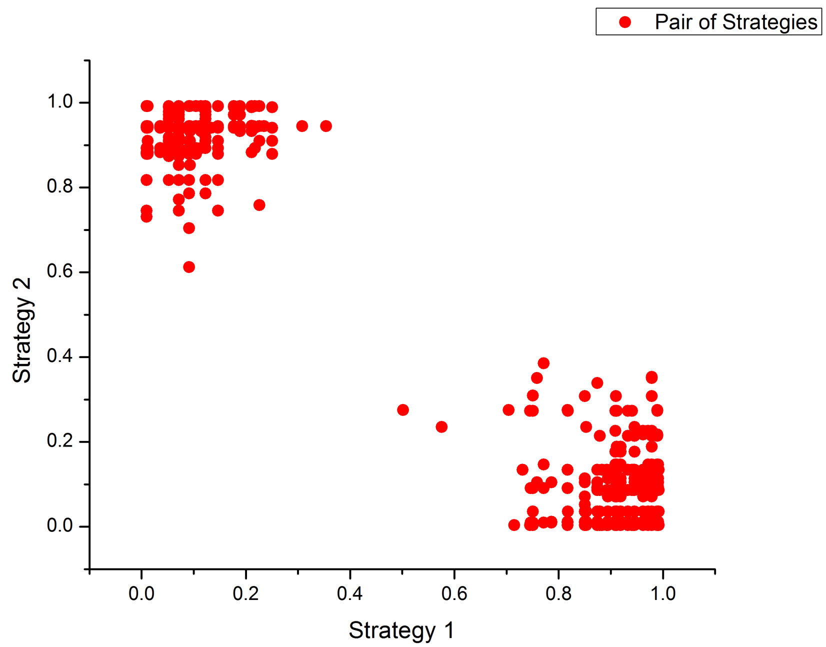


Figure 4.11: Strategy pairs at stationary state, with parameters of Table 4.3 and $N = 99$.

stationary state and in the third column the corresponding population fraction playing them.

Index	Strategies	Population Portion
1	(0.3, 0.85)	(0.62, 0.36)
2	(0.07 , 0.90)	(0.48, 0.50)
3	(0.003, 0.94)	(0.52, 0.47)
4	(0.01, 0.88)	(0.44, 0.54)
5	(0.17, 0.97)	(0.58, 0.41)

Table 4.4: Strategy and average fractions of population playing them, with $N = 99$ and Table 4.3 parameters.

Notably, regardless of the strategy pair, the fractions of the populations playing each strategy are such that expectedly half of the agents will play action $\{1\}$ (and the rest action $\{0\}$). Imitation pushes the agents to be equal and tries to settle to the symmetric NE. What is more, this type of behavior is consistent for all population sizes. Therefore, as stated above, social imitation is a more

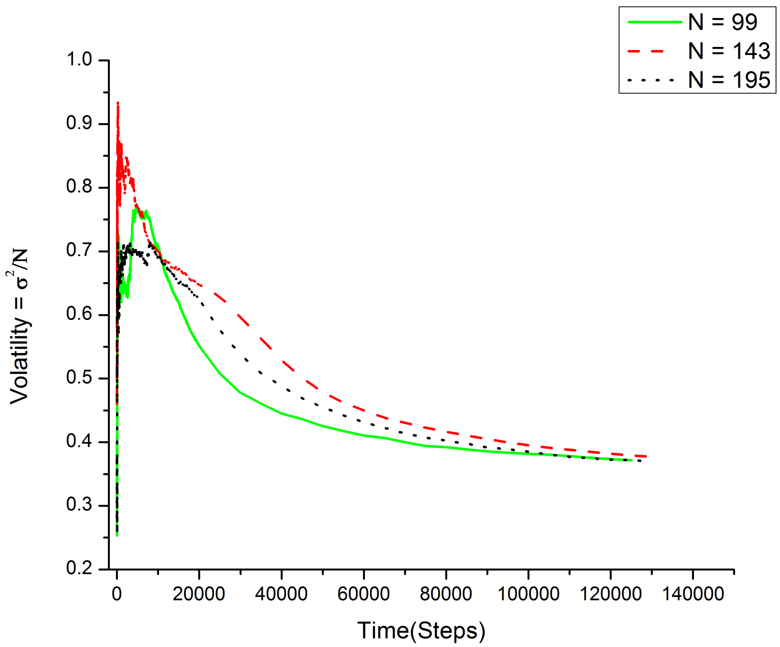


Figure 4.12: Joint volatilities w.r.t time of the experiment with parameters of Table 4.3. $N = 99, 143$ and 195 are represented by a solid, dashed and dotted line respectively.

democratic platform of learning than individual learning. However equality has as a consequence, the lost efficiency in resource allocation. Continuing on studying the robustness of the pure strategies in the MG using the pairwise imitation protocol, preliminary tests are performed using simplistic initial strategy distributions. Specifically in Figure 4.16, it is shown a typical plot of strategies evolution, when the game is played using only three strategies, namely the two pure ones and the symmetric mixed strategy. As we observe in Figure4.16 , the mixed strategy does not survive, which is a consistent behavior. The mixed strategy never survives the two pure ones. However, adding more mixed strategy closer to the extreme ones, there is a likelihood that pure strategies diminish in the evolution process. Specifically in Figure 4.17, $p = 0$ disappears and $p = 0.2$ survives. Intuitively, the preference of a mixed strategy over the pure ones is a function of the number of agents using each strategy. If the initial distribution of players favors the mixed strategy over the extreme one, this can lead to the pure strategy disappearance. Furthermore, adding more initial strategies to the system enhances the chaotic behavior of the convergence of strategies, where no definite conclusions can be drawn. Nevertheless, in order to acquire a clearer view of the convergence phenomena, further experimentation and dedi-

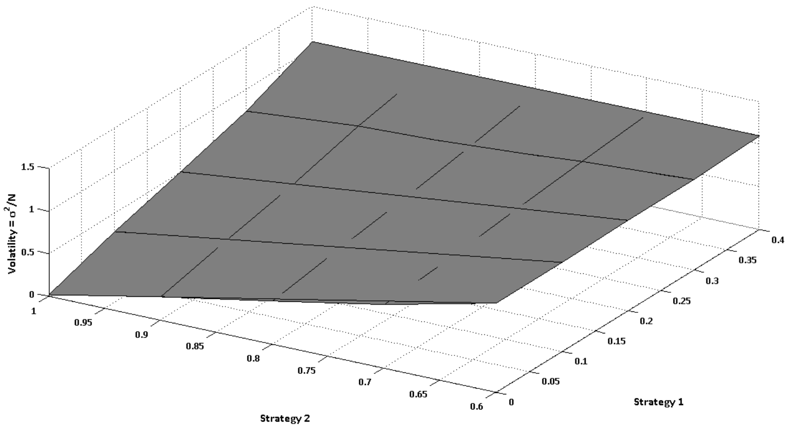


Figure 4.13: Volatility w.r.t pair of strategies (Strategy 1, Strategy 2) in the stationary state for parameters of Table 4.3 ($N = 99$).

cated analysis is required. Specifically, there is further questions regarding the microfoundations of the model, i.e having a population of agents and only a certain number of them play the MG in each turn. Such a setting can provide more intuition on the behavior of the social imitation process.

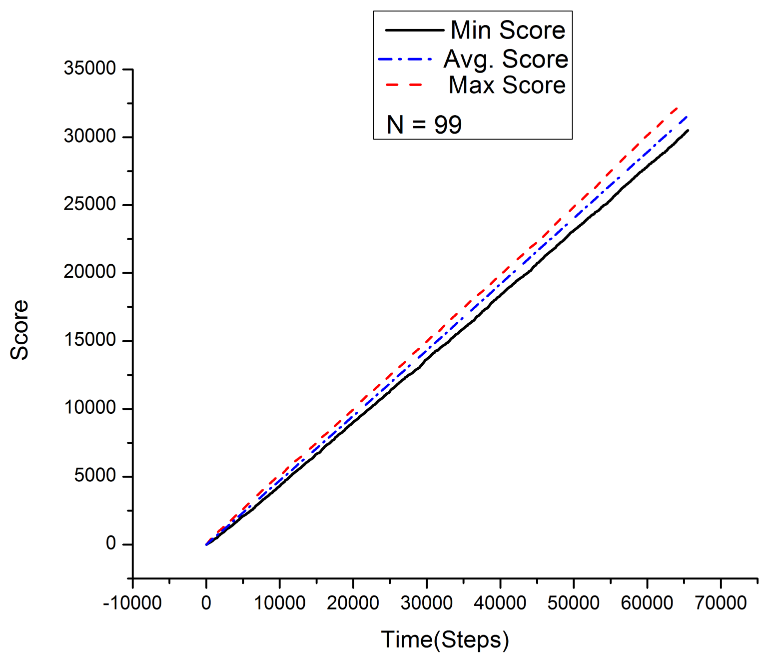


Figure 4.14: Minimum, maximum and average scores w.r.t time of an instance of $N = 99$ with parameters of Table 4.3. Minimum, maximum and average are represented by a black solid, red dashed and blue dot-dashed line respectively.

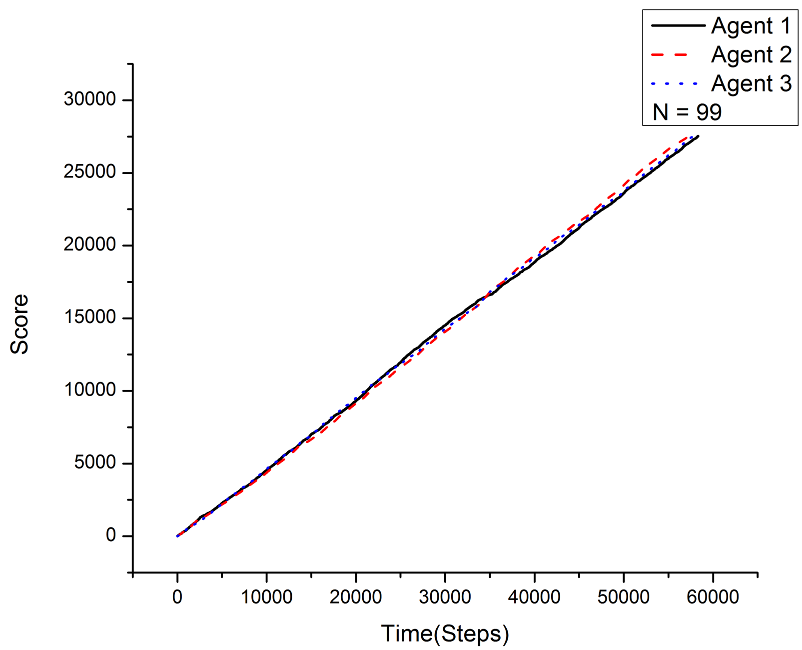


Figure 4.15: Scores of three random agents w.r.t time of an instance of $N = 99$ with parameters of Table 4.3.

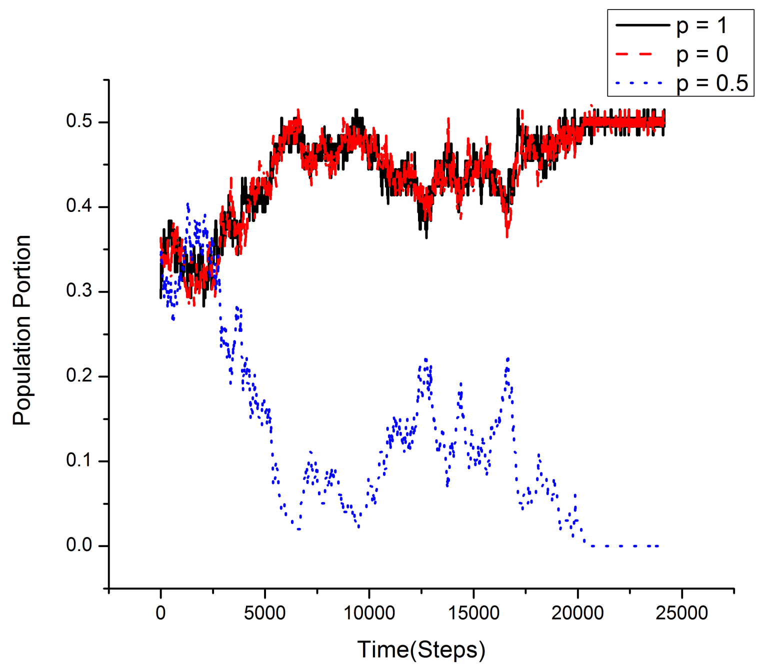


Figure 4.16: Strategies evolution, using initial values $p = 0, 1$ and 0.2.

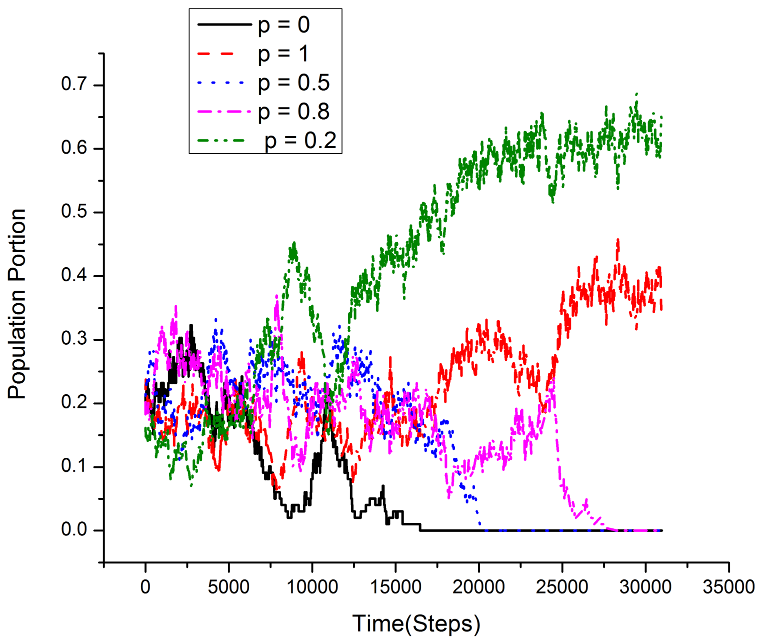


Figure 4.17: Strategies evolution, using initial values $p = 0, 1, 0.5, 0.8$ and 0.2.

Chapter 5

Conclusions

In this thesis individual and social learning was investigated through a simple congestion game with many (odd number) participants, the Minority Game. The Minority Game, originally designed to model bounded rational agents, proved to be a fruitful research platform of evolutionary game theory learning methods. On one hand individual learning, representing the behavioral reinforcement adaptive methods, is modeled through Multi-Population replicator dynamics. In the memoryless MG, reinforcement learning leads to optimality in terms of resource allocation of the game, i.e. minimization of the systems volatility of attendance. Optimality is reached through the asymptotically stable Nash Equilibrium of the replicator dynamics, where player populations split equally between the two pure strategies and one using a mixed strategy. Nevertheless, optimality in attendance levels comes at the price of inequality between agents performance. Social imitation on the other hand, representing social learning in a single population of agents, does provide a more egalitarian environment for the players. Agent-based simulations showed that social imitative agents, playing the MG, keep the mean attendance to zero, much like the original algorithm results. The system does not reach a Nash Equilibrium, but rather continuously fluctuates around the zero attendance level, much like the original MG. Further research should be performed to provide analytical treatment on the imitation process of agents in a single population. Furthermore, simulations using various initial strategy distributions and different number of participants from the whole population, should yield more insight on the evolution of strategies and the convergence properties of social learning in the Minority Game.

Bibliography

[1] *The Cambridge Handbook of the Learning Sciences*. Cambridge University Press, 2006.

[2] Minority-game-based resource allocation for run-time reconfigurable multi-core processors. *Design, Automation & Test in Europe Conference & Exhibition*, 0:1–6, 2011.

[3] John Robert Anderson. *Learning and memory*, volume 86. John Wiley New York, 2000.

[4] M. Andrecut and M. K. Ali. Q learning in the minority game. *Phys. Rev. E*, 64:067103, Nov 2001.

[5] James Andreoni and John H. Miller. Rational cooperation in the finitely repeated prisoner's dilemma: Experimental evidence. *The Economic Journal*, 103(418):pp. 570–585, 1993.

[6] Ricardo M. Araujo and Luis C. Lamb. On the use of memory and resources in minority games. *ACM Trans. Auton. Adapt. Syst.*, 4(2):11:1–11:23, May 2009.

[7] W. Brian Arthur. Inductive Reasoning and Bounded Rationality. *The American Economic Review*, 84(2):406–411, 1994.

[8] RobertJ. Aumann. Rationality and bounded rationality. In Sergiu Hart and Andreu Mas-Colell, editors, *Cooperation: Game-Theoretic Approaches*, volume 155 of *NATO ASI Series*, pages 219–231. Springer Berlin Heidelberg, 1997.

[9] Robert Axelrod and William D Hamilton. The evolution of cooperation. *Science*, 211(4489):1390–1396, 1981.

[10] J.Neil Bearden and Terry Connolly. On optimal satisficing: How simple policies can achieve excellent results. In Tamar Kugler, J.Cole Smith, Terry Connolly, and Young-Jun Son, editors, *Decision Modeling and Behavior in Complex and Uncertain Environments*, volume 21 of *Springer Optimization and Its Applications*, pages 79–97. Springer New York, 2008.

[11] Tilman Borgers and Rajiv Sarin. Learning through reinforcement and replicator dynamics. *Journal of Economic Theory*, 77(1):1 – 14, 1997.

[12] G. Bottazzi, G. Devetag, and G. Dosi. Adaptive learning and emergent coordination in minority games. *Simulation Modelling Practice and Theory*, 10(5–7):321 – 347, 2002. ¡ce:title¿Organisational Processes¡/ce:title¿.

[13] Giulio Bottazzi and Giovanna Devetag. A laboratory experiment on the minority game. *Physica A: Statistical Mechanics and its Applications*, 324(1–2):124 – 132, 2003. ¡ce:title¿Proceedings of the International Econophysics Conference¡/ce:title¿.

[14] Giulio Bottazzi and Giovanna Devetag. Competition and coordination in experimental minority games. *Journal of Evolutionary Economics*, 17(3):241–275, 2007.

[15] R.B. Burns. *The adult learner at work: a comprehensive guide to the context, psychology and methods of learning for the workplace.* Business and Professional Publishing, 1995.

[16] Colin Camerer. *Behavioral game theory: Experiments in strategic interaction.* Princeton University Press, 2003.

[17] David Catteeuw and Bernard Manderick. Heterogeneous populations of learning agents in the minority game. In Peter Vrancx, Matthew Knudson, and Marek Grześ, editors, *Adaptive and Learning Agents*, volume 7113 of *Lecture Notes in Computer Science*, pages 100–113. Springer Berlin Heidelberg, 2012.

[18] Andrea Cavagna. Irrelevance of memory in the minority game. *Physical Review E*, 59(4):R3783–R3786, April 1999.

[19] D. Challet. http://www3.unifr.ch/econophysics/minority/papers.html. webpage. retrieved 16-09-2013.

[20] D. Challet, M. Marsili, and R. Zecchina. Statistical Mechanics of Systems with Heterogeneous Agents: Minority Games. *Physical Review Letters*, 84:1824–1827, February 2000.

[21] D. Challet and Y.-C. Zhang. Emergence of cooperation and organization in an evolutionary game. *Physica A: Statistical Mechanics and its Applications*, 246(3–4):407 – 418, 1997.

[22] Damien Challet, Matteo Marsili, and Yi-Cheng Zhang. *Minority Games: Interacting agents in financial markets.* Oxford University Press, 2004.

[23] Damien Challet and Yi-Cheng Zhang. On the minority game: Analytical and numerical studies. *Physica A: Statistical Mechanics and its Applications*, 256(3):514–532, 1998.

[24] Thorsten Chmura and Werner Guth. The minority of three-game: An experimental and theoretical analysis. *Games*, 2(3):333–354, 2011.

[25] Franco Cicirelli, Angelo Furfaro, Andrea Giordano, and Libero Nigro. Performance of a multi-agent system over a multi-core cluster managed by terracotta. In *Proceedings of the 2011 Symposium on Theory of Modeling & Simulation: DEVS Integrative M&S Symposium*, TMS-DEVS '11, pages 125–133, San Diego, CA, USA, 2011. Society for Computer Simulation International.

[26] Kerstin Dautenhahn and Chrystopher L. Nehaniv, editors. *Imitation in animals and artifacts.* MIT Press, Cambridge, MA, USA, 2002.

[27] A. De Martino and M. Marsili. Replica symmetry breaking in the minority game. *eprint arXiv:cond-mat/0007397*, July 2000.

[28] Giovanna Maria Devetag, Francesca Pancotto, and Thomas Brenner. The minority game unpacked: Coordination and competition in a team-based experiment. Working papers, Dipartimento Scienze Economiche, Universita' di Bologna, 2011.

[29] Gary L Drescher. A mechanism for early piagetian learning. In *AAAI*, pages 290–294, 1987.

[30] Gary L Drescher. *Made-up minds: a constructivist approach to artificial intelligence.* The MIT Press, 1991.

[31] Ido Erev and Alvin E Roth. Predicting how people play games: Reinforcement learning in experimental games with unique, mixed strategy equilibria. *American economic review*, pages 848–881, 1998.

[32] Nick Feltovich. Reinforcement-Based vs. Belief-Based Learning Models in Experimental Asymmetric-Information. *Econometrica*, 68(3):605–642, May 2000.

[33] Fernando F Ferreira, Gerson Francisco, Birajara S Machado, and Paulsamy Muruganandam. Time series analysis for minority game simulations of financial markets. *Physica A: Statistical Mechanics and its Applications*, 321(3–4):619 – 632, 2003.

[34] D. Fudenberg and D.K. Levine. *The Theory of Learning in Games.* Economics Learning and Social Evolution Series. MIT Press, 1998.

[35] D. Fudenberg and J. Tirole. *Game Theory.* MIT Press, 1991.

[36] Jay R. Galbraith. Organization design: An information processing view. *Interfaces*, 4(3):28–36, 1974.

[37] Syed Md. Galib and Irene Moser. Road traffic optimisation using an evolutionary game. In *Proceedings of the 13th annual conference companion on Genetic and evolutionary computation*, GECCO '11, pages 519–526, New York, NY, USA, 2011. ACM.

[38] Aram Galstyan, Shashikiran Kolar, and Kristina Lerman. Resource allocation games with changing resource capacities. In *Proceedings of the second international joint conference on Autonomous agents and multiagent systems*, AAMAS '03, pages 145–152, New York, NY, USA, 2003. ACM.

[39] G. Gigerenzer and R. Selten. *Bounded Rationality: The Adaptive Toolbox*. Dahlem Konferenzen. MIT Press, 2002.

[40] Gerd Gigerenzer. How to Make Cognitive Illusions Disappear: Beyond Heuristics and Biases. *European Review of Social Psychology*, 2:83–115, 1991.

[41] Gerd Gigerenzer and Wolfgang Gaissmaier. Heuristic decision making. *Annual review of psychology*, 62:451–482, 2011.

[42] Gerd Gigerenzer, Ralph Hertwig, and Thorsten Pachur. *Heuristics: The Foundations of Adaptive Behavior*. Oxford University Press, Inc., New York, NY, USA, 1st edition, 2011.

[43] Gerd Gigerenzer and Peter M. Todd. *Simple heuristics that make us smart*. Evolution and cognition. Oxford Univ. Press, New York, NY [u.a.], 1999.

[44] Luca Grilli, MassimoAlfonso Russo, and Angelo Sfrecola. Financial time series and neural networks in a minority game context. In Marco Corazza and Claudio Pizzi, editors, *Mathematical and Statistical Methods for Actuarial Sciences and Finance*, pages 153–162. Springer Milan, 2010.

[45] Luca Grilli and Angelo Sfrecola. A neural networks approach to minority game. *Neural Computing and Applications*, 18(2):109–113, 2009.

[46] Peter E Hart, Nils J Nilsson, and Bertram Raphael. A formal basis for the heuristic determination of minimum cost paths. *Systems Science and Cybernetics, IEEE Transactions on*, 4(2):100–107, 1968.

[47] E.R. Hilgard and G.H. Bower. *Theories of learning*. Number v. 644; v. 1966 in Theories of Learning. Appleton-Century-Crofts, 1966.

[48] Shahar Hod and Ehud Nakar. Self-segregation versus clustering in the evolutionary minority game. *Phys. Rev. Lett.*, 88:238702, May 2002.

[49] Josef Hofbauer and Karl Sigmund. *Evolutionary Games and Population Dynamics*. Cambridge University Press, Cambridge, 1998.

[50] N. F. Johnson, P. M. Hui, R. Jonson, and T. S. Lo. Self-organized segregation within an evolving population. *Phys. Rev. Lett.*, 82:3360–3363, Apr 1999.

[51] L. P. Kælbling, M. L. Littman, and A. W. Moore. Reinforcement Learning: A Survey. *eprint arXiv:cs/9605103*, April 1996.

[52] John H. Kagel and Alvin E. Roth, editors. *The Handbook of Experimental Economics*. Princeton University Press, Princeton, N.J., 1995.

[53] W. Kets and M. Voorneveld. *Congestion, Equilibrium and Learning: the Minority Game*. Discussion paper. Tilburg University, 2007.

[54] Willemien Kets. Learning with fixed rules: The minority game. *Journal of Economic Surveys*, 26(5):865–878, 2012.

[55] W. Kinzel, R. Metzler, and I. Kanter. Dynamics of Interacting Neural Networks. *eprint arXiv:cond-mat/9906058*, June 1999.

[56] H.W. Kuhn, J. von Neumann, O. Morgenstern, and A. Rubinstein. *Theory of Games and Economic Behavior (60th Anniversary Commemorative Edition)*. Princeton Classic Editions. Princeton University Press, 2007.

[57] P. Laureti, P. Ruch, J. Wakeling, and Y. C. Zhang. The Interactive Minority Game: a Web-based investigation of human market interactions. *Physica A*, 331:651–659, 2004.

[58] G.R. Lefrançois. *Psychological Theories and Human Learning*. Core books in psychology series. Brooks/Cole, 1982.

[59] Y. Li, A. VanDeemen, and R. Savit. The minority game with variable payoffs. *Physica A Statistical Mechanics and its Applications*, 284:461–477, September 2000.

[60] Petri Mahonen and Marina Petrova. Minority game for cognitive radios: Cooperating without cooperation. *Physical Communication*, 1(2):94 – 102, 2008.

[61] Radu Manuca, Yi Li, Rick Riolo, and Robert Savit. The structure of adaptive competition in minority games. *Physica A: Statistical Mechanics and its Applications*, 282(3–4):559 – 608, 2000.

[62] M. Marsili, D. Challet, and R. Zecchina. Exact solution of a modified El Farol's bar problem: Efficiency and the role of market impact. *Physica A Statistical Mechanics and its Applications*, 280:522–553, June 2000.

[63] Stephen Marsland. *Machine learning: an algorithmic perspective*. CRC Press, 2011.

[64] Panayotis Mertikopoulos and Aris L. Moustakas. The simplex game: can selfish users learn to operate efficiently in wireless networks? In *Proceedings of the 2nd international conference on Performance evaluation methodologies and tools*, ValueTools '07, pages 7:1–7:10, ICST, Brussels, Belgium, Belgium, 2007. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering).

[65] John Mingers. An empirical comparison of pruning methods for decision tree induction. *Machine Learning*, 4(2):227–243, 1989.

[66] M. Minsky. *The Emotion Machine: Commonsense Thinking, Artificial Intelligence, and the Future of the Human Mind*. Simon & Schuster, 2007.

[67] Dov Monderer and Lloyd S Shapley. Potential games. *Games and economic behavior*, 14(1):124–143, 1996.

[68] E. Moro. The Minority Game: an introductory guide. *eprint arXiv:cond-mat/0402651*, February 2004.

[69] R.B. Myerson. *Game Theory: Analysis of Conflict*. Harvard University Press, 2013.

[70] Roger B. Myerson. Nash equilibrium and the history of economic theory. *Journal of Economic Literature*, 37(3):pp. 1067–1082, 1999.

[71] Jr. John F. Nash. Equilibrium points in n-person games. *Proceedings of the National Academy of Sciences of the United States of America*, 36(1):48–49, January 1950.

[72] Noam Nisan, Tim Roughgarden, Eva Tardos, and Vijay V. Vazirani. *Algorithmic Game Theory*. Cambridge University Press, New York, NY, USA, 2007.

[73] Ann Nowe, Peter Vrancx, and Yann-Michael Hauwere. Game theory and multi-agent reinforcement learning. In Marco Wiering and Martijn Otterlo, editors, *Reinforcement Learning*, volume 12 of *Adaptation, Learning, and Optimization*, pages 441–470. Springer Berlin Heidelberg, 2012.

[74] Oracle. Java programming language. http://www.oracle.com/technetwork/java/index.html#943.

[75] J.E. Ormrod. *Human Learning*. Pearson College Division, 2011.

[76] Tadeusz Patkowski and Ramsza. Playing minority game. *Physica A: Statistical Mechanics and its Applications*, 323(0):726 – 734, 2003.

[77] J. Pearl. *Heuristics: Intelligent Search Strategies for Computer Problem Solving*. The Addison-Wesley Series in Artificial Intelligence. Addison-Wesley, 1984.

[78] Jean Piaget. *The moral judgement of the child*. SimonandSchuster. com, 1997.

[79] Jean Piaget and Margaret Trans Cook. The origins of intelligence in children. 1952.

[80] Giovanni Ponti. Continuous-time evolutionary dynamics: theory and practice. *Research in Economics*, 54(2):187–214, 2000.

[81] Michael J Prince and Richard M Felder. Inductive teaching and learning methods: Definitions, comparisons, and research bases. *Journal of Engineering Education*, 95(2):123–138, 2006.

[82] Steven F Railsback, Steven L Lytinen, and Stephen K Jackson. Agent-based simulation platforms: Review and development recommendations. *Simulation*, 82(9):609–623, 2006.

[83] C.M. Reigeluth. *Instructional Design Theories and Models: An Overview of Their Current Status*. Taylor & Francis, 2013.

[84] RobertW. Rosenthal. A class of games possessing pure-strategy nash equilibria. *International Journal of Game Theory*, 2(1):65–67, 1973.

[85] Denise M. Rousseau. Schema, promise and mutuality: The building blocks of the psychological contract. *Journal of Occupational and Organizational Psychology*, 74(4):511–541, 2001.

[86] Stuart Russell and Peter Norvig. *Artificial Intelligence: A Modern Approach*. Prentice Hall Press, Upper Saddle River, NJ, USA, 3rd edition, 2009.

[87] Stuart J. Russell. Rationality and intelligence. *Artificial Intelligence*, 94(1–2):57 – 77, 1997. ¡ce:title¿Economic Principles of Multi-Agent Systems¡/ce:title¿.

[88] W.H. Sandholm. *Population Games and Evolutionary Dynamics*. Economic Learning and Social Evolution. MIT Press, 2010.

[89] William H. Sandholm. Potential games with continuous player sets. *Journal of Economic Theory*, 97(1):81 – 108, 2001.

[90] William H Sandholm. Stochastic evolutionary game dynamics: Foundations, deterministic approximation, and equilibrium selection. *American Mathematical Society*, 201(1):1–1, 2010.

[91] Robert Savit, Katia Koelle, Wendy Treynor, and Richard Gonzalez. Man and superman: Human limitations, innovation, and emergence in resource competition. In Kagan Tumer and David Wolpert, editors, *Collectives and the Design of Complex Systems*, pages 199–211. Springer New York, 2004.

[92] Robert Savit, Radu Manuca, and Rick Riolo. Adaptive competition, market efficiency, and phase transitions. *Physical Review Letters*, 82(10):2203, 1999.

[93] Yoav Shoham and Kevin Leyton-Brown. *Multiagent systems: Algorithmic, game-theoretic, and logical foundations*. Cambridge University Press, 2009.

[94] H.A. Simon. A behavioral model of rational choice. *Quarterly Journal of economics*, 69:99–118, 1954.

[95] Richard S. Sutton and Andrew G. Barto. *Reinforcement Learning: An Introduction*. MIT Press, 1998.

[96] M. Sysi-Aho, A. Chakraborti, and K. Kaski. Intelligent minority game with genetic crossover strategies. *European Physical Journal B*, 34:373–377, August 2003.

[97] Olivier Tercieux and Mark Voorneveld. The cutting power of preparation. *Mathematical Methods of Operations Research*, 71(1):85–101, 2010.

[98] Carsten Ullrich. Descriptive and prescriptive learning theories. In *Pedagogically Founded Courseware Generation for Web-Based Learning*, volume 5260 of *Lecture Notes in Computer Science*, pages 37–42. Springer Berlin Heidelberg, 2008.

[99] David Vernon and Dermot Furlong. Philosophical foundations of ai. In Max Lungarella, Fumiya Iida, Josh Bongard, and Rolf Pfeifer, editors, *50 Years of Artificial Intelligence*, volume 4850 of *Lecture Notes in Computer Science*, pages 53–62. Springer Berlin Heidelberg, 2007.

[100] Joseph Wakeling and Per Bak. Intelligent systems in the context of surrounding environment. *Phys. Rev. E*, 64:051920, Oct 2001.

[101] J.W. Weibull. *Evolutionary Game Theory*. MIT Press, 1997.

[102] Duncan Whitehead. The El Farol Bar Problem Revisited: Reinforcement Learning in a Potential Game. ESE Discussion Papers 186, Edinburgh School of Economics, University of Edinburgh, September 2008.

[103] U. Wilensky. Netlogo. http://ccl.northwestern.edu/netlogo/. center for connected learning and computer-based modeling, northwestern university, evanston, il., 1999.

[104] Chun Zhang, Wei Wu, Hantao Huang, and Hao Yu. Fair energy resource allocation by minority game algorithm for smart buildings. In *Proceedings of the Conference on Design, Automation and Test in Europe*, DATE '12, pages 63–68, San Jose, CA, USA, 2012. EDA Consortium.