UNIVERSITY OF TRENTO

CIFREM

INTERDEPARTMENTAL CENTRE FOR RESEARCH TRAINING
IN ECONOMICS AND MANAGEMENT

DOCTORAL SCHOOL IN ECONOMICS AND MANAGEMENT

INTERACTIVE LEARNING AND GENERALIZATION IN REPEATED GAMES: THEORIES, MODELS, AND EXPERIMENTS

A DISSERTATION

SUBMITTED TO THE DOCTORAL SCHOOL OF ECONOMICS AND MANAGEMENT

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DOCTORAL DEGREE

(PH.D.)

IN ECONOMICS AND MANAGEMENT

Davide Marchiori

January 2010

THESIS SUPERVISORS

Professor Massimo Warglien
Department of Business Economics and Management
& Advanced School of Economics
Ca' Foscari University of Venice

Professor Alessandro Rossi
Department of Computer and Management Science
University of Trento, Italy

REFEREE COMMITTEE

Professor Giovanna Devetag
Department of Law and Management
University of Perugia, Italy

Professor Ido Erev
William Davidson Faculty of Industrial Engineering and Management
Technion - Israel Institute of Technology, Israel

Acknowledgements

Thanks to Judith Avrahami, Ido Erev, and Thorsten Chmura for having kindly provided me with their experimental datasets.

I am particularly grateful to Massimo Warglien, who supervised my work and funded my experiments. I am also indebted to Marco LiCalzi, Paolo Pellizzari, Paola Manzini, Marco Mariotti, Scott Page, John Miller, Shu-Heng Cheng, Werner Güth, and Armin Falk for their insightful and helpful suggestions and comments.

The organizational support of the CEEL laboratory staff (in particular of Marco Tecilla) and of its director Professor Luigi Mittone is also acknowledged.

Finally, special thanks go to my wife, to whom this work is dedicated, for her invaluable support, encouragement, and patience.

CONTENTS

1. INTRODUCTION 1

1.1 OVERVIEW OF THE THESIS 2

1.1.1 Part One (Chapter 2) 2

1.1.2 Part Two (Chapter 3) 3

1.1.3 Part Three (Chapter 4) 4

1.2 ON LEARNING IN REPEATED, COMPLETELY MIXED GAMES 5

1.2.1 Learning: Empirical Findings 7

1.3 QUANTITATIVE MODELS OF LEARNING 9

1.3.1 Reinforcement Learning Models 9

1.3.2 Beliefs Learning Models 13

1.4 REGRET AND CHOICE BEHAVIOR 16

1.4.1 Psychology of Regret 17

1.4.2 Regret and Decision-Making Modeling 20

1.5 BEST RESPONSE AND BEHAVIORAL MODELS OF EQUILIBRIUM 26

1.6 SIMILARITY, CATEGORIZATION, AND GENERALIZATION 30

1.7 MODELING CATEGORIZATION AND GENERALIZATION WITH NEURAL NETWORKS 35

1.8 METHODOLOGICAL APPENDIX 40

2. PREDICTING HUMAN BEHAVIOR BY REGRET-DRIVEN NEURAL NETWORKS 47

2.1 MODELS OF LEARNING 48

2.2 THE PB MODEL 48

2.3 METHODS 54

2.4 THE DATA 56

2.5 SIMULATION RESULTS: ACTUAL PAYOFFS 59

2.6 SIMULATION RESULTS: RESCALED PAYOFFS 64

2.7 CONCLUSIONS 67

2.8 APPENDIX A. SUPPORTING MATERIAL 70

2.9 APPENDIX B. THE DATASET DESCRIPTION 82

2.9.1 Suppes and Atkinson (1960) 82

2.9.2 Malcolm and Lieberman (1965) 82

2.9.3 O'Neill (1987) *83*
2.9.4 Rapoport and Boebel (1992) *83*
2.9.5 Ochs (1995) *84*
2.9.6 Rosenthal, Shachat, and Walker (2003) *85*
2.9.7 Avrahami, Güth and Kareev (2005) *86*
2.9.8 Erev, Roth, Slonim and Barron (2007) *88*
2.9.9 Selten and Chmura (2008) *90*

3. NET REWARD ATTRACTIONS EQUILIBRIUM FOR STRATEGIC FORM GAMES AND ITS EXPERIMENTAL TEST 97

3.1 INTRODUCTION 98
3.2 THE NRA EQUILIBRIUM 99
3.2.1 Theoretical Framework *100*
3.2.2 Parametric NRA *106*
3.2.3 Convergence to NRA Equilibrium *107*
3.3 RELATED WORK 109
3.4 MODEL COMPARISON METHODOLOGY 111
3.5 THE DATA 113
3.6 RESULTS 115
3.6.1 First 50 Trials *116*
3.6.2 Last 50 Trials *117*
3.6.3 All Trials *119*
3.7 SUMMARY AND CONCLUSIONS 120

4. LEARNING IN MULTI-GAME EXPERIMENTS 133

4.1 ECONOMIC MODELS OF GENERALIZATION AND EMPIRICAL EVIDENCE 134
4.2 THE GENERALIZING PB MODEL 136
4.3 EXPERIMENTAL DESIGN 137
4.4 EXPERIMENTAL RESULTS 139
4.5 METHODS 142
4.6 SIMULATION RESULTS 143
4.7 WHAT DO SUBJECTS LEARN? 147
4.8 CONCLUSIONS AND FURTHER RESEARCH 152
4.9 APPENDIX A. SUPPORTING MATERIALS AND TABLES 155
4.10 APPENDIX B. EXPERIMENTAL INSTRUCTIONS 157

5. OVERALL CONCLUSIONS AND FURTHER RESEARCH 159

5.1 PART ONE (CHAPTER 2) 160

5.2 PART TWO (CHAPTER 3) 161

5.3 PART THREE (CHAPTER 4) 164

5.4 FURTHER RESEARCH 165

6. BIBLIOGRAPHY 167

CHAPTER 1

1. INTRODUCTION

The topic of my Ph.D. thesis spans different disciplines, as it is in the fields of behavioral game theory, experimental economics, and agent-based modeling. Specifically, my research addresses issues of *learning* in repeated games and *generalization* i.e., how human beings generalize and apply their acquired strategic skills to new strategic situations, and it heavily relies on and makes use of the tools of computational social sciences.

This work provides further evidence that insights from psychology and neuroscience can be successfully used to design agent-based models that help improve understanding of human decision-making processes and that these models can far outperform (neoclassical) standard economic theory in describing and predicting human choice behavior.

The aim of my Ph.D. thesis is to advance understanding of human choice behavior in repeated strategic interactions. This is potentially important, since it would help explain empirical phenomena that cannot be accounted for by standard economic theory, such as overbidding in auctions and overtrading in financial markets (Selten, Abbink, and Cox, 2005). A further confirmation of the relevance of this topic comes from Erev and Haruvy (2005:359): "[it] is our conviction that some of the most promising directions for learning research lie in the investigation of "small" repeated decisions that are made with little information and little deliberation", and "Though small decisions are of small consequence to the individual making them, they are potentially of tremendous importance to firms and society". In economics, interactive strategic situations are commonly modeled as games, in which the gain (or *payoff*) of an agent (or *player*) depends upon its own choice and the choices of the other players. All throughout my thesis I devote my attention to a particular class of games i.e., the class of two-person 2x2, completely mixed[1] games. This choice is coherent with an established paradigm of analysis in the behavioral and experimental economics fields, and is not only necessary to disentangle the effects of reciprocation and adaptation

[1] With *completely mixed games* I mean here games with a unique equilibrium in mixed strategies (MSE). In the remainder, I will use interchangeably these two terms.

processes (Erev and Roth, 1998), but also particularly interesting for reasons explained below.

This project is well divided into three, recognizably distinct parts (addressed in Chapters 2-4, respectively) that constitute a wider, unitary, and coherent research project on how past experience affects current behavior in interactive decision tasks, as well as the formal modeling of this behavioral process. Section 1 offers an overview of the main argument by providing a brief summary of each part. In Section 2, I illustrate the motivations for which the topic of learning in repeated games is important and, specifically, why repeated games with a unique equilibrium in mixed strategies are noteworthy. Section 3 reviews the most important models of learning proposed in the behavioral game theory and experimental economics literature. Section 4 provides a background for the role of regret in models and theories of choice behavior. Section 5 illustrates some of the most popular concepts of equilibrium, alternative to the theory proposed by Nash (1950) and based on very different assumptions. These stationary concepts can be grouped into two main classes: best-response and behavioral models. Section 6 introduces the concepts of similarity, categorization, and generalization, reviewing some of the most important contributions on these topics in the field of cognitive psychology. Section 7 provides a short introduction to neural networks and their important properties as models of information categorization and generalization. Finally, a section on methodological issues related to model comparison and selection criteria concludes.

1.1 Overview of the Thesis

1.1.1 Part One (Chapter 2)

This part of my thesis deals with *interactive learning* in repeated decision tasks. In a paper coauthored with Professor Massimo Warglien (Marchiori and Warglien, 2008), I propose a new model of learning, the Perceptron-based (PB) model, which embeds the basic principles of Learning Direction Theory (Selten and Stoecker, 1986) and translates them into a neural network model.

The basic assumption of the PB model is that learning is driven by an *ex-post* rationalizing process: individuals modify their behavior by looking backward to what might have been their best moves, once they know others' moves; then, they adjust in the direction of such *ex-post* best response, and it is assumed that the intensity of such

directional change is proportional to a measure of regret i.e., how much they have missed by not making this move. This is coherent with recent neuroscience research on individual decision making, according to which regret affects learning, and both neuro-physiological and behavioral responses to the experience of regret are correlated to its magnitude (Coricelli et al., 2005 and Daw et al., 2006).

Further extending and improving the methodology adopted by Marchiori and Warglien (2008), I test the PB model on a set of 35 different datasets drawn from different experiments on games with unique equilibria in mixed strategies in which the participants received a complete description of the payoff matrix and of their opponents' choices. In addition, I compare the performance of the PB model with those of other six popular models of learning in the behavioral game theory literature. As a result, the PB model outperforms in accuracy Nash equilibrium and all other models of learning, with the exception of a model (Normalized Fictitious Play proposed by Ert and Erev, 2007) similarly based on regret.

1.1.2 Part Two (Chapter 3)

In the second part of my thesis I propose and analyze the formal properties and the predictive power of a new concept of equilibrium I call Net Reward Attractions (NRA) Equilibrium.

The NRA Equilibrium is a stationary concept designed for strategic form games and is based on behavioral assumptions about human choice behavior, rather than on the principle of full rationality. It is assumed that, in equilibrium, agents are not expected utility maximizers, but that, for a player, the propensity of choosing an action is proportional to its corresponding expected net reward – net reward being defined as the difference between the actual payoff and the minimum obtainable one, given other players' moves. I simply assume here that players are attracted by actions, and that this attraction can be quantified in terms of how much, on average, an action is perceived as better than the others. I propose also a parameterized version of NRA I call Parametric NRA (pNRA), obtained by introducing a parameter $\lambda > 0$, which tunes players' sensitivity to expected net rewards.

The concept of net reward, as introduced here, is very similar to Loomes and Sugden's (1982) concept of *rejoicing* i.e., a measure of the additional pleasure associated to the awareness of having chosen the best action. In this vein, the approach based on net rewards, which I adopt to model choice behavior in the long run, is

complementary, although not equivalent (as I show in Chapter 3), to that based on regret. In Loomes and Sugden's (1982) *regret theory*, these two complementary aspects are fused together in the *Rejoice/Regret* function (see Section 4.2 of Introduction), and I show in Chapters 2 and 3 of my thesis that these two components can be separately used to successfully design models of choice behavior.

The intuition at the basis of the NRA model, that relative rewards are what matters in determining choice behavior rather than absolute payoffs, is coherent with recent neuroeconomic research (Tremblay and Schultz, 1999; Tobler, Fiorillo, and Schultz, 2005; Daw et al., 2006).

In part two of my thesis, I test the predictive accuracy of the NRA equilibrium on data from experiments on 26 repeated, completely mixed games run under full-feedback condition. In addition, I compare NRA's predictive power with that of other five equilibrium concepts and eight models of learning, representing cutting-edge research on interactive decision making modeling. As a result, NRA turns out to be always among the best predictors of empirical data, performing significantly better than Nash equilibrium, self-tuning EWA, and reinforcement-based models.

1.1.3 Part Three (Chapter 4)

The third part of my thesis stands as a first attempt to investigate how do human subjects generalize their past experience when facing new strategic situations i.e., it addresses issues of *conditional behavior* and *generalization*.

With *generalization* I mean here the set of cognitive mechanisms and rules according to which subjects extract from past experience some general knowledge to deal with new, never encountered strategic situations. Issues of generalization and conditional behavior (different responses to different inputs) are relevant because most human interactive learning happens in contexts where tasks do not repeat themselves identically over time, contrary to the typical patterns of interaction that have been empirically studied until now. Generalizing from examples and learning of conditional behavior are natural features of human behavior.

I designed and ran some preliminary multi-game experiments in which subjects played sequences of different two-person 2x2 games with a unique equilibrium in mixed strategies. Each game in a sequence was obtained multiplying by a randomly drawn positive constant the payoffs of two completely mixed games.

I use my experimental data to test the predictive power of the Perceptron-Based (PB) model and compare it with that of other popular learning and equilibrium models of interactive choice behavior. It is worth noting here that conventional "attractions and stochastic choice rule" models of economic learning cannot capture such features of human behavior, since they are designed only for fitting and predicting data from situations in which subjects repeatedly play the same stage game. On the contrary, the architecture of the PB model accounts for this kind of dependence of behavior from the perception of changes in game payoffs.

As a result, the PB model outperforms in accuracy Nash equilibrium and all other models of learning as well. Further, I do not observe learning spillover effects in my experiments, which means that subjects are able to discriminate the different strategic situations and act accordingly. This fact might provide an explanation for why non-standard equilibrium models turn out to be the best predictors of my experimental data.

1.2 On Learning in Repeated, Completely Mixed Games

Despite their apparent simple structure, games with a unique mixed-strategy equilibrium (MSE) are worthy of particular consideration. Zero-sum games, which model a situation in which a player's win corresponds to an opponent's loss and vice-versa, are perhaps the most known and extreme example. In general, constant-sum games, of which zero-sum ones are a particular case, model situations of conflict, since players' interests are opposed: in other words, players cannot help their opponents without being damaged. In this way, feelings such as fairness, reciprocity, and cooperation are almost completely excluded from this kind of interactions. Given their nature, these games faithfully portray situations of everyday life in which strict competitiveness is the most salient feature.

In games with MSE, equilibrium play requires players to randomize their actions: if a player behaves predictably – for example always choosing the same action – an opponent can anticipate his moves and then win. For this reason, as can be intuitively understood, an equilibrium can be established if and only if players behave unpredictably i.e., if they randomize their actions. However, this says nothing about the processes that induce players to introduce randomness into their behavior, and theorists do not yet agree on a unique interpretation of MSE. According to one interpretation (Osborne and Rubinstein, 1994), an equilibrium in mixed strategies can be seen as a

profile of *common* beliefs on the players' moves and each player will choose an action that best responds to those beliefs; in this vein, a player chooses an action rather than a mixed strategy and an equilibrium is a steady state of players' beliefs. Another interpretation is that proposed by Harsanyi (1973), who provides the proof that almost any MSE is the limit of pure strategy strict equilibria of opportunely chosen games whose payoffs are affected by random perturbations; therefore, players merely choose among their possible pure strategies, being the random fluctuations of the payoffs that lead players to use their pure strategies with the right frequencies.

From an experimental and behavioral point of view, however, this class of games represents a serious challenge for the predictive power of Nash equilibrium. Indeed, two strong – but behaviorally weak – assumptions stand at the core of the concept of Nash equilibrium. First, players are assumed to act in accordance with the theory of rational choice: they only care about the maximization of their own expected payoff, given their beliefs about the other players' moves. Second, these beliefs are correct – in that sense players are said to be *experienced*.

In the realistic case of human, bounded-rational, and non-experienced players, it is not clear that an MSE can be learned and the question of how an equilibrium of play (if any) arises is still unanswered. There are at least four problems.

First, the ability itself of providing a series of random independently drawn numbers has been proved to be quite unnatural for human beings (Neuringer, 1986; Camerer, 2003), and stochastic behavior, in this context a major source of cognitive complexity, makes equilibrium strategy hard to be learnt. Second, as it has been shown in Crawford (1985), in games with MSE, learning dynamics that assume players move toward strategies with higher expected payoffs are known not to converge to MSE, at least in finite repetitions. Third, in equilibrium, all mixed strategies yield to each player the same expected payoff (given others are playing the equilibrium mixed strategy), and hence they are all best responses: as a consequence, in equilibrium players have no positive incentives to play the predicted mixed strategy. A fourth problem arises in games repeatedly played by randomly matched subjects of a population, as pointed out in Camerer (2003). In this case, an MSE can be reached at a population level, even if individuals play one of their pure strategies with probability equal to one. As an example, consider the case of the matching pennies game (see figure 1) repeatedly played by random matched individuals of a population; if in the population 50 percent of the individuals always choose Head and the other 50 percent always choose Tail,

then, when two individuals from that population are randomly matched, it is impossible for them to guess the moves of their opponents, making this situation identical to that in which subjects choose randomly their pure strategies with probability 0.5.

Player 2 / **Player 1**	Head	Tail
Head	(1,-1)	(-1,1)
Tail	(-1,1)	(1,-1)

Figure 1. The *matching pennies* game, one of the most popular examples of zero-sum games.

1.2.1 Learning: Empirical Findings

Since the late 1950s, the experimental game theory literature on repeated games has provided significant departures from Nash equilibrium behavior (Erev and Roth, 1998) and especially data from experiments involving repeated games with unique MSE seem to contradict the predictions of standard game theory. In this specific context, indeed, Nash equilibrium not only fails to approximate laboratory observed behavior in the early rounds, but often it is also a poor predictor of the stable behavior emerging in the long run (Erev and Roth, 1998; Erev, Roth, Slonim, and Barron, 2007). As noted in Erev and Roth (1998:851) "in 5 of the 12 games equilibrium predicts badly: average choice probabilities, pooled over all rounds, are closer to random choices than to the equilibrium predictions". The unsatisfactory performances of Nash equilibrium have led researchers to find alternative theories and models of learning to better explain and justify experimentally observed human behavior.

As a result, most of the models of learning proposed in the behavioral game theory literature outperform standard equilibrium theory in the tasks of fitting and predicting experimental data and these models attribute to other factors the role of drivers of choice behavior (Camerer, 2003; Erev and Roth, 1998; Erev, Bereby-Meyer, and Roth, 1999; Erev, Roth, Slonim, and Barron, 2002; Erev et al., 2007).

On the other hand, a growing body of empirical literature that addresses the evaluation of the descriptive and predictive power of MSE for real life, on-field situations, has provided contrasting results with those obtained in the laboratory. The

contributions by Walker and Wooders (2001), Chiappori, Levitt, and Groseclose (2002), Palacios-Huerta (2003), Palacios-Huerta and Volij (2006a, 2006b) show that the behavior of sport and chess professionals is "largely consistent with the minimax hypothesis" (Walker and Wooders, 2001:1521) and "remarkably consistent with equilibrium play in every respect" (Palacios-Huerta, 2003:395). One of the reasons of the discrepancy between on-field and lab-observed behaviors is that in the two cases players have different levels of experience with the situation they are facing. Indeed, as Walker and Wooders (2010) point out, "MSE is effective for explaining and predicting behavior in strategic situations at which the competitors are experts and less effective when the competitors are novices, as experimental subjects typically are". Selten and Chmura (2008) propose another explanation: when a game is repeatedly played with random matching by two populations, subjects' behavior can be quite different from that observed when the same game is played repeatedly by the same two individuals. In the latter case, playing hundreds of times against each other makes players focus on not being predictable by the other, which should reasonably push their behavior to minimax play. However, this explanation seems to be rather weak, since significant departures from MSE have been observed also in many experiments with fix-pairing protocol.

Could context be a further explanation for professionals' behavior? Empirical evidence provides a negative answer to this question. Palacios-Huerta and Volij (2006a), indeed, observed the behavior of students and soccer professionals playing in laboratory settings a 2x2 game, formally identical to the typical strategic interactive situation of a penalty kick. The authors find that while professionals continue to play consistently with Nash theory, even in settings that entirely differ from those they are familiar with, college students perform quite poorly in terms of equilibrium play. This can be interpreted as evidence that professionals are able to transfer their strategic skills across different environments and that context has a negligible role in pushing subjects' behavior to equilibrium play.

In my view, the fact that experienced players tend to conform to Nash play surely adds important insights on human choice behavior, but does not invalidate the results obtained in labs. Indeed, in most of everyday contexts, we do not repeatedly face the same identical strategic situations – and we do not perceive them as identical, either. Thus, in many interesting applications to everyday life, it seems reasonable to assume

that individuals' behavior is closer to that of college students than to professionals, and this justifies the need for experiments on repeated games.

1.3 Quantitative Models of Learning

Standard game theory does not provide a theory of learning and is limited to describing a steady state situation. On the contrary, experimentally observed behavior provides overwhelming evidence of the existence of a process – i.e. *learning* – after which past experience dramatically affects subjects' current strategic choices (Camerer, 2003). Specifically, interactive learning differs from individual learning in that given N agents, each agent adapts to a strategic environment which is continuously modified by the concurrent learning of the other $N-1$ agents.

Learning models try to replicate artificially the process in which past experience affects agents' current behavior; more specifically, they establish how the probabilities with which future actions will be chosen are affected by information about the outcomes produced by actions chosen in the past. In order to do this, quantitative theories assume that, for a player, all his possible actions are associated with numerical evaluations, called *attractions* or *propensities* (these two terms will be used interchangeably), which are mapped, according to opportune rules, into choice probabilities. Propensities can be interpreted as a measure of the propensity of a player to choose the actions they are associated with, while learning rules determine how these attractions are updated in response to past experience.

There is a wide variety of different approaches for modeling learning (for a comprehensive review of these models and theories see Camerer, 2003), but the most successful learning theories proposed so far are those of *reinforcement learning*, *beliefs learning*, hybrid models combining both (Ho, Camerer, and Chong, 2007) and, finally, theories which emphasize the role of post-decision regret as the driver of human behavior (Erev et al., 1999; Ert and Erev, 2007).

1.3.1 Reinforcement Learning Models

Reinforcement learning models are based on the following assumptions about human choice behavior (Erev and Roth, 1998):

1. *The Law of Effect*: choices that have led to good outcomes in the past are more likely to be repeated in the future (Thorndike, 1898). This law implicitly assumes that choice behavior is probabilistic.
2. *The power law of practice*: learning curves tend to be steep initially, and then flatter (Blackburn, 1936).
3. *Experimentation (or Generalization)*: not only the choices which were successful in the past are more likely to be employed in the future, but also similar choices will be employed more often (Erev and Roth, 1998).
4. *Recency*: recent experience plays a larger role than past experience in determining behavior (Erev and Roth, 1998).

Erev and Roth's Reinforcement Learning (REL), the standard Reinforcement Learning (RL), and the Normalized Reinforcement Learning (NRL) models embed in their structure these four principles.

In reinforcement models, agents are assumed to have a very simple cognitive structure: they do not know anything about foregone or historical payoffs from strategies they did not choose, and occasionally experiment with the effects of similar choices. Here, only the actually played actions are reinforced. Typically, these models underestimate the empirical rate of learning, although correctly predicting its direction, being in the majority of the cases too slow to adapt to the observed dynamics. This seems to be due to the fact that in experiments in which subjects are provided with complete information about payoffs, they actually use that information in forming their strategies, while those models, by design, do not.

The REL Model

This model was first proposed in Erev et al. (1999) and further considered and developed in Erev et al. (2002). Here, I describe the REL model as reported in the latter contribution.

Attractions updating. The propensity of player i to play her k-th pure strategy at period $t+1$ is given by:

$$a_{ij}(t+1) = \begin{cases} \dfrac{a_{ij}(t)\cdot\left[N(1)+C_{ij}(t)-1\right]+x}{N(1)+C_{ij}(t)} & \text{if } k=j \\ a_{ij}(t) & \text{otherwise,} \end{cases}$$

where $C_{ij}(t)$ indicates the number of times that strategy *j* has been chosen in the first *t* rounds, *x* is the obtained payoff, and $N(1)$ a parameter of the model determining the weight of the initial attractions.

Stochastic choice rule. Player *i*'s attractions are mapped into choice probabilities by the following logistic rule:

$$p_{ik}(t) = \frac{\exp\left[\lambda \cdot a_{ik}(t)/S(t)\right]}{\sum_j \exp\left[\lambda \cdot a_{ij}(t)/S(t)\right]},$$

where λ is a parameter tuning the sensitivity to payoff values, and $S(t)$ gives a measure of payoff variability.

Initial attractions. The value $S(1)$ is defined as the expected absolute distance between the payoff from random choices and the expected payoff given random choices, denoted as $A(1)$. For period $t > 1$, the authors define:

$$S(t+1) = \frac{S(t) \cdot \left[t + m \cdot N(1)\right] + \left|A(t) - x\right|}{t + m \cdot N(1) + 1},$$

where *x* is the received payoff, *m* the number of player *i*'s pure strategies, and $A(t+1)$ is defined as:

$$S(t+1) = \frac{A(t) \cdot \left[t + m \cdot N(1)\right] + x}{t + m \cdot N(1) + 1}.$$

The authors fix initial attractions as follows:

$a_{ij}(1) = A(1)$, for all *i* and *j*.

Thus, this model has two free parameters, namely λ and $N(1)$.

The RL Model

This model has been proposed in Erev et al. (2007) and enriches the Basic Reinforcement model described in Erev and Roth (1999); the main difference between the two models is that in the latter, propensities are mapped into choice probability by simple normalization, while, in the former, this mapping is operated by a logit function.

Initial propensities. At time period $t = 1$, player *i*-th associates to the propensity of playing his pure strategy *j*, the value corresponding to the expected payoff from random choice (denoted by $A(1)$). Thus:

$a_{ij}(1) = A(1)$, for all *i* and *j*.

Attractions updating. At each time step, propensities are updated according to the following:

$$a_{ij}(t+1) = \begin{cases} (1-w)\cdot a_{ij}(t) + w\cdot v_{ik}(x) & \text{if } j = k \\ a_{ij}(t) & \text{otherwise}, \end{cases}$$

where $v_{ij}(t)$ is the realized payoff and w one of the two parameters of the model (sensitivity to foregone payoffs). The updating rule implies agents' insensitivity to foregone payoffs.

Stochastic choice rule. Attractions at time t are mapped into choice probabilities according to the rule:

$$p_{ik}(t) = \frac{\exp\left[\lambda \cdot a_{ik}(t)\right]}{\sum_j \exp\left[\lambda \cdot a_{ij}(t)\right]},$$

where λ is a free parameter tuning sensitivity to payoffs. In the first period, the authors suggest setting $recent_i = A(1)$.

The NRL Model

This model, described in Erev et al. (2007), is quite similar to REL and differs from RL in the fact that here payoff sensitivity is assumed to decrease with payoff variability.

Initial propensities. At time period $t = 1$, player *i*-th associates to the propensity of playing his pure strategy *j*, the value corresponding to the expected payoff from random choice (denoted by $A(1)$). Thus:

$a_{ij}(1) = A(1)$, for all *i* and *j*.

Attractions updating. At each time step, propensities are updated according to the following:

$$a_{ij}(t+1) = \begin{cases} (1-w)\cdot a_{ij}(t) + w\cdot v_{ik}(x) & \text{if } j = k \\ a_{ij}(t) & \text{otherwise}, \end{cases}$$

where $v_{ij}(t)$ is the realized payoff and w one of the two parameters of the model (sensitivity to foregone payoffs). The updating rule implies agents' insensitivity to foregone payoffs.

Stochastic choice rule. Attractions at time t are mapped into choice probabilities according to the rule:

$$p_{ik}(t) = \frac{\exp\left[\lambda \cdot a_{ik}(t)/S(t)\right]}{\sum_j \exp\left[\lambda \cdot a_{ij}(t)/S(t)\right]},$$

where $S(t)$ gives a measure of payoff variability and λ is a free parameter tuning sensitivity to payoffs.

$$S(t+1) = (1-w)\cdot S(t) + w\left|\max\{recent_1, recent_2\} - v_{ij}(t)\right|,$$

where $recent_i$ is the most recent experienced payoff from action *i*. In the first period, the authors suggest setting $recent_i = A(1)$; in addition, the initial value $S(1)$ is set equal to λ. Similarly to the case of the NFP model, payoff sensitivity (the ratio $\lambda/S(t)$) is assumed to decrease with payoff variability.

1.3.2 Beliefs Learning Models

The models of this class embed the principles of the *beliefs learning* theory and are generally much more sophisticated than reinforcement models. According to this theory, players are assumed to keep track of the history of all other players' moves and form their beliefs about what other players will do based on this past information. The strategy that will be chosen is that which maximizes the expected payoff given the beliefs about other players' actions.

Two very popular models derived from this theory are the *fictitious play* and *weighted fictitious play* models. In the first, players keep track of the relative frequency with which other players have employed each strategy in the past, and then calculate the expected payoff given these beliefs and choose that with the highest expected value. While in this model all previous observations are equally salient, in the *weighted fictitious play* model distant experiences in the past are less salient than recent ones (*recency effect*).

The Normalized Fictitious Play (NFP), the Stochastic Fictitious Play (SFP), and the Self-Tuning Experience Weighted Attraction (stEWA) models belong to this class of models. The last model, however, would be better described as a hybrid model, blending the main features of reinforcement and fictitious play models; indeed, if parameters are constrained to specific values, it reduces to a simple version of the reinforcement model in which only chosen strategies are reinforced and if parameters are set in a different way, stEWA reduces exactly to *weighted fictitious play*.

The weakness of these models (and of most of the reinforcement ones), however, stands in the logit response function which operates the mapping of propensities into

choice probabilities; by construction, this function is extremely sensitive to how initial propensities are defined, and different approaches can dramatically affect the performances of these models.

The NFP Model

This model has been proposed by Ert and Erev (2007) and described in Erev et al. (2007).

Initial propensities. At time period $t = 1$, player *i*-th associates to the propensity of playing his pure strategy *j* the value corresponding to the expected payoff from random choice (denoted by $A(1)$). Thus:

$a_{ij}(1) = A(1)$, for all *i* and *j*.

Attractions updating. At each time step, propensities are updated according to the following:

$a_{ij}(t+1) = (1-w) \cdot a_{ij}(t) + w \cdot v_{ij}(t)$, for all *i* and *j*,

where $v_{ij}(t)$ is the expected payoff in the selected cell and *w* is one of the two parameters of the model (sensitivity to foregone payoffs).

Stochastic choice rule. Attractions at time t are mapped into choice probabilities according to the rule:

$$p_{ik}(t) = \frac{\exp\left[\lambda \cdot a_{ik}(t)/S(t)\right]}{\sum_j \exp\left[\lambda \cdot a_{ij}(t)/S(t)\right]},$$

where $S(t)$ gives a measure of payoff variability and λ is a free parameter tuning sensitivity to payoffs.

$S(t+1) = (1-w) \cdot S(t) + w\left|\max\{recent_1, recent_2\} - v_{ij}(t)\right|$,

where $recent_i$ is the last experienced payoff from action *i*. In the first period, the authors suggest setting $recent_i = A(1)$; in addition, the initial value $S(1)$ is set equal to λ.

The SFP Model

This model, described in Erev et al. (2007), is identical to NFP with the exception that here stable payoff sensitivity is assumed.

Initial propensities. At time period $t = 1$, player *i*-th associates to the propensity of playing his pure strategy *j* the value corresponding to the expected payoff from random choice (denoted by $A(1)$). Thus:

$a_{ij}(1) = A(1)$, for all *i* and *j*.

Attractions updating. At each time step, propensities are updated according to the following:

$a_{ij}(t+1) = (1-w) \cdot a_{ij}(t) + w \cdot v_{ij}(t)$, for all *i* and *j*,

where $v_{ij}(t)$ is the expected payoff in the selected cell and *w* one of the two parameters of the model (sensitivity to foregone payoffs).

Stochastic choice rule. Attractions at time t are mapped into choice probabilities according to the rule:

$$p_{ik}(t) = \frac{\exp\left[\lambda \cdot a_{ik}(t)\right]}{\sum_j \exp\left[\lambda \cdot a_{ij}(t)\right]},$$

where λ is a free parameter tuning sensitivity to payoffs. In the first period, the authors suggest setting $recent_i = A(1)$.

The stEWA Model

Self-tuning Experience Weighted Attraction is a one-parameter model of learning in games proposed by by Ho, Camerer, and Chong (2007). It replaces part of the 5 parameters in an earlier model called EWA (Camerer and Ho, 1999) with functions of experience that operate a self-tuning over time.

Attractions updating. At time *t*, player *i* associates to his *j*-th pure strategy the attraction $a_{ij}(t)$, given by:

$$a_{ij}(t) = \frac{\phi_i(t) \cdot N(t-1) \cdot a_{ij}(t-1) + \left[\delta_{ij}(t) + \left(1 - \delta_{ij}(t)\right) \cdot I\left(s_{ij}, s_i(t)\right)\right] \cdot \pi_i\left(s_{ij}, s_{-i}(t)\right)}{N(t-1) \cdot \phi_i(t) + 1},$$

where are parameters, $s_i(t)$ and $s_{-i}(t)$ are the strategies played by player *i* and his opponents, respectively, and $\pi_i\left(s_{ij}, s_{-i}(t)\right)$ is the ex-post payoff deriving from playing strategy *j*. The function $I(\cdot)$ is defined as:

$$I(x,y) = \begin{cases} 0 & if\ x \neq y \\ 1 & if\ x = y, \end{cases}$$

while the functions $\delta_{ij}(t)$ and $\phi_i(t)$ are called, respectively, the *attention function* end the *change detector function*. The second depends primarily on the difference between the relative frequencies of chosen strategies in the most recent periods and the relative frequencies calculated on the entire series of actions. The attention function essentially tunes the importance that players associate to past payoffs (see Camerer, Ho and Chong, 2007 for details). Thus, attractions on time t depend on the attractions on time *t-1* multiplied by an experience weight $N(t-1)$, on received and foregone payoffs, and are pseudo normalized by the quantity $N(t) = N(t-1) \cdot \phi_i(t) + 1$ ($N(0) = 1$).

Stochastic choice rule. Attractions are mapped into choice probabilities by the following equation:

$$p_{ij}(t+1) = \frac{\exp\left(\lambda \cdot a_{ij}(t)\right)}{\sum_j \exp\left(\lambda \cdot a_{ij}(t)\right)},$$

where λ is the unique free parameter of the model.

Initial attractions. The authors do not provide a unique method to define initial attractions $a_{ij}(0)$ and suggest at least four ways it might be done. In this specific case, I define initial attractions according to the method adopted for reinforcement models, which leads to first-period uniformly distributed choices.

1.4 Regret and Choice Behavior

The unsatisfactory performances of Nash equilibrium have led researchers to find alternative theories and models to better explain and justify experimentally observed interactive choice behavior.

As a result, no matter the methodology adopted, most of the models proposed in the behavioral game theory literature outperform standard equilibrium theory in both the tasks of fitting and predicting experimental data, and attribute to other factors the role of drivers of choice behavior (Camerer, 2003; Erev and Roth, 1998; Erev et al., 1999, 2002, 2007; Selten and Chmura, 2008). Specifically, some recent contributions have shown that regret-based models are the best predictors of data from experiments on interactive repeated choice tasks, thus suggesting that regret for foregone payoffs must play a central role in shaping human choice behavior. Before proceeding further with the description of the most important economic theories of decision based on regret, I

will provide a short review of the principal contributions on regret proposed in the psychology literature, in order to more precisely define its meaning and nature.

1.4.1 Psychology of Regret

Behavioral economics, experimental economics, and psychology have devoted much attention to the effects of emotions on decision-making, and the literature on this topic is vast. If we consider all contributions on emotions and the role they play in shaping human choice behavior, regret has been the most studied. I will present some of the most important contributions, mainly from the field of psychology, investigating nature and properties of this *counterfactual* emotion.

Regret is generally defined as the emotion that a decision maker experiences whenever the outcome of his action is worse than the one he would have received, had he acted in a different way. A first distinction has to be done between *regret* and *disappointment*; generally, these two emotions are reputed to be different and have been shown to produce different behaviors (Mellers, Schwartz, and Ritov, 1999; Zeelenberg, van Dijk, and Manstead, 1998). Disappointment arises whenever the received outcome is worse than the outcome one would have obtained in another state of the world. Therefore, the difference between these two negative emotions relies on the decision maker's intervention (*agency*); for regret to occur, not only the actual outcome must be worse than foregone ones, but the decision maker must also consider himself as directly responsible for it by having chosen a specific course of action.

As said, regret is a *counterfactual* emotion i.e., it arises in those situations in which we make comparisons between the reality and what might have been, had we acted differently. Then, regret can be seen as a consequence of the natural humans' attitude to think counterfactually (Zeelenberg at al., 1998). As Roese (1994:805) writes, "The ability to imagine alternative, or *counterfactual*, versions of actual events appears to be pervasive, perhaps even essential, feature of human consciousness." Counterfactual thoughts are precisely structured: they can be represented as conditional sentences with an antecedent of the form "If only I had done X", and a consequent of the form "Y would have happened"; in other words, one alters some factual antecedent and evaluates the consequences of that alteration. The question that arises is, then, why do we reason counterfactually? It has been shown (Roese, 1994) that counterfactual reasoning serves two functions: *affective* and *preparative*. As for the first, people might think to how things might have been different to make themselves feel better (e.g., rape

victims sometimes generate positive feeling by noting that they could have been more seriously injured or killed). As for the latter function, comparisons with better alternatives (called *upward counterfactuals*) can serve to develop patterns of future actions. Indeed, as Roese (1997) points out, counterfactual thinking is triggered when our choices have a negative effect i.e., in those situations in which corrective thinking is most important.

Two main factors have been shown to determine regret and its intensity: the first is the degree of availability of possible alternatives and, second, the active versus inactive attitude of the decision maker. Seta, Seta, McElroy, and Hatz's (2008) experimental results show that the salience of counterfactuals is positively correlated with the intensity of experienced regret, coherently with Kahnemann and Miller's (1986) norm theory. In addition, also mutability of events or states can affect the intensity of regret. The underlying idea is that if events can be changed in many ways, it is also true that some modifications are more natural than others as well as some attributes are easier to be changed than others. Kahnemann and Tversky (1982) show that exceptional features are more mutable than routine ones since the former explicitly provide alternative scenarios to the occurred state. Kahnemann and Miller (1986) further investigate the role of mutability and find that "an event is more likely to be undone by altering exceptional than routine aspects of the causal chain that led to it" (Kahnemann and Miller, 1986:143). From this point of view, when agents' decisions involve active behavior, they are likely to be considered as exceptional features and generate regret ("If only I did not do that…"). On the opposite, when agents' behavior is inertial (i.e., they do not act to change things), their choices are more naturally interpreted as routine features in the causal chain, and are less likely to generate regret.

A series of empirical studies have shown not only that post-decisional regret experienced in the past plays a crucial role in determining our behavior in current decision tasks, but also that this emotion is conditional to the knowledge of the outcomes from unchosen actions. However, regret has also been proved to have an anticipatory dimension, and studies by Bar-Hillel and Neter (1996), Zeelenberg (1999), and Hetts et al. (2000) have shown that the anticipation of regret does influence current decisions and behavior. These contributions provide empirical evidence supporting the hypothesis that individuals are able to anticipate counterfactual regret, by imagining the consequences of each action, and choose the action that would produce the lowest level of regret. This kind of behavior is in accordance with a large body of literature showing

that people not only anticipate emotions, but also take them into account when deciding (Larrick and Boles, 1995; Ritov, 1996; Bar-Hillel and Neter, 1996; Zeelenberg, Beattie, van der Pligt, and de Vries 1996; Zeelenberg and Beattie, 1997). In addition, Zeelenberg (1999) provides an explanation of how anticipated regret can lead to relatively risk seeking behavior, as previously experimentally shown by Larrick and Boles (1995) and Ritov (1996). Zeelenberg's argument starts from the reasonable assumption that people are *regret averse*; regret is a negative and unpleasant emotion, and then people tend to make choices as to minimize it. Now, regret-minimizing choices can be either safe or risky; indeed, it can happen that risky options are those to which there corresponds the lowest level of regret. As a simple example, consider a situation in which an individual has to choose between two choices, one riskier than the other. Assume also that the riskier option will always be resolved, whereas the safer will only be resolved if chosen. Then, if the decision maker chooses the safer choice, he runs the risk of learning that the riskier option turned out to be better and then experiences regret.

Zeelenberg (1999) mentions five conditions, not yet experimentally tested, that might determine occurrence and intensity of anticipated counterfactual regret. First, regret is likely to be anticipated when available actions have similar degree of dominance. If an action is evidently dominant (for some particularly salient reason) with respect to the others, an agent will choose it without spending too much time thinking about its consequences, and he would not consider himself as particularly responsible for a possible bad or suboptimal outcome (of course, he would be *disappointed*). On the contrary, if available actions are of equal attractiveness, then an agent would consider them more thoroughly and anticipate the feeling of regret he might feel for not having chosen the best one, and a bad outcome would be easily interpreted as the consequence of a wrong choice. Second, the shorter the time interval between an action and its consequence, the more intense the anticipated regret; if consequences are delayed in time, agents might discount the associated regret. Third, the relative importance of actions plays a central role in the anticipation of regret; it is reasonable to assume that if an action has important consequences, then it will result in a more intense feeling of regret. The fourth factor is the availability of feedback about unchosen options. Zeelenberg shows that when post-decisional feedback is available, people anticipate regret; on the opposite, when this feedback is not available, regret plays almost no role in the process of decision making. Lastly, the social dimension of

the decision making process might affect the level of anticipated regret, particularly high in situations in which people that are important to the agent expect him to carefully evaluate all alternatives or delay his choice.

As noted by Zeelenberg (1999), these five aspects deserve further empirical investigation, as their understanding would help design a psychological theory of regret aversion and determine its scope of applicability.

1.4.2 Regret and Decision-Making Modeling

As far as I know, Savage (1951 and 1954) was the first to formally introduce regret in a theory of decision-making. His theory of (statistical) decision-making applies to situations in which the utility of an individual depends upon his own choice and the occurrence of one of *n* mutually exclusive states of the world. It is assumed that agents know how their own utilities depend jointly upon their choices and the (unknown) state of the world that will occur, but they do not know the probabilities that are associated to each state of the world. Savage defined the loss associated to action *a* and state *s* as the difference between the best outcome over all possible actions (given state *s*) and the outcome from action *a*. Let us consider the following example proposed by Savage in which the decision maker has to decide whether or not to carry with him his umbrella. Two states of the world can take place: it might be rainy or shiny, and the decision maker does not know the probabilities of the two states. Suppose that the utility of the decision maker for each possible combination (action, state) is as reported in the following matrix:

State / Action	Rain	Shine
Carry	4	5
Do not carry	-10	10

The corresponding matrix of losses will be then:

State / Action	Rain	Shine
Carry	0	5
Do not carry	14	0

Savage proposed as a decision rule the *minimax principle*, according to which the decision maker chooses the action that minimizes the maximum loss. This theory of choice allows for violations of the axiom of *independence of irrelevant alternatives* and is quite pessimistic, since the decision maker looks only at the worst possible state for each of his actions. For these reasons both normative and descriptive validity of the *minimax regret choice rule* were criticized (Mellers, Schwartz, and Ritov, 1999).

The tendency to anticipate regret and avoid post-decision regret shown by humans was first incorporated in an economic model of individual decision-making by Bell (1982) and by Loomes and Sugden (1982). These two contributions independently introduced and developed the *regret theory* to account for empirical systematic violations of some of the axioms of expected utility theory (von Neumann and Morgenstern, 1947). In particular, the aim of Loomes and Sugden (1982) was that of proposing a new, alternative theory of individual choice under uncertainty much simpler and intuitively more appealing than Kahnemann and Tversky's (1979) *prospect theory*. The starting point of Loomes and Sugden's (1982) theory is the systematic violation of some of the axioms of the conventional expected utility theory, observed in the experiments on choice between pairs of prospects (i.e., probability distributions over consequences) described by Kanhemann and Tversky (1979). Specifically, three different kinds of *paradoxical* behavior emerged (nowadays widely known) that cannot be accommodated by conventional theory of choice without dropping one or more of its axioms: the *common ratio effect*, *Allais paradox*, and the *isolation effect*. Regret theory was formulated to account for these *irrational* behaviors. According to this theory, utility is interpreted in its classical, Bernoullian sense i.e., as the psychological experience of pleasure associated to satisfaction of desire. In this view, it is clear that when deciding, psychological factors other than the sole income can modify our utility, and regret for foregone gains and rejoicing for foregone losses are perhaps the most

important. This means that our utility is determined not only by the outcome from our choice (as assumed by von Neumann and Morgenstern's theory), but also by outcomes corresponding to unchosen actions. As an example, if foregone outcomes are better than the obtained one, we would experience regret for not having chosen differently, with a consequent decrease in the utility level. On the opposite, if foregone outcomes are worse than the obtained one, we would then experience rejoicing for having made the best decision, and this would translate in an increase of utility. The concept of regret as illustrated above was not new in the early 1980s, but closely resembles the argument exposed in Savage (1951) in the ambit of the theory of statistical decision, with the difference that in regret theory probabilities associated to outcomes are known. Of course, the importance of *what might have been* can be assessed only in those cases in which all outcomes are known i.e., in those situations in which agents receive feedback about their actions. In the light of these considerations about utility, Loomes and Sugden (1982) and Bell (1982) proposed a model in which agents are supposed to maximize a *modified utility function*, which explicitly takes into account the role of regret. In a restricted version of this model formulated by Loomes and Sugden, the utility function is of the form:

$$m_{ij}^{k} = c_{ij} + R\left(c_{ij} - c_{kj}\right), \qquad (1)$$

where m_{ij}^{k} is the modified utility when action *i* has been chosen and the *j*-th state of the world has occurred, with respect to the consequence of action *k*; similarly, c_{ij} represents the *choiceless utility*, defined as the utility that the individual would derive from outcome *x without having chosen* it – as if it were exogenously assigned to the individual. This assumption about c_{ij} is quite important because, in contrast with the concept of utility provided by von Neumann and Morgenstern, it provides a sort of utility measure free from any psychological implication. Psychological aspects are introduced in (1) through the real valued *regret-rejoice function* $R(\cdot)$, which weights the difference between obtained and foregone utility. Obviously, $R(\cdot)$ is supposed to be non-decreasing. In the limiting case in which $R(c) = 0 \ \forall c$, (1) is equivalent to standard expected utility theory. In terms of (1), action A_k is non-preferred to action A_i if and only if:

$$\sum_{j=1}^{n} p_j \left[c_{ij} - c_{kj} + R\left(c_{ij} - c_{kj}\right) - R\left(c_{kj} - c_{ij}\right) \right] \geq 0, \qquad (2)$$

given that each of the n states of the world occurs with probability p_j. As its authors suggested, equation (2) can be reformulated in terms of a function $Q(\cdot)$ such that $Q(c) = c + R(c) - R(-c)$, obtaining:

$$\sum_{j=1}^{n} p_j \left[Q\left(c_{ij} - c_{kj}\right) \right] \geq 0 .$$

Now, individuals for which $Q(\cdot)$ is non-linear behave in such a way that might violate "consistently and knowingly the axioms of transitivity and equivalence without ever accepting, even after the most careful reflection, that they have made a mistake" (Loomes and Sugden, 1982:820). With this sentence, the authors challenge the assumption that choice behavior under uncertainty can be defined as *rational* if and only if it conforms to the axioms of von Neumann and Morgenstern's expected utility theory. On the contrary, Loomes and Sugden propose the idea that agents whose choice behavior violates some of the axioms of expected utility theory is not necessarily *ir*rational, as it can still be described (as in the case of *regret theory*) in terms of a behavior that maximizes an opportunely defined (or, better, *modified*) utility function.

Loomes and Sugden (1987a) compare regret theory with the *skew-symmetric bilinear utility theory* (SSB) proposed by Fishburn (1982 and 1983). The approach followed by Fishburn is essentially axiomatic rather than psychologically based. The two theories are similar in that they both drop the transitivity axiom. However, Fishburn's model cannot account for the isolation effect, as it is presented in terms of prospects rather than actions. On the other hand, if we consider the particular case of (statistically) independent prospects, regret theory and SSB are equivalent.

Loomes and Sugden (1987b) provide further empirical evidence of violations of the axioms of expected utility theory, supporting the hypothesis that individuals' capacity to anticipate feelings of regret and rejoicing heavily affect choice behavior and confirming the need for a theory that takes explicitly into account this psychological aspect.

The contribution by Selten and Stoeker (1986) first generalized the concept of post-decision regret to the context of interactive strategic situations (*games*), building the foundations of *Learning Direction Theory* (LDT) – successively developed in Selten and Buchta (1999). The approach adopted in LDT is quite different from that of regret theory: in the former case post-decision regret is emphasized, whereas in the latter it is the anticipation of regret and its avoidance that conditions choice behavior.

LDT is a qualitative theory of learning in repeated decision tasks and assumes that agents decide on the basis of the *ex-post rationality* principle: one looks at what might have been better in the previous instance of decision making and adjusts in this direction. The central point is that agents' behavior is based on a *qualitative* and *causal* representation of their environment. The feedback about actions chosen in the previous trial is a necessary condition for a qualitative and causal representation of the context in which the new decisions are taken. Such a representation of the world and feedback about previous choices are the two fundamental assumptions of LDT.

LDT is not a complete explanation of adaptive behavior and does not postulate that ex-post rationality is always sufficient in the explanation of the experimentally observed behavior. Sometimes other factors may influence the decisional process, leading to adjustments in the "wrong" direction. However, this theory assumes that *ex-post rationality* is more important than the other factors. These considerations lead to the following prediction: more frequently than randomly, changes in the parameters are in the direction suggested by *ex-post rationality*.

Due to its qualitative nature, LDT does not specify the probabilities with which changes will occur, and hence we cannot use it to make quantitative predictions. However, this theory provides important insights – largely supported by experimental data (Selten, Abbink, and Cox, 2001) – whose basic principles can be incorporated into other quantitative models.

Recently, also some game theorists have devoted their attention to regret. Contributions by Hart and Mas-Colell (2000 and 2003) and by Hart (2005) show the existence of some adaptive procedures of choice behavior, defined in discrete time and based on regret, that can be proved to converge to the set of correlated equilibria of a game (the notion of correlated equilibrium was first introduced by Aumann, 1974). The approach followed by the authors is almost exclusively theoretical, leaving no room for empirical tests of their models. The most important is the *regret matching* procedure introduced by Hart and Mas-Colell (2000) and defined by the following, simple rule:

> "Switch next period to a different action with a probability that is *proportional* to the *regret* for that action, where *regret* is defined as the increase in payoff had such a change always been made in the past." (Hart, 2005)

The mathematical formulation of the rule above is as follows. Consider player *i* at time $T+1$. The average obtained payoff over the first *T* periods is:

$$U = \frac{1}{T}\sum_{t=1}^{T} u^i(s_t),$$

and denote with $j = s_T^i$ the action chosen by player *i* at time *T*. For each available alternative action $k \neq j$, consider the average payoff that *i* would have obtained had he always played *k* instead of *j* in all previous trials:

$$V(k) = \frac{1}{T}\sum_{t=1}^{T} v_t$$

where

$$v_t = \begin{cases} u^i\left(k, s_t^{-i}\right) & \text{if } s_t^i = j \\ u^i\left(s_t^i, s_t^{-i}\right) & \text{if } s_t^i \neq j. \end{cases}$$

The regret associated to action *k* is then defined as:

$$R(k) = V(k) - U,$$

if the difference is positive and zero otherwise. According to *regret matching*, the probability $p_{T+1}(k)$ with which action *k* is played at time $T+1$ is proportional to $R(k)$ according to the following:

$$p_{T+1}(k) = \begin{cases} c \cdot R(k) & \text{if } k \neq j \\ 1 - \sum_{k \neq j} c \cdot R(k) & \text{if } k \neq j, \end{cases}$$

where *c* is an opportune positive constant. Therefore, if at time $T+1$, before choosing his action, player *i* has no regret (i.e., all $R(k) = 0$ for all $k \neq j$), then he will play action *j* for sure. If instead there are some actions *k* for which $R(k)$ is positive, then the probability for player *i* to choose those actions will be different than zero and proportional to their corresponding regret.

The *unconditional regret matching* model (Hart, 2005) is obtained by slightly changing function $V(k)$ and replacing it with the following:

$$\tilde{V}(k) = \frac{1}{T}\sum_{t=1}^{T} u^i\left(k, s_t^{-i}\right).$$

In this case $R(k)$ correspond to an increase in the average payoff, if any, were one to replace all past plays, and not the *j*-plays, by *k*. Of course, in the case of 2x2 games (and in general in any strategic situation in which all sets of actions have two elements), *regret matching* and *unconditional regret matching* are equivalent.

The most important results proved by Hart and Mas-Colell (2003) is that if all players play regret matching strategies, then the joint distribution of play converges to the set of correlated equilibria of the stage game. This result, known as the *Regret Matching Theorem*, is important as it shows that behavior of bounded rational agents can nonetheless converge to a *rational* outcome i.e., a correlated equilibrium. This result must then be seen as an effort to reconcile bounded rationality and rational behavior.

1.5 Best Response and Behavioral Models of Equilibrium

In spite of what reported at the beginning of Section 3, a stream of economic literature on non-standard equilibrium models has shown that also stationary concepts based on psychological considerations about human behavior are good predictors of data from experiments on auctions and repeated, completely mixed games (Selten, Abbink, and Cox, 2005; Ockenfels and Selten, 2005; Avrahami, Güth, and Kareev, 2005; Neugebauer and Selten, 2006; Selten and Chmura, 2008). In particular, I am referring to the Impulse Balance Equilibrium (IBE) model proposed by Ockenfels and Selten (2005). This stationary concept incorporates the principles of Learning Direction Theory (see previous section) in a quantitative theory. Specifically, its authors define *upward* and *downward impulses*: we have an upward impulse if a higher parameter would have yielded a higher payoff, and a downward impulse in the opposite case. The decision maker is assumed to act in the direction of impulses. In the context of normal form games, impulses are determined as the expected payoff in a transformed game obtained subtracting to all payoffs above the pure strategy maximin payoffs (regarded as a "*natural aspiration level*", Selten and Chmura, 2008:947) one half of the difference between original payoffs and the maximin payoffs. The rationale for this kind of rescaling is that, as in Kahnemann and Tversky's (1979) prospect theory, losses (evaluated with respect to the *natural level of aspiration* embodied by the maximin payoff) are weighted double in the computation of impulses. An equilibrium of play is established when probabilities are such that the downward impulse equals the upward impulse. Impulse Balance Equilibrium is quite important because agents are not supposed to be neither expected utility maximizers nor best responders to some kind of partial information. In other non-standard stationary models as Payoff-sampling equilibrium (Osborne and Rubinstein, 1998) and Action-sampling equilibrium

(described in Selten and Chmura, 2008, but previously formulated by Selten), agents are supposed to choose optimally with respect to the information from *n* samples of equal size (one for each available action), and from a sample of seven observations of the strategies played by their opponents, respectively. In both cases, the size of the sample can be interpreted as the unique free parameter of the two models. The concept of Quantal Response (QRE) equilibrium proposed by (McKelvey and Palfrey, 1995) can be considered as a generalization of the equilibrium model proposed by Nash (1950). It is based on the idea that players give quantal best responses to the behavior of the others i.e., players make mistakes and assume other players to do so as well; players are still supposed to be maximizers, departing from Nash's theory in that perfectly rational expectations are replaced with noisy, imperfect ones. Assuming a particular distribution of errors, along the theoretical framework proposed by McFadden (1976), McKelvey and Palfrey designed the *Logit Equilibrium*, which converges to Nash equilibrium as the free parameter of the *logistic quantal response function* tends to infinite.

Selten and Chmura (2008) show that the free parameters IBE model is the best predictor of the data from experiments on twelve 2x2 repeated, completely mixed games, if compared with the other stationary concepts mentioned in this Section. However, the authors raise two important, yet unanswered, questions: first, it is not clear why in some games equilibrium models with so different theoretical foundations provide equivalently accurate predictions; second, they do not test models on more general patterns of strategic interaction (e.g., games with more than two players and more than two actions available to each player).

I provide here a short description of each model of equilibrium in the particular context of two-person 2x2 games with a unique equilibrium in mixed strategies. A detailed description and a comparative analysis of these models is reported in Selten and Chmura (2008).

Any game in this particular class can be described by the following payoff structure (Selten and Chmura, 2008), the other possible case being obtained by just switching its rows and columns:

Player 2 / Player 1	L	R
U	$(a_L + c_L; b_U)$	$(a_R; b_U + d_U)$
D	$(a_L; b_D + d_D)$	$(a_R + c_R; b_D)$

where the constants c_L, c_R, d_U, and d_D are strictly bigger than zero. Let us assume that Player 1 will choose action U with probability *p* and that Player 2 will choose action L with probability *q*.

Quantal Response Equilibrium (QRE)

Quantal Response Equilibrium was first introduced by McKelvey and Palfrey (1995). Equilibrium probabilities are determined as follows:

$$p = \frac{e^{\lambda E_U(q)}}{e^{\lambda E_U(q)} + e^{\lambda E_D(q)}} \text{ and } q = \frac{e^{\lambda E_L(p)}}{e^{\lambda E_L(p)} + e^{\lambda E_R(p)}},$$

where $\lambda \geq 0$ is the unique free parameter of the model.

Action-Sampling Equilibrium (7-sampling)

Proposed by Reinhard Selten, this stationary concept assumes that players sample 7 actions made by their opponents and best respond based to that sample. Formally, choice probabilities are defined as follows:

$$p = \sum_{k=0}^{7} \binom{7}{k} q^k (1-q)^{7-k} \alpha_U(k) \text{ and } q = \sum_{m=0}^{7} \binom{7}{m} (1-p)^m p^{7-m} \alpha_L(m),$$

where $\alpha_U(k)$ is the probability with which Player 1 will choose U given *k* Ls in the sample by his opponent, and $\alpha_L(m)$ the probability with which Player 2 will choose L given *m* Us in the sample. Those are defined as:

$$\alpha_U(k) = \begin{cases} 1 & \text{if } \frac{k}{7} > \frac{c_R}{c_L + c_R} \\ \frac{1}{2} & \text{if } \frac{\mathrm{k}}{7} = \frac{c_R}{c_L + c_R} \\ 0 & \text{otherwise} \end{cases} \text{ and } \alpha_L(m) = \begin{cases} 1 & \text{if } \frac{m}{7} > \frac{d_U}{d_U + d_D} \\ \frac{1}{2} & \text{if } \frac{\mathrm{m}}{7} = \frac{d_U}{d_U + d_D} \\ 0 & \text{otherwise} \end{cases}.$$

Payoff-Sampling Equilibrium

This parametric stationary concept was introduced by Osborne and Rubinstein (1998). According to it, players are assumed to play each of their available actions for *n* (the parameter of the model) times, record their opponents' moves, and best respond to those samples. In the case of 2x2 games, suppose that k_U and k_D are the number of Ls in the two samples of Player 1, whereas m_L and m_R are the number of Us in the two samples of Player 2. Then, the probabilities with which Player 1 chooses U and Player 2 chooses R are, respectively:

$$\beta(k_U,k_D)=\begin{cases}1 & \text{if } k_U(a_L+c_L)+(n-k_U)a_R > k_D a_L+(n-k_D)(a_R+c_R)\\ \frac{1}{2} & \text{if } k_U(a_L+c_L)+(n-k_U)a_R = k_D a_L+(n-k_D)(a_R+c_R)\\ 0 & \text{otherwise}\end{cases}, \text{ and}$$

$$\gamma(m_L,m_R)=\begin{cases}1 & \text{if } m_L b_U+(n-m_L)(b_D+d_D) > m_R(b_U+d_U)+(n-m_R)b_D\\ \frac{1}{2} & \text{if } m_L b_U+(n-m_L)(b_D+d_D) = m_R(b_U+d_U)+(n-m_R)b_D\\ 0 & \text{otherwise}\end{cases},$$

Choice probabilities are defined as the expectation of the β and γ functions:

$$p=\sum_{k_U=0}^{n}\sum_{k_D=0}^{n}\binom{n}{k_U}\binom{n}{k_D}q^{k_U+k_D}(1-q)^{2n-k_U-k_D}\beta(k_U,k_D)$$

$$q=\sum_{m_L=0}^{n}\sum_{m_R=0}^{n}\binom{n}{m_L}\binom{n}{m_R}(1-p)^{m_L+m_R}p^{2n-m_L-m_R}\gamma(m_L,m_R).$$

Impulse Balance Equilibrium (IBE)

This concept of equilibrium is based on the qualitative Learning Direction Theory (LDT), proposed by Selten and Buchta (1999). According to Impulse Balance Equilibrium (Selten, Abbink, and Cox, 2005; Ockenfels and Selten, 2005), equilibrium probabilities are obtained as follows:

$$p=\frac{qc_L^*}{qc_L^*+(1-q)c_R^*} \text{ and } q=\frac{(1-p)d_D^*}{pd_U^*+(1-p)d_D^*},$$

Constants c_L^*, c_R^*, d_U^*, and d_D^* are the payoff differences the transformed game obtained, for each of the players, leaving unchanged the payoffs below or equal to the pure strategy minimax value and adding to payoffs above that value half of the surplus.

1.6 Similarity, Categorization, and Generalization

Issues of similarity, categorization, and generalization have been deeply and systematically investigated in the field of cognitive psychology in the 1970s and 1980s (Holland, Holyoak, Nisbett, and Thagard, 1986). These concepts are intimately linked, as similarity judgments about objects or events affect the way in which they are categorized (but the other way around holds true, too), and our responses, as human beings, depend upon past learning and categorization. In this vein, that of categorization can be considered as one of the most fundamental functions of all living creatures.

Categorization takes place whenever two or more stimuli are treated equivalently and this can happen in many different ways e.g., by associating to different objects the same name or by responding to different situations with the same actions. All environmental stimuli are unique, but humans (and, more in general, most of all living creatures) tend to partition them in subsets and consider as equivalent those belonging to same set (Mervis and Rosch, 1981).

Rosch (1973) and Rosch and Mervis (1975) report experimental evidence supporting the hypothesis that both artificial and natural categories are constructed around some (naturally) focal, prototypical objects in terms of degree of family resemblance, thus associating to different objects different degrees of membership. Prototypes of a category are those objects for which family resemblance with the other members of their own category is maximal, and the overlap with members of other categories is minimized. From a probabilistic point of view, prototypical objects can be defined as those items that are the best predictors of a given category or, equivalently, as those with the highest cue validity. These findings lead us to reject the Aristotelian interpretation of categories as logical and clearly bounded entities, according to which objects (once again, in the broadest meaning of the word) are unambiguously classified on the basis of the presence or absence of some specific attributes (Rosch, 1975), and that all objects are equally representative of their category.

The degree of family resemblance is not the unique variable driving categorization, as other factors such that frequency of stimuli and salience of particular attributes can significantly affect the process of prototype formation. However, empirical results presented by Rosch and Mervis (1975) support the hypothesis that family resemblance is the most important factor conditioning the way in which categories of natural and

artificial objects are created. Nosofsky (1990) provides a model that considers the joint effect of similarity and frequency on the process of category formation. His experiments on classification learning showed that classification accuracy and typicality ratings increase for objects presented with high frequency and for members of the target category that are similar to the high frequency objects, and decrease for members of the contrast category that are similar to the high frequency objects. Nosofsky's model provides a good quantitative account of the classification learning and typicality data and relies on the assumption that people learn categories by storing individual objects in memory. According to this approach, the process of category formation is based on similarity comparisons to the stored patterns, a principle embedded in many successive economic models of choice behavior based on similarity comparisons with previous experience (for example, Gilboa and Schmeidler's, 1995 *Case-Based Decision Theory*).

Nor the system of categories is arbitrarily structured. On the contrary, natural categories are highly determined as they reflect the structure of the environment, intended as the correlation in the occurrence of attributes (Rosch et al., 1976). The same authors have also shown that there exists one level of abstraction (intended as degree of inclusiveness) at which the most category cuts are made. This is the *basic level* of abstraction to which there correspond basic categories, such as *chair* and *car*; above it, there is the level of *superordinate* categories, such as *furniture* and *vehicle*, and below that one of *subordinate* categories, such as *kitchen chair* or *sports car*. Basic natural categories have been shown to be optimal in the sense that they maximize the cue validity of attributes: superordinate categories have lower cue validity because they have fewer common attributes within category, and subordinate categories have lower cue validity because they share many attributes with other categories (Rosch and Mervis, 1975). In this sense, basic category partitions are the most informative as they provide the best compromise between the need for an accurate description of the environment and the necessity to operate a reduction of the potentially infinite environmental stimuli to facilitate further processing.

Other than optimality, basic categories have been shown to possess other important properties. Indeed, at this level, given a set of objects belonging to the same category, persons use similar motor actions for interacting with them; objects have similar shapes; it is possible for people to create an overall mental representation of these objects (Rosch et al., 1976). Not only, the same contribution reports that objects are

more easily and readily recognized as members of basic categories than as members of superordinate or subordinate categories.

The importance of the role played by similarity judgments in the process of category formation is evident, since the most important factor driving categorization is family resemblance (Rosch and Mervis, 1975). As Tversky (1977:327) put it "Similarity serves as an organizing principle by which individuals classify objects, form concepts, and make generalizations".

At a first glance, it might seem that the most natural way to define and conceptualize similarity is in terms of the Euclidean distance (or one of its generalized formulation e.g., the Minkowsky *r*-metric) between two objects, represented as points in the multidimensional space of attributes (Shepard, 1962; Kruskal, 1964; Hutchinson and Lockhead, 1977). Indeed, many abstract models of similarity rely on this assumption and, indirectly, ascribe to similarity all the metric properties of a function of distance (namely, minimality, symmetry, and the triangle inequality) (Carroll and Wish, 1974; Shepard, 1974). However, contributions by Tversky (1977) and Tversky and Gati (1982) show that metric models of similarity are not adequate, providing empirical evidence that similarity judgments violate minimality, symmetry, and the triangle inequality. Indeed, recognition experiments have shown that an object is more likely to be recognized as another one rather than as itself, thus violating minimality. Moreover, similarity assessments have an intrinsic asymmetric nature: a statement of the form "*a* is like *b*" is not equivalent to "*b* is like *a*". Experimental results show that, given two objects *a* and *b*, we tend to select as *referent* (i.e., the object of the sentence) the most prototypical or salient of the two, and the other, less salient, as the *variant* (i.e., the subject of the sentence). As an example, we say that "an ellipse is like a circle" and not that "a circle is like an ellipse". It is also important to note that asymmetries in similarity have been observed not only in *comparative* tasks, but also in *production* tasks (e.g. pattern recognition and stimulus identification) in which a subject is given a stimulus and is asked to respond with the most similar response. We conclude that similarity is asymmetric, that the direction of the asymmetry is determined by the relative salience of stimuli, and that we generally choose as referents the most prototypical stimuli. As for the triangle inequality, it cannot be so easily tested, as it cannot be expressed in ordinal terms (contrary to minimality and similarity features). However, if *a* and *b* are *quite* similar to *c*, then it must be the case that *a* is not so dissimilar from *c*. For an example, we can say that Jamaica is similar to Cuba because

of geographical proximity and that Cuba is similar to China because of their common political regime; on the other hand, Jamaica and China are very dissimilar from one another. This example suggests that similarity is not transitive and that the perceived distance between Jamaica and China much exceeds the sum of the distances between Jamaica and Cuba and Cuba and China (thus violating the triangle inequality). Tverky (1977) proposes a model of similarity (called *contrast model*) that accounts for experimental data, according to which objects (or more in general stimuli) are described as a set of features and similarity is described as a process of feature matching.

Now, let us analyze more in dept the interplay between similarity and categorization and the effects of a change of the context on similarity. As said, similarity judgments are not symmetric but depend on the salience of some attributes and objects/stimuli are grouped so that similarity is maximized within and minimize between categories. Tversky (1977) argues that attribute salience has two components: *intensity* and *diagnosticity*. The first has to do with the frequency with which an attribute occurs and is rather stable across contexts, the second with the frequency with which an attribute is employed as a criterion of classification and varies across contexts. The effects of context on similarity can be explained in terms of a change in the diagnostic salience of attributes induced by different groupings of objects, and have been verified experimentally; given a set of objects, the addition or deletion of some objects alters the diagnostic salience of the attributes of the remainder objects, leading to a corresponding change in the perception of their similarity. It follows that if classification is determined by similarity among objects, it is also true that the similarity among objects depends on how they are grouped. This means that there is a bidirectional relationship between categorization and similarity, in the sense described above.

Tversky's *contrast model* of similarity is quite important not only because it accounts for empirical data, but also for the fact that it unifies under the same framework the intimately linked concepts of similarity, family resemblance, and prototipicality, interpreting them as linear combinations of the measures of the sets of common and distinctive features.

Generalization is defined as the tendency to react in the same way to stimuli that are similar (but not equal) to a stimulus experienced in the past. In other words, a response conditioned to one stimulus tends also to occur to other stimuli, and the

correlation stimulus-response is function of the degree of similarity between that stimulus and that one to which the response was originally conditioned. As Shepard (1958:242) puts it "The principle of stimulus generalization is of such fundamental importance that any quantitative theory of behavior that fails to deal with it explicitly can only be regarded as incomplete". It is evident from its definition that generalization is intimately linked with similarity.

The legitimacy of the statement "All objects *A* have property *B*" based on the knowledge that some observed objects *A* have property *B*, has been discussed and questioned, since Aristotle, by many logicians and philosophers.

In particular, instance-based generalization can be defined as a process of inductive inference by which we add new rules to existing concepts (Holland et al., 1986). According to these authors, generalization can be made about abstract categories (of the kind "If X is a dog, then X barks") and individuals (of the type "If you do this, he will do that"). In both cases, generalization produces the expectation of certain properties of the object or individual; however, whereas these properties are rather stable (in terms of variability) for objects, the behavioral properties of individuals can be quite instable.

The question that arises is then what are the factors that lead us to consider our generalizations as valid or, more precisely, what are the factors we consider to assess the acceptability of sentences of the type "every object (in the broadest sense of the term) *F* has property *G*". Holland et al. (1986) argue that these factors are two. The first is the number of items *F* that have been observed to have property *G*. The second is our knowledge about statistical properties of the population about which we are generalizing a concept; if we know that *F* and *G* are highly invariant across the population, the generalization from few instances will be legitimate; on the contrary, if we know that *F* and *G* exhibit high variability, then we will consider our generalization as acceptable only after having observed a greater number of instances. This principle of acceptability can also be expressed equivalently saying that acceptability of generalization is proportional to the cue validity of observed objects. However, we do not have to forget that this cue validity depends upon the category that we select as *reference class* to assess the degree of variability of the objects in consideration. This last point shows also that organization of categories has a direct influence on generalization.

1.7 Modeling Categorization and Generalization with Neural Networks

From 1943, when McCulloch and Pitts first introduced their neuron model, neural networks have always been intended and used as tools for classifying data. As Hertz, Krogh, and Palmer (1991:9) point out "The reason for much of the excitement about neural networks is their ability to generalize to new situations". The term *neural networks* refers to a large family of models that implement a computational paradigm alternative to the usual one, taking inspiration from neuroscience and the structure of *real* neurons in the brain. McCulloch and Pitts's model is a computationally powerful device, as it has been shown that an assembly of such artificial neurons is capable in principle of *universal computation* i.e., it can perform any computation that an ordinary digital computer can.

I will briefly illustrate here two ways for modeling the processes of category formation and generalization with neural tools. Detailed descriptions about these models can be found in Hertz, Krogh, and Palmer (1991), Bishop (1995), and Ripley (1996).

The first is that followed by the model of *associative memory* proposed by Hopfield (1982). This model gives a solution to the following problem:

> Store a set of patterns in such a way that when presented with a new pattern *P*, the network responds by producing whichever one of the stored patterns that most closely resembles *P*.

Hopfield provides a procedure that allows to store in the connection weights of a network some binary vectors. These vectors (also referred to as *patterns*) are generally interpreted as binary codes of the possible true states the world can assume (or some ideal prototypical objects); in a pattern, each component (*bit*) corresponds to a feature of that state or object, and the value of each bit can be interpreted as the presence/absence of that feature. In this kind of networks, units can then assume only binary values, and the *state* of a network is defined as a particular configuration of values of its units.

Whenever a network is fed with an arbitrarily chosen input, after an iterated process of updating of its unit activation states that minimizes a measure of energy associated

to network states (*relaxation*), it produces as an output the stored pattern that most resembles that particular input (usually in terms of Hamming distance). It is worth noting that connection weights remain unchanged and what changes is the activation state of units. Here inputs are commonly interpreted as noisy stimuli from the environment; according to this view, the task is then that of reconstructing the right true state from the information contained in the stimulus.

From another perspective, in the space of all possible states of the network (called *configuration space*), stored patterns behave as attractors. The dynamics of the system carries starting point (inputs) into one of the attractors (the response).

This approach of modeling associative memory is very similar to Gilboa and Schmeidler's (1995) *Case-Based Decision Theory*, which seems rather a particular case of the Hopfield model.

The main limitation of the Hopfield's model is that classes are constructed around prototypes or true states that are known a priori i.e., the patterns stored in a network. This assumption is quite strong and is reasonably applicable only to a restricted class of real world situations. However, there are neural models according to which networks can be thought a classification task and generalize this knowledge to never seen before mapping tasks.

As explained in Hertz, Krogh, and Palmer (1991), according to the procedure called *supervised learning* (or *learning with a teacher*), network outputs are compared with known correct answers, and networks receive feedback about any errors. Usually, networks with separate input and output units are considered. More specifically, networks can be taught to operate classification tasks on some representative pairs of input-target (*training set*), changing their connection weights in response to each training pair as to minimize the difference between networks' and desired outputs (Rumelhart, Hinton, and Williams, 1986a and 1986b; McClelland, Rumelhart, and the PDP Research Group, 1986). What a neural network learns is stored in its connection weights, and this *knowledge* can be used to deal with new classification tasks for which, in most of the cases, an *a priori* solution is not known. The association task learnt with supervised learning is more general than that in the associative memory problem; there we wanted the stored patterns to reproduce themselves when used as inputs (*auto-association*), in contrast to the *hetero-association* task, in which output patterns differ from input ones.

Supervised learning is particularly straightforward in the case of networks with a particular structure i.e., *layered feed forward networks*. These models were called *perceptrons* when they where first studied in detail by Rosenblatt (1962).

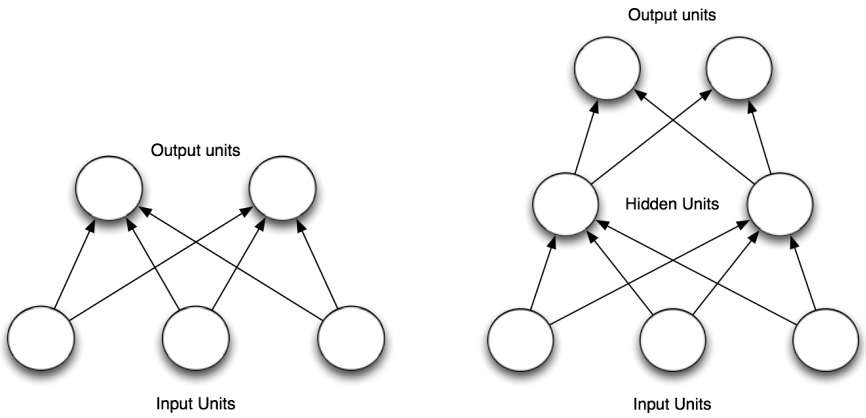


Figure 2. Perceptrons. On the left hand side, a simple perceptron, which has (by definition) only one layer of connections. On the right hand side, an example of two-layer perceptron.

Figure 2 shows two examples of percetrons; there is a set of input units whose role is that of feeding the network with external stimuli and on which no computation is performed. After this, one or more intermediate layers of units can come (called *hidden units*), followed by a final layer of output units that yield the result of the computation. From a unit, there are neither connections pointing to units in previous layers, nor to other units in the same layer, nor to units in more than one layer ahead. In virtue of the described structure, connections in feed-forward networks are asymmetric i.e., all connections are unidirectional. This fact is of great importance, as it implies that, in general, the existence of an *energy function* defined over network states is not guaranteed; only symmetric connections guarantee the existence of such a function – as in the Hopfield's model, wherein connections are all bidirectional and the relaxation process relies on energy function minimization.

Perceptrons are particularly powerful models and can virtually learn any classification task, as two-layer networks have the important property that they can approximate arbitrarily well any functional (*one-to-one* or *many-to-many*) continuous mapping from one finite-dimensional space to another, provided the number of hidden units is sufficiently large (see Bishop, 1995).

One-layer feed-forward networks are called *simple perceptrons*. They have a layer of input units, one layer of output units, and a layer of connections in between. Therefore, there are not hidden layers. If output units are labeled with y_i ($i = 1,2,...,n$), output units with x_j ($j = 1,2,...,m$), and connection weight from input unit *j* to output unit *i* with w_{ij}, then the computation of outputs is simply:

$$y_i = g(h_i) = g\left(\sum_j w_{ij} x_j\right),$$

where $g(\cdot)$ is the *activation function* computed by the units, sometimes referred to also as *gain function* or *squashing function*. Usually $g(\cdot)$ is taken to be non linear and differentiable (in which case we have continuous output units), but also threshold functions (binary output units) and linear functions (linear output units) are often used.

Supervised learning is implemented in the context of simple perceptrons as follows. In response to each training pair, the following error or cost function is evaluated:

$$E(w) = \sum_{i,\mu}\left[t_i^\mu - y_i^\mu\right]^2 = \sum_{i,\mu}\left[t_i^\mu - g\left(\sum_j w_{ij} x_j\right)\right]^2, \quad (1)$$

where $t^\mu = \left(t_1^\mu,...,t_i^\mu,...,t_n^\mu\right)$ represents the target vector ($\mu = 1,2,...,N$). Function (1) provides a measure of the divergence between the output provided by the network and the desired output (*target*), and depends upon weights. The gradient descent algorithm provides a procedure that allows us to find a set of weights, which produces exactly the desired outputs from each input pattern, by successive improvement from a point arbitrarily chosen. Specifically, this procedure suggests changing all weights by a quantity Δw_{ij} given by:

$$\Delta w_{ij} = -\eta \frac{\partial E}{\partial w_{ij}}, \quad (2)$$

where the parameter $\eta \in (0,1)$ (called *rate of learning*) tunes the speed of learning i.e., how rapidly the network adapts. If we consider the error function (1), then we can write:

$$\frac{\partial E}{\partial w_{ij}} = -\sum_\mu \left[t_i^\mu - g\left(h_i^\mu\right)\right] \cdot g'\left(h_i^\mu\right) \cdot x_j^\mu,$$

leading to

$$\Delta w_{ij} = \eta \cdot \left[t_i^\mu - y_i^\mu\right] \cdot g'\left(h_i^\mu\right) \cdot x_j^\mu. \quad (3)$$

A sufficient condition, although not necessary, for the network to be able to learn successfully the association tasks in the training set is linear independence of input patterns. However, assuming that a solution exists, the gradient descent algorithm might not be able to converge to it; if for example the targets lie outside the range of $g(\cdot)$ (e.g., ± 1 targets with $g(h) = tahn(h)$), the cost function might have local minima besides the global one at which $E = 0$. The gradient descent can then become stuck in such a minimum. In order to overcome this problem, some simulations techniques have been developed (e.g., *simulated annealing*).

Nor the quadratic cost function (1) is the only possible one. Other error functions have been proposed in the literature. The relative entropy function has received particular attention (Kullback, 1959; Hopfield, 1987; Baum and Wilczek, 1988), and is defined as follows:

$$E(w) = \sum_{i,\mu}\left[\frac{1}{2}\left(1 + t_i^\mu\right)\log\frac{1 + t_i^\mu}{1 + y_i^\mu} + \frac{1}{2}\left(1 - t_i^\mu\right)\log\frac{1 - t_i^\mu}{1 - y_i^\mu}\right]. \qquad (4)$$

This error function can be naturally interpreted in terms of learning the correct probabilities of a set of hypothesis represented by the output units; provided that the range of the activation function $g(\cdot)$ is $(-1,1)$, then $\frac{1}{2}\left(1 + y_i^\mu\right)$ can be interpreted as the guess of the network, and $\frac{1}{2}\left(1 + t_i^\mu\right)$ as the correct probability.

The use of the entropy cost function (4) has been shown to solve some learning problems that cannot be solved through the use of the quadratic cost function. Moreover, using (4) as a measure of error and taking $g(h) = \tanh(\beta h)$ then,

$$\Delta w_{ij} = \eta \cdot \beta \cdot \left[t_i^\mu - y_i^\mu\right] \cdot x_j^\mu. \qquad (5)$$

As we can see, (4) is equal to (5), except for the term $g'\left(h_i^\mu\right)$. The updating rule (5) is identical to the rule that can be derived for linear output units i.e., when the activation function is of the form $g(h) = h$.

Marchiori and Warglien (2008), in defining their Perceptron-Based (PB) learning model (which I describe in Chapter 2), multiply the updating rule (5) by a term of regret; this fact adds further insights and meaning to the fundamental assumptions and mechanisms of this model of learning.

1.8 Methodological Appendix

Erev and Haruvy (2005) note that literature on learning model comparisons has provided contrasting results on which model best fits and/or predicts empirical data. This is the consequence of the fact that the methodological approaches adopted are different and serve different purposes. Specifically, contributions by Erev and Roth (1998) and Sarin and Vahid (2001) show that simple reinforcement learning models provide the best approximation of empirical data. On the opposite, Camerer and Ho (1999) suggest that the model of learning that best approximate data is the EWA model, a hybrid model merging reinforcement learning and beliefs learning. Along a third line of research, Stahl (1999) shows that a simple logit best reply model with inertia and adaptive expectations outperforms both EWA and reinforcement models.

The above-mentioned results appear to contradict each other. However, this is consequence of two facts: first, the methodologies adopted are different and, second, models of learning are intrinsically misspecified.

As for the methodologies adopted, they can be grouped into two main classes: *one-period-ahead* and *T-period-ahead* techniques. The former class of techniques is focused on *within-game* predictions i.e., observed information from past trials is used to predict the behavior in the next period. Accordingly, the following likelihood function is maximized:

$$L(\theta) = \prod_{i=1}^{N}\prod_{t=1}^{T} p_{it}\left(x_{it} \mid x_1,...,x_{t-1},\theta\right),$$

where i indexes players, t time periods, x_t is the choice of player i at time t, θ is the vector of parameters, and p_{it} is the estimated probability of choice on period t. Some authors allow for different set of parameters in different games (Camerer and Ho, 1999), whereas some others (Cheung and Friedman, 1998; Stahl, 1999) suggest a single parameter set for all games.

According to the *T-period-ahead* techniques, researchers simulate the entire path of interaction and compare it with the observed one. It is then evident that this approach heavily relies on simulations. Usually, the best model is that one that most accurately predicts observed data in terms of *Mean Squared Deviation* (MSD) rather than in terms of likelihood. Indeed, in this case, likelihood estimation would require the computation of the following (t-1)-fold integral:

$$L(\theta) = \prod_{i=1}^{N} \int \ldots \int f_{it}(x_1,\ldots,x_{t-1},x_{it},\theta) dx_1 \cdot \ldots \cdot dx_{t-1},$$

where $f_{it}(\cdot)$ id the density function of choices for player *i* at time *t*. On the one hand, the calculus of the integrals above would be computationally infeasible, and, on the other, MSD can be shown to have nice properties not shared with other scoring rules (Selten, 1998).

One might consider the T-periods-ahead technique as rather inefficient, as information on past periods of play is not taken into consideration. Nonetheless, this approach can be successfully adopted to predict data in those cases in which there are no previous available observations for that particular game or class of games. Contributions by Roth and Erev (1995), Erev and Roth (1998), and Sarin and Vahid (2001) insist on a single set of parameters for all games when adopting this kind of analysis.

If models were well specified, then the two described techniques would provide the same ranking over models. However, this is not the case as models, in general, provide only an approximation of phenomena and, perhaps more important, it has been shown that subjects are sufficiently heterogeneous that just pooling them together can be a source of model misspecification. Moreover, misspecification can arise also considering a unique set of parameters across different games (this aspect concerns mainly the *new-game* analytical approach). Due to model misspecification, maximizing one-period-ahead likelihood might be different than minimizing MSD, then resulting in different rankings of models. Although all models of learning are likely to be misspecified (particularly those which assert the same set of parameters for different games), they can nonetheless provide some useful approximation of observed behavior. Indeed, in certain classes of games, it has been shown that *new-game* predictions of some learning models are more accurate than those derived using Nash equilibrium (see Section 2).

Erev and Haruvy (2005) show that differences in the responses of *new-game* and *within-game* approaches are mainly due to action inertia i.e., the tendency of players to repeat past actions – independently from their beliefs. Clearly, in those cases in which the inertial component of behavior is important, *within-game* analyses based on the one-period-ahead technique will favor models that take inertia (either explicitly or implicitly) into consideration, as in the case, for example, of the EWA model. On the

opposite, the informative value of inertia is negligible within the *new-game* analytical framework, as individuals' past history of play is not considered. In this case, the analysis is likely to favor reinforcement models, as shown in Erev and Roth (1998), Erev et al. (1999), and Sarin and Vahid (2001).

The conclusion is that there does not exist the "right" procedure to compare models of learning and that different methodological approaches can lead to different rankings of the models. However, noting that different analyses serve different purposes, the contradictory results proposed in the literature can be interpreted as "a result of ignoring the effect of the type of available information on a model's success" (Erev and Haruvy, 2005:369).

Individual Versus Pooled Data-Driven Predictions

As said in the previous paragraph, generalizing model parameters across games can be a source of model misspecification. On the other hand, as noted in Erev and Haruvy (2001), also pooling subjects together and describe their behavior, within the same game, with the same set of parameters can produce serious forms of model misspecification; indeed, there is no valid reason to exclude a priori that different subjects behave differently in the same strategic situation, and that these differences in behavior are not due to random factors, but rather to systematic characteristics of subjects.

However, introducing agents' heterogeneity, models "can easily get out of hand and lose robustness" (Erev and Haruvy, 2001:4), with the final result that in spite of the evidence against agents' homogeneity (Cheung and Friedman, 1997; Camerer and Ho, 1998; Busemeyer and Stout, 2002), most part of the analyses proposed adopt the parsimonious approach of describing behavior of all agents with the same set of parameters.

Scoring Rules

Scoring rules provide a measure of divergence between observed data and data estimated with the use of some probabilistic model that fully specifies a probability distribution over outcomes or actions. More specifically, in the case of repeated games, a score is computed in each period of play on the basis of observed actions and

estimated probabilities; eventually, scores are summed up yielding a measure of model performance.

The two concepts of *distance* and *scoring rule* are different but intimately related, as it will be made clear later. Before giving the definition of scoring rule some notation is in order. Let X be a random variable with *cumulative distribution function* (cdf) F and *probability density function* (pdf) f defined over the range $D \subseteq R$. A scoring rule is a real valued functional (possibly equal to $-\infty$) $S(g,x)$ defined for all densities g in D (Friedman, 1983). In other words, a scoring rule provides an assessment of the quality of the forecast in terms of a real number (possibly $-\infty$), based on the estimated, or hypothesized, density g and the realization x of the random variable X (whose pdf f is obviously unknown to the researcher). A scoring rule is said to be *proper*, or *incentive compatible*, if the expected value:

$$E_f[S(g)] = \int_D S(g,x) f(x) dx,$$

is maximized on D at $g = f$. Incentive compatibility implies that the correct theory is the only one that obtains the highest score and is a minimal requirement for a scoring rule. A scoring rule is said to be *effective* with respect to distance d if for all densities f, g, and h in D we have:

$$E_f[S(g)] > E_f[S(h)] \Leftrightarrow d(f,g) < d(f,h).$$

Effectiveness establishes a precise relationship between the scoring rule $S(g,x)$ and distance d; the expected score is a monotone decreasing function of the distance between the true and the estimated distribution. It is worth noting that if $S(g,x)$ is effective, then it is also proper.

Let us now consider the measure d_2 defined as:

$$d_2(f,g) = \left\{ \int_D |f(x) - g(x)| dx \right\}^{1/2},$$

which in the case of discrete distributions assumes the well known form:

$$d_2(f,g) = \left\{ \frac{1}{n} \sum_{i=1}^{n} (p_i - q_i)^2 \right\}^{1/2}.$$

The norm of order two of a function f defined over range D is:

$$\|f\|_2 = \left\{ \int_D |f(x)|^2 dx \right\}^{1/2}.$$

Now, it can be proved that the quadratic scoring rule, defined up to a linear transformation, $Q(g,x) = 2g(x) - \|g\|_2^2$ is proper and effective with respect to distance d_2 above defined. As an example, the linear scoring rule $N(g,x) = g(x)$ is not proper and effective with respect to any measure, whereas the logarithmic one $L(g,x) = \log g(x)$ has shown to be proper.

The family of *Mean Square Deviation (MSD)*-based rules (whose elements differ up to a linear transformation) satisfies some nice properties other than those of *incentive compatibility* and *effectiveness* above mentioned (Selten, 1998).

The Information-Theoretic Approach

MSD can be used as a criterion for comparing models, provided that the number of parameters of the models in consideration is the same; in other words, MSD provides a very good measure of data fitting, but it cannot be interpreted as a measure of model predictive power and robustness, as more complex models are intrinsically favored. Indeed, an arbitrary increase in the number of parameters cannot correspond to a decrease in the accuracy of fit. Hence, if the analysis is focused on measuring model predictive power, other methods have to be adopted that explicitly penalize model *statistical* complexity.

In the *new-game* analytical framework, one of the possible alternatives is to adopt the *information-theoretic* approach. According to this approach, models are compared on the basis of Akaike's Information Criterion (AIC) defined as follows (Akaike, 1973; Burnham and Anderson, 2003):

$$AIC = -2\log\left(L\left(\hat{\theta} \mid data\right)\right) + 2K,$$

where $\log\left(L\left(\hat{\theta} \mid data\right)\right)$ is the log-likelihood evaluated at the MLE $\hat{\theta}$ of the true parameter (or parameter set) θ, given the observed data y, and K is the number of estimable parameters. A second order information criterion for small samples is also defined (see Burnham and Anderson, 2003 for further details). In the case of *Least Squares* estimation, under the assumption that residuals are normally distributed with constant variance, then AIC can be easily calculated as:

$$AIC = n\log\left(\hat{\sigma}^2\right) + 2K,$$

where,

$$\hat{\sigma}^2 = \frac{\sum \hat{\varepsilon}_i^2}{n},$$

is the MLE of σ^2, and $\hat{\varepsilon}_i$ are the estimated residuals for a particular model. In this particular case, K must include also σ^2 and all other parameters. Despite its simplicity, this approach does not seem to be applicable to measure the performance of models of learning in games. Indeed, computing AIC to obtain an overall assessment of predictions over all games necessarily implies to pool data from different experiments and hypothesize that residuals are normally distributed with common variance; however, there is no valid reason for considering the variance of residuals as constant across different games.

Cross Validation and Generalization Criteria

The cross validation criterion was first formalized by Mosier (1951) and then subsequently elaborated by a number of researchers. The essential idea is that of dividing the total sample of N observations into two independent subsamples of sizes N_1 and N_2. During the *calibration stage*, model parameters are estimated as to minimize the discrepancy between predictions and observed statistics based on N_1 (called *in-sample data*). In the *validation stage*, the estimated parameters are used to make predictions over statistics based on the second set of observations N_2 (*out-of-sample data*). The model that best performs in this second stage is to be preferred. This process can be repeated using different divisions of the total sample into *in-sample* and *out-of-sample* data.

As noted in Busemeyer and Wang (1999), the usefulness of this criterion is limited to the case in which the sample size is small because as the sample size increases, the target statistics from in-sample and out-sample data tend to the same value; in this case, a lower discrepancy in the calibration stage is very likely to produce a small discrepancy in the validation stage as well.

The Generalization Criterion for model comparison was first formalized by Busemeyer and Wang (1999) and is quite similar to the cross-validation procedure. Whereas cross-validation employs data from the *same* design for both the calibration and the validation stages, generalization employs data from an entirely *new design* for the validation stage. The latter criterion distinguishes for the emphasis placed on

extrapolations to new experimental conditions in the second stage, and is useful also when the sample size is large.

However, the use of Busemeyer and Wang's criterion does not seem appropriate in the context of learning model comparison. Specifically, for a meaningful use of the generalization criterion, the different conditions (i.e., games) should be randomly drawn; on the opposite, a typical compound dataset groups together data from different experiments in which games are usually chosen ad hoc, reflecting the different purposes of experimenters. For this reason, it is not clear at all how to partition datasets in the two subsamples for calibrating and validating models.

The Equivalent Number of Observations (ENO) Measure

ENO was first proposed by Erev et al. (2002) to measure model predictive accuracy of choice behavior in experiments on repeated games. The idea proposed by the authors is that rather looking at whether a theory can or cannot be rejected based on the data, we should ask ourselves how good is the approximation of the data provided by that theory. One intuitive way to formalize this concept is then to determine the number of empirical observations that are needed to provide a prediction as accurate as that of the model. This number is called *equivalent number of observations* (ENO) (Erev et al., 2007). Therefore, the higher the ENO, the better the model. ENO is defined as follows:

$$ENO = \frac{S^2}{\left(M - S^2\right)},$$

where S^2 is the pooled variance (across games in the experiment), and M is the mean square error associated to the model.

I do not use ENO in my analyses because in order to compute S^2 and M, data for each independent observation are needed. For the datasets I consider in Chapter 2, individual data are not available in all cases, whereas for the datasets I consider in Chapter 3, individual data are available. Nonetheless, I adopt the approach based on MSD Prediction scores (see the method section of Chapter 2), as it can be applied in both situations (with or without data on individual observations), thus providing a uniform methodology for evaluating and comparing model performances.

CHAPTER 2

2. PREDICTING HUMAN BEHAVIOR BY REGRET-DRIVEN NEURAL NETWORKS

Abstract. The surge of interest in the neural bases of economic behavior raises the question of how well neural networks can model human interactive decision-making. Experimental game theory has provided a large set of laboratory data on human interactive learning in repeated games, often contradicting the predictions of standard game theory and justifying the search for new explanatory models. Here, I use datasets from 35 experiments on repeated games with a unique equilibrium in mixed strategies to compare the descriptive and predictive performances of the Perceptron-Based (PB) learning model (Marchiori and Warglien, 2008) with some of the most popular learning models in the behavioral game theory literature. As a result, the PB model turns out to be the best predictor of empirical data with respect to all other models of learning, with the exception of a model proposed by Ert and Erev (2007), similarly based on regret.

2.1 Models of Learning

Standard game theory does not provide a theory of learning and is limited to describing a steady state situation. On the contrary, experimentally observed behavior provides overwhelming evidence of the existence of a process, commonly known as *learning*, after which past experience dramatically affects subjects' current strategic choices. Specifically, interactive learning differs from individual learning in that given *N* agents, each agent adapts to a strategic environment which is continuously modified by the concurrent learning of the other *N-1* agents.

Learning models try to replicate artificially the process in which past experience affects agents' current behavior; more specifically, they establish how the probabilities with which future actions will be chosen are affected by information about the outcomes produced by actions chosen in the past. In order to do this, many quantitative theories assume that, for a player, all his possible actions are associated with numerical evaluations, called *attractions* or *propensities* (these two terms will be used interchangeably), which are mapped, according to opportune rules, into choice probabilities. Propensities can be interpreted as a measure of the propensity of a player to choose the actions they are associated with, while learning rules determine how these attractions are updated in response to past experience.

There is a wide variety of different approaches for modeling learning (for a comprehensive review of these models and theories see Camerer, 2003), but the most successful learning theories proposed so far are those of *reinforcement learning*, *beliefs learning*, hybrid models combining both (Ho, Camerer, and Chong, 2007), and, finally, theories which emphasize the role of post-decision regret as the driver of human behavior (Erev et al., 1999; Ert and Erev, 2007).

2.2 The PB Model

Departing from the mainstream behavioral game theory literature on learning, Marchiori and Warglien (2008) propose a model of interactive learning that embeds the basic principles of Learning Direction Theory (LDT) (Selten and Stoecker, 1986), and translate them into a neural network-based model which is potentially more flexible than the traditional "attractions and stochastic choice rule" models, because it is responsive to the structure of the game and capable of modeling the transfer of learning to new games.

The basic assumption of the Perceptron-Based (PB) model is that learning is driven by a sort of ex-post rationalizing process: individuals modify their behavior by looking backward to what might have been their best move, once they know their opponents' moves. They adjust in the direction of such ex-post best response, and it is assumed that the intensity of such directional change is proportional to a measure of regret – how much they have missed by not playing that move. This is consistent with recent neuroscience research on individual decision making, showing that regret affects learning, and that both neuro-physiological and behavioral responses to the experience of regret are correlated to its amplitude (Coricelli et al., 2005 and Daw et al., 2006).

The PB model consists of a *simple analog perceptron* (i.e., a one-layer feed-forward neural network) fed back by a measure of regret for foregone payoffs. Input units (labeled with x_j) receive relevant information about the structure of the strategic environment (i.e., game payoffs) and propagate such information to higher-level units in the network. Output units (labeled with y_i) are in a one-to-one mapping with the elements in each agent's set of actions and have a sigmoid (hyperbolic tangent) activation function. The activation state of output units represents the propensity (or attraction) demonstrated by an agent to play the corresponding action. Propensities are turned into choice's probabilities by a simple normalization.

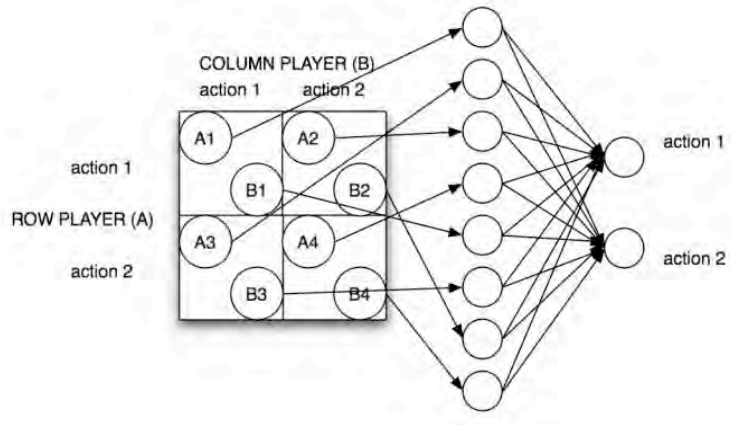


Figure 1. How strategic information (the payoff matrix) is mapped into neural net structured-agents in the PB model.

For an agent, the propensity of playing its action *i*-th, is given by:

$$y_i = \tanh\left(\lambda \sum_j w_{ij} x_j\right), \qquad (1)$$

where w_{ij} is the weight of the connection from the *j*-th input unit to the *i*-th output unit. For a generic *k*-th player, propensities' updating rule is given by:

$$w_{ij}^{t} = w_{ij}^{t-1} + \Delta w_{ij}, \quad (2)$$

with:

$$\Delta w_{ij} = -\lambda^{2}\left[t_i\left(a^{-k}\right) - y_i\right]R^{k}\left(a_i^{k}, a^{-k}\right)x_j. \quad (3)$$

In (1) and (3), λ is the unique free parameter of the model; $t_i\left(a^{-k}\right)$ is the *ex-post* best response to other players' a^{-k} action profile; the term $R^{k}\left(a_i^{k}, a^{-k}\right)$ represents the *regret* of player *k*-th given the action profile $\left(a_i^{k}, a^{-k}\right)$; x_j (the activation state of *j*-th input unit) could be interpreted as the strength of the input to the node (*payoff saliency factor*). Regret is defined as the difference between the maximum obtainable payoff given other players' actions and the payoff actually received. Thus, the psychological intuition underlying (3) is that connection weight adjustment is driven by a series of factors that can be summarized as follows:

Adjustment = Learning rate × Distance from ex-post best response × Regret × Input saliency

It is important to note that in the PB model past experience affects future attractions only indirectly, through changes in connection weights. No explicit track of past experiences needs to be kept and it is indirectly store in the configuration of the weights of connections.

The free parameter λ has two roles in the model: it determines, in (3), agents' learning rate and, in (1), the steepness of the activation function in a neighborhood of the origin. Marchiori and Warglien (2008) also propose a zero-parameters version of the model, called PB0, where λ is, in (1) and (3), replaced by a deterministic function. The value of this self-tuning function is defined at each time-step as the ratio between the actual cumulated regret and the maximum cumulated regret, as follows:

$$\lambda_t^{k} = \frac{\sum_{t} R_t^{k}}{\sum_{t} \max\left(R_t^{k}\right)}, \quad (4)$$

where t is the number of iterations, R_t^k is the actually experienced regret by player k at round t, and $\max\left(R_t^k\right)$ is the maximum possible regret player k could experience at time t (of course in a repeated game this value is constant).

A highly simplified example will clarify the mechanics of the model. To make things easier, I will consider a network with only two inputs and two output nodes.

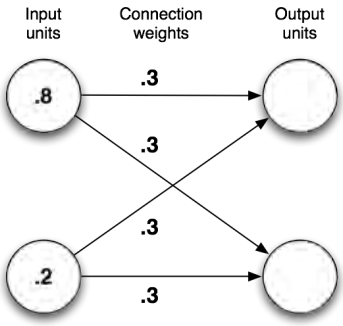


Figure 2. An example of how the PB model works. Time step 0: initialization of agents.

For simplicity, let's start from an initial state in which all connection weights are equal to 0.3 (usually, they are initialized randomly). In the first run, output units are activated, assuming values that are the sum of the inputs, weighted by connection weights, and transformed by the hyperbolic tangent function (here we assume $\beta = 1$). The activation state of both output units will be $\tanh(0.24 + 0.06)$, that is 0.291. In practice, this implies that both actions will be played with equal probability after normalization.

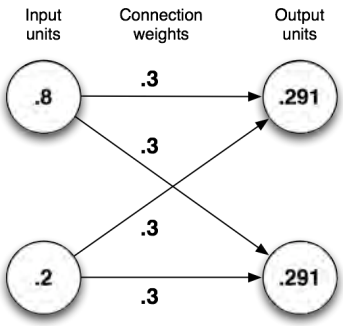


Figure 3. An example of how the PB model works. Time step 0: calculation of attractions.

Imagine that the network plays the "low node" action. It turns out to be the wrong move, and a regret of 0.6 is experienced. After the updating, weights will be as in figure 4.

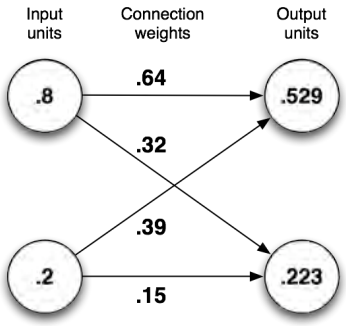


Figure 4. An example of how the PB model works. Time step 1: ex-post rationalization.

The output vector (0.529, -0.223), after normalization, implies a 0.667 probability to play the "high node" action.

Notice that changing the input weights will change the learning trajectory even if the inputs always repeat themselves. Once more, a simple numeric example will clarify the point. Consider the following network, which is identical to that one in figure 2 and against the same environment, except for the input vector, which has now been modified to (0.5,0.5). The initial output will be exactly the same:

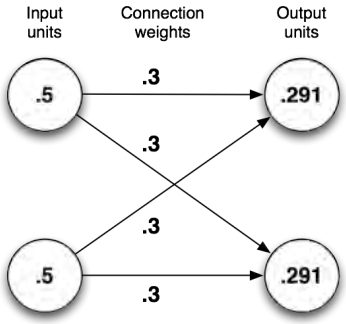


Figure 5. An example of how the PB model works. Time step 0: initialization of agents with new inputs.

If the network plays also in this case the "low node" action and receives a same amount of regret 0.6, the network will change as follows:

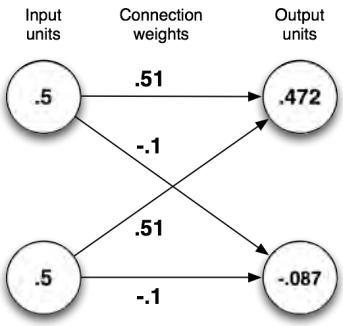


Figure 6. An example of how the PB model works. Time step 0: calculation of attractions.

which, after normalization, will imply a 0.617 probability to play the "high node" action.

Will differences in learning paths persist in the long run, and with full-fledged networks? I conduced a further analysis over the 10 Erev et al.'s (2007) games, comparing the learning trajectories of pairs of networks with complete and correct payoff inputs with those of pairs of networks with "flat" inputs (e.g., representing all inputs as 0.5), which is equivalent to a simpler "attraction and choice rule" architecture. All the rest was kept the same. I observed that in general the two versions produce average behaviors that are significantly different. Moreover, the more diverse the payoffs, the more the trajectories tend to differ (up to a 4% difference in the predicted frequencies of play). Therefore, the Perceptron could be well approximated by a more conventional learning architecture only in the cases in which all payoffs are similar enough.

A theoretical issue remains open: is it possible to obtain an explicit form the error function *E* that changes in weights defined in (3) try to minimize? In order to do that, the differential equation:

$$\Delta w_{ij} = -\lambda \frac{\partial E}{\partial w_{ij}} = -\lambda^2 \left[t_i\left(a^{-k}\right) - y_i \right] R^k\left(a_i^k, a^{-k}\right) x_j, \quad (5)$$

may be integrated, and we obtain the following:

$$E = \lambda \int \left[t_i\left(a^{-k}\right) - y_i \right] R^k\left(a_i^k, a^{-k}\right) x_j \partial w_{ij}. \quad (6)$$

Now, the question that arises is that whether or not in equation (6) it is possible to consider the term $R^k\left(a_i^k, a^{-k}\right)$ independent of w_{ij}. In the affirmative case, we could easily derive the explicit analytical form of *E*. Yet, this does not seems to be possible

because the regret term $R^k\left(a_i^k, a^{-k}\right)$ depends on the profile of actions that in its turn depends on the distribution of probability over actions, and then, although indirectly, on w_{ij}.

2.3 Methods

Drawing from the approach pioneered by Erev and Roth (1998), for each model I determine three different measures of data fitting and prediction, based on data from experiments on 35 different, repeatedly played games (described in the next section). All three types of scores are based on Mean Squared Deviation (MSD) and are: *By Game*, *Best Fit*, and *Prediction scores*.

MSD is a suitable way to measure the divergence between estimated and observed vectors of choice frequencies (Selten, 1998; Erev and Roth, 1995; 1998). Labeling with y the vector of observed choice frequencies (of length N) and with $y'(\theta)$ the vector of estimated choice frequencies, given the parameter configuration $\theta = (\theta_1, ..., \theta_k)$, $MSD(\theta)$ is defined as follows:

$$MSD(\theta) = \frac{\sum_{i=1}^{N} \left(y_i'(\theta) - y_i\right)^2}{N}.$$

By Game scores are obtained by selecting the lowest value MSD for each dataset, whereas *Best Fit* scores correspond to the performance of the models obtained when the parameters minimize MSD across all datasets. Finally, *Prediction scores* are computed according to the *leave-one-out* estimation procedure: for each experimental dataset, the remaining datasets are used to estimate the free parameter values that minimize MSD over the remaining datasets. The ensuing parameter estimates are subsequently used to generate predictions over the left-out condition.

Since all models I consider here are stochastic, I compute for each condition and for each parameter configuration the MSD value for the estimated frequency of choice averaged over 150 simulations. Simulations reproduced the structure of laboratory conditions, including subject pairing protocol, information feedback available to agents, number of agents, and payoffs. To make simulation results comparable, the initialization of all models is set to assure equal probabilities of choosing each action at the first round of the simulation.

Put all together, the experimental data I gathered provide 35 independent observations. I compute MSD scores for each model in correspondence to each independent experimental observation, and store them in a vector of length 35. In order to assess the significance of model pairwise comparisons, I use a *Mann-Whitney-Wilcoxon match-paired signed-rank (two-tailed) test*, as done in Selten and Chmura (2008). For each pair of models, I test the null hypothesis that the corresponding vectors of scores have the same mean.

I compare the performances of the six different models of learning described in the Introduction (namely, REL, RL, NRL, NFP, SFP, and stEWA), together with those of the PB1 and PB0 models, and Nash equilibrium.

Furthermore, I also consider in my analysis three benchmark models. The first is the model of random behavior (labeled Random). It is a desirable property for a good model of learning that of being able to produce more accurate predictions than those of blind random behavior. The other two models of benchmark are called NNET2 and NNET. The NNET2 model is nothing but a traditional one-layer analog perceptron, where output units are fed back, as usual, by a measure of target-output error. Thus, the regret factor is dropped from the updating rule (3) illustrated in Section 2, resulting in the following equation:

$$\Delta w_{ij} = \lambda \cdot \beta \cdot \left[t_i\left(a^{-k}\right) - y_i \right] \cdot x_j . \qquad (1)$$

NNET2 is a two-parameter model and further differs from PB1 in that steepness and learning rate parameters (labeled λ and β, respectively) are allowed to vary independently one another. The NNET model is identical to the NNET2 one, with the exception that in (1) I set $\lambda = \beta$, thus obtaining a model with only one free parameter.

I include the NNET2 and NNET models in my analysis in order to better assess in what measure incorporating regret in a one-layer perceptron results in an improvement of its predictive and descriptive performances.

I test also the effects of payoff rescaling on the performance of models. Thus, the same methodology described above is applied to test model accuracy when payoffs are rescaled according to *Diminishing Sensitivity* i.e., transformed via the following function (called *value function*):

$$v(x) = \begin{cases} x^{\alpha} & \text{if } x \geq 0 \\ -\lambda(-x)^{\beta} & \text{otherwise,} \end{cases}$$

where α, β, and λ are positive constants. Tversky and Kahneman (1992) provide estimates for these parameters using their experimental data. The obtained values are: $\alpha = \beta = 0.88$ and $\lambda = 2.25$.

For each model, the grid search for optimal parameter values was conducted on broad parameter spaces, summarized in Table 1. The portions of parameter spaces I investigated were suggested by the authors of the models in previous works (Erev at al., 2007; Ho, Camerer, and Chong, 2007).

Table 1. Values of model free parameters used in my simulations.

Model	**Free Parameter: [Interval of variation] – Increment**	
NFP	λ: $[1.5,4.0]$ - 0.25	w: $[0.1,0.9]$ - 0.1
NNET	λ: $[0.05,1.00]$ - 0.05	
NNET2	λ: $[0.1,1.0]$ - 0.1	β: $[0.1,1.0]$ - 0.1
NRL	λ: $[3.0,7.0]$ - 0.5	w: $[0.10,0.90]$ - 0.05
PB1	λ: $[0.05,1.00]$ - 0.05	
REL	λ: $[2.2,3.4]$ - 0.1	$N(1)$: $[27,34]$ - 1
RL	λ: $[6.0,10.0]$ - 0.5	w: $[0.10,0.90]$ - 0.05
SFP	λ: $[10.0,14.0]$ - 0.5	w: $[0.05,0.90]$ - 0.05
stEWA	λ: $[1,9]$ - 1	

2.4 The Data

I collected datasets from different experiments on games with unique equilibria in mixed strategies, wherein the participants received a complete description of the payoff matrix and feedback about their choices and those of their opponents. These experiments have been conducted in a range of more than 50 years, under a variety of experimental conditions, and by different researchers in different fields. The games involved have a number of actions available to each player ranging from 2 to 5 (Appendix B provides a detailed description of all datasets).

Out of the 35 games considered, 24 are constant-sum, while in the remainders players could find incentive to reciprocate; in other words, in 24 experiments, subjects had to learn strategies of pure conflict, while in the other 11 the conflictual aspect does

not necessarily exclude a sort of cooperative (or fair) behavior, as in the non-constant sum games reported in Selten and Chmura (2008).

In order to let the learning processes fully unfold, I selected experiments with a minimum of 100 iterations of the stage game; this allows for the testing of the descriptive and predictive power of the different models on subjects' behavior not only in the early rounds, but also in the long run.

The reasons why I chose to test the models on data gathered by other experimenters are twofold. First, as Erev and Roth (1998:851) observed, "there is a danger that investigators will treat the models they propose like as their toothbrushes, and each will use his own model only on his own data"; thus, I considered a number as large as possible of datasets from experiments concerning long runs of games with unique equilibria in mixed strategies. Secondly, experimenters may unconsciously make some decisions, when designing their experiments, which in some way could favor their starting hypotheses. As a consequence, testing models on datasets from experiments conducted by other researchers helps to reduce these methodological biases.

The experimental datasets I gathered are in the form of sequences of average choice frequencies, computed over blocks of repetitions of the stage game. Relative frequencies are calculated for the two typologies of players (*row* and *column* players). For example, Malcolm and Lieberman (1965) present choice frequencies in 8 blocks of 25 trials each (the game was repeated 200 times).

Table 2. Summary of the datasets. The first column indicates the name of the researchers and the second one the year of publication of the experiment. The third column reports the number of times that the stage game was played. The fourth column specifies the number of blocks of trials over which the average choice frequencies are calculated. The fifth column indicates the number of subject who participated to the experiments. The sixth column reports additional important features (if any) for each experiment. Finally, the seventh column reports whether or not subjects were randomly paired at each trial.

Experimenters	Year	Games	Rounds #	Treatments/ Games	Blocks of Iterations	Subjects #	Other	Random Matching
Suppes and Atkinson	1960	2x2	210	1	7	20 pairs	No monetary reward – only "correct" feedback	No
Malcolm and Liberman	1965	2x2	200	1	8	9 pairs		No
O'Neill	1987	4x4	105	1	7	25 pairs	15 practice rounds were run	No
Rapoport and Boebel	1992	5x5	120	2	4	10 pairs for each treatment	10 practice rounds were run	No
Ochs	1995	2x2	56x10, 64x10, and 64x10	3	7 in the first treatment, 8 in the other two	8 pairs for each treatment		Yes
Rosenthal, Shachat, and Walker	2003	2x2	100 and 200	2	10	20 pairs for each treatment		No
Avrahami, Guth, and Kareev	2005	2x2	100	3	3	6 pairs in the first treatment and 12 pairs in the other two	Only the "Known" treatment is considered	No
Erev, Roth, Slonim, and Barron	2007	2x2	500	10	5	9 pairs for each treatment		No
Selten and Chmura	2008	2x2	200	12	8	16 pairs for each treatment		Yes

2.5 Simulation Results: Actual Payoffs

I will focus on the comparison of model performances based on *Prediction* scores (see the Methods Section), which penalize, although not directly, model complexity and provide a measure of model predictive power. Best Fit and By Game Scores provide a mere measure of data fitting and, for that reason, heavily favor more complex models.

Model performances are tested on a combined dataset from 35 experiments on repeated games with a unique equilibrium in mixed strategies. These experiments share the feature that the games were played repeatedly 100 or more times, in order to let the learning processes fully unfold. The present analysis is conducted over a larger number of models and datasets and further improves the methodology adopted in Marchiori and Warglien (2008).

Table 3 reports, for each condition (or game) and model, the corresponding Prediction Score multiplied by 100; models are then accordingly ranked from the best (top line) to the worst (bottom line). The average Prediction score is a summary statistic by which models with a different number of free parameters can be roughly compared, since model complexity is opportunely weighted. The significance of pairwise model comparisons is tested with a Mann-Whytney-Wilcoxon Match-Pairs Signed-Rank Test; p-values are reported in Tables 4, whereas the estimates of the differences between average Prediction scores for all pairs of models are reported in Table 5. In the analysis that follows, I will consider the 5% level of significance of the test i.e., two models are considered as equivalent if the null hypothesis of no differences in their respective average prediction scores cannot be rejected at a 5% level.

The main result of my analysis is that regret-based models are the best predictors of observed frequencies of play. The crucial importance of post-decisional regret is evident if we compare the performance of the PB model with that of the simple perceptrons NNET and NNET2; this shows that the introduction in the updating rule of a term accounting for regret dramatically improves the reliability of predictions.

The NFP and SFP models can be considered as the best predictors because, reading from Table 4, they provide, on average, equivalent predictions. These two fictitious play models are both based on regret and differ for a minor structural detail. SFP and NFP perform significantly better than all other models, with the exception of RL. This result, surprising at a first glance, can be explained by looking at the high variability of

RL's performance: even though the average prediction score for RL is three times as bigger as that of SFP, the distribution of the average ranked scores of RL does not stochastically dominate the ones corresponding to NFP and SFP. Nonetheless, it seems reasonable to claim that SFP and NFP are the best performing models.

The PB0 and PB1 models are equivalent predictors of the data and are not able to predict data significantly better than do Nash and RL models. Also this result might appear surprising at a first sight. Let us consider, for example, the PB0 model: if we look at the estimated differences between average prediction scores (reported in Table 5), we can see that its predictions are two times more accurate than those provided by Nash equilibrium. However, once again, a signed rank test fails to reject the null of no differences between average scores due to the high variance in one of the two distributions. These two examples of highly counterintuitive responses of the ranked test might constitute a point of weakness of the present analysis; however, they should rather be interpreted as the lack of sufficient information that might be either obtained considering new games, or disaggregating existing data at the lowest unit of observation (i.e., individuals or groups of players, depending on whether experiments were run using fix-pairing or random-matching protocol). It is also worth noting that the PB0 model does significantly better that stEWA.

At the ideal third place there is the stEWA model. It is outperformed by NFP, SFP, and PB0, but it performs equivalently to PB1, Nash equilibrium, and the RL model. As mentioned in the introduction of this thesis, stEWA is a hybrid model merging beliefs and reinforcement learning theories. Most probably, its lower level of accuracy is due to its reinforcement component.

All models based on reinforcement, with the exception of the RL model (most probably due to the high variability of its prediction accuracy), are outperformed by Nash equilibrium. The NRL model is even a poorer predictor of data if compared with the random choice model. This is another important result of my analysis, as it goes against more than one study in the extant literature on learning (Erev and Roth, 1998; Sarin and Vahid, 2001; Erev et al., 2007). This result might be due to reasons of two different natures. For first, reinforcement based models do very poorly in the last six games proposed by Selten and Chmura (2008). Even though completely mixed, these games are not constant-sum and players might found incentive to cooperate and/or reciprocate. Indeed, observed choice frequencies in the first (constant-sum) and last six (non constant-sum) games described by Selten and Chmura are quite different, in spite

of the fact that the equilibria of the first six games are correspondingly equal to those of the last six. The difference is then due to the fact that in non constant-sum games cooperative behavior is not necessarily excluded. Also the games proposed by Avrahami, Guth, and Kareev (2005) are not constant-sum. Reinforcement learning models, by design, do not take into account these cooperative features of human behavior, even indirectly, and are not able to predict behavior in such psychologically richer interactive situations. Another reason for the failure of reinforcement models could be that testing models on a large dataset would require the exploration of broader regions of the parameter spaces than those suggested by the authors of the models in previous works, where smaller datasets were considered.

Table 3. MSD and Prediction Scores. *Actual* Payoffs. In the first column, between parentheses, the number of model free parameters is reported.

	AGK50	AGK67	AGK75	ERSB G1	ERSB G2	ERSB G3	ERSB G4	ERSB G5	ERSB G6	ERSB G7
SFP (2)	0.045	1.169	2.248	0.408	0.318	0.374	0.689	0.204	1.098	0.893
NFP (2)	0.045	1.165	2.240	0.389	0.302	0.350	0.862	0.108	1.032	0.856
PB0 (0)	0.050	0.112	0.523	0.780	1.253	0.503	2.078	0.848	0.120	0.323
PB1 (1)	0.048	0.140	0.813	0.581	1.821	0.271	1.967	0.368	0.189	0.384
stEWA (1)	0.045	0.152	0.172	1.832	5.078	3.010	2.772	1.953	1.065	1.700
NE (0)	0.045	1.903	4.515	1.919	10.582	6.728	1.078	1.117	5.261	5.804
RL (2)	0.043	0.522	1.565	0.377	0.244	0.205	2.150	0.111	0.128	0.249
NNET (1)	0.046	0.355	0.345	2.220	6.962	4.309	2.827	2.469	1.433	2.236
NNET2 (2)	0.051	0.369	0.338	2.234	6.978	4.314	2.824	2.482	1.427	2.248
REL (2)	0.043	0.390	0.366	2.230	6.990	4.340	2.841	2.483	1.420	2.248
Random (0)	0.045	0.373	0.348	2.228	6.972	4.310	2.824	2.472	1.429	2.241
NRL (2)	0.063	0.639	1.770	2.182	1.229	4.042	1.124	1.383	2.105	2.347

	ERSB G8	ERSB G9	ERSB G10	M&L	Oc1	Oc4	Oc9	On	R&B10	R&B15
SFP (2)	0.323	0.730	1.038	0.593	0.434	0.790	1.149	0.166	0.269	0.587
NFP (2)	0.553	0.319	1.047	0.613	0.437	0.777	1.110	0.163	0.199	0.448
PB0 (0)	1.962	0.138	3.062	0.924	0.450	1.219	2.139	0.302	0.155	0.495
PB1 (1)	1.529	0.140	2.053	0.704	0.435	1.256	1.973	0.309	0.183	0.541
stEWA (1)	4.163	0.302	5.381	1.106	0.418	2.026	4.492	0.135	0.219	0.357
NE (0)	1.866	0.668	1.233	2.114	0.435	1.366	2.240	0.136	0.354	0.865
RL (2)	0.880	0.216	0.998	7.016	0.423	1.687	1.525	0.130	0.102	0.331
NNET (1)	4.612	0.418	6.183	2.459	0.446	1.779	3.897	0.301	0.181	0.519
NNET2 (2)	4.612	0.418	6.186	2.457	0.436	1.781	3.902	0.313	0.180	0.542
REL (2)	4.613	0.414	6.190	2.491	0.433	1.875	3.954	2.237	1.067	1.400
Random (0)	4.611	0.418	6.188	2.458	0.435	1.778	3.897	1.236	5.041	5.428
NRL (2)	0.516	1.298	0.934	10.602	0.471	3.334	3.695	0.091	0.350	0.279

	RSW D	RSW S	S&A3K	S&C G1	S&C G2	S&C G3	S&C G4	S&C G5	S&C G6	S&C G7
SFP (2)	1.118	2.610	1.621	1.150	0.061	0.899	0.605	0.207	0.088	0.468
NFP (2)	1.404	2.624	1.415	1.039	0.055	0.940	0.599	0.455	0.088	0.572
PB0 (0)	1.905	4.252	1.610	0.478	0.651	0.486	0.703	0.656	0.219	1.191
PB1 (1)	1.397	3.893	1.665	2.847	0.798	0.913	0.912	0.774	0.232	0.924
stEWA (1)	3.764	5.629	6.690	8.390	0.491	0.283	0.077	0.114	0.165	0.793
NE (0)	0.397	0.610	7.327	2.546	2.137	1.331	0.672	0.309	0.113	6.520
RL (2)	0.236	1.007	9.912	2.661	4.952	1.843	1.112	0.720	0.217	12.741
NNET (1)	3.767	3.756	2.551	10.810	4.116	9.975	5.094	2.869	0.645	7.077
NNET2 (2)	3.789	3.797	2.550	10.813	4.116	9.979	5.096	2.868	0.645	7.073
REL (2)	3.661	5.493	2.592	10.954	4.145	10.078	5.159	2.886	0.666	7.131
Random (0)	3.763	5.496	2.550	10.809	4.117	9.977	5.098	2.869	0.645	7.073
NRL (2)	3.799	2.985	21.957	3.431	7.541	2.002	2.264	1.664	0.414	7.421

	S&C G8	S&C G9	S&C G10	S&C G11	S&C G12	Mean	sd
SFP (2)	0.323	1.289	0.443	0.177	0.090	**0.705**	**0.596**
NFP (2)	0.438	1.514	0.721	0.187	0.098	**0.719**	**0.598**
PB0 (0)	1.604	1.453	0.955	0.570	0.237	**0.983**	**0.913**
PB1 (1)	1.773	1.874	0.992	0.656	0.256	**1.017**	**0.865**
stEWA (1)	0.491	0.266	0.575	0.176	0.149	**1.841**	**2.237**
NE (0)	1.388	0.418	0.785	0.410	0.095	**2.151**	**2.530**
RL (2)	1.405	15.210	2.227	1.284	0.918	**2.153**	**3.602**
NNET (1)	3.668	8.434	2.971	2.648	0.758	**3.232**	**2.798**
NNET2 (2)	3.668	8.437	2.972	2.656	0.758	**3.237**	**2.798**
REL (2)	3.696	8.465	2.985	2.656	0.773	**3.410**	**2.753**
Random (0)	3.668	8.436	2.972	2.651	0.758	**3.589**	**2.730**
NRL (2)	9.672	5.345	10.357	10.652	14.959	**4.083**	**4.863**

Table 4. This table reports the P-values for pair-wise model comparisons in terms of Prediction scores. Simulations were run feeding models with *actual* payoffs. The null of no differences in the mean scores was tested with a Mann-Whitney-Wilcoxon match-paired signed-rank (two-tailed) test. Shaded cells refer to the cases in which the null is not rejected at a 5% level of significance.

	SFP (2)	NFP (2)	PB0 (0)	PB1 (1)	stEWA (1)	NE (0)	RL (2)	NNET (1)	NNET2 (2)	REL (2)	Random (0)	NRL (2)
SFP (2)		0.918	0.028	0.006	0.010	0.000	0.063	0.000	0.000	0.000	0.000	0.000
NFP (2)	0.918		0.018	0.004	0.009	0.000	0.052	0.000	0.000	0.000	0.000	0.000
PB0 (0)	0.028	0.018		0.676	0.045	0.056	0.298	0.000	0.000	0.000	0.000	0.000
PB1 (1)	0.006	0.004	0.676		0.052	0.076	0.528	0.000	0.000	0.000	0.000	0.000
stEWA (1)	0.010	0.009	0.045	0.052		0.193	0.762	0.000	0.000	0.000	0.000	0.017
NE (0)	0.000	0.000	0.056	0.076	0.193		0.915	0.047	0.043	0.023	0.017	0.028
RL (2)	0.063	0.052	0.298	0.528	0.762	0.915		0.012	0.012	0.009	0.008	0.000
NNET (1)	0.000	0.000	0.000	0.000	0.000	0.047	0.012		0.004	0.000	0.026	0.851
NNET2 (2)	0.000	0.000	0.000	0.000	0.000	0.043	0.012	0.004		0.000	0.567	0.863
REL (2)	0.000	0.000	0.000	0.000	0.000	0.023	0.009	0.000	0.000		0.003	0.812
Random (0)	0.000	0.000	0.000	0.000	0.000	0.017	0.008	0.026	0.567	0.003		0.800
NRL (2)	0.000	0.000	0.000	0.000	0.017	0.028	0.000	0.851	0.863	0.812	0.800	

Table 5. Pair-wise model comparisons in terms of Prediction scores. Simulations were run feeding models with actual payoffs. Each cell of this table reports the estimate for the difference of the location parameters of *x* and *y*, where *x* is row model's vector of scores, and *y* that one of the column model.

	SFP (2)	NFP (2)	PB0 (0)	PB1 (1)	stEWA (1)	NE (0)	RL (2)	NNET (1)	NNET2 (2)	REL (2)	Random (0)	NRL (2)
SFP (2)		-0.001	-0.274	-0.300	-0.858	-0.803	-0.432	-2.231	-2.244	-2.395	-2.639	-1.976
NFP (2)	0.001		-0.249	-0.252	-0.851	-0.786	-0.422	-2.180	-2.180	-2.348	-2.596	-1.897
PB0 (0)	0.274	0.249		-0.016	-0.513	-0.554	-0.208	-1.751	-1.759	-1.863	-2.116	-1.835
PB1 (1)	0.300	0.252	0.016		-0.605	-0.472	-0.102	-1.909	-1.910	-1.981	-2.232	-1.715
stEWA (1)	0.858	0.851	0.513	0.605		-0.259	0.056	-0.943	-0.953	-1.189	-1.350	-1.072
NE (0)	0.803	0.786	0.554	0.472	0.259		0.011	-1.007	-1.014	-1.220	-1.514	-0.982
RL (2)	0.432	0.422	0.208	0.102	-0.056	-0.011		-1.176	-1.178	-1.421	-1.628	-1.287
NNET (1)	2.231	2.180	1.751	1.909	0.943	1.007	1.176		-0.004	-0.030	-0.002	-0.036
NNET2 (2)	2.244	2.180	1.759	1.910	0.953	1.014	1.178	0.004		-0.027	0.000	-0.024
REL (2)	2.395	2.348	1.863	1.981	1.189	1.220	1.421	0.030	0.027		0.017	0.081
Random (0)	2.639	2.596	2.116	2.232	1.350	1.514	1.628	0.002	0.000	-0.017		0.107
NRL (2)	1.976	1.897	1.835	1.715	1.072	0.982	1.287	0.036	0.024	-0.081	-0.107	

2.6 Simulation Results: Rescaled Payoffs

I ran the same simulations described in the methods section feeding models with payoff rescaled according to Kahnemann and Tversky's (1979) *prospect theory*. The estimates of the parameters of the transformation (*value function*) I used for simulations are those reported in Kahnemann and Tversky (1992).

As done in the previous section, the present analysis is focused on Prediction scores and the significance of pairwise comparisons is assessed through the use of a Mann-Whitney-Wilcoxon test. For each pair of models, the null hypothesis of no difference in their average Prediction scores (over the 35 independent observations) is tested. Table 6 reports, for each model and for each game the corresponding Prediction score and models are ranked accordingly from the best (on the top line) to the worst (on the bottom line). Tables 7 and 8 report, respectively, the p-values and the estimated differences between average scores, for all possible model pairwise comparisons.

Comparing numbers reported in Tables 3 and 6, the first result is that regret based models are those for which the increase in prediction accuracy due to the introduction of rescaled payoff is the largest. The improvement of the predictive power is marginal for stEWA and all reinforcement based models. If the transformation of payoffs

according to prospect theory does not improve model accuracy in the same measure for all models, it is also true that it preserves the ranking of the models. Indeed, comparing Tables 3 and 6, it is evident that the models are ranked in the same way with the exception of minor changes.

Even though with rescaled payoffs NFP is a better predictor than SFP, the difference is not significant and, once again, they turn out to be the best models of learning. Furthermore, both the PB1 and PB0 models perform significantly better than stEWA, Nash equilibrium, and all other models, with the exception of the RL model.

The stEWA model performs equivalently to Nash equilibrium and the RL model. Lastly, the simple perceptron (in both its versions NNET and NNET2), NRL, and REL are all outperformed by Nash equilibrium and provide equivalently accurate predictions.

Table 6. MSD and Prediction Scores. *Rescaled* Payoffs. In the first column, between parentheses, the number of model free parameters is reported.

	AGK50	AGK67	AGK75	ERSB G1	ERSB G2	ERSB G3	ERSB G4	ERSB G5	ERSB G6	ERSB G7
NFP (2)	0.046	1.116	1.277	0.349	0.283	0.411	0.999	0.092	0.977	0.810
SFP (2)	0.043	1.073	1.315	0.407	0.313	0.458	0.901	0.096	1.085	0.887
PB0 (0)	0.037	0.079	0.399	0.713	1.225	0.446	2.066	0.720	0.113	0.255
PB1 (1)	0.046	0.105	0.633	0.541	1.765	0.229	1.960	0.234	0.201	0.312
stEWA (1)	0.035	0.080	0.428	1.791	5.081	3.020	2.524	1.911	0.382	1.723
RL (2)	0.038	0.394	1.296	0.383	0.236	4.451	2.119	0.103	0.101	0.196
NE (0)	0.045	1.903	4.515	1.919	10.582	6.728	1.078	1.117	5.261	5.804
NNET2 (2)	0.050	0.385	0.348	2.237	6.978	4.309	2.817	2.485	1.439	2.246
NNET (1)	0.046	0.365	0.363	2.255	6.976	4.319	2.829	2.469	1.434	2.246
REL (2)	0.040	0.395	0.374	2.245	7.031	4.337	2.795	2.464	1.443	2.248
NRL (2)	0.039	0.462	1.377	2.423	1.199	3.945	0.933	1.696	2.251	2.323

	ERSB G8	ERSB G9	ERSB G10	M&L	Oc1	Oc4	Oc9	On	R&B10	R&B15
NFP (2)	0.355	0.302	0.877	0.451	0.431	0.623	0.864	0.148	0.132	0.433
SFP (2)	0.338	0.442	0.671	0.595	0.440	0.640	0.942	0.167	0.301	0.562
PB0 (0)	1.822	0.126	2.869	0.606	0.435	1.005	1.530	0.303	0.170	0.485
PB1 (1)	1.343	0.152	2.154	0.359	0.434	0.866	1.459	0.311	0.181	0.536
stEWA (1)	4.110	0.123	5.450	0.315	0.433	1.474	3.478	0.126	0.171	0.403
RL (2)	0.792	0.222	1.029	12.469	0.425	1.220	1.339	0.166	0.115	0.321
NE (0)	1.866	0.668	1.233	2.114	0.435	1.366	2.240	0.136	0.354	0.865
NNET2 (2)	4.615	0.422	6.175	2.462	0.438	1.781	3.892	0.329	0.181	0.535
NNET (1)	4.609	0.422	6.203	2.458	0.459	1.820	3.898	0.319	0.186	0.517
REL (2)	4.639	0.424	6.168	1.677	0.446	1.750	3.915	2.299	1.077	1.383
NRL (2)	0.496	1.528	0.979	11.089	0.488	2.606	3.538	0.071	0.347	0.295

	RSW D	RSW S	S&A3K	S&C G1	S&C G2	S&C G3	S&C G4	S&C G5	S&C G6	S&C G7
NFP (2)	0.837	1.788	1.448	0.636	0.056	0.422	0.201	0.150	0.049	0.587
SFP (2)	0.739	2.604	1.646	0.421	0.084	0.337	0.144	0.108	0.046	0.539
PB0 (0)	1.524	3.860	1.633	0.304	0.422	0.351	0.476	0.416	0.115	1.345
PB1 (1)	1.622	2.957	1.612	2.221	0.494	0.586	0.580	0.504	0.136	0.905
stEWA (1)	2.829	4.824	6.460	5.809	1.524	0.743	1.134	0.546	0.247	0.863
RL (2)	0.199	0.569	9.340	2.205	4.792	1.804	1.245	0.904	0.387	5.882
NE (0)	0.397	0.610	7.327	2.546	2.137	1.331	0.672	0.309	0.113	6.520
NNET2 (2)	3.767	3.771	2.551	10.807	4.115	9.987	5.111	2.874	0.646	7.086
NNET (1)	3.774	3.774	2.557	10.809	4.132	9.978	5.108	2.871	0.646	7.074
REL (2)	3.607	5.348	2.607	10.877	4.115	9.986	5.112	2.850	0.640	7.152
NRL (2)	2.135	3.355	20.221	3.602	7.527	1.980	2.509	2.481	0.670	7.419

	S&C G8	S&C G9	S&C G10	S&C G11	S&C G12	Mean	sd
NFP (2)	0.394	1.088	0.363	0.142	0.047	**0.548**	**0.440**
SFP (2)	0.348	0.836	0.297	0.122	0.041	**0.571**	**0.524**
PB0 (0)	1.305	1.256	0.759	0.422	0.112	**0.849**	**0.847**
PB1 (1)	1.517	1.724	0.788	0.487	0.131	**0.860**	**0.751**
stEWA (1)	1.178	0.485	1.167	0.395	0.066	**1.752**	**1.883**
RL (2)	1.129	9.397	1.822	1.523	1.263	**1.996**	**2.972**
NE (0)	1.388	0.418	0.785	0.410	0.095	**2.151**	**2.530**
NNET2 (2)	3.688	8.466	2.984	2.653	0.765	**3.240**	**2.798**
NNET (1)	3.669	8.441	2.974	2.662	0.758	**3.241**	**2.796**
REL (2)	3.701	8.448	2.958	2.668	0.740	**3.370**	**2.751**
NRL (2)	9.890	5.080	8.531	6.779	15.998	**3.893**	**4.600**

Table 7. This table reports the P-values for pair-wise model comparisons in terms of Prediction scores. Simulations were run feeding models with *rescaled* payoffs. The null hypothesis of no differences in the mean scores was tested with a Mann-Whitney-Wilcoxon match-paired signed-rank (two-tailed) test. Shaded cells refer to the cases in which the null hypothesis is not rejected at a 5% level of significance.

	NFP (2)	SFP (2)	PB0 (0)	PB1 (1)	stEWA (1)	RL (2)	NE (0)	NNET2 (2)	NNET (1)	REL (2)	NRL (2)
NFP (2)		0.688	0.006	0.003	0.000	0.003	0.000	0.000	0.000	0.000	0.000
SFP (2)	0.688		0.019	0.012	0.001	0.008	0.000	0.000	0.000	0.000	0.000
PB0 (0)	0.006	0.019		0.700	0.001	0.081	0.004	0.000	0.000	0.000	0.000
PB1 (1)	0.003	0.012	0.700		0.003	0.114	0.012	0.000	0.000	0.000	0.000
stEWA (1)	0.000	0.001	0.001	0.003		0.993	0.417	0.000	0.000	0.000	0.013
RL (2)	0.003	0.008	0.081	0.114	0.993		0.456	0.002	0.002	0.001	0.000
NE (0)	0.000	0.000	0.004	0.012	0.417	0.456		0.043	0.045	0.028	0.047
NNET2 (2)	0.000	0.000	0.000	0.000	0.000	0.002	0.043		1.000	0.147	0.980
NNET (1)	0.000	0.000	0.000	0.000	0.000	0.002	0.045	1.000		0.229	0.980
REL (2)	0.000	0.000	0.000	0.000	0.000	0.001	0.028	0.147	0.229		0.749
NRL (2)	0.000	0.000	0.000	0.000	0.013	0.000	0.047	0.980	0.980	0.749	

Table 8. Pair-wise model comparisons in terms of Prediction scores. Simulations were run feeding models with *rescaled* payoffs. Each cell of this table reports the estimate for the difference of the location parameters of *x* and *y*, where *x* is row model's vector of scores, and *y* that one of the column model.

	NFP (2)	SFP (2)	PB0 (0)	PB1 (1)	stEWA (1)	RL (2)	NE (0)	NNET2 (2)	NNET (1)	REL (2)	NRL (2)
NFP (2)		-0.007	-0.245	-0.264	-0.907	-0.634	-0.966	-2.328	-2.336	-2.483	-2.073
SFP (2)	0.007		-0.236	-0.249	-0.905	-0.617	-1.000	-2.298	-2.302	-2.437	-2.119
PB0 (0)	0.245	0.236		-0.014	-0.594	-0.436	-0.748	-1.923	-1.933	-1.996	-1.927
PB1 (1)	0.264	0.249	0.014		-0.682	-0.298	-0.652	-2.052	-2.057	-2.143	-1.835
stEWA (1)	0.907	0.905	0.594	0.682		0.002	-0.156	-1.108	-1.114	-1.212	-1.029
RL (2)	0.634	0.617	0.436	0.298	-0.002		-0.176	-1.208	-1.219	-1.364	-1.136
NE (0)	0.966	1.000	0.748	0.652	0.156	0.176		-1.019	-1.013	-1.160	-0.945
NNET2 (2)	2.328	2.298	1.923	2.052	1.108	1.208	1.019		0.000	-0.009	-0.007
NNET (1)	2.336	2.302	1.933	2.057	1.114	1.219	1.013	0.000		-0.008	-0.006
REL (2)	2.483	2.437	1.996	2.143	1.212	1.364	1.160	0.009	0.008		0.167
NRL (2)	2.073	2.119	1.927	1.835	1.029	1.136	0.945	0.007	0.006	-0.167	

2.7 Conclusions

Simulation results show that the two regret-based models Normalized Fictitious Play (NFP) and Stochastic Fictitious Play (SFP) are the best predictors of the data. Compared to the other models, they provide the smallest Prediction scores and their performances are substantially equivalent. Indeed, these two models are identical except for a minor structural detail.

The second best model is the PB model. The parameter free version PB0 performs, surprisingly, better that the parametric version PB1 in terms of average Prediction scores, even though this difference is not statistically significant. In general, the Perceptron shows very fast convergence to rather stable frequencies of choice (similarly to stEWA and NFP, and differently from the reinforcement learning models, which are slower to adapt). Its advantage, then, is not in mimicking well the experimental speed of learning, which is, generally, slower. On the contrary, its success is due to a large extent to its ability to fit the experimental average behavior in the long run. It is well known from the literature on the "learning direction" that subjects tend to adjust in the direction of the ex-post best response in a large variety of experimental games. The PB model seems to capture an empirically valid quantification of this qualitative tendency by tuning the intensity of the adjustment proportionally to the regret (and conditionally to the salience of the payoffs). The NFP model follows

similar learning principles and performs in a very similar way. Why then do other models do worse? As well known, reinforcement learning models respond mainly to experienced payoffs, and thus adapt slowly and fail to capture all the relevance of foregone payoffs even in the long run. This seems to be a source of empirical relative weakness in the experiments considered here. It is less clear what may be the source of relative weakness of stEWA (which, by the way, performs very well). The stEWA model tends to preserves a weight for reinforcement learning, that may keep it away from the long run average behavior of experimental subjects. The model has one free parameter and adjusts via its self-tuning mechanism at least two additional parameters. I conjecture that this might lead to some form of over-fitting that makes it comparatively less robust; this is consistent with the fact that the model is much stronger in fitting separately single games, while its performance deteriorates in the cross-prediction task.

As perhaps the most important results of my thesis, simulation results show that regret-based models fare better than Nash equilibrium, self-tuning EWA and reinforcement based models, thus supporting the hypothesis that regret for foregone payoffs plays a central role in shaping human choice behavior.

Another important result is that reinforcement based models turn out to be very poor predictors of empirical data; no matters whether with actual or rescaled payoffs, the NRL model is less accurate than the model of random choice behavior. This result is in contradiction with some recent contributions in the learning literature (Erev and Roth, 1998; Sarin and Vahid, 2001; Erev et al., 2007) and the motivations might be of two different natures. First, testing models on Selten and Chmura's (2008) games seems to particularly penalize reinforcement based models. In particular, the last six games, even though completely mixed, are not constant-sum and, for that reason, might have provided some incentive for cooperative or reciprocating behaviors. Reinforcement learning models do not take into account these cooperative features of human behavior, even indirectly, and are not able to predict behavior in such richer interactive situations. Another reason for the failure of reinforcement models could be that testing models on a large dataset would require the exploration of broader regions of the parameter spaces than those suggested by the authors of the models in previous works, where smaller datasets were considered.

In this chapter I also analyze the predictive power of models when fed with payoffs rescaled according to Kahnemann and Tversky's (1979 and 1992) *prospect theory*.

Results show two facts as we pass from actual to rescaled payoffs: first, the ranking of the models remains unaltered; second, the increase in accuracy is significant for regret-based models (NFP, SFP, and PB) and marginal for all the others (stEWA and reinforcement learning models).

2.8 Appendix A. Supporting Material

Table A1. MSD and By Game Scores. *Actual* Payoffs. In the first column, between parentheses, the number of model free parameters is reported.

	AGK50	AGK67	AGK75	ERSB G1	ERSB G2	ERSB G3	ERSB G4	ERSB G5	ERSB G6	ERSB G7
SFP (2)	0.043	0.369	0.344	0.221	0.229	0.145	0.439	0.092	0.169	0.156
NFP (2)	0.042	0.366	0.344	0.150	0.214	0.137	0.479	0.102	0.174	0.137
stEWA (1)	0.034	0.074	0.171	0.216	0.349	0.183	0.944	0.195	0.094	0.209
PB1 (1)	0.037	0.085	0.219	0.581	1.443	0.271	1.967	0.368	0.118	0.305
PB0 (0)	0.050	0.112	0.523	0.780	1.253	0.503	2.078	0.848	0.120	0.323
RL (2)	0.027	0.065	0.377	0.176	0.238	0.205	1.542	0.111	0.074	0.206
NE (0)	0.045	1.903	4.515	1.919	10.582	6.728	1.078	1.117	5.261	5.804
NRL (2)	0.021	0.204	0.653	2.170	0.734	2.924	0.876	1.344	1.371	1.737
NNET2 (2)	0.035	0.354	0.327	2.211	6.946	4.284	2.812	2.457	1.416	2.225
NNET (1)	0.038	0.355	0.334	2.216	6.962	4.302	2.814	2.469	1.423	2.222
REL (2)	0.035	0.351	0.340	2.196	6.942	4.287	2.784	2.451	1.409	2.223
Random (0)	0.045	0.373	0.348	2.228	6.972	4.310	2.824	2.472	1.429	2.241

	ERSB G8	ERSB G9	ERSB G10	M&L	Oc1	Oc4	Oc9	On	R&B10	R&B15
SFP (2)	0.306	0.235	0.330	0.289	0.394	0.624	0.762	0.037	0.131	0.234
NFP (2)	0.322	0.123	0.397	0.287	0.401	0.584	0.773	0.061	0.114	0.133
stEWA (1)	0.476	0.106	0.113	0.638	0.377	1.151	1.940	0.044	0.140	0.342
PB1 (1)	1.529	0.126	2.053	0.689	0.421	1.134	1.646	0.303	0.175	0.527
PB0 (0)	1.962	0.138	3.062	0.924	0.450	1.219	2.139	0.302	0.155	0.495
RL (2)	0.284	0.100	0.138	7.016	0.388	0.437	0.766	0.130	0.087	0.281
NE (0)	1.866	0.668	1.233	2.114	0.435	1.366	2.240	0.136	0.354	0.865
NRL (2)	0.149	1.265	0.826	10.602	0.308	0.473	0.809	0.063	0.086	0.279
NNET2 (2)	4.590	0.402	6.164	2.448	0.405	1.741	3.835	0.298	0.176	0.520
NNET (1)	4.598	0.415	6.179	2.444	0.407	1.743	3.879	0.295	0.178	0.406
REL (2)	4.558	0.410	6.135	2.172	0.397	1.708	3.816	2.196	1.050	1.347
Random (0)	4.611	0.418	6.188	2.458	0.435	1.778	3.897	1.236	5.041	5.428

	RSW D	RSW S	S&A3K	S&C G1	S&C G2	S&C G3	S&C G4	S&C G5	S&C G6	S&C G7
SFP (2)	0.290	0.716	0.312	0.049	0.060	0.045	0.020	0.048	0.061	0.384
NFP (2)	0.421	1.031	0.434	0.141	0.055	0.053	0.014	0.042	0.060	0.405
stEWA (1)	3.451	5.581	3.642	3.806	0.491	0.283	0.077	0.114	0.165	0.619
PB1 (1)	1.397	3.893	0.072	2.732	0.755	0.860	0.864	0.754	0.168	0.884
PB0 (0)	1.905	4.252	1.610	0.478	0.651	0.486	0.703	0.656	0.219	1.191
RL (2)	0.185	0.445	6.305	2.661	2.315	1.843	1.112	0.720	0.217	7.814
NE (0)	0.397	0.610	7.327	2.546	2.137	1.331	0.672	0.309	0.113	6.520
NRL (2)	3.151	2.783	6.330	3.431	4.013	1.814	2.264	1.664	0.345	7.421
NNET2 (2)	3.724	3.696	2.543	10.803	4.106	9.965	5.074	2.851	0.637	7.066
NNET (1)	3.748	3.703	2.548	10.794	4.104	9.958	5.089	2.860	0.640	7.071
REL (2)	3.568	5.356	2.554	10.763	4.079	9.914	5.038	2.801	0.624	7.026
Random (0)	3.763	5.496	2.550	10.809	4.117	9.977	5.098	2.869	0.645	7.073

	S&C G8	S&C G9	S&C G10	S&C G11	S&C G12	Mean	sd
SFP (2)	0.306	0.309	0.428	0.079	0.046	**0.249**	**0.191**
NFP (2)	0.430	0.391	0.476	0.072	0.049	**0.269**	**0.229**
stEWA (1)	0.436	0.266	0.575	0.172	0.066	**0.787**	**1.303**
PB1 (1)	1.698	1.829	0.989	0.652	0.191	**0.907**	**0.861**
PB0 (0)	1.604	1.453	0.955	0.570	0.237	**0.983**	**0.913**
RL (2)	1.405	1.320	2.227	1.284	0.918	**1.240**	**1.948**
NE (0)	1.388	0.418	0.785	0.410	0.095	**2.151**	**2.530**
NRL (2)	7.782	1.278	4.227	4.980	2.347	**2.306**	**2.507**
NNET2 (2)	3.661	8.430	2.962	2.644	0.753	**3.216**	**2.797**
NNET (1)	3.657	8.380	2.949	2.648	0.756	**3.217**	**2.798**
REL (2)	3.620	8.302	2.912	2.595	0.725	**3.334**	**2.719**
Random (0)	3.668	8.436	2.972	2.651	0.758	**3.589**	**2.730**

Table A2. This table reports the P-values for pair-wise model comparisons in terms of By Game scores. Simulations were run feeding models with *actual* payoffs. The null hypothesis of no differences in the mean scores was tested with a Mann-Whitney-Wilcoxon match-paired signed-rank (two-tailed) test. Shaded cells refer to the cases in which the null hypothesis is not rejected at a 5% level of significance.

	SFP (2)	NFP (2)	stEWA (1)	PB1 (1)	PB0 (0)	RL (2)	NE (0)	NRL (2)	NNET2 (2)	NNET (1)	REL (2)	Random (0)
SFP (2)		0.000	0.000	0.017	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
NFP (2)	0.000		0.016	0.025	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
stEWA (1)	0.000	0.016		0.052	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
PB1 (1)	0.017	0.025	0.052		0.073	0.056	0.041	0.004	0.000	0.005	0.000	0.000
PB0 (0)	0.000	0.000	0.000	0.073		0.466	0.399	0.902	0.166	0.928	0.004	0.008
RL (2)	0.000	0.000	0.000	0.056	0.466		0.107	0.000	0.003	0.000	0.000	0.000
NE (0)	0.000	0.000	0.000	0.041	0.399	0.107		0.259	0.008	0.000	0.000	0.000
NRL (2)	0.000	0.000	0.000	0.004	0.902	0.000	0.259		0.048	0.126	0.000	0.000
NNET2 (2)	0.000	0.000	0.000	0.000	0.166	0.003	0.008	0.048		0.147	0.001	0.001
NNET (1)	0.000	0.000	0.000	0.005	0.928	0.000	0.000	0.126	0.147		0.000	0.000
REL (2)	0.000	0.000	0.000	0.000	0.004	0.000	0.000	0.000	0.001	0.000		0.199
Random (0)	0.000	0.000	0.000	0.000	0.008	0.000	0.000	0.000	0.001	0.000	0.199	

Table A3. Pair-wise model comparisons in terms of By Game scores. Simulations were run feeding models with *actual* payoffs. Each cell of this table reports the estimate for the difference of the location parameters of *x* and *y*, where *x* is row model's vector of scores, and *y* that one of the column model.

	SFP (2)	**NFP (2)**	**stEWA (1)**	**PB1 (1)**	**PB0 (0)**	**RL (2)**	**NE (0)**	**NRL (2)**	**NNET2 (2)**	**NNET (1)**	**REL (2)**	**Random (0)**
SFP (2)		0.046	0.018	1.514	2.207	2.116	2.351	2.315	2.576	2.484	3.070	3.071
NFP (2)	-0.046		-0.019	1.121	1.892	1.785	2.013	1.968	2.152	2.195	2.641	2.672
stEWA (1)	-0.018	0.019		0.999	1.770	1.740	1.999	1.904	2.128	2.192	2.562	2.581
PB1 (1)	-1.514	-1.121	-0.999		0.421	0.554	0.514	0.748	0.845	0.694	1.160	1.156
PB0 (0)	-2.207	-1.892	-1.770	-0.421		-0.068	-0.062	0.020	0.295	0.028	0.492	0.523
RL (2)	-2.116	-1.785	-1.740	-0.554	0.068		0.067	0.121	0.276	0.188	0.612	0.608
NE (0)	-2.351	-2.013	-1.999	-0.514	0.062	-0.067		0.043	0.186	0.142	0.551	0.554
NRL (2)	-2.315	-1.968	-1.904	-0.748	-0.020	-0.121	-0.043		0.144	0.059	0.464	0.474
NNET2 (2)	-2.576	-2.152	-2.128	-0.845	-0.295	-0.276	-0.186	-0.144		-0.092	0.127	0.125
NNET (1)	-2.484	-2.195	-2.192	-0.694	-0.028	-0.188	-0.142	-0.059	0.092		0.393	0.386
REL (2)	-3.070	-2.641	-2.562	-1.160	-0.492	-0.612	-0.551	-0.464	-0.127	-0.393		0.007
Random (0)	-3.071	-2.672	-2.581	-1.156	-0.523	-0.608	-0.554	-0.474	-0.125	-0.386	-0.007	

Table A4. MSD and Best Fit Scores. *Actual* Payoffs. In the first column, between parentheses, the number of model free parameters is reported.

	AGK50	AGK67	AGK75	ERSB G1	ERSB G2	ERSB G3	ERSB G4	ERSB G5	ERSB G6	ERSB G7
SFP (2)	0.045	1.169	2.248	0.408	0.318	0.374	0.671	0.204	1.098	0.893
NFP (2)	0.045	1.165	2.240	0.389	0.302	0.350	0.862	0.108	1.032	0.856
PB0 (0)	0.050	0.112	0.523	0.780	1.253	0.503	2.078	0.848	0.120	0.323
PB1 (1)	0.048	0.140	0.813	0.581	1.821	0.271	1.967	0.368	0.189	0.384
stEWA (1)	0.045	0.152	0.172	1.832	5.078	3.010	2.772	1.953	1.065	1.700
NE (0)	0.045	1.903	4.515	1.919	10.582	6.728	1.078	1.117	5.261	5.804
RL (2)	0.043	0.522	1.565	0.377	0.244	0.205	2.150	0.111	0.128	0.249
NNET (1)	0.046	0.355	0.345	2.220	6.962	4.309	2.827	2.469	1.433	2.236
NNET2 (2)	0.042	0.369	0.338	2.225	6.978	4.314	2.824	2.469	1.427	2.248
REL (2)	0.043	0.390	0.366	2.230	6.990	4.326	2.820	2.462	1.420	2.248
NRL (2)	0.063	0.639	1.770	2.182	1.229	4.042	1.124	1.383	2.105	2.347
Random (0)	0.045	0.373	0.348	2.228	6.972	4.310	2.824	2.472	1.429	2.241

	ERSB G8	ERSB G9	ERSB G10	M&L	Oc1	Oc4	Oc9	On	R&B10	R&B15
SFP (2)	0.323	0.293	0.435	0.593	0.434	0.790	1.134	0.166	0.269	0.587
NFP (2)	0.338	0.319	0.556	0.613	0.437	0.777	1.110	0.163	0.199	0.448
PB0 (0)	1.962	0.138	3.062	0.924	0.450	1.219	2.139	0.302	0.155	0.495
PB1 (1)	1.529	0.140	2.053	0.704	0.435	1.256	1.973	0.309	0.183	0.541
stEWA (1)	4.163	0.302	5.381	1.106	0.418	2.026	4.492	0.135	0.219	0.357
NE (0)	1.866	0.668	1.233	2.114	0.435	1.366	2.240	0.136	0.354	0.865
RL (2)	0.880	0.216	0.998	7.016	0.423	1.687	1.525	0.130	0.102	0.331
NNET (1)	4.612	0.418	6.183	2.459	0.423	1.779	3.897	0.301	0.181	0.519
NNET2 (2)	4.612	0.418	6.186	2.457	0.411	1.744	3.902	0.304	0.180	0.526
REL (2)	4.613	0.414	6.190	2.251	0.433	1.785	3.954	2.237	1.067	1.400
NRL (2)	0.516	1.298	0.934	10.602	0.471	3.334	3.695	0.091	0.350	0.279
Random (0)	4.611	0.418	6.188	2.458	0.435	1.778	3.897	1.236	5.041	5.428

	RSW D	RSW S	S&A3K	S&C G1	S&C G2	S&C G3	S&C G4	S&C G5	S&C G6	S&C G7
SFP (2)	0.659	1.143	1.621	0.571	0.061	0.425	0.227	0.207	0.088	0.468
NFP (2)	0.529	1.277	1.415	0.624	0.055	0.451	0.246	0.222	0.088	0.572
PB0 (0)	1.905	4.252	1.610	0.478	0.651	0.486	0.703	0.656	0.219	1.191
PB1 (1)	1.397	3.893	1.665	2.847	0.798	0.913	0.912	0.774	0.232	0.924
stEWA (1)	3.764	5.629	3.642	3.806	0.491	0.283	0.077	0.114	0.165	0.793
NE (0)	0.397	0.610	7.327	2.546	2.137	1.331	0.672	0.309	0.113	6.520
RL (2)	0.236	1.007	9.912	2.661	4.952	1.843	1.112	0.720	0.217	12.741
NNET (1)	3.767	3.756	2.551	10.810	4.116	9.975	5.094	2.869	0.645	7.077
NNET2 (2)	3.777	3.743	2.550	10.807	4.116	9.979	5.096	2.868	0.645	7.073
REL (2)	3.661	5.493	2.592	10.774	4.120	9.967	5.159	2.874	0.645	7.034
NRL (2)	3.799	2.985	21.957	3.431	7.541	2.002	2.264	1.664	0.414	7.421
Random (0)	3.763	5.496	2.550	10.809	4.117	9.977	5.098	2.869	0.645	7.073

	S&C G8	S&C G9	S&C G10	S&C G11	S&C G12	Mean	sd
SFP (2)	0.323	0.974	0.443	0.177	0.090	**0.569**	**0.477**
NFP (2)	0.438	1.130	0.476	0.187	0.098	**0.575**	**0.473**
PB0 (0)	1.604	1.453	0.955	0.570	0.237	**0.983**	**0.913**
PB1 (1)	1.773	1.874	0.992	0.656	0.256	**1.017**	**0.865**
stEWA (1)	0.491	0.266	0.575	0.176	0.149	**1.623**	**1.792**
NE (0)	1.388	0.418	0.785	0.410	0.095	**2.151**	**2.530**
RL (2)	1.405	15.210	2.227	1.284	0.918	**2.153**	**3.602**
NNET (1)	3.668	8.434	2.971	2.648	0.758	**3.232**	**2.798**
NNET2 (2)	3.668	8.437	2.972	2.651	0.758	**3.232**	**2.800**
REL (2)	3.696	8.452	2.985	2.656	0.773	**3.386**	**2.731**
NRL (2)	9.672	5.345	10.357	5.042	2.354	**3.563**	**4.328**
Random (0)	3.668	8.436	2.972	2.651	0.758	**3.589**	**2.730**

Table A5. This table reports the P-values for pair-wise model comparisons in terms of Best Fit scores. Simulations were run feeding models with *actual* payoffs. The null hypothesis of no differences in the mean scores was tested with a Mann-Whitney-Wilcoxon match-paired signed-rank (two-tailed) test. Shaded cells refer to the cases in which the null hypothesis is not rejected at a 5% level of significance.

	SFP (2)	NFP (2)	PB0 (0)	PB1 (1)	stEWA (1)	NE (0)	RL (2)	NNET (1)	NNET2 (2)	REL (2)	NRL (2)	Random (0)
SFP (2)		0.550	0.007	0.008	0.017	0.000	0.000	0.000	0.000	0.000	0.000	0.000
NFP (2)	0.550		0.980	0.001	0.035	0.013	0.000	0.000	0.000	0.000	0.000	0.000
PB0 (0)	0.007	0.980		0.012	0.050	0.000	0.000	0.000	0.000	0.000	0.000	0.000
PB1 (1)	0.008	0.001	0.012		0.915	0.889	0.528	0.298	0.103	0.087	0.011	0.014
stEWA (1)	0.017	0.035	0.050	0.915		0.122	0.076	0.056	0.010	0.004	0.000	0.000
NE (0)	0.000	0.013	0.000	0.889	0.122		0.056	0.045	0.016	0.010	0.003	0.004
RL (2)	0.000	0.000	0.000	0.528	0.076	0.056		0.676	0.000	0.016	0.001	0.002
NNET (1)	0.000	0.000	0.000	0.298	0.056	0.045	0.676		0.009	0.000	0.003	0.004
NNET2 (2)	0.000	0.000	0.000	0.103	0.010	0.016	0.000	0.009		0.700	0.007	0.015
REL (2)	0.000	0.000	0.000	0.087	0.004	0.010	0.016	0.000	0.700		0.014	0.021
NRL (2)	0.000	0.000	0.000	0.011	0.000	0.003	0.001	0.003	0.007	0.014		0.682
Random (0)	0.000	0.000	0.000	0.014	0.000	0.004	0.002	0.004	0.015	0.021	0.682	

Table A6. Pair-wise model comparisons in terms of Best Fit scores. Simulations were run feeding models with *actual* payoffs. Each cell of this table reports the estimate for the difference of the location parameters of *x* and *y*, where *x* is row model's vector of scores, and *y* that one of the column model.

	SFP (2)	**NFP (2)**	**PB0 (0)**	**PB1 (1)**	**stEWA (1)**	**NE (0)**	**RL (2)**	**NNET (1)**	**NNET2 (2)**	**REL (2)**	**NRL (2)**	**Random (0)**
SFP (2)		0.268	0.002	1.628	1.514	1.464	2.232	2.116	2.415	2.315	2.797	2.769
NFP (2)	-0.268		-0.011	1.151	0.885	1.018	1.537	1.682	1.697	1.823	1.982	1.998
PB0 (0)	-0.002	0.011		1.176	1.007	1.000	1.909	1.751	2.080	1.922	2.317	2.353
PB1 (1)	-1.628	-1.151	-1.176		-0.011	-0.015	0.102	0.208	0.296	0.335	0.500	0.535
stEWA (1)	-1.514	-0.885	-1.007	0.011		0.330	0.472	0.554	0.652	0.748	0.934	0.943
NE (0)	-1.464	-1.018	-1.000	0.015	-0.330		0.504	0.510	0.685	0.639	0.898	0.874
RL (2)	-2.232	-1.537	-1.909	-0.102	-0.472	-0.504		0.016	0.141	0.122	0.405	0.391
NNET (1)	-2.116	-1.682	-1.751	-0.208	-0.554	-0.510	-0.016		0.116	0.121	0.337	0.330
NNET2 (2)	-2.415	-1.697	-2.080	-0.296	-0.652	-0.685	-0.141	-0.116		0.014	0.220	0.217
REL (2)	-2.315	-1.823	-1.922	-0.335	-0.748	-0.639	-0.122	-0.121	-0.014		0.198	0.192
NRL (2)	-2.797	-1.982	-2.317	-0.500	-0.934	-0.898	-0.405	-0.337	-0.220	-0.198		0.003
Random (0)	-2.769	-1.998	-2.353	-0.535	-0.943	-0.874	-0.391	-0.330	-0.217	-0.192	-0.003	

Table A7. MSD and By Game Scores. *Rescaled* Payoffs. In the first column, between parentheses, the number of model free parameters is reported.

	AGK50	AGK67	AGK75	ERSB G1	ERSB G2	ERSB G3	ERSB G4	ERSB G5	ERSB G6	ERSB G7
SFP (2)	0.043	0.239	0.065	0.146	0.232	0.154	0.474	0.096	0.169	0.152
NFP (2)	0.040	0.366	0.342	0.141	0.214	0.152	0.489	0.092	0.174	0.143
stEWA (1)	0.035	0.064	0.162	0.194	0.399	0.196	1.022	0.145	0.085	0.178
PB1 (1)	0.038	0.071	0.199	0.541	1.377	0.229	1.960	0.234	0.111	0.246
PB0 (0)	0.037	0.079	0.399	0.713	1.225	0.446	2.066	0.720	0.113	0.255
RL (2)	0.028	0.051	0.331	0.166	0.233	2.931	1.540	0.103	0.073	0.188
NE (0)	0.045	1.903	4.515	1.919	10.582	6.728	1.078	1.117	5.261	5.804
NRL (2)	0.024	0.147	0.493	2.305	0.722	2.850	0.852	1.604	1.307	1.644
NNET (1)	0.036	0.362	0.335	2.220	6.957	4.301	2.816	2.462	1.417	2.229
NNET2 (2)	0.032	0.354	0.332	2.215	6.955	4.287	2.805	2.458	1.412	2.228
REL (2)	0.032	0.357	0.333	2.196	6.928	4.279	2.795	2.452	1.414	2.207

	ERSB G8	ERSB G9	ERSB G10	M&L	Oc1	Oc4	Oc9	On	R&B10	R&B15
SFP (2)	0.315	0.111	0.485	0.306	0.383	0.562	0.726	0.038	0.133	0.234
NFP (2)	0.328	0.108	0.528	0.295	0.390	0.543	0.696	0.059	0.116	0.134
stEWA (1)	0.444	0.108	0.077	0.290	0.378	0.852	1.582	0.115	0.128	0.359
PB1 (1)	1.343	0.120	1.867	0.313	0.415	0.860	1.319	0.302	0.176	0.528
PB0 (0)	1.822	0.126	2.869	0.606	0.435	1.005	1.530	0.303	0.170	0.485
RL (2)	0.238	0.098	0.123	12.469	0.358	0.435	0.791	0.166	0.090	0.283
NE (0)	1.866	0.668	1.233	2.114	0.435	1.366	2.240	0.136	0.354	0.865
NRL (2)	0.157	1.464	0.948	11.089	0.353	0.504	0.999	0.064	0.089	0.295
NNET (1)	4.598	0.415	6.178	2.428	0.394	1.750	3.848	0.292	0.177	0.435
NNET2 (2)	4.591	0.407	6.162	2.456	0.406	1.721	3.802	0.300	0.175	0.517
REL (2)	4.571	0.407	6.141	1.411	0.396	1.729	3.821	2.208	1.050	1.351

	RSW D	RSW S	S&A3K	S&C G1	S&C G2	S&C G3	S&C G4	S&C G5	S&C G6	S&C G7
SFP (2)	0.397	1.015	0.392	0.080	0.058	0.055	0.021	0.038	0.041	0.389
NFP (2)	0.562	1.288	0.435	0.098	0.052	0.082	0.029	0.041	0.041	0.411
stEWA (1)	2.762	4.488	3.290	0.666	0.351	0.213	0.045	0.064	0.095	0.148
PB1 (1)	1.546	2.872	0.102	1.952	0.426	0.586	0.575	0.478	0.110	0.676
PB0 (0)	1.524	3.860	1.633	0.304	0.422	0.351	0.476	0.416	0.115	1.345
RL (2)	0.185	0.340	6.305	2.205	2.356	1.750	1.245	0.904	0.387	5.882
NE (0)	0.397	0.610	7.327	2.546	2.137	1.331	0.672	0.309	0.113	6.520
NRL (2)	1.679	3.141	7.717	3.602	3.301	1.829	2.509	2.370	0.619	7.414
NNET (1)	3.737	3.744	2.549	10.776	4.101	9.948	5.066	2.852	0.642	7.053
NNET2 (2)	3.737	3.699	2.544	10.795	4.099	9.949	5.084	2.848	0.642	7.064
REL (2)	3.570	5.313	2.542	10.723	4.072	9.898	5.035	2.819	0.622	7.033

	S&C G8	S&C G9	S&C G10	S&C G11	S&C G12	Mean	sd
SFP (2)	0.342	0.217	0.296	0.063	0.032	**0.243**	**0.221**
NFP (2)	0.354	0.306	0.328	0.060	0.031	**0.270**	**0.255**
stEWA (1)	0.220	0.258	0.344	0.083	0.027	**0.568**	**0.991**
PB1 (1)	1.392	1.359	0.776	0.487	0.111	**0.734**	**0.697**
PB0 (0)	1.305	1.256	0.759	0.422	0.112	**0.849**	**0.847**
RL (2)	1.129	1.309	1.822	1.500	1.263	**1.408**	**2.420**
NE (0)	1.388	0.418	0.785	0.410	0.095	**2.151**	**2.530**
NRL (2)	7.836	1.310	4.235	5.750	3.408	**2.418**	**2.642**
NNET (1)	3.643	8.429	2.943	2.632	0.747	**3.215**	**2.796**
NNET2 (2)	3.662	8.427	2.962	2.641	0.754	**3.215**	**2.795**
REL (2)	3.615	8.352	2.930	2.617	0.730	**3.313**	**2.729**

Table A8. This table reports the P-values for pair-wise model comparisons in terms of By Game scores. Simulations were run feeding models with *rescaled* payoffs. The null hypothesis of no differences in the mean scores was tested with a Mann-Whitney-Wilcoxon match-paired signed-rank (two-tailed) test. Shaded cells refer to the cases in which the null hypothesis is not rejected at a 5% level of significance.

	SFP (2)	NFP (2)	stEWA (1)	PB1 (1)	PB0 (0)	RL (2)	NE (0)	NRL (2)	NNET (1)	NNET2 (2)	REL (2)
SFP (2)		0.054	0.004	0.000	0.000	0.002	0.000	0.000	0.000	0.000	0.000
NFP (2)	0.054		0.018	0.000	0.000	0.003	0.000	0.000	0.000	0.000	0.000
stEWA (1)	0.004	0.018		0.001	0.001	0.030	0.000	0.000	0.000	0.000	0.000
PB1 (1)	0.000	0.000	0.001		0.126	0.688	0.005	0.001	0.000	0.000	0.000
PB0 (0)	0.000	0.000	0.001	0.126		0.737	0.004	0.004	0.000	0.000	0.000
RL (2)	0.002	0.003	0.030	0.688	0.737		0.045	0.000	0.000	0.000	0.000
NE (0)	0.000	0.000	0.000	0.005	0.004	0.045		0.390	0.054	0.052	0.038
NRL (2)	0.000	0.000	0.000	0.001	0.004	0.000	0.390		0.061	0.061	0.045
NNET (1)	0.000	0.000	0.000	0.000	0.000	0.000	0.054	0.061		0.647	0.003
NNET2 (2)	0.000	0.000	0.000	0.000	0.000	0.000	0.052	0.061	0.647		0.016
REL (2)	0.000	0.000	0.000	0.000	0.000	0.000	0.038	0.045	0.003	0.016	

Table A9. Pair-wise model comparisons in terms of By Game scores. Simulations were run feeding models with *rescaled* payoffs. Each cell of this table reports the estimate for the difference of the location parameters of *x* and *y*, where *x* is row model's vector of scores, and *y* that one of the column model.

	SFP (2)	**NFP (2)**	**stEWA (1)**	**PB1 (1)**	**PB0 (0)**	**RL (2)**	**NE (0)**	**NRL (2)**	**NNET (1)**	**NNET2 (2)**	**REL (2)**
SFP (2)		-0.009	-0.075	-0.412	-0.501	-0.613	-1.161	-1.659	-2.598	-2.586	-2.661
NFP (2)	0.009		-0.060	-0.401	-0.484	-0.602	-1.142	-1.640	-2.549	-2.557	-2.630
stEWA (1)	0.075	0.060		-0.215	-0.224	-0.531	-1.020	-1.404	-2.197	-2.191	-2.273
PB1 (1)	0.412	0.401	0.215		-0.059	-0.081	-0.694	-1.136	-2.156	-2.165	-2.189
PB0 (0)	0.501	0.484	0.224	0.059		-0.116	-0.748	-1.161	-1.902	-1.906	-1.963
RL (2)	0.613	0.602	0.531	0.081	0.116		-0.454	-0.766	-1.684	-1.665	-1.733
NE (0)	1.161	1.142	1.020	0.694	0.748	0.454		-0.314	-0.996	-0.999	-1.108
NRL (2)	1.659	1.640	1.404	1.136	1.161	0.766	0.314		-0.544	-0.544	-0.883
NNET (1)	2.598	2.549	2.197	2.156	1.902	1.684	0.996	0.544		0.001	0.019
NNET2 (2)	2.586	2.557	2.191	2.165	1.906	1.665	0.999	0.544	-0.001		0.015
REL (2)	2.661	2.630	2.273	2.189	1.963	1.733	1.108	0.883	-0.019	-0.015	

Table A10. MSD and Best Fit Scores. *Rescaled* Payoffs. In the first column, between parentheses, the number of model free parameters is reported.

	AGK50	AGK67	AGK75	ERSB G1	ERSB G2	ERSB G3	ERSB G4	ERSB G5	ERSB G6	ERSB G7
NFP (2)	0.046	1.116	1.277	0.349	0.283	0.411	0.999	0.092	0.900	0.810
SFP (2)	0.043	1.073	1.315	0.407	0.313	0.458	0.846	0.096	1.085	0.887
PB0 (0)	0.037	0.079	0.399	0.713	1.225	0.446	2.066	0.720	0.113	0.255
PB1 (1)	0.046	0.105	0.633	0.541	1.765	0.229	1.960	0.234	0.201	0.312
stEWA (1)	0.035	0.080	0.428	0.872	1.808	0.665	2.524	0.845	0.382	0.491
RL (2)	0.038	0.394	1.296	0.383	0.236	4.451	2.119	0.103	0.101	0.196
NE (0)	0.045	1.903	4.515	1.919	10.582	6.728	1.078	1.117	5.261	5.804
NNET2 (2)	0.038	0.373	0.334	2.220	6.978	4.309	2.817	2.472	1.432	2.236
NNET (1)	0.046	0.365	0.363	2.221	6.976	4.301	2.829	2.469	1.422	2.246
REL (2)	0.040	0.395	0.374	2.245	7.031	4.337	2.795	2.464	1.443	2.248
NRL (2)	0.039	0.462	1.377	2.423	1.199	3.945	0.933	1.696	2.251	2.323

	ERSB G8	ERSB G9	ERSB G10	M&L	Oc1	Oc4	Oc9	On	R&B10	R&B15
NFP (2)	0.355	0.302	0.877	0.451	0.431	0.623	0.864	0.148	0.132	0.264
SFP (2)	0.338	0.442	0.663	0.595	0.440	0.630	0.904	0.167	0.301	0.562
PB0 (0)	1.822	0.126	2.869	0.606	0.435	1.005	1.530	0.303	0.170	0.485
PB1 (1)	1.343	0.152	1.867	0.359	0.434	0.866	1.459	0.311	0.181	0.536
stEWA (1)	2.843	0.123	3.055	0.315	0.433	1.474	2.420	0.126	0.171	0.403
RL (2)	0.792	0.222	1.029	12.469	0.425	1.220	1.339	0.166	0.115	0.321
NE (0)	1.866	0.668	1.233	2.114	0.435	1.366	2.240	0.136	0.354	0.865
NNET2 (2)	4.606	0.422	6.175	2.462	0.413	1.721	3.892	0.307	0.181	0.535
NNET (1)	4.609	0.422	6.184	2.458	0.435	1.769	3.898	0.298	0.186	0.517
REL (2)	4.639	0.424	6.168	1.413	0.446	1.750	3.915	2.299	1.077	1.383
NRL (2)	0.496	1.528	0.979	11.089	0.488	2.606	3.538	0.071	0.347	0.295

	RSW D	RSW S	S&A3K	S&C G1	S&C G2	S&C G3	S&C G4	S&C G5	S&C G6	S&C G7
NFP (2)	0.837	1.788	1.252	0.636	0.056	0.422	0.201	0.150	0.049	0.587
SFP (2)	0.731	1.610	1.646	0.421	0.084	0.337	0.144	0.108	0.046	0.513
PB0 (0)	1.524	3.860	1.633	0.304	0.422	0.351	0.476	0.416	0.115	1.345
PB1 (1)	1.622	2.957	1.612	2.221	0.494	0.586	0.580	0.504	0.136	0.905
stEWA (1)	2.829	4.824	6.460	5.809	1.524	0.743	1.134	0.546	0.247	0.863
RL (2)	0.199	0.569	9.340	2.205	4.792	1.804	1.245	0.904	0.387	5.882
NE (0)	0.397	0.610	7.327	2.546	2.137	1.331	0.672	0.309	0.113	6.520
NNET2 (2)	3.767	3.771	2.551	10.807	4.115	9.987	5.094	2.874	0.646	7.075
NNET (1)	3.754	3.744	2.549	10.809	4.116	9.978	5.097	2.871	0.646	7.074
REL (2)	3.607	5.348	2.607	10.877	4.115	9.986	5.112	2.850	0.640	7.152
NRL (2)	2.135	3.355	20.221	3.602	7.527	1.980	2.509	2.481	0.670	7.419

	S&C G8	S&C G9	S&C G10	S&C G11	S&C G12	Mean	sd
NFP (2)	0.394	1.088	0.363	0.142	0.047	**0.535**	**0.429**
SFP (2)	0.348	0.836	0.297	0.122	0.041	**0.539**	**0.427**
PB0 (0)	1.305	1.256	0.759	0.422	0.112	**0.849**	**0.847**
PB1 (1)	1.517	1.724	0.788	0.487	0.131	**0.851**	**0.738**
stEWA (1)	1.178	0.485	1.167	0.395	0.066	**1.365**	**1.609**
RL (2)	1.129	9.397	1.822	1.523	1.263	**1.996**	**2.972**
NE (0)	1.388	0.418	0.785	0.410	0.095	**2.151**	**2.530**
NNET2 (2)	3.669	8.439	2.966	2.653	0.765	**3.231**	**2.800**
NNET (1)	3.669	8.434	2.974	2.650	0.758	**3.232**	**2.797**
REL (2)	3.701	8.448	2.958	2.668	0.740	**3.363**	**2.756**
NRL (2)	9.890	5.080	8.531	6.779	3.408	**3.534**	**4.090**

Table A11. This table reports the P-values for pair-wise model comparisons in terms of Best Fit scores. Simulations were run feeding models with *rescaled* payoffs. The null hypothesis of no differences in the mean scores was tested with a Mann-Whitney-Wilcoxon match-paired signed-rank (two-tailed) test. Shaded cells refer to the cases in which the null hypothesis is not rejected at a 5% level of significance.

	NFP (2)	SFP (2)	PB0 (0)	PB1 (1)	stEWA (1)	RL (2)	NE (0)	NNET2 (2)	NNET (1)	REL (2)	NRL (2)
NFP (2)		0.863	0.005	0.002	0.001	0.002	0.000	0.000	0.000	0.000	0.000
SFP (2)	0.863		0.019	0.009	0.003	0.005	0.000	0.000	0.000	0.000	0.000
PB0 (0)	0.005	0.019		0.700	0.003	0.081	0.004	0.000	0.000	0.000	0.000
PB1 (1)	0.002	0.009	0.700		0.007	0.107	0.010	0.000	0.000	0.000	0.000
stEWA (1)	0.001	0.003	0.003	0.007		0.617	0.147	0.000	0.000	0.000	0.003
RL (2)	0.002	0.005	0.081	0.107	0.617		0.456	0.003	0.002	0.001	0.000
NE (0)	0.000	0.000	0.004	0.010	0.147	0.456		0.050	0.048	0.033	0.050
NNET2 (2)	0.000	0.000	0.000	0.000	0.000	0.003	0.050		0.925	0.004	0.967
NNET (1)	0.000	0.000	0.000	0.000	0.000	0.002	0.048	0.925		0.026	0.954
REL (2)	0.000	0.000	0.000	0.000	0.000	0.001	0.033	0.004	0.026		0.676
NRL (2)	0.000	0.000	0.000	0.000	0.003	0.000	0.050	0.967	0.954	0.676	

Table A12. Pair-wise model comparisons in terms of Best Fit scores. Simulations were run feeding models with *rescaled* payoffs. Each cell of this table reports the estimate for the difference of the location parameters of *x* and *y*, where *x* is row model's vector of scores, and *y* that one of the column model.

	NFP (2)	**SFP (2)**	**PB0 (0)**	**PB1 (1)**	**stEWA (1)**	**RL (2)**	**NE (0)**	**NNET2 (2)**	**NNET (1)**	**REL (2)**	**NRL (2)**
NFP (2)		0.004	-0.268	-0.287	-0.575	-0.637	-0.967	-2.328	-2.329	-2.484	-1.983
SFP (2)	-0.004		-0.245	-0.270	-0.597	-0.629	-1.004	-2.346	-2.340	-2.457	-2.016
PB0 (0)	0.268	0.245		-0.014	-0.235	-0.436	-0.748	-1.918	-1.923	-1.996	-1.848
PB1 (1)	0.287	0.270	0.014		-0.282	-0.298	-0.652	-2.066	-2.085	-2.156	-1.737
stEWA (1)	0.575	0.597	0.235	0.282		-0.083	-0.278	-1.542	-1.534	-1.633	-1.442
RL (2)	0.637	0.629	0.436	0.298	0.083		-0.176	-1.205	-1.205	-1.364	-1.106
NE (0)	0.967	1.004	0.748	0.652	0.278	0.176		-1.004	-1.011	-1.154	-0.915
NNET2 (2)	2.328	2.346	1.918	2.066	1.542	1.205	1.004		0.000	-0.020	0.016
NNET (1)	2.329	2.340	1.923	2.085	1.534	1.205	1.011	0.000		-0.016	0.010
REL (2)	2.484	2.457	1.996	2.156	1.633	1.364	1.154	0.020	0.016		0.185
NRL (2)	1.983	2.016	1.848	1.737	1.442	1.106	0.915	-0.016	-0.010	-0.185	

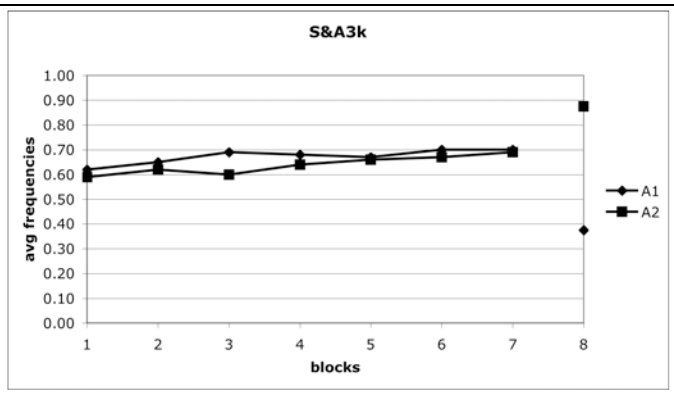
2.9 Appendix B. The Dataset Description

2.9.1 Suppes and Atkinson (1960)

In this experiment, 20 pairs of subjects played this game for 210 times. The payoff matrix was known to the subjects. The authors presented the data they gathered in 7 blocks of 30 repetitions of the stage game. Table B1 shows the matrix of payoffs and experimentally observed frequencies of choice.

Table B1. Suppes and Atkinson's (1960) game and empirical data.

Player 2 / Player 1	A2	B2
A1	(3,7)	(8,2)
B1	(4,6)	(1,9)

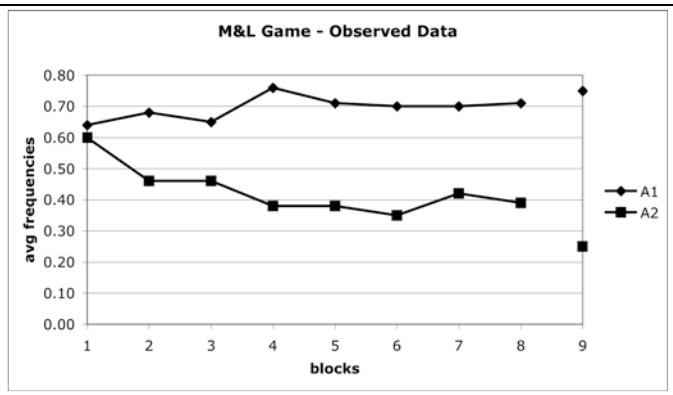


2.9.2 Malcolm and Lieberman (1965)

Nine pairs of subjects played this game for 200 rounds. The payoff matrix was known to the subjects. The authors presented the choices' relative frequencies in 8 blocks of 25 trials. Table B2 shows the matrix of payoffs and reports experimentally observed frequencies of choice.

Table B2. Malcolm and Lieberman's (1965) game and empirical data.

Player 2 / Player 1	A2	B2
A1	(3,-3)	(-1,1)
B1	(-9,9)	(3,-3)



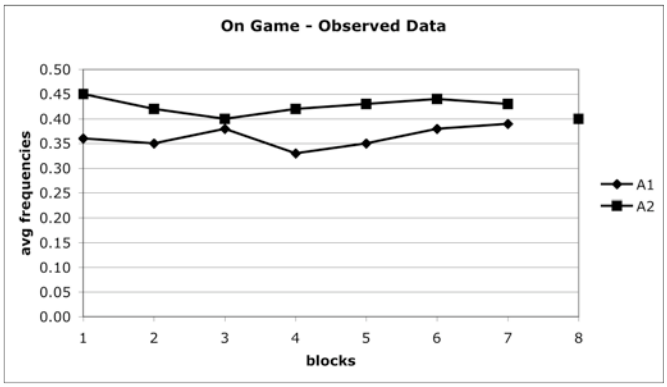
2.9.3 O'Neill (1987)

Experimental settings and data: 20 pairs of subjects participated in 105 replications of this zero-sum 4x4 game. Strategies B, C and D are symmetrical for both players. The author presented the relative frequency with which strategy A was played in 7 blocks of 15 iterations. Table B3 shows the matrix of payoffs and reports experimentally observed frequencies of choice.

Equilibrium prediction: According to the unique mixed strategy equilibrium of this game, both players are expected to choose A with probability 0.4 and each of the other strategies with probability 0.2.

Table B3. O'Neill's (1987) game and empirical data.

Player 2 / Player 1	A2	B2	C2	D2
A1	+5	-5	-5	-5
B1	-5	-5	+5	+5
C1	-5	+5	-5	+5
D1	-5	+5	+5	-5



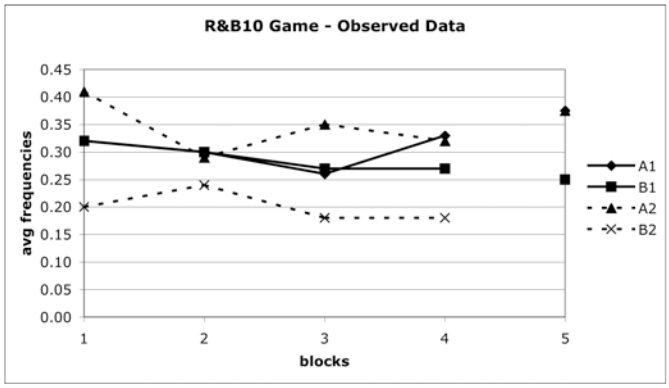
2.9.4 Rapoport and Boebel (1992)

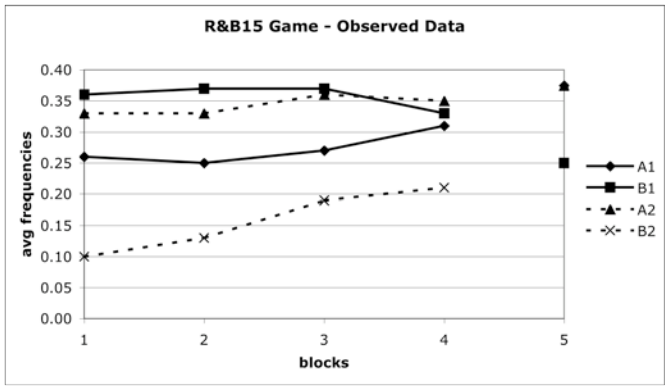
Experimental settings and data: Both those constant-sum games were played in two sessions. In the first session, 10 pairs of subjects played the game for 120 rounds, while in the second session subjects exchanged roles and played another 120 rounds of the game. Here are presented only data gathered in the first sessions. The authors presented the proportions of A and B choices in 4 blocks of 30 trials. Table B4 shows the matrix of payoffs and reports experimentally observed frequencies of choice.

Equilibrium prediction: According to the unique mixed strategy equilibrium of this game, both players are expected to choose A with probability $3/8$, B with probability $2/8$ and each of the other (symmetrical) strategies with equal probability $(1/8)$.

Table B4. Rapoport and Boebel's (1992) games and empirical data.

Player 2 / Player 1	A2	B2	C2	D2	E2
A1	W	L	L	L	L
B1	L	L	W	W	W
C1	L	W	L	L	W
D1	L	W	L	W	L
E1	L	W	W	L	L





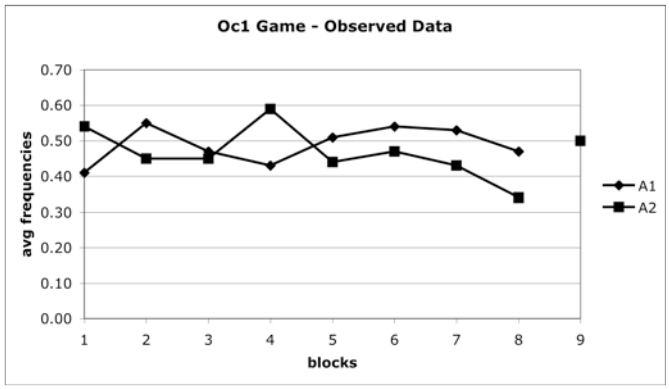
2.9.5 Ochs (1995)

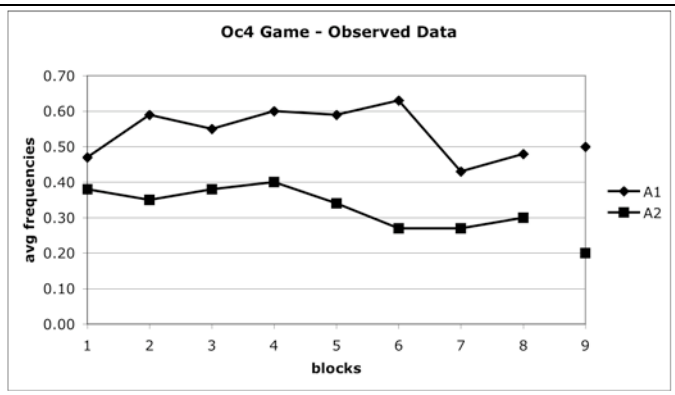
Experimental settings and data: This experiment by Ochs developed in three sessions in which subjects were asked to state at each round the frequency of A choices that they wished to make in the next 10 games. In the first, 8 pairs of subjects played for 64 rounds game Oc1 with $a = b = 1$; in the second, 8 pairs of subjects played for 16 rounds game Oc1 and for 56 rounds game Oc9, with $a = 9$ and $b = 1$; finally, in section three, 8 pairs of subjects played for 16 rounds game Oc1 and for 64 rounds game Oc4, with $a = 4$ and $b = 1$. The author presented the average frequencies of choice A in blocks of 80 trials (80 games). Table B5 shows the matrix of payoffs and reports experimentally observed frequencies of choice.

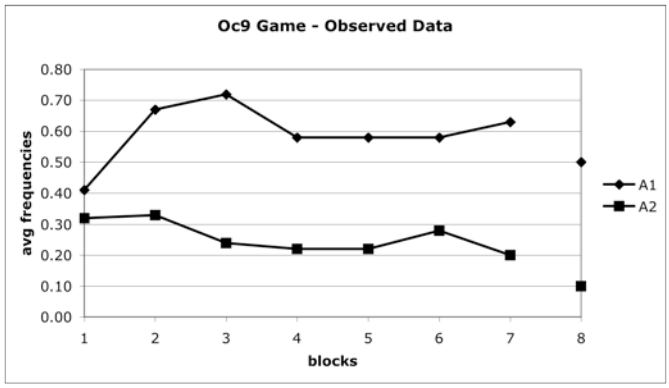
Equilibrium prediction: In the unique mixed strategy equilibrium of the game, player 1 chooses A1 with probability $p = 1/2$ and player 2 chooses A2 with probability $q = b/(a+b)$.

Table B5. Ochs's (1995) games and empirical data.

Player 2 / Player 1	A2	B2
A1	(a,0)	(0,b)
B1	(0,b)	(b,0)





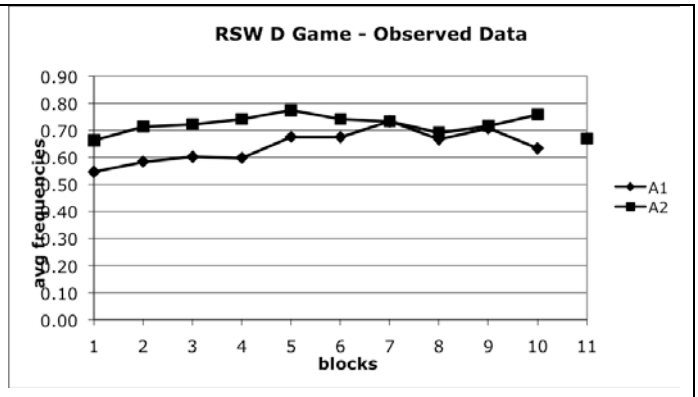


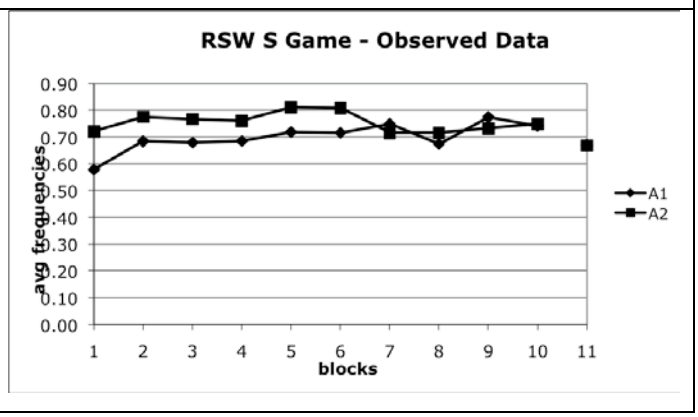
2.9.6 Rosenthal, Shachat, and Walker (2003)

The authors considered two versions of a zero sum game with the same matrix representation, called Deterministic Game (D) and Stochastic Game (S), respectively. In the first version game payoffs are deterministic, in the second stochastic. Six pairs of subjects played repeatedly game D 200 times and other twenty pairs played game S 200 times in a fix pairing protocol. Players were informed about the structure of the game and, at each round, were given full feedback about their own actions and their opponents'. Table B6 shows the matrix of payoffs and reports experimentally observed frequencies of choice.

Table B6. Rosenthal, Shachat, and Walker's (2003) games and empirical data.

Player 2 / Player 1	A2	B2
A1	(1,-1)	(0,0)
B1	(0,0)	(2,-2)





2.9.7 Avrahami, Güth and Kareev (2005)

In three different experimental sessions, three groups of subjects (for a total of 60 participants) played for 100 times a game that the authors named Parasite Game.

This game involves two players and an indifferent nature. Nature moves first and decides where a resource for player 1 becomes available (H or T, which stands for two different locations). Then, T and H are also the locations where player 1 can search for the resource and where player 2 can steal it from player 1. So, success for player 1 means to guess nature's move but not to be outguessed by player 2; for player 2, instead, it means to outguess player 1 when player 1 has guessed nature.

Two different protocols were considered: one in which w was initially declared to participants (labeled with KNOWL=1) and one in which w was unknown by participants (labeled with KNOWL=0). We used data gathered in the three session using protocol KNOWL=0.

In the first session, 6 pairs of subjects played the Parasite Game with $w = 1/2$ (we labeled with AGK50); in the second, 12 pairs played the same game, but with $w = 2/3$ (AGK67); finally, in the third session 12 pairs played the Parasite Game with $w = 3/4$

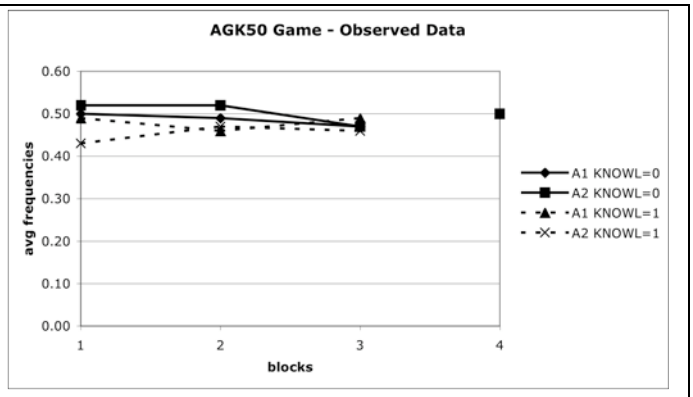
(AGK75). In each group half of the pairs were in the known and the other half in the unknown treatment.

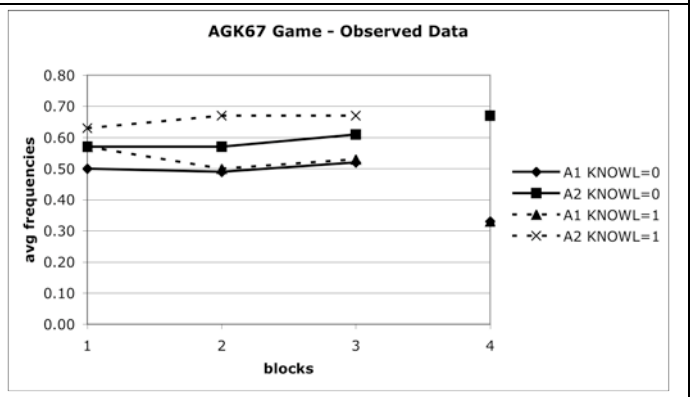
The authors presented the data in 3 blocks of 30 iterations – the first and the last 5 rounds were excluded from analysis.

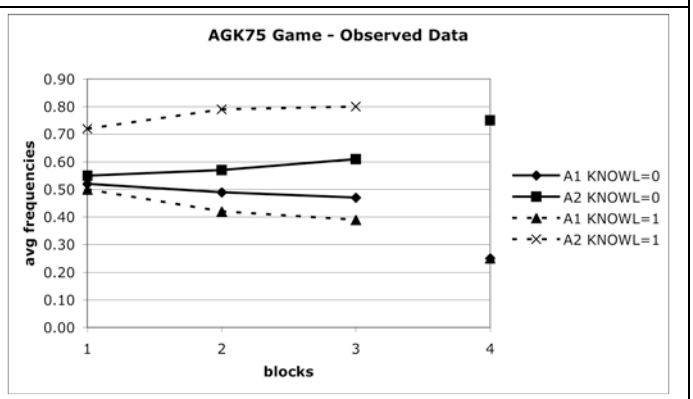
Equilibrium prediction: The unique mixed strategy equilibrium of the game predicts that player 1 chooses strategy H with probability $p = 1 - w$ and player 2 chooses H with probability $q = w$. Table B7 shows the payoffs of the normal form games and reports experimentally observed frequencies of choice.

Table B7. Avrahami, Güth, and Kareev's (2005) games and empirical data.

Player 2 / Player 1	A2	B2
A1	(0,w)	(w,0)
B1	(1-w,0)	(0,1-w)





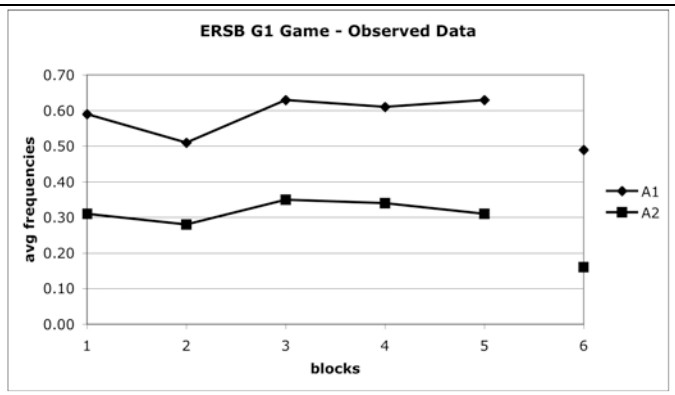


2.9.8 Erev, Roth, Slonim and Barron (2007)

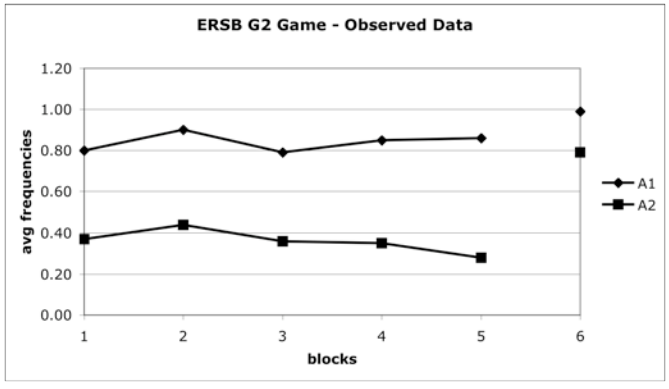
Nine pairs of subjects played for 500 rounds ten randomly selected games. The numbers in each matrix represents probabilities that the players will win a fixed amount w on each trial. For example, if on a certain trial both players choose A, then player 1 will win w with the specified probability p_1 and player 2 will win w with probability $1-p_1$. A player who does not win w earns zero for that period. Subjects knew the probabilities that define the game. The authors presented the average frequencies of choice A in 5 blocks of 100 trials. Table B8 shows the matrices of payoffs and reports experimentally observed frequencies of choice.

Table B8. Erev, Roth, Slonim, and Barron's (2007) games and empirical data.

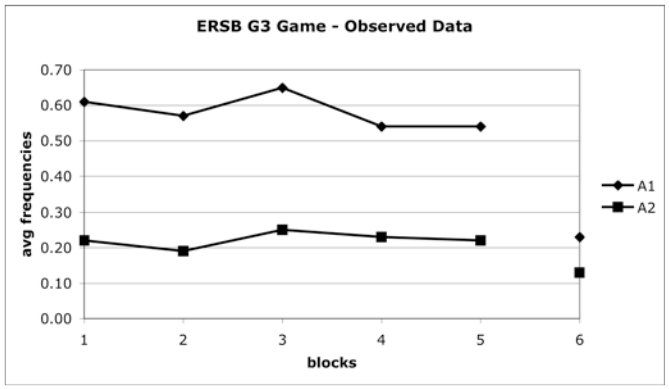
Player 2 / Player 1	A2	B2
A1	0.77	0.35
B1	0.08	0.48



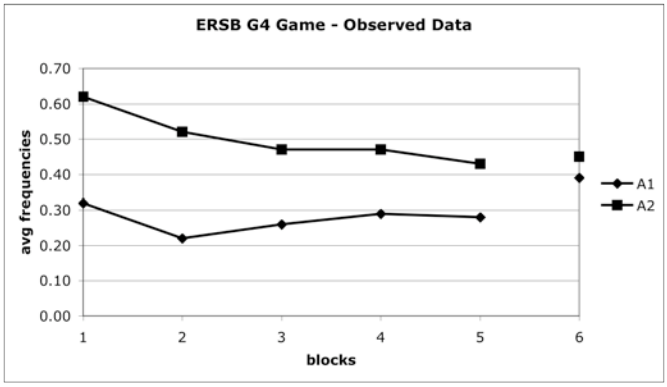
Player 2 / Player 1	A2	B2
A1	0.73	0.74
B1	0.87	0.20



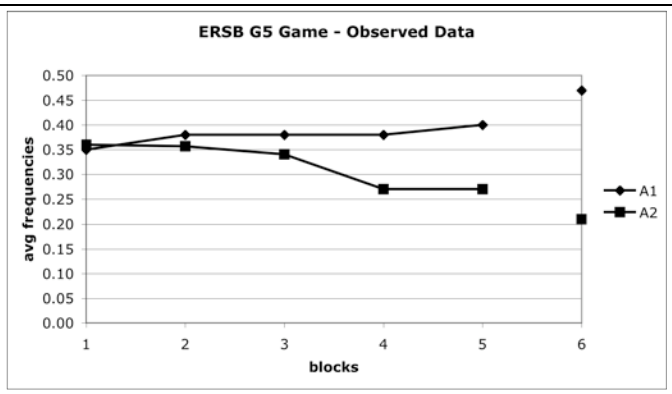
Player 2 / Player 1	A2	B2
A1	0.63	0.08
B1	0.01	0.17



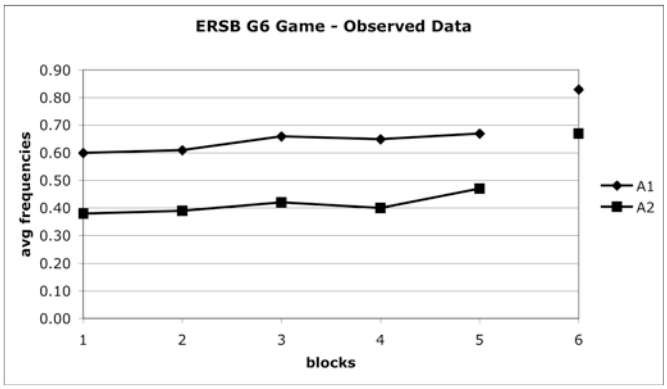
Player 2 / Player 1	A2	B2
A1	0.55	0.75
B1	0.73	0.60



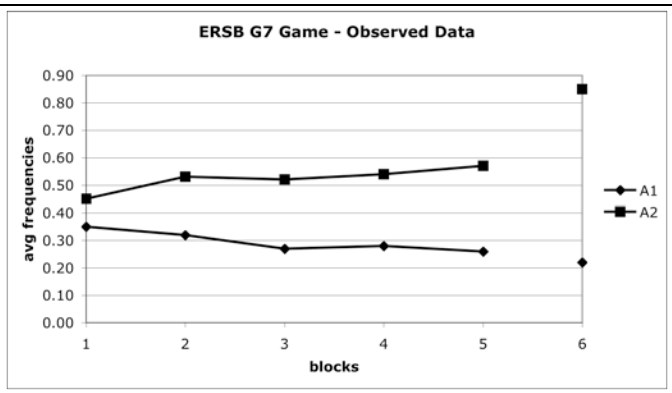
Player 2 / Player 1	A2	B2
A1	0.05	0.64
B1	0.93	0.40



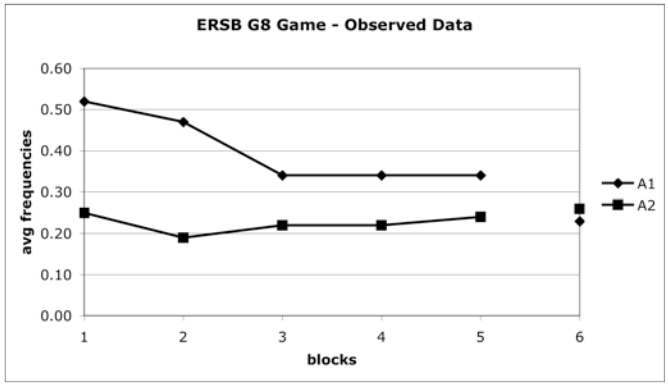
Player 2 / Player 1	A2	B2
A1	0.46	0.54
B1	0.61	0.23



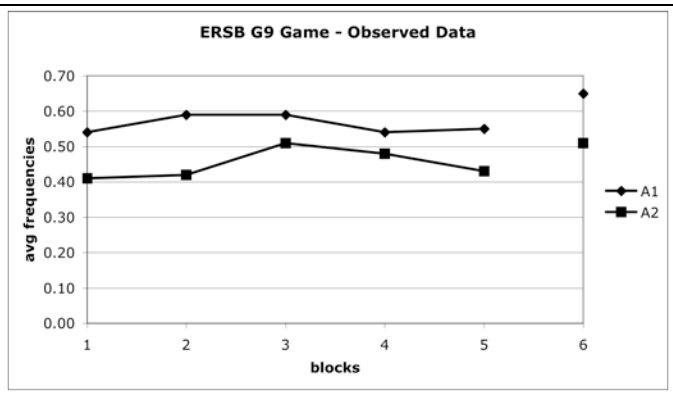
Player 2 / Player 1	A2	B2
A1	0.89	0.53
B1	0.82	0.92



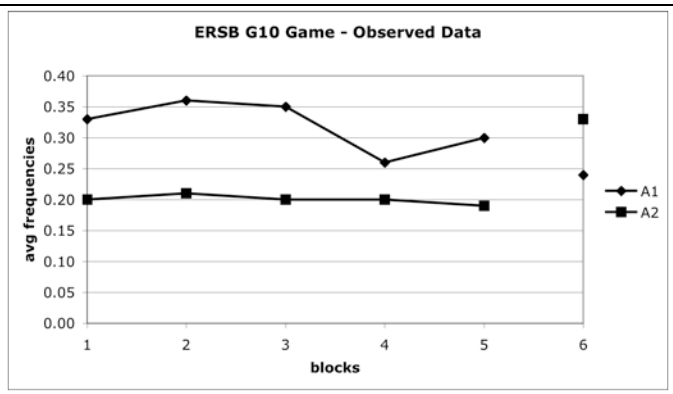
Player 2 / Player 1	A2	B2
A1	0.88	0.38
B1	0.40	0.55



Player 2 / Player 1	A2	B2
A1	0.40	0.76
B1	0.91	0.23



Player 2 / Player 1	A2	B2
A1	0.69	0.05
B1	0.13	0.33

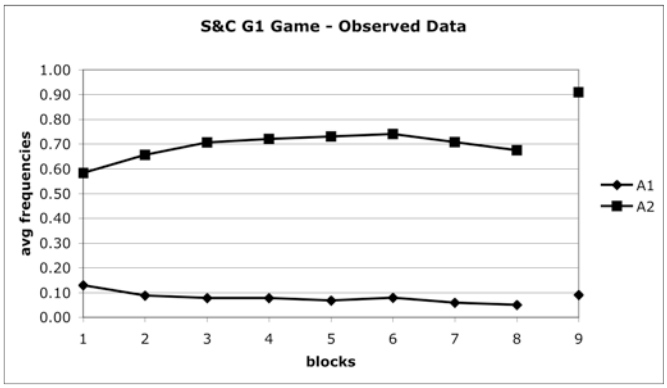


2.9.9 Selten and Chmura (2008)

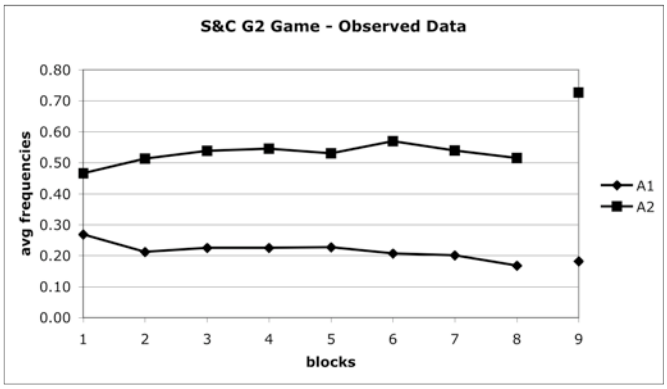
The authors ran experiments on 6 constant sum games and 6 non-constant sum games. Each of the first 6 constant sum games was played for 200 times by 12 groups of 8 subjects each (random matching protocol) and each of the other non-constant sum games was played for 200 times by 6 groups of 8 subjects each (random matching protocol). The average frequencies of choice A are presented in 8 blocks of 25 trials each. Table B9 shows the matrices of payoffs and reports experimentally observed frequencies of choice.

Table B9. Selten and Chmura's (2008) games and empirical data.

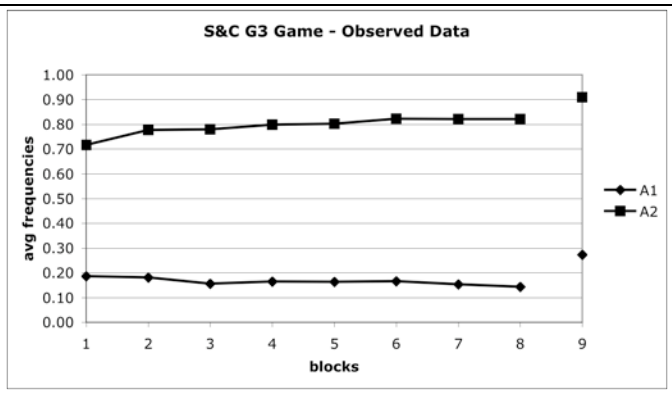
Player 2 / Player 1	A2	B2
A1	(10,8)	(0,18)
B1	(9,9)	(10,8)



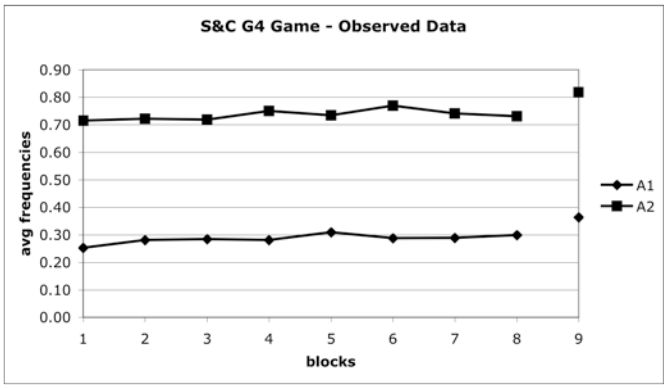
Player 2 / Player 1	A2	B2
A1	(9,4)	(0,13)
B1	(6,7)	(8,5)



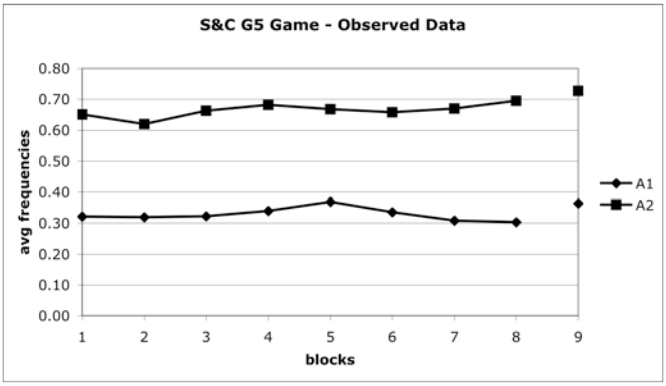
Player 2 / Player 1	A2	B2
A1	(8,6)	(0,14)
B1	(7,7)	(10,4)



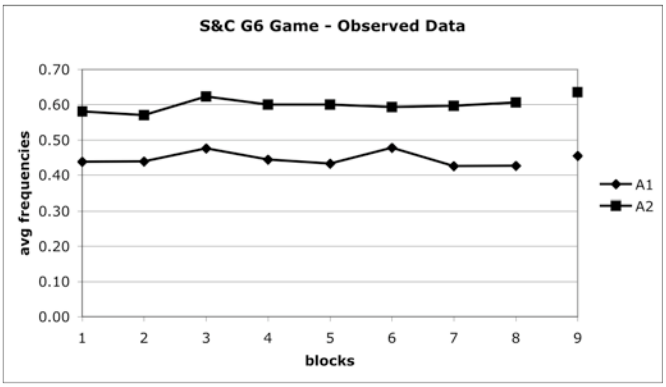
Player 2 / Player 1	A2	B2
A1	(7,4)	(0,11)
B1	(5,6)	(9,2)



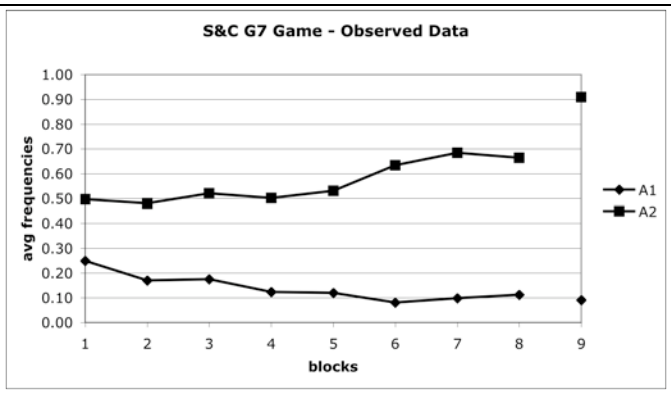
Player 2 / Player 1	A2	B2
A1	(7,2)	(0,9)
B1	(4,5)	(8,1)



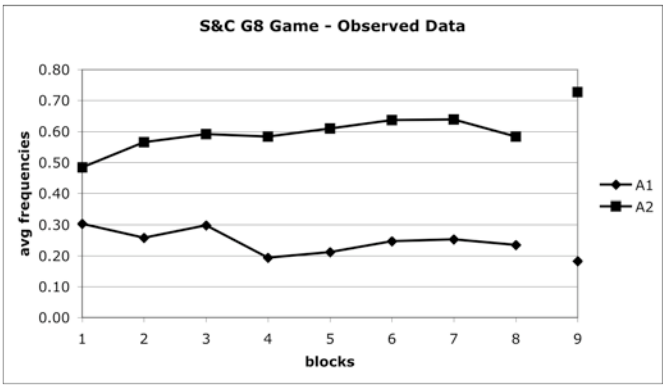
Player 2 / Player 1	A2	B2
A1	(7,1)	(1,7)
B1	(3,5)	(8,0)



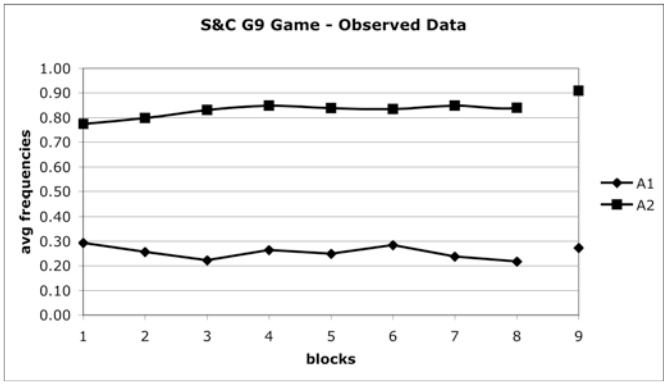
Player 2 / Player 1	A2	B2
A1	(10,12)	(4,22)
B1	(9,9)	(14,8)



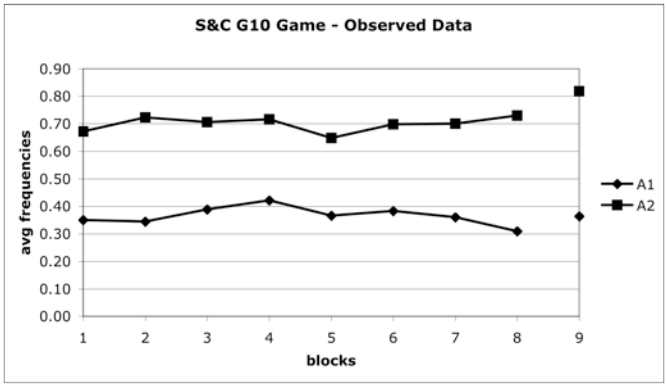
Player 2 / Player 1	A2	B2
A1	(9,7)	(3,16)
B1	(6,7)	(11,5)



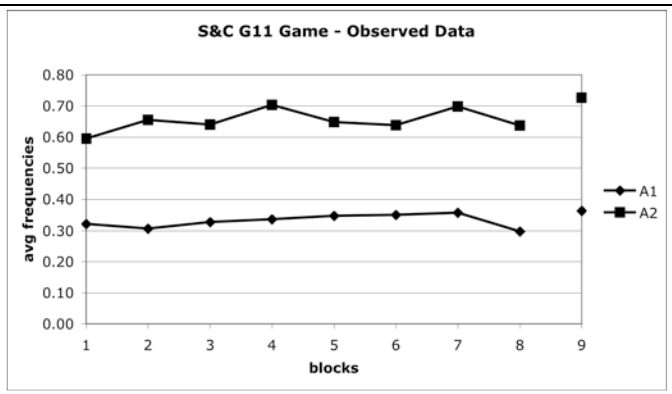
Player 2 / Player 1	A2	B2
A1	(8,9)	(3,17)
B1	(7,7)	(13,4)



Player 2 / Player 1	A2	B2
A1	(7,6)	(2,13)
B1	(5,6)	(11,2)



Player 2 / Player 1	A2	B2
A1	(7,4)	(2,11)
B1	(4,5)	(10,1)



Player 2 / Player 1	A2	B2
A1	(7,3)	(3,9)
B1	(3,5)	(10,0)

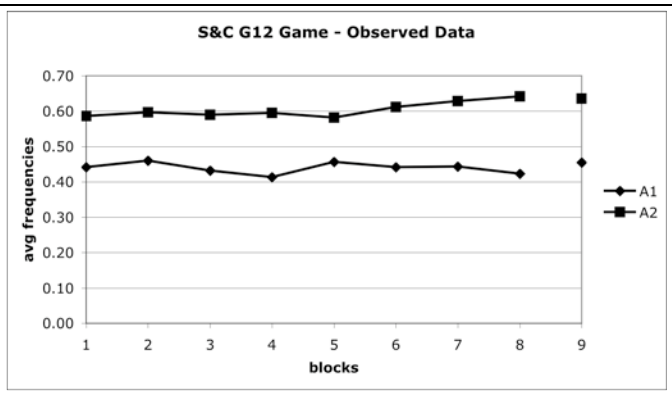


Table B10. Summary of empirical frequencies of play in the 35 games.

		BLOCKS										
GAME	**ACTION**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**NE**
AGK50	A1	0.50	0.49	0.47								**0.500**
AGK50	A2	0.52	0.52	0.47								**0.500**
AGK67	A1	0.50	0.49	0.52								**0.330**
AGK67	A2	0.57	0.57	0.61								**0.670**
AGK75	A1	0.52	0.49	0.47								**0.250**
AGK75	A2	0.55	0.57	0.61								**0.750**
AGK50k	A1	0.49	0.46	0.49								**0.500**
AGK50k	A2	0.43	0.47	0.46								**0.500**
AGK67k	A1	0.57	0.50	0.53								**0.330**
AGK67k	A2	0.63	0.67	0.67								**0.670**
AGK75k	A1	0.50	0.42	0.39								**0.250**
AGK75k	A2	0.72	0.79	0.80								**0.750**
ERSB G1	A1	0.59	0.51	0.63	0.61	0.63						**0.490**
ERSB G1	A2	0.31	0.28	0.35	0.34	0.31						**0.160**
ERSB G2	A1	0.80	0.90	0.79	0.85	0.86						**0.990**
ERSB G2	A2	0.37	0.44	0.36	0.35	0.28						**0.790**
ERSB G3	A1	0.61	0.57	0.65	0.54	0.54						**0.230**
ERSB G3	A2	0.22	0.19	0.25	0.23	0.22						**0.130**
ERSB G4	A1	0.32	0.22	0.26	0.29	0.28						**0.390**
ERSB G4	A2	0.62	0.52	0.47	0.47	0.43						**0.450**
ERSB G5	A1	0.35	0.38	0.38	0.38	0.40						**0.470**
ERSB G5	A2	0.36	0.36	0.34	0.27	0.27						**0.210**
ERSB G6	A1	0.60	0.61	0.66	0.65	0.67						**0.830**
ERSB G6	A2	0.38	0.39	0.42	0.40	0.47						**0.670**
ERSB G7	A1	0.35	0.32	0.27	0.28	0.26						**0.220**
ERSB G7	A2	0.45	0.53	0.52	0.54	0.57						**0.850**
ERSB G8	A1	0.52	0.47	0.34	0.34	0.34						**0.230**
ERSB G8	A2	0.25	0.19	0.22	0.22	0.24						**0.260**
ERSB G9	A1	0.54	0.59	0.59	0.54	0.55						**0.650**
ERSB G9	A2	0.41	0.42	0.51	0.48	0.43						**0.510**
ERSB G10	A1	0.33	0.36	0.35	0.26	0.30						**0.240**
ERSB G10	A2	0.20	0.21	0.20	0.20	0.19						**0.330**
M&L	A1	0.64	0.68	0.65	0.76	0.71	0.70	0.70	0.71			**0.750**
M&L	A2	0.60	0.46	0.46	0.38	0.38	0.35	0.42	0.39			**0.250**
Oc1	A1	0.41	0.55	0.47	0.43	0.51	0.54	0.53	0.47			**0.500**
Oc1	A2	0.54	0.45	0.45	0.59	0.44	0.47	0.43	0.34			**0.500**
Oc4	A1	0.47	0.59	0.55	0.60	0.59	0.63	0.43	0.48			**0.500**
Oc4	A2	0.38	0.35	0.38	0.40	0.34	0.27	0.27	0.30			**0.200**
Oc9	A1	0.41	0.67	0.72	0.58	0.58	0.58	0.63				**0.500**
Oc9	A2	0.32	0.33	0.24	0.22	0.22	0.28	0.20				**0.100**
On	A1	0.36	0.35	0.38	0.33	0.35	0.38	0.39				**0.400**
On	A2	0.45	0.42	0.40	0.42	0.43	0.44	0.43				**0.400**
R&B10	A1	0.32	0.30	0.26	0.33							**0.375**
R&B10	B1	0.32	0.30	0.27	0.27							**0.250**
R&B10	A2	0.41	0.29	0.35	0.32							**0.375**
R&B10	B2	0.20	0.24	0.18	0.18							**0.250**
R&B15	A1	0.26	0.25	0.27	0.31							**0.375**
R&B15	B1	0.36	0.37	0.37	0.33							**0.250**

R&B15	A2	0.33	0.33	0.36	0.35							**0.375**
R&B15	B2	0.10	0.13	0.19	0.21							**0.250**
RSW D	A1	0.55	0.58	0.60	0.60	0.68	0.68	0.73	0.67	0.71	0.63	**0.670**
RSW D	A2	0.66	0.71	0.72	0.74	0.77	0.74	0.73	0.69	0.72	0.76	**0.670**
RSW S	A1	0.58	0.69	0.68	0.69	0.72	0.72	0.75	0.68	0.78	0.74	**0.670**
RSW S	A2	0.72	0.78	0.77	0.76	0.81	0.81	0.72	0.72	0.73	0.75	**0.670**
S&A3k	A1	0.62	0.65	0.69	0.68	0.67	0.70	0.70				**0.375**
S&A3k	A2	0.59	0.62	0.60	0.64	0.66	0.67	0.69				**0.875**
S&C G1	A1	0.13	0.09	0.08	0.08	0.07	0.08	0.06	0.05			**0.091**
S&C G1	A2	0.58	0.66	0.71	0.72	0.73	0.74	0.71	0.67			**0.909**
S&C G2	A1	0.27	0.21	0.23	0.22	0.23	0.21	0.20	0.17			**0.182**
S&C G2	A2	0.47	0.51	0.54	0.55	0.53	0.57	0.54	0.51			**0.727**
S&C G3	A1	0.19	0.18	0.16	0.16	0.16	0.17	0.15	0.14			**0.273**
S&C G3	A2	0.72	0.78	0.78	0.80	0.80	0.82	0.82	0.82			**0.909**
S&C G4	A1	0.25	0.28	0.28	0.28	0.31	0.29	0.29	0.30			**0.364**
S&C G4	A2	0.72	0.72	0.72	0.75	0.74	0.77	0.74	0.73			**0.818**
S&C G5	A1	0.32	0.32	0.32	0.34	0.37	0.34	0.31	0.30			**0.364**
S&C G5	A2	0.65	0.62	0.66	0.68	0.67	0.66	0.67	0.69			**0.727**
S&C G6	A1	0.44	0.44	0.48	0.44	0.43	0.48	0.43	0.43			**0.455**
S&C G6	A2	0.58	0.57	0.62	0.60	0.60	0.59	0.60	0.61			**0.636**
S&C G7	A1	0.25	0.17	0.17	0.12	0.12	0.08	0.10	0.11			**0.091**
S&C G7	A2	0.50	0.48	0.52	0.50	0.53	0.63	0.68	0.66			**0.909**
S&C G8	A1	0.30	0.26	0.30	0.19	0.21	0.25	0.25	0.24			**0.182**
S&C G8	A2	0.48	0.57	0.59	0.58	0.61	0.64	0.64	0.58			**0.727**
S&C G9	A1	0.29	0.26	0.22	0.26	0.25	0.28	0.24	0.22			**0.273**
S&C G9	A2	0.78	0.80	0.83	0.85	0.84	0.84	0.85	0.84			**0.909**
S&C G10	A1	0.35	0.35	0.39	0.42	0.37	0.38	0.36	0.31			**0.364**
S&C G10	A2	0.67	0.72	0.71	0.72	0.65	0.70	0.70	0.73			**0.818**
S&C G11	A1	0.32	0.31	0.33	0.34	0.35	0.35	0.36	0.30			**0.364**
S&C G11	A2	0.60	0.66	0.64	0.70	0.65	0.64	0.70	0.64			**0.727**
S&C G12	A1	0.44	0.46	0.43	0.41	0.46	0.44	0.44	0.42			**0.455**
S&C G12	A2	0.59	0.60	0.59	0.60	0.58	0.61	0.63	0.64			**0.636**

CHAPTER 3

3. NET REWARD ATTRACTIONS EQUILIBRIUM FOR STRATEGIC FORM GAMES AND ITS EXPERIMENTAL TEST

Abstract. Data from experiments on repeated, completely mixed games show that Nash equilibrium is a poor predictor of observed human choice behavior. Here I propose the concept of Net Reward Attractions (NRA) equilibrium and test its predictive accuracy on data from experiments on 26 repeated, completely mixed games run under full-feedback condition. Moreover, I compare NRA's predictive power with that of other five equilibrium concepts and eight models of learning, representing cutting-edge research on interactive decision making modeling. NRA turns out to be among the best predictors of empirical data, performing significantly better than Nash equilibrium, self-tuning EWA, and reinforcement-based models.

3.1 Introduction

Since the late 1950s, the experimental game theory literature on repeated games has provided significant departures from Nash equilibrium behavior and especially data from experiments on repeated games with a unique equilibrium in mixed strategy (MSE) seem to contradict the predictions of standard game theory (Erev and Roth, 1998; Erev et al., 2007; Selten and Chmura, 2008). The unsatisfactory performances of Nash equilibrium have led researchers to find alternative theories and models to better explain and justify experimentally observed interactive choice behavior.

As a result, most of the models proposed in the behavioral game theory literature outperform standard equilibrium theory in both the tasks of fitting and predicting experimental data, and attribute to other factors the role of drivers of choice behavior (Camerer, 2003; Erev and Roth, 1998; Erev et al., 1999, 2002, 2007; Selten and Chmura, 2008). Specifically, some recent contributions have shown that regret-based models are the best predictors of data from experiments on interactive repeated choice tasks, thus suggesting that regret for foregone payoffs must play a central role in shaping human choice behavior.

In the recent literature on repeated strategic interaction, two patterns of analysis are usually adopted, one focusing on dynamic models and the other on stationary ones. According to the first approach, authors are mainly interested in comparing the accuracy of a bunch of models of learning, considering the performance of one or very few equilibrium concepts merely as a benchmark (Erev and Roth, 1998; Erev et al., 2007). According to the second approach, only equilibrium models are tested and compared (Selten and Chmura, 2008). However, an overall and systematic comparison of the predictive power of both equilibrium and learning models on many different experimental datasets has not yet been proposed.

In the first place, such an overall analysis involving both equilibrium and learning models would shed light on the gap (if any) between these two approaches. In general, equilibrium models are less complex than learning ones from at least three points of view: statistically (i.e., number of free parameters), analytically, and computationally. Stationary concepts are designed to predict and describe choice behavior emerging in the long run, once play has converged to a steady state, whereas learning models (in virtue of their higher degree of complexity) are expected to be more flexible and capable to capture learning dynamics also in the early trials. If this were not the case,

that would constitute a strong argument in favor of the less complex and analytically more tractable stationary concepts, in accordance with Occam's principle of parsimony.

In this paper, I propose a new behavioral equilibrium concept and compare its predictive accuracy with that of other five equilibrium concepts and eight models of learning, among the most popular in the literature on interactive decision making modeling.

I test models on a large compound dataset: I collected data from experiments on 26 two-person 2x2 games played more than 100 times, run under full feedback condition, conducted by other researchers other than me, and for which data for each independent observation were available.

3.2 The NRA Equilibrium

The Net Reward Attractions (NRA) Equilibrium is a stationary concept designed for strategic form games and is based on behavioral assumptions about human choice behavior, rather than on the principle of full rationality. It is assumed that, in equilibrium, agents do not maximize their expected utility function, but that, for a player, the propensity of choosing an action is proportional to its corresponding expected net reward – net reward being defined as the difference between the actual payoff and the minimum obtainable one, given other players' moves. I simply assume here that players are attracted by actions, and that this attraction can be quantified in terms of how much, on average, an action is perceived as better than the others.

The concept of net reward, as introduced here, is very similar to Loomes and Sugden's (1982) concept of *rejoicing* i.e., a measure of the additional pleasure associated to the awareness of having chosen the best action. In this vein, the approach based on net rewards, which I adopt to model choice behavior in the long run, is complementary, although not equivalent (see Section 3), to that based on regret. In Loomes and Sugden's (1982) *regret theory*, these two complementary aspects are fused together in the *Rejoice/Regret* function (see the Introduction), and I show in Chapters 2 and 3 of my thesis that these two components can be separately used to successfully design models of choice behavior.

The intuition at the basis of the NRA model that relative rewards rather than absolute payoffs are what matters in determining choice behavior, is coherent with recent neuroeconomic research. In a pioneering study, Tremblay and Schultz (1999)

report that neurons in the orbitofrontal area of the brain of primates do not encode absolute values of reward objects, but just relative preferences over a set of few objects available at a certain time. In the same vein, Tobler, Fiorillo, and Schultz (2005) show results supporting the hypothesis that dopamine neurons in Macaque monkeys encode a measure of prediction error, defined as the difference between the value of the reward and its expected value. Brain's limited computational resources offer an explanation for these results: whereas in the brain there is a finite number of neurons that can code a finite number of objects, absolute reward values of objects are potentially infinite. For this reason, coding relative rewards of a small number of available objects at a time would not only reduce the complexity of the task of object evaluation, but also increase the accuracy of the process of discrimination between objects. Daw et al. (2006) report data from an experiment with human subjects, consisting in a gambling task. The authors tested the predictive power of the (reinforcement) *softmax* model, in which choice probabilities are determined on the basis of actions' relative expected values. As a result, the softmax model turns out to be the best predictor of observed behavior, in comparison with other two models of reinforcement. In addition, the authors find that there is a positive correlation between softmax predicted probabilities and the activation of the medial and lateral orbitofrontal cortex areas of the brain.

In interactive, conflictual decision tasks, wherein feelings such as fairness, reciprocity, and cooperation are almost completely excluded, the problem of choosing a strategy in a game can be interpreted as an individual choice problem under uncertainty (Brandenburger, 1992), and it seems reasonable to assume that the neural mechanisms involved in those interactive strategic situations are the same as those triggered in the contest of individual choice problems. This gives credit to the idea that relative preferences between rewards are at the basis of human choice behavior, at least in the class of strategic situations above described. Of course, it is a well-known fact that in general interactive decision-making is a psychologically richer process than individual decision-making (Camerer, 2003), as social comparison issues (see for example Fliessbach et al., 2007), inequality aversion feelings, and reciprocating behaviors (see for example Fehr and Schmidt, 1999) are involved.

3.2.1 Theoretical Framework

Before providing the formal definition of NRA equilibrium and describing its properties, I introduce some notation. Consider a finite n-person strategic game G,

defined by a set of players $N=\{1,...,n\}$, a non-empty set $A_i=\{a_{i1},...,a_{im_i}\}$ of available actions for each player $i\in N$, and a payoff function $u_i:\times_{i\in N}A_i\to R$. Denote the elements of $A=\times_{i\in N}A_i$ (action profiles) with a. Let $\Delta_i=\Delta(A_i)$ be the set of probability measures on A_i; all elements $p_i=(p_{i1},...,p_{im_i})$ of $\Delta(A_i)$ are such that $\sum_{j=1}^{m_i}p_{ij}=1$ and $p_{ij}\geq 0\ \ \forall j=1,...,m_i$, so that $\Delta(A_i)$ is isomorphic to the m_i-dimensional simplex $\Delta(A_i)=\left\{p_i\left|\sum_{j=1}^{m_i}p_{ij}=1,p_{ij}\geq 0\right.\right\}$. Elements in $\Delta=\times_{i\in N}\Delta_i$ will be denoted by $p=(p_1,...,p_n)$.

Consider now the transformed payoff function $u_i'(a):A\to R$, defined as the difference between the payoff received and the minimum obtainable payoff, given other players' actions; indicating with (a_i',a_{-i}) the action profile in which player *i* chooses a_i' and all other players play the profile a_{-i}, we write:

$u_i'(a)=u_i(a)-\min_{a_i'\in A_i}\{u_i(a_i',a_{-i})\}$ and $u_i'(a)=(u_i'(a),...,u_i'(a))$.

It is possible to extend the payoff function $u'(a)$ to the domain Δ by writing:

$$u_i'(p)=\sum_{a\in A}p_i(a)\cdot u_i'(a)=\sum_{a\in A}p_i(a)\cdot\left(u_i(a)-\min_{a_i'\in A}\{u_i(a_i',a_{-i})\}\right);$$

denote with $E_{ij}(p)=u_i'(a_{ij},p_{-i})$, player i's expected net reward from action j.

Definition 1. Let $G=(N,A,u)$ be a finite *n*-person strategic game. A vector $p=(p_1,...,p_n)\in\Delta$ is said to be a *Net Reward Attractions* (NRA) equilibrium if

$$p_{ij}=\frac{E_{ij}(p)}{\sum_{k=1}^{m_i}E_{ik}(p)},\qquad \text{for } \forall\ i=1,...,n;\ j=1,...,m_i,$$

provided that $\sum_{k=1}^{m_i}E_{ik}(p)>0$; otherwise, p_{ij} can assume all values in $[0,1]$.

■

As an illustration, consider the case of an *mxn* two-person game represented by the following payoff matrix.

Player 2 / Player 1	A_{21}	A_{22}	…	A_{2n}
A_{11}	a_{11},b_{11}	a_{12},b_{12}	…	a_{1n},b_{1n}
A_{12}	a_{21},b_{21}	a_{22},b_{22}	…	a_{2n},b_{2n}
…	…	…	…	…
A_{1m}	a_{m1},b_{m1}	a_{m2},b_{m2}	…	a_{mn},b_{mn}

The NRA equilibrium is computed on the transformed payoff matrix:

Player 2 / Player 1	A_{21}	A_{22}	…	A_{2n}
A_{11}	a'_{11},b'_{11}	a'_{12},b'_{12}	…	a'_{1n},b'_{1n}
A_{12}	a'_{21},b'_{21}	a'_{22},b'_{22}	…	a'_{2n},b'_{2n}
…	…	…	…	…
A_{1m}	a'_{m1},b'_{m1}	a'_{m2},b'_{m2}	…	a'_{mn},b'_{mn}

where $a'_{ij} = a_{ij} - \min_k \{a_{ik}\}$, for $i = 1,...,m$, and $b'_{ij} = b_{ij} - \min_l \{b_{il}\}$, for $j = 1,...,n$. In equilibrium, provided that the denominators are positive, choice probabilities are the solutions of the following two systems of equations:

$$\begin{cases} p_1 = \dfrac{E_{11}(q)}{\sum_{j=1}^{m} E_{1j}(q)} \\ ... \\ p_m = \dfrac{E_{1m}(q)}{\sum_{j=1}^{m} E_{1j}(q)} \end{cases} \quad \text{and} \quad \begin{cases} q_1 = \dfrac{E_{21}(p)}{\sum_{j=1}^{n} E_{2j}(p)} \\ ... \\ q_n = \dfrac{E_{2n}(p)}{\sum_{j=1}^{n} E_{2j}(p)} \end{cases},$$

whith $p = (p_1,...,p_m)$ and $q = (q_1,...,q_n)$, and E_{1i} and E_{2j} are the expected net reward (average, transformed utilities) from actions i and j for row and column players, respectively. Therefore, the above-defined equations tell us that the larger the relative reward associated to an action, the larger the probability that that action will be chosen.

The following theoretical results hold.

Theorem 1. Every finite *n*-person strategic game has a NRA equilibrium.

Proof. If $\sum_{k=1}^{m_i} E_{ik}(p) > 0$, the vector function f (with $f_{ij}(p) = \dfrac{E_{ij}(p)}{\sum_{k=1}^{m_i} E_{ik}(p)}$) is continuous in the simplex Δ and then, by Brouwer's Theorem, has at least one fixed-point.

If instead $\sum_{k=1}^{m_i} E_{ik}(p) = 0$, this implies that $E_{ij}(p) = 0 \quad \forall j = 1,...,m_i$ because all expectations are, by definition, non-negative. In this latter case, any vector $p_i = (p_{i1},...,p_{im_i})$ of $\Delta(A_i)$ in $p = (p_1,...,p_n) \in \Delta$ is a NRA equilibrium.

■

Theorem 2. In every finite two-person 2x2 strategic game,

- If there is a unique Nash equilibrium in mixed strategies (MSE), then there is a unique NRA equilibrium.
- Every pure strategy Nash equilibrium is a pure strategy NRA equilibrium.

Proof of a). Without loss of generality, we can consider the payoff matrix reported below as the general structure of a 2x2 strategic game with a unique MSE (Selten and Chmura, 2008), being the other possible case obtained by switching its rows and columns.

Player 1 \ Player 2	L	R
U	$(a_L + c_L; b_U)$	$(a_R; b_U + d_U)$
D	$(a_L; b_D + d_D)$	$(a_R + c_R; b_D)$

The constants c_L, c_R, d_U, and d_D are supposed to be strictly bigger than zero.
For player 1, the transformed payoff matrix will assume the form:

c_L	0
0	c_R

Let us indicate with p the probability with which player 1 plays U and with q the probability with which player 2 plays L; then we can write:

$$p(q) = \frac{q \cdot c_L}{q \cdot c_L + (1-q) \cdot c_R}. \quad (1)$$

Then, p is a continuous, differentiable function in q. In fact, the denominator of (1) is always strictly positive (it is a linear combination of strictly positive numbers).

Analogously, for player 2 the transformed payoffs will be as those in the following payoff matrix:

0	d_U
d_D	0

and we can write:

$$q(p) = \frac{(1-p)\cdot d_D}{(1-p)\cdot d_D + p\cdot d_U}. \qquad (2)$$

Also in this case, q is a continuous, differentiable function in p. In fact, the denominator of (2) is always strictly positive (it is a linear combination of strictly positive numbers).

In addition, we have that the derivative of (1) is always strictly positive, whereas the derivative of (2) is always strictly negative.

In the $p \times q = [0,1] \times [0,1]$ space, the two functions $p(q)$ and $q(p)$ either cross each other once or do not cross at all, since $p(q)$ is strictly increasing and $q(p)$ is strictly decreasing. But then, by Theorem 1, the two curves must cross once.

Proof of b). In the pure strategy Nash equilibrium (a_1^*, a_2^*), NRA players' rescaled payoffs are either bigger or equal than zero, whereas the rescaled payoffs from the other choice are always equal to zero. In the first case, the probability associated to that action is either equal to 1 or can assume any value in $[0,1]$. In both cases, (a_1^*, a_2^*) is a NRA equilibrium.

■

Theorem 3. Every pure strategy NRA equilibrium is a pure strategy Nash equilibrium.

Proof. Consider a finite *n*-person strategic game. A strategy profile $(a_1^*, ..., a_n^*)$ is a pure strategy NRA equilibrium if and only if for all players the corresponding (rescaled) outcomes are bigger or equal than zero and the outcomes from the other strategies are equal to zero. But then $(a_1^*, ..., a_n^*)$ must be a Nash equilibrium, since each player cannot be better off acting differently.

■

Theorem 4. The NRA equilibrium for a 2x2 strategic game with a unique MSE is:

$$\left(p^* = \frac{c_L \cdot q^*}{c_L \cdot q^* + c_R \cdot (1 - q^*)}, q^* \right),$$

where q^* is the unique solution in the interval $[0,1]$ of the polynomial

$$t(q) = (c_L d_U - c_R d_D) \cdot q^2 + 2 c_R d_D q - c_R d_D.$$

Proof. The NRA equilibrium for a game with a unique MSE is the solution of the following system of equations (see the proof of Theorem 2a):

$$\begin{cases} p = \dfrac{q \cdot c_L}{q \cdot c_L + (1-q) \cdot c_R} \\ q = \dfrac{(1-p) \cdot d_D}{(1-p) \cdot d_D + p \cdot d_U} \end{cases},$$

and substituting the first equation in the second we get:

$$\begin{cases} p = \dfrac{q \cdot c_L}{q \cdot c_L + (1-q) \cdot c_R} \\ (c_L d_U - c_R d_D) q^2 + 2 c_R d_D q - c_R d_D = 0 \end{cases}.$$

Now, consider the polynomial $t(q) = (c_L d_U - c_R d_D) q^2 + 2 c_R d_D q - c_R d_D$. In the case in which $c_L d_U - c_R d_D = 0$, the unique root of $t(q)$ is $q = \dfrac{1}{2}$.

If instead $c_L d_U - c_R d_D \neq 0$, then $t(q)$ has always two roots. Indeed, $\Delta = b^2 - 4ac = 4 c_L d_U c_R d_D > 0$ (remember that the constants c_L, c_R, d_U, and d_D are supposed to be strictly bigger than zero).

Now, the first derivative of $t(q)$ is $t'(q) = 2(c_L d_U - c_R d_D) q + 2 c_R d_D$. We can distinguish two possible cases.

I) $c_L d_U - c_R d_D > 0$.

We have that $t(q)$ is always increasing in $[0,1]$ since $t'(q) > 0 \quad \forall \quad q \in (0,1)$. Moreover, we have that $t(0) < 0$ and that $t(1) > 0$ and then, by the Intermediate Value Theorem, $t(q)$ has one and only one root in $[0,1]$.

II) $c_L d_U - c_R d_D < 0$.

We have that $t(q)$ is always increasing in $[0,1]$ since $t'(q) > 0 \quad \forall \quad q \in (0,1)$. Moreover, it is $t(0) < 0$ and $t(1) > 0$, and then, by the Intermediate Value Theorem, $t(q)$ has one and only one root in $[0,1]$.

We have then shown that the polynomial $t(q)$ has always a unique solution in the interval $[0,1]$, denote it with q^*, which concludes the proof.

■

Reading together points a) and b) of Theorem 2, we can conclude that there is a one to one correspondence between Nash pure strategy equilibria and NRA pure strategy equilibria. However, this holds true only in the case of 2x2 two-person games. In general, the set of pure strategy NRA equilibria is a subset of that one of pure strategy Nash equilibria.

3.2.2 Parametric NRA

I provide also a parameterized version of NRA I call Parametric NRA (pNRA). It is obtained introducing in the calculus of expectations in Definition 1 a parameter $\lambda > 0$, tuning players' sensitivity to net rewards. In my analysis, I consider both NRA and pNRA and it turns out that the introduction of a parameter λ leads to a significant increase in the accuracy of predictions.

The pNRA equilibrium is defined as follows:

Definition 2. Let $G = (N, A, u)$ be a finite *n*-person strategic game and λ be a positive constant. A vector $p = (p_1, ..., p_n) \in \Delta$ is said to be a *Parametric Net Reward Attractions* (pNRA) equilibrium if

$$p_{ij} = \frac{E_{ij}^{\lambda}(p)}{\sum_{k=1}^{m_i} E_{ik}^{\lambda}(p)}, \qquad \text{for } \forall\ i = 1,...,n;\ j = 1,...,m_i,$$

provided that $\sum_{k=1}^{m_i} E_{ik}(p) > 0$; otherwise p_{ij} can assume all values in $[0,1]$.

■

Theorems 1 to 3 still hold true if we replace NRA with pNRA, with obvious, minor adaptations of the proofs. The formulation of Parametric NRA might resemble a simple linearization of *logit* equilibrium (McKelvey and Palfrey, 1995), but this is not the case. The two concepts are profoundly different for two reasons: first, according to pNRA, equilibrium probabilities are determined via the linear probabilistic choice rule

of Definition 1, and not via the logistic quantal response function; second, and more importantly, even though logit equilibrium is invariant to additive changes of the payoffs, in pNRA expectations are taken with respect to *net rewards*, which are not obtained by simply adding a constant to all payoffs. We must then conclude that the two concepts cannot correspond.

3.2.3 Convergence to NRA Equilibrium

One of the most important questions associated to stationary concepts is that of how an equilibrium (if any) emerges in a population of players (Camerer, 2003). In order to complete the theoretical framework described in the previous section, I propose a procedure, which I call *Counterfactual Reinforcement Learning* (CRL), according to which players' choice behavior approaches to the set of NRA equilibria. Although I do not use CRL as a model to predict empirical data, it will turn useful to understand the NRA equilibrium in depth.

The CRL procedure is defined as follows:

1. *Initial propensities.* At time period $T = 1$, before the game has been played for the first time, player *i*'s propensity of playing his pure strategy *j* is equal to the expected payoff from random choice (denoted by $A(0)$). Then, $a_{ij}(1) = A(0)$ for all *i* and *j*.
2. *Attractions updating.* At each time step $T > 1$, player *i*'s propensities are updated according to the rule:

   $$a_{ij}(T) = a_{ij}(T-1) + NR_{-i}^{T-1}(j), \quad \forall j \qquad (1)$$

   where $NR_{-i}^{T}(j)$ is the net reward at time *T* associated to player *i*'s action *j*. The net reward $NR_{-i}^{T}(j)$ of action *j* is defined as the difference between the corresponding payoff and the minimum obtainable payoff, given other player' moves.
3. *Stochastic choice rule.* Attractions at time step $T > 1$ are mapped into choice probabilities according to the linear probabilistic response rule:

   $$p_{ik}(T) = \frac{a_{ik}(T-1)}{\sum_j a_{ij}(T-1)}. \qquad (2)$$

The motivation for the adjective *counterfactual* is that CRL implicitly assumes that players know their own payoffs, receive full feedback about their choices, and make comparisons between actual and foregone payoffs. At a first glance, one might consider

these assumptions as too strong; however, especially in the context of repeated games, when the payoff matrix is initially not known to players, it is reasonable to assume that players can easily, after few trials, figure out the structure of their own payoffs and use this information when forming their strategies. Empirical evidence supports this conjecture. Indeed, the experiment described in Erev et al. (2007) replicated a former experiment on the same set of ten repeated games (Erev et el., 1999), with the exception that subjects were given complete information about the payoff structure of each game. In the former version of the experiment, subjects received information only about their actually experienced payoffs. It seems reasonable to assume that players in the former experiment could, after some trials, infer the structure of their own payoffs, as the two treatments produced only minor differences on the observed average choice behavior.

It can be shown analytically that CRL converges to NRA equilibrium. Indeed, according to (1) and (2), we can write:

$$p_{ik}(T) = \frac{a_{ik}(T-1)}{\sum_j a_{ij}(T-1)} = \frac{A(0) + \sum_{t=1}^{T-2} NR_{-i}^{t}(k)}{\sum_j \left[A(0) + \sum_{t=1}^{T-2} NR_{-i}^{t}(j) \right]}$$

provided that the denominator is positive; multiplying and dividing both numerator and denominator by $T-2$, we get:

$$p_{ik}(T) = \frac{\frac{A(0)}{T-2} + < NR_{-i}(k) >}{\sum_j \left[\frac{A(0)}{T-2} + < NR_{-i}(j) > \right]}, \qquad (3)$$

where $< NR_{-i}(k) >= \frac{\sum_{t=1}^{T-2} NR_{-i}^{t}(k)}{T-2}$. Now, taking the limit of (3) as $T \to +\infty$, we have that:

$$p_{ik} \to \frac{E_{ik}(p)}{\sum_j E_{ij}(p)}, \ \forall\ i = 1,...,n \ and \ j = 1,...,m_i$$

as E_{ij} is defined as the expected net reward for player *i* from his action *j*, showing that choice probabilities converge to NRA equilibria.

3.3 Related Work

CRL produces dynamics that are different from those produced by any beliefs-based model. According to fictitious play models, players are assumed to keep track of the relative frequency with which other players have employed each strategy in the past, and then calculate the expected payoff given these beliefs and choose that corresponding to the highest expected value. On the contrary, in CRL, players are not supposed to be maximizers, but simply choose an action with probability proportional to its expected net reward.

In particular, the Experience Weighted Attraction (EWA) model (Camerer and Ho, 1999; Ho, Camerer, and Chong, 2007) cannot capture CRL dynamics. EWA is a hybrid model, blending the main features of the reinforcement and beliefs based models; indeed, if parameters are constrained to specific values, it reduces to the (average or cumulative) reinforcement model in which only chosen strategies are reinforced and if parameters are set in a different way, EWA reduces exactly to *(weighted) fictitious play*. More specifically, according to the EWA model, attractions are updated as follows:

$$a_{ij}(T) = \frac{\phi \cdot N(T-1) \cdot a_{ij}(T-1) + \left[\delta + (1-\delta) \cdot I(s_{ij}, s_i(T))\right] \cdot \pi_i(s_{ij}, s_{-i}(T))}{N(T-1) \cdot \phi \cdot (1-\kappa) + 1} \quad (4)$$

and choice probabilities determined by the (*logit*) stochastic choice rule:

$$p_{ij}(T) = \frac{\exp(\lambda \cdot a_{ij}(T-1))}{\sum_j \exp(\lambda \cdot a_{ij}(T-1))}. \quad (5)$$

As one can easily see, CRL dynamics cannot be replicated by setting in (4) $\delta = 0$; in this case, EWA corresponds to a reinforcement learning model in which only propensities corresponding to played actions are updated. The best approximation to CRL is obtained by setting in (4) $\delta = 1$, $\kappa = 0$, and $\phi = 1$; in this case, EWA corresponds to the *weighted fictitious play* model where distant experiences in the past are less salient than recent ones (*recency effect*), and propensities are reinforced by their corresponding payoffs (with weight $\delta = 1$). Moreover, propensities are mapped into choice probabilities by the *logit* response function (5) and not by the simple normalization operated by (2). Then, it can be can easily seen that there is no parameter configuration allowing EWA to capture CRL dynamics.

There are also some similarities between CRL and the *stochastic fictitious play* (SFP) model proposed by Erev et al. (2007) (see Introduction). According to the SFP model, at each time step, propensities are updated according to the following:

$a_{ij}(T) = (1-w)\cdot a_{ij}(T-1) + w\cdot v_{ij}(T-1)$, for all i and j,

where $v_{ij}(T)$ is the expected payoff in the selected cell and w one of the two parameters of the model tuning *sensitivity to foregone payoffs*. However, SFP cannot replicate the behavior of CRL, because past experience is decayed, propensities are reinforced with the payoffs, and choice probabilities are determined via the *logit* choice rule (5).

The CRL procedure closely resembles *Unconditional Regret Matching* (URM), as described in Young (2004) and in Hart (2005). However, these two procedures, though very similar, are not equivalent. According to URM, the regret for not having played action k is defined as:

$$\tilde{R}(k) := \left[\tilde{V}(k) - U\right]_+, \qquad (6)$$

where

$$\tilde{V}(k) := \frac{1}{T}\sum_{t=1}^{T} u^i\left(k, s_t^{-i}\right) \text{ and } U := \frac{1}{T}\sum_{t=1}^{T} u^i\left(s^t\right).$$

Choice probabilities are determined via:

$$\sigma_{T+1}(k) := \frac{\tilde{R}(k)}{\sum_{l=1}^{m}\tilde{R}(l)} \qquad \text{for each } k = 1,2,...,m.$$

For simplicity, let us consider the particular class of 2x2 completely mixed games. According to CRL, at each round the probability of playing an action is always positive and the reference point is set equal to the (average) minimum obtainable payoff given other player's choices; in URM, choice probabilities are not necessarily always positive and the reference point is the average received payoff. Specifically, let us consider the quantities $\tilde{R}(k)*T$ and $a_{ik}(T)$, as defined in (1) and (6), respectively. We have that $a_{ik}(T) = A(0) + \sum_{t=1}^{T-1} NR_{-i}^{t}(k)$, which means that the attraction associated to action k at time T is equal to the summation of initial propensities and received net rewards, and for T large we can write (with some abuse of notation) $a_{ik}(T) = \sum_{t=1}^{T} NR_{-i}^{t}(k)$. Now, we can write, according to the definition of CRL:

$$a_{ik}(T) = \sum_{t=1}^{T} NR_{-i}^{t}(k) = \sum_{t=1}^{T} u^i\left(k, s_t^{-i}\right) - \sum_{t=1}^{T}\min_j\left\{u^i\left(j, s_{-i}^{t}\right)\right\}.$$

Compared with the following:

$$\tilde{R}(k) * T := \left[\tilde{V}(k) - U\right]_+ * T = \left[\sum_{t=1}^{T} u^i\left(k, s_t^{-i}\right) - \sum_{t=1}^{T} u^i\left(s^t\right)\right]_+ ,$$

it is clear that the two quantities $\tilde{R}(k) * T$ and $a_{ik}(T)$ are different, then leading to different dynamics of choice behavior.

In addition, experimental data do not support the hypothesis of convergence to correlated equilibria, as joint distributions of play can be easily verified to be to the product of the marginal distributions in all experiments I consider in this study (and which I describe in Section 5).

A stationary concept similar to NRA is Impulse Balance Equilibrium (IBE) (Selten and Chmura, 2008). According to these two solution concepts, equilibrium probabilities are calculated considering a transformed game whose payoffs quantify players' propensities (or impulses) to choose actions. Nonetheless, these two models are deeply different, as can be easily seen considering the simple case of 2x2 two person games. Indeed, in the former model, all payoffs are rescaled by subtracting the minimum obtainable payoff given other players' moves, whereas, in the latter, only payoffs above the pure strategy maximin payoff ("*a natural aspiration level*", Selten and Chmura, 2008:947) are rescaled by subtracting one half of the difference between these payoffs and the maximin payoff.

3.4 Model Comparison Methodology

For each parametric model, I determine the corresponding *Prediction scores* based on data from experiments on 26 different, repeatedly played games. I computed these scores according to the *leave-one-out estimation procedure*, as described in the Methods section of Chapter 2.

As for non-parametric models, I simply determine the corresponding Mean Squared Deviation, as no parameters are to be estimated.

Considering all 26 experiments together, I gathered a total of 234 independent observations. For each model, I calculate the MSD (or Prediction) scores corresponding to each independent observation, and store them in a vector of length 234. In order to assess the significance of pairwise comparisons of models' accuracy, I use a *Mann-Whytney-Wilcoxon match-paired signed-rank (two-tailed) test*, as done in Selten and Chmura (2008). For each pair of models, the null hypothesis that the vectors of scores

have the same mean is tested. As already said, for the testing I use here a dataset of experiments on 26 different games, smaller than that I use in the second chapter (which counts 35 games). Indeed, I consider here only datasets for which data for each independent observation were available (either at the individual or group level, depending on whether fix-pairing or random-matching protocol was used in the experiment). This allows me to gather a large number of independent conditions on which to test each model, also giving the Mann-Whitney-Wilcoxon test more chances to compare models more precisely.

My analysis can be divided into three, distinct parts; I compare models' accuracy in predicting observed choice behavior in the short run (i.e., choice frequencies averaged over the first 50 trials), in the long run (i.e., choice frequencies averaged over the last 50 trials), and choice behavior averaged over all periods.

Since learning models are stochastic, the estimated frequency of choice was obtained as the average over 150 simulations, which were run for each experiment and for each parameter configuration. Moreover, in order to make simulation results comparable, the initialization of all dynamic models was set to assure equal probabilities of choosing each action at the first round of the simulation.

I compare the performances of eight different models of learning and five stationary concepts. I consider in my analysis the following models of learning: Normalized Fictitious Play (NFP) (Erev et al., 2007); Normalized Reinforcement Learning (NRL) (Erev et al., 2007); Perceptron-Based (PB0 and PB1 with, respectively, zero and one free parameters) (Marchiori and Warglien, 2008); Reiforcement Learning (REL) (Erev and Roth, 1998); Reinforcement Learning (RL) (Erev et al., 2007); Stochastic Fictitious Play (Erev at al., 2007); and Self-tuning Experience Weighted Attraction (stEWA) (Ho, Camerer, and Chong, 2007). The equilibrium concepts I consider are: Nash Equilibrium; Quantal Response Equilibrium (QRE) (McKelvey and Palfrey, 1995); Impulse Balance Equilibrium (IBE) (Ockenfels and Selten, 2005); Action-Sampling Equilibrium (Sample-7) (Selten, 2000); Payoff-Sampling Equilibrium (Osborne and Rubinstein, 1998); Net Reward Attractions (NRA) Equilibrium and its parametric version pNRA.

The grid search for optimal parameter values was conducted on broad parameter spaces, summarized in Table 1. The portions of parameter spaces that have been investigated were suggested by the authors of the models in previous works (Erev at al., 2007; Ho, Camerer, and Chong, 2007).

Table 1. Values of model free parameters used in my simulations.

Model	Free Parameter: [Interval of variation] – Increment	
NFP	λ: $[1.5,4.0]$ - 0.25	w: $[0.1,0.9]$ - 0.1
NRL	λ: $[3.0,7.0]$ - 0.5	w: $[0.10,0.90]$ - 0.05
PB1	λ: $[0.05,1.00]$ - 0.05	
QRE	λ: $[0.01,18]$ - 0.01	
REL	λ: $[2.2,3.4]$ - 0.1	$N(1)$: $[27,34]$ - 1
RL	λ: $[6.0,10.0]$ - 0.5	w: $[0.10,0.90]$ - 0.05
SFP	λ: $[10.0,14.0]$ - 0.5	w: $[0.05,0.90]$ - 0.05
stEWA	λ: $[1,9]$ - 0.1	

3.5 The Data

I collected datasets from different experiments on two-person 2x2 games with a unique equilibrium in mixed strategies (also known as "completely mixed games"), run under full feedback condition, and for which data for each independent observation were available. These experiments have been conducted under a variety of experimental conditions and by different researchers. Out of the 26 games considered, 16 are constant-sum, while in the remainder players could find incentive to reciprocate; in other words, in 16 experiments, subjects had to learn strategies of pure conflict, while in the other 10 the conflict aspect did not exclude a priori a sort of cooperative (or fair) behavior, as in the non-constant sum games reported in Selten and Chmura (2008). In order to let the learning processes fully unfold, I selected experiments with a minimum of 100 iterations of the stage game; this allows for the testing of the descriptive and predictive power of the different models on subjects' behavior not only in the early rounds, but also in the long run (Erev and Roth, 1998). I labeled games with the initials of the authors who conducted the experiments (AGK = Avrahami, Guth, and Kareev, 2005; ERSB = Erev et al., 2007; RSW = Rosenthal, Shachat, and Walker, 2003; S&C = Selten and Chmura, 2008).

Table 2. Summary of the Datasets. The first column of Table 2 indicates the name of the researchers and the second one the year of publication of the experiment. The third column reports the number of times that the stage game was played. The fourth column indicates how many different games experimenters considered. The fifth column indicates the number of subject who participated to the experiments. The sixth column reports additional important features (if any) for each experiment. Finally, the seventh column reports whether or not subjects were randomly paired at each trial.

Experimenters	Year	Rounds #	Treatments/ Games	Subjects #	Independent Observations	Other	Random Matching
Rosenthal, Shachat, and Walker	2003	100 and 200	1	20 pairs for each treatment	6		No
Avrahami, Guth, and Kareev	2005	100	3	6 pairs in the first treatment and 12 pairs in the other two	6 + 12x2	Only the "Known" treatment is considered	No
Erev, Roth, Slonim, and Barron	2007	500	10	9 pairs for each treatment	9x10		No
Selten and Chmura	2008	200	12	16 pairs for each treatment	12x6 + 6x6		Yes

Table 3. Observed Frequencies of Play.

Game	First 50 Trials		Last 50 Trials		All Trials	
	Row	Column	Row	Column	Row	Column
AGK50	0.460	0.440	0.480	0.467	0.470	0.454
AGK67	0.567	0.515	0.620	0.569	0.593	0.542
AGK75	0.492	0.640	0.460	0.663	0.476	0.652
ERSB G1	0.598	0.289	0.642	0.318	0.591	0.318
ERSB G2	0.731	0.362	0.876	0.256	0.840	0.361
ERSB G3	0.633	0.220	0.502	0.244	0.583	0.222
ERSB G4	0.340	0.627	0.309	0.436	0.274	0.502
ERSB G5	0.342	0.344	0.411	0.296	0.378	0.320
ERSB G6	0.536	0.409	0.631	0.489	0.638	0.410
ERSB G7	0.362	0.480	0.198	0.602	0.295	0.522
ERSB G8	0.518	0.222	0.351	0.220	0.400	0.226
ERSB G9	0.504	0.404	0.571	0.369	0.562	0.449
ERSB G10	0.313	0.178	0.304	0.198	0.320	0.202
RSW D	0.583	0.713	0.660	0.730	0.659	0.737
S&C G1	0.109	0.620	0.055	0.691	0.079	0.690
S&C G2	0.240	0.490	0.185	0.527	0.217	0.527
S&C G3	0.183	0.747	0.148	0.822	0.164	0.793
S&C G4	0.267	0.719	0.294	0.736	0.286	0.736
S&C G5	0.320	0.635	0.306	0.682	0.327	0.664
S&C G6	0.439	0.575	0.426	0.602	0.445	0.596

S&C G7	0.210	0.488	0.105	0.673	0.141	0.564
S&C G8	0.281	0.524	0.244	0.611	0.250	0.586
S&C G9	0.275	0.787	0.228	0.844	0.254	0.827
S&C G10	0.348	0.698	0.335	0.715	0.366	0.699
S&C G11	0.314	0.625	0.328	0.668	0.331	0.652
S&C G12	0.451	0.592	0.433	0.635	0.439	0.604

3.6 Results

I tested all models on three prediction tasks, measuring their accuracy in predicting observed choice behavior averaged over the first 50 trials, the last 50 trials, and all trials. Tables 4-6 report a summary of the results of my analysis; specifically, they report, for each prediction task, the ranking of the models according to MSD (non parametric models) or Prediction scores (computed for parametric models), and the significance of pairwise comparisons with respect to the best performing model. Tables 7-9 report more detailed data of model performances. Finally, Tables 10-12 show the p-values of all possible model pairwise comparisons (for each pair of models, the null hypothesis of no difference between their average scores is tested).

The models I consider are based on quite different theories and hypotheses about human choice behavior in repeated, interactive decision tasks. As a premise to the following analysis, it is worth noting that when we test a model, we obtain a joint evaluation of the validity of the theory it relies on and of how that theory is implemented. Therefore, we cannot reject a theory based only on a bad performance of the corresponding model. However, whenever different implementations of the same theory perform well (bad), we are then allowed to conclude that that theory is (is not) supported by empirical evidence.

In light of their lower degree of complexity, one would expect models of equilibrium to perform better than models of learning in predicting behavior in the long run, when play has converged to a stable state and initial effects have been washed out. On the opposite, in virtue of their higher degree of complexity, learning models should be more suitable to capture the dynamic of learning in the early periods of play if compared to stationary models. I will test these conjectures in the following paragraphs under the light of my simulation results.

For what concerns the statistical significance of model pairwise comparisons, I will refer to the standard, widely accepted, 5% level of significance.

3.6.1 First 50 Trials

Reading from Table 4, if we compare the Prediction score of each of the first eleven best performing models with that of the best one (in this case, NFP), these differences are not significant, with the sole exception of the SFP model. This can be explained by noting that the NFP and SFP models differ only for a minor structural detail (see Introduction) and generate predictions that are, on average, almost equal; however, NFP is systematically more accurate than SFP, which justifies the statistical significance of the difference of their performances.

In the group of the best performing models, whose accuracy of predictions is on average equivalent to that of NFP, we count both learning and equilibrium models. Predictions generated by NRA, IBE, pNRA, and Payoff-sampling equilibria are equivalent to those of the NFP and PB0 models. This is one of the most important results of the third chapter of my thesis: there is no significant gap between the best equilibrium and learning models, even in the task of predicting behavior in the early trials. Simulation results deny the conjecture I made in the previous section, as models of learning, in spite of their higher complexity, are not able to predict observed behavior in the early trials significantly better than stationary concepts. Fast convergence of play to stable behavior might provide a possible explanation for that, since in this case averaging choice frequencies over the first 50 periods would be sufficient to wash out initial effects. However, such an explanation is not satisfactory. Indeed, if we look at Table 3 above, showing the empirical frequencies of play, we can see that the behavior in the first 50 trials is, in most games, different from that in the last 50 trials. It seems rather that the models considered in my analysis are, on average, better predictors of the behavior in the first trials rather than of that emerging in the long run, as it is clear if we compare the Prediction scores reported in Tables 4 and 5.

QRE, stEWA, Nash equilibrium, and REL are less accurate models; in particular, if we look at Table 10, we can see that the Nash equilibrium and REL models perform significantly worse than all the others.

Reinforcement learning models capture well behavior in the short run, but, as I will illustrate in the next section, they are very poor predictors of long run behavior. This seems to confirm the hypothesis according which reinforcement models suffer of inertia i.e., they are too slow in adapting their behavior.

Table 4. Summary of simulation results for the first 50 trials of play. Models are ranked from the best to the worst (from the left to the right) according to average MSD or Prediction Scores (third row). The fourth row reports in percentage how worse is the accuracy of a certain model with respect to the best performing one. Each cell in the fifth row reports the p-value of the test of the null hypothesis of no difference between the average score of the corresponding model and that of the best performing model. Shaded cells refer to the cases in which the null hypothesis is *not* rejected at a 5% level.

Ranking	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Model (# of parameters)	**NFP (2)**	**NRA (0)**	**PB0 (0)**	**PB1 (1)**	**SFP (2)**	**IBE (0)**	**pNRA (1)**	**Payoff-sampling (1)**
Avg. Scores	0.0473	0.0498	0.0499	0.0501	0.0503	0.0504	0.0506	0.0508
Gap to the best (%)	-	5.45%	5.54%	6.00%	6.32%	6.70%	7.11%	7.48%
Comparison significance		0.939	0.928	0.183	0.003	0.341	0.874	0.073

Ranking	**9**	**10**	**11**	**12**	**13**	**14**	**15**
Model (# of parameters)	**NRL (2)**	**7-sampling (0)**	**RL (2)**	**QRE (1)**	**stEWA (1)**	**Nash (0)**	**REL (2)**
Avg. Scores	0.0514	0.0538	0.0540	0.0595	0.0659	0.0975	0.0986
Gap to the best (%)	8.74%	13.83%	14.30%	25.93%	39.38%	106.32%	108.70%
Comparison significance	0.397	0.312	0.084	0.030	0.001	0.000	0.000

3.6.2 Last 50 Trials

Reading from Table 5, we note two things: first, the set of best predictors (the set of models predicting equivalently well to the best one) shrinks a lot with respect to the short run prediction task; second, out of the five best predictors, three are equilibrium models.

IBE turns out to be the most accurate model in terms of Prediction scores, but its performance is statistically equivalent to that of pNRA, NFP, Action-sampling, and Payoff-sampling. The IBE model is here particularly advantaged by the inclusion in the dataset of the games described by Selten and Chmura (2008); indeed, six games among those considered by these authors, though completely mixed, are not constant-sum, thus leaving room for cooperative and reciprocating behaviors. IBE, by design (see Introduction), takes indirectly into account this kind of behavior, thus resulting particularly favored with respect to the other models.

As for the other models, self-tuning EWA and Nash equilibrium are equivalent predictors of the behavior in the long run (with rather high Prediction scores), whereas reinforcement based models (REL, NRL, and RL) do very poorly, with Prediction scores much larger than IBE's.

QRE and Nash equilibria are the two stationary models excluded from the set of best performing models. In addition, it is worth noting that only IBE and pNRA predict observed data significantly better than QRE.

The NFP model is the sole model of learning whose predictions are equivalent to those of IBE and pNRA. The PB0 and PB1 models predict better than Nash equilibrium, and their scores are about 20% larger than NFP's.

Table 5. Summary of simulation results for the last 50 trials play. Models are ranked from the best to the worst (from the left to the right) according to average MSD or Prediction Scores (third row). The fourth row reports in percentage how worse is the accuracy of a certain model with respect to the best performing one. Each cell in the fifth row reports the p-value of the test of the null hypothesis of no difference between the average score of the corresponding model and that of the best performing model. Shaded cells refer to the cases in which the null hypothesis is *not* rejected at a 5% level.

Ranking	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Model (# of parameters)	**IBE (0)**	**pNRA (1)**	**NFP (2)**	**7-sampling (0)**	**SFP (2)**	**Payoff-sampling (1)**	**QRE (1)**	**NRA (0)**
Avg. Scores	0.0505	0.0525	0.0536	0.0545	0.0548	0.0566	0.0576	0.0599
Gap to the best (%)	-	3.93%	6.21%	7.87%	8.60%	12.08%	14.09%	18.60%
Comparison significance		0.253	0.084	0.488	0.004	0.125	0.005	0.000

Ranking	**9**	**10**	**11**	**12**	**13**	**14**	**15**
Model (# of parameters)	**PB1 (1)**	**PB0 (0)**	**Nash (0)**	**stEWA (1)**	**REL (2)**	**NRL (2)**	**RL (2)**
Avg. Scores	0.0636	0.0653	0.0828	0.0865	0.1303	0.1311	0.1749
Gap to the best (%)	25.91%	29.24%	63.94%	71.19%	158.07%	159.51%	246.32%
Comparison significance	0.000	0.000	0.000	0.000	0.000	0.000	0.000

3.6.3 All Trials

The conclusions I draw in this section are quite similar to those in the previous one. It is worth noting that the best performing models (see Table 6) do much better in predicting average behavior over all trials than they do in the other two prediction tasks.

The most accurate model is, in this third prediction task, the SFP model. The set of best performing models counts six models, of which four are equilibrium concepts. IBE performs significantly better than pNRA, although the estimated difference between their Prediction scores is quite small. Moreover, the accuracy of pNRA is equivalent to that of SFP.

The reinforcement models REL and NRL do once again very poorly, providing predictions of the 180% and 275% less accurate than those of SFP. Also in this case, Self-tuning EWA gives predictions statistically equivalent to those of Nash equilibrium and RL model.

Surprisingly, the QRE model is among the best performing models. This is quite interesting as this means that QRE is not able to capture neither short nor long run behavior, but is able to capture behavior that is in the middle of the previous two.

Table 6. Summary of simulation results for all trials of play. Models are ranked from the best to the worst (from the left to the right) according to average MSD or Prediction Scores (third row). The fourth row reports in percentage how worse is the accuracy of a certain model with respect to the best performing one. Each cell in the fifth row reports the p-value of the test of the null hypothesis of no difference between the average score of the corresponding model and that of the best performing model. Shaded cells refer to the cases in which the null hypothesis is *not* rejected at a 5% level.

Ranking	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Model (# of parameters)	**SFP (2)**	**IBE (0)**	**NFP (2)**	**7-sampling (0)**	**pNRA (1)**	**Payoff-sampling (1)**	**QRE (1)**	**NRA (0)**
Avg. Scores	0.0360	0.0363	0.0365	0.0385	0.0389	0.0418	0.0419	0.0419
Gap to the best (%)	-	0.88%	1.46%	7.15%	8.05%	16.12%	16.39%	16.59%
Comparison significance		0.718	0.088	0.910	0.070	0.015	0.208	0.000

Ranking	**9**	**10**	**11**	**12**	**13**	**14**	**15**
Model (# of parameters)	**PB0 (0)**	**PB1 (1)**	**stEWA (1)**	**RL (2)**	**Nash (0)**	**REL (2)**	**NRL (2)**
Avg. Scores	0.0437	0.0448	0.0634	0.0644	0.0716	0.1008	0.1350
Gap to the best (%)	21.63%	24.44%	76.18%	78.96%	98.99%	180.34%	275.27%
Comparison significance	0.005	0.000	0.000	0.000	0.000	0.000	0.000

3.7 Summary and Conclusions

In all three prediction tasks (early trials, long run, and average behavior), the performance pNRA is equivalent to that of the most accurate model (at a 5% level of significance). NRA is outperformed in the long run and average behavior tasks by pNRA, showing that the introduction of a parameter tuning sensitivity to net rewards significantly improves accuracy. NRA and pNRA are very accurate predictors of empirical data, performing always significantly better than Nash equilibrium, stEWA, and reinforcement models.

In each prediction task, we can define a set of best performing models i.e., models whose performance is statistically equivalent to that of the model with the smallest Prediction score. In the three prediction tasks, the model that provides the smallest Prediction score is not always the same: NFP in the short run, IBE in the long run, and SFP in all trials. These results not only confirm the robustness and reliability of regret-based learning models (particularly those of SFP and NFP), but also show that some

stationary models are very good predictors of behavior in the early periods of play as well as in the long run.

If it is clear that on average regret based models outperform reinforcement-based ones (confirming the results reported in the second chapter of my thesis), the analysis concerning equilibrium models is less straightforward, and it is not clear why models based on so different assumptions provide in some cases equivalently accurate predictions (as also pointed out in Selten and Chmura, 2008).

Another important result is that behavioral stationary concepts (IBE and NRA) are never outperformed by QRE, Action-sampling, and Payoff-sampling (i.e., best response models), although in some cases the two classes of solution concepts are equivalent in predicting data. For this reason, I think that it would be important to include in the set of criteria for model selection the plausibility of the assumptions on which models are based, at least as a tie breaking rule, since the causal relationship between assumptions and model accuracy is, in this context, of particular interest (Burnham and Anderson, 2003). We do not have to forget that best-response models are to be interpreted as "*as if*" models: they do not aim at replicating the *real* mechanisms at the basis of the decision-making process, but merely its effects. In other words, from this point of view, what matters is whether or not models are able to predict data, *as if* agents would act according to them. Trivially, the fact that none of us is able to think rationally (i.e., as prescribed by standard theory of choice) and act accordingly is not new, and any argument against standard theory based on this objection would be rather poor. The point here is that if we have to choose between two models which perform almost equivalently, why should not we privilege the use of that one that embeds principles about the *real* mechanisms of choice behavior? This approach I suggest would be much more informative, as it would allow us to infer the *real* bases of choice behavior. Of course, the judgment about plausibility of assumptions must be cautiously done because there are no principles that can guide us in this kind of task, and caution is primarily needed in those cases in which we are interested to judge whether certain assumptions are more plausible than others.

The NRA and pNRA models are analytically tractable, straightforwardly generalizable to *n*-person games, and based on assumptions validated by recent research on the neural mechanism at the basis of human choice behavior. These features make the NRA and pNRA models particularly appealing.

My analysis confirms the poor predictive power of Nash equilibrium, as reported in many other contributions. Compared to the most accurate model, standard theory provides predictions that are worse of the 106% in the first 50 trials, of the 64% in the last 50 trials, and of the 99% over all periods.

In the long run, reinforcement models provide significantly less accurate predictions than Nash equilibrium (with the exception of RL in average prediction task). I also find confirmation of the result shown in Marchiori and Warglien (2008), according to which regret-based learning models are better than reinforcement-based ones; indeed, NFP, SFP and PB0 always perform significantly better than stEWA, NRL, REL, and RL.

Among models of learning, NFP and SFP are the best predictors: their predictive accuracy is statistically equivalent to that of PB0 and PB1 only in the short run, and predict always significantly better than stEWA and reinforcement models. It is worth noting that out of the eight models of equilibrium I consider, only four (NFP, SFP, PB0, and PB1) perform always significantly better than Nash equilibrium, whereas all equilibrium models give more accurate predictions than does standard theory.

If compared to learning models, stationary concepts are, in general, less complex (statistically, analytically, and computationally). Nonetheless, with the exception of QRE, their predictions of short run behavior are as accurate as that of the best performing learning model, which constitutes a strong argument in favor of equilibrium concepts.

As stated in Selten and Chmura (2008), two-person 2x2 completely mixed games constitute a small set of games for testing models of interactive choice behavior and it would be interesting and important to gather data also from more general patterns of strategic interaction.

Table 7. MSD and Prediction Scores in the First 50 Trials.

Model (# of parameters)	NFP (2)	NRA (0)	PB0 (0)	PB1 (1)	SFP (2)	IBE (0)	pNRA (1)	Payoff-sampling (1)
Avg. Score	**0.047**	**0.050**	**0.050**	**0.050**	**0.050**	**0.050**	**0.051**	**0.051**
Gap to the best (%)	**-**	**5.45%**	**5.54%**	**6.00%**	**6.32%**	**6.70%**	**7.11%**	**7.48%**
AGK50	0.014	0.014	0.013	0.014	0.014	0.014	0.014	0.014
AGK67	0.103	0.075	0.051	0.056	0.103	0.075	0.079	0.109
AGK75	0.087	0.090	0.085	0.083	0.095	0.090	0.092	0.090
ERSB G1	0.105	0.114	0.131	0.121	0.105	0.115	0.113	0.111
ERSB G2	0.061	0.120	0.054	0.080	0.072	0.116	0.121	0.059
ERSB G3	0.094	0.094	0.104	0.091	0.106	0.110	0.098	0.112
ERSB G4	0.105	0.107	0.122	0.114	0.107	0.106	0.107	0.110
ERSB G5	0.037	0.035	0.047	0.041	0.038	0.040	0.035	0.044
ERSB G6	0.079	0.095	0.048	0.059	0.098	0.127	0.103	0.098
ERSB G7	0.078	0.094	0.060	0.064	0.096	0.126	0.105	0.075
ERSB G8	0.097	0.089	0.111	0.101	0.098	0.100	0.090	0.094
ERSB G9	0.036	0.026	0.021	0.023	0.051	0.039	0.027	0.042
ERSB G10	0.069	0.069	0.122	0.107	0.067	0.050	0.067	0.063
RSW D	0.032	0.031	0.083	0.068	0.028	0.023	0.030	0.026
S&C G1	0.026	0.026	0.015	0.043	0.028	0.015	0.025	0.040
S&C G2	0.013	0.015	0.017	0.023	0.013	0.017	0.015	0.013
S&C G3	0.021	0.014	0.011	0.013	0.018	0.007	0.012	0.014
S&C G4	0.023	0.016	0.010	0.021	0.025	0.006	0.015	0.021
S&C G5	0.015	0.016	0.015	0.020	0.015	0.010	0.015	0.012
S&C G6	0.008	0.008	0.008	0.008	0.008	0.008	0.008	0.009
S&C G7	0.011	0.022	0.048	0.017	0.012	0.040	0.027	0.037
S&C G8	0.014	0.023	0.026	0.032	0.013	0.014	0.023	0.012
S&C G9	0.052	0.052	0.048	0.047	0.045	0.031	0.050	0.047
S&C G10	0.032	0.032	0.031	0.034	0.036	0.021	0.031	0.045
S&C G11	0.008	0.009	0.007	0.013	0.008	0.006	0.008	0.016
S&C G12	0.008	0.008	0.009	0.009	0.007	0.005	0.008	0.010

Model (# of parameters)	NRL (2)	7-sampling (0)	RL (2)	QRE (1)	stEWA (1)	Nash (0)	REL (2)
Avg. Score	**0.051**	**0.054**	**0.054**	**0.060**	**0.066**	**0.098**	**0.099**
Gap to the best (%)	**8.74%**	**13.83%**	**14.30%**	**25.93%**	**39.38%**	**106.32%**	**108.70%**
AGK50	0.013	0.014	0.015	0.014	0.013	0.014	0.013
AGK67	0.084	0.084	0.066	0.061	0.049	0.126	0.052
AGK75	0.091	0.121	0.086	0.082	0.086	0.148	0.098
ERSB G1	0.119	0.105	0.121	0.136	0.156	0.134	0.162
ERSB G2	0.050	0.109	0.048	0.071	0.091	0.298	0.121
ERSB G3	0.090	0.135	0.096	0.136	0.158	0.265	0.192
ERSB G4	0.121	0.099	0.116	0.126	0.133	0.124	0.134
ERSB G5	0.035	0.043	0.036	0.054	0.071	0.069	0.087
ERSB G6	0.057	0.110	0.053	0.047	0.050	0.200	0.056
ERSB G7	0.061	0.107	0.058	0.060	0.068	0.213	0.077
ERSB G8	0.093	0.107	0.103	0.132	0.152	0.171	0.168
ERSB G9	0.029	0.033	0.024	0.022	0.024	0.053	0.029
ERSB G10	0.077	0.053	0.096	0.146	0.160	0.071	0.188
RSW D	0.025	0.025	0.022	0.028	0.069	0.031	0.078
S&C G1	0.017	0.016	0.026	0.054	0.132	0.096	0.183
S&C G2	0.081	0.032	0.148	0.052	0.029	0.072	0.081
S&C G3	0.021	0.008	0.048	0.020	0.021	0.040	0.177
S&C G4	0.016	0.008	0.023	0.013	0.008	0.025	0.109
S&C G5	0.024	0.013	0.019	0.014	0.011	0.020	0.062
S&C G6	0.009	0.008	0.009	0.009	0.008	0.010	0.017
S&C G7	0.045	0.062	0.068	0.148	0.087	0.199	0.096
S&C G8	0.044	0.030	0.039	0.048	0.036	0.063	0.066
S&C G9	0.035	0.043	0.037	0.033	0.058	0.038	0.161
S&C G10	0.039	0.022	0.032	0.027	0.029	0.033	0.081
S&C G11	0.043	0.007	0.011	0.009	0.004	0.017	0.060
S&C G12	0.017	0.005	0.007	0.006	0.009	0.007	0.017

Table 8. MSD and Prediction Scores in the Last 50 Trials.

Model (# of parameters)	IBE (0)	pNRA (1)	NFP (2)	7-sampling (0)	SFP (2)	Payoff-sampling (1)	QRE (1)	NRA (0)
Avg. Score	**0.051**	**0.052**	**0.054**	**0.054**	**0.055**	**0.057**	**0.058**	**0.060**
Gap to the best (%)	**-**	**3.93%**	**6.21%**	**7.87%**	**8.60%**	**12.08%**	**14.09%**	**18.60%**
AGK50	0.009	0.009	0.009	0.009	0.009	0.009	0.009	0.009
AGK67	0.101	0.125	0.135	0.118	0.142	0.149	0.119	0.101
AGK75	0.142	0.150	0.167	0.162	0.155	0.154	0.146	0.142
ERSB G1	0.167	0.166	0.166	0.159	0.170	0.177	0.156	0.159
ERSB G2	0.068	0.074	0.074	0.138	0.086	0.116	0.073	0.068
ERSB G3	0.137	0.139	0.141	0.149	0.138	0.144	0.130	0.135
ERSB G4	0.105	0.108	0.106	0.113	0.106	0.103	0.127	0.116
ERSB G5	0.069	0.068	0.065	0.065	0.066	0.067	0.068	0.074
ERSB G6	0.062	0.060	0.058	0.045	0.067	0.059	0.036	0.050
ERSB G7	0.093	0.097	0.099	0.089	0.095	0.102	0.118	0.091
ERSB G8	0.053	0.052	0.050	0.052	0.050	0.050	0.082	0.067
ERSB G9	0.081	0.079	0.083	0.080	0.083	0.087	0.078	0.075
ERSB G10	0.056	0.057	0.057	0.057	0.068	0.059	0.080	0.075
RSW D	0.019	0.020	0.015	0.021	0.015	0.013	0.009	0.036
S&C G1	0.025	0.013	0.028	0.014	0.030	0.019	0.044	0.051
S&C G2	0.010	0.009	0.010	0.017	0.011	0.011	0.040	0.019
S&C G3	0.013	0.012	0.022	0.015	0.021	0.013	0.023	0.037
S&C G4	0.009	0.010	0.007	0.010	0.007	0.007	0.014	0.023
S&C G5	0.018	0.019	0.017	0.015	0.015	0.016	0.018	0.032
S&C G6	0.008	0.008	0.007	0.007	0.007	0.007	0.008	0.010
S&C G7	0.010	0.010	0.010	0.011	0.015	0.011	0.048	0.039
S&C G8	0.012	0.023	0.016	0.012	0.016	0.030	0.023	0.040
S&C G9	0.024	0.026	0.025	0.032	0.029	0.027	0.018	0.054
S&C G10	0.009	0.014	0.014	0.011	0.011	0.024	0.016	0.023
S&C G11	0.006	0.008	0.006	0.007	0.005	0.010	0.008	0.018
S&C G12	0.007	0.009	0.008	0.006	0.008	0.007	0.005	0.014

Model (# of parameters)	PB1 (1)	PB0 (0)	Nash (0)	stEWA (1)	REL (2)	NRL (2)	RL (2)
Avg. Score	**0.064**	**0.065**	**0.083**	**0.086**	**0.130**	**0.131**	**0.175**
Gap to the best (%)	**25.91%**	**29.24%**	**63.94%**	**71.19%**	**158.07%**	**159.51%**	**246.32%**
AGK50	0.009	0.009	0.009	0.009	0.010	0.010	0.009
AGK67	0.087	0.084	0.171	0.086	0.096	0.102	0.093
AGK75	0.143	0.143	0.185	0.151	0.163	0.138	0.140
ERSB G1	0.159	0.163	0.205	0.200	0.211	0.222	0.156
ERSB G2	0.067	0.071	0.362	0.205	0.254	0.094	0.092
ERSB G3	0.141	0.142	0.213	0.172	0.189	0.221	0.131
ERSB G4	0.126	0.130	0.101	0.135	0.135	0.108	0.125
ERSB G5	0.077	0.088	0.074	0.105	0.109	0.125	0.072
ERSB G6	0.039	0.037	0.098	0.040	0.042	0.064	0.032
ERSB G7	0.112	0.115	0.150	0.174	0.187	0.128	0.100
ERSB G8	0.092	0.101	0.066	0.143	0.149	0.055	0.097
ERSB G9	0.075	0.077	0.099	0.088	0.095	0.162	0.079
ERSB G10	0.097	0.113	0.071	0.160	0.182	0.078	0.070
RSW D	0.075	0.102	0.011	0.084	0.085	0.195	0.194
S&C G1	0.054	0.038	0.062	0.158	0.247	0.111	0.109
S&C G2	0.030	0.026	0.049	0.028	0.112	0.091	0.267
S&C G3	0.021	0.021	0.033	0.025	0.235	0.064	0.055
S&C G4	0.019	0.020	0.019	0.011	0.102	0.077	0.082
S&C G5	0.031	0.032	0.020	0.016	0.089	0.125	0.209
S&C G6	0.010	0.010	0.009	0.009	0.021	0.210	0.249
S&C G7	0.042	0.027	0.064	0.121	0.198	0.131	0.709
S&C G8	0.059	0.057	0.026	0.057	0.087	0.161	0.168
S&C G9	0.038	0.038	0.022	0.034	0.216	0.059	0.606
S&C G10	0.021	0.021	0.019	0.017	0.078	0.094	0.090
S&C G11	0.017	0.018	0.010	0.009	0.070	0.122	0.154
S&C G12	0.013	0.014	0.006	0.010	0.027	0.460	0.460

Table 9. MSD and Prediction Scores in All Trials.

Model (# of parameters)	SFP (2)	IBE (0)	NFP (2)	7-sampling (0)	pNRA (1)	Payoff-sampling (1)	QRE (1)	NRA (0)
Avg. Score	**0.036**	**0.036**	**0.036**	**0.039**	**0.039**	**0.042**	**0.042**	**0.042**
Gap to the best (%)	**-**	**0.88%**	**1.46%**	**7.15%**	**8.05%**	**16.12%**	**16.39%**	**16.59%**
AGK50	0.009	0.009	0.009	0.009	0.009	0.009	0.009	0.009
AGK67	0.107	0.077	0.107	0.090	0.094	0.123	0.085	0.077
AGK75	0.112	0.106	0.112	0.131	0.114	0.122	0.107	0.106
ERSB G1	0.072	0.083	0.072	0.069	0.081	0.090	0.073	0.077
ERSB G2	0.037	0.065	0.037	0.069	0.062	0.051	0.039	0.073
ERSB G3	0.075	0.082	0.074	0.101	0.085	0.096	0.070	0.072
ERSB G4	0.059	0.063	0.063	0.066	0.069	0.062	0.091	0.074
ERSB G5	0.030	0.031	0.028	0.030	0.030	0.032	0.030	0.031
ERSB G6	0.043	0.056	0.041	0.049	0.052	0.068	0.022	0.037
ERSB G7	0.073	0.092	0.072	0.079	0.095	0.109	0.062	0.071
ERSB G8	0.043	0.042	0.043	0.044	0.042	0.044	0.068	0.049
ERSB G9	0.037	0.039	0.037	0.035	0.035	0.037	0.034	0.034
ERSB G10	0.056	0.055	0.059	0.056	0.058	0.059	0.089	0.071
RSW D	0.029	0.028	0.026	0.031	0.033	0.023	0.020	0.046
S&C G1	0.031	0.021	0.029	0.010	0.015	0.011	0.037	0.045
S&C G2	0.007	0.010	0.007	0.016	0.009	0.010	0.038	0.015
S&C G3	0.016	0.006	0.019	0.008	0.007	0.006	0.019	0.023
S&C G4	0.011	0.005	0.012	0.007	0.008	0.004	0.011	0.019
S&C G5	0.015	0.012	0.016	0.011	0.014	0.010	0.012	0.022
S&C G6	0.003	0.005	0.003	0.003	0.005	0.003	0.004	0.006
S&C G7	0.002	0.008	0.002	0.019	0.014	0.025	0.096	0.008
S&C G8	0.012	0.006	0.013	0.011	0.018	0.026	0.024	0.029
S&C G9	0.035	0.022	0.043	0.033	0.031	0.028	0.018	0.050
S&C G10	0.011	0.011	0.013	0.013	0.020	0.032	0.019	0.024
S&C G11	0.007	0.004	0.007	0.006	0.006	0.003	0.007	0.013
S&C G12	0.004	0.004	0.004	0.003	0.005	0.004	0.004	0.007

Model (# of parameters)	PB0 (0)	PB1 (1)	stEWA (1)	RL (2)	Nash (0)	REL (2)	NRL (2)
Avg. Score	**0.044**	**0.045**	**0.063**	**0.064**	**0.072**	**0.101**	**0.135**
Gap to the best (%)	**21.63%**	**24.44%**	**76.18%**	**78.96%**	**98.99%**	**180.34%**	**275.27%**
AGK50	0.009	0.009	0.008	0.010	0.009	0.009	0.009
AGK67	0.057	0.061	0.057	0.076	0.138	0.063	0.078
AGK75	0.104	0.103	0.109	0.104	0.156	0.119	0.099
ERSB G1	0.080	0.077	0.102	0.071	0.103	0.109	0.104
ERSB G2	0.056	0.066	0.131	0.035	0.243	0.171	0.059
ERSB G3	0.077	0.073	0.128	0.071	0.206	0.155	0.144
ERSB G4	0.088	0.085	0.102	0.087	0.067	0.101	0.071
ERSB G5	0.043	0.034	0.065	0.029	0.048	0.076	0.055
ERSB G6	0.023	0.024	0.041	0.023	0.128	0.051	0.069
ERSB G7	0.063	0.062	0.090	0.061	0.170	0.100	0.098
ERSB G8	0.075	0.068	0.117	0.055	0.072	0.128	0.049
ERSB G9	0.033	0.033	0.037	0.035	0.045	0.039	0.056
ERSB G10	0.108	0.091	0.154	0.068	0.073	0.171	0.064
RSW D	0.108	0.088	0.099	0.018	0.020	0.094	0.183
S&C G1	0.014	0.057	0.162	0.055	0.057	0.226	0.068
S&C G2	0.020	0.027	0.029	0.132	0.048	0.087	0.211
S&C G3	0.009	0.013	0.018	0.043	0.031	0.208	0.045
S&C G4	0.014	0.018	0.006	0.016	0.017	0.106	0.079
S&C G5	0.021	0.024	0.012	0.017	0.015	0.065	0.099
S&C G6	0.006	0.007	0.006	0.004	0.004	0.014	0.173
S&C G7	0.011	0.012	0.063	0.199	0.124	0.135	0.195
S&C G8	0.038	0.045	0.044	0.050	0.030	0.077	0.361
S&C G9	0.035	0.039	0.032	0.265	0.021	0.180	0.071
S&C G10	0.024	0.024	0.022	0.113	0.021	0.064	0.285
S&C G11	0.012	0.015	0.006	0.023	0.010	0.055	0.335
S&C G12	0.008	0.008	0.008	0.014	0.004	0.018	0.446

Table 10. Significance of model pairwise comparisons based on prediction scores in the First 50 Trials. The null hypothesis of no differences in Row and Column Model average scores is tested (Mann-Whitney-Wilcoxon test). Shaded cells indicate comparisons for which we fail to reject the null at the 5% level.

	NFP (2)	NRA (0)	PB0 (0)	PB1 (1)	SFP (2)	IBE (0)	pNRA (1)	Payoff-sampling (1)
NFP (2)		0.939	0.928	0.183	0.003	0.341	0.874	0.073
NRA (0)	0.939		0.772	0.032	0.067	0.679	0.795	0.044
PB0 (0)	0.928	0.772		0.929	0.443	0.137	0.558	0.186
PB1 (1)	0.183	0.032	0.929		0.929	0.220	0.042	0.838
SFP (2)	0.003	0.067	0.443	0.929		0.049	0.094	0.849
IBE (0)	0.341	0.679	0.137	0.220	0.049		0.471	0.003
pNRA (1)	0.874	0.795	0.558	0.042	0.094	0.471		0.015
Payoff-sampling (1)	0.073	0.044	0.186	0.838	0.849	0.003	0.015	
NRL (2)	0.397	0.006	0.085	0.438	0.877	0.000	0.009	0.957
7-sampling (0)	0.312	0.194	0.545	0.945	0.788	0.007	0.244	0.932
RL (2)	0.084	0.013	0.039	0.143	0.462	0.000	0.014	0.400
QRE (1)	0.030	0.016	0.000	0.013	0.111	0.001	0.017	0.111
stEWA (1)	0.001	0.000	0.000	0.000	0.004	0.000	0.000	0.010
Nash (0)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
REL (2)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

	NRL (2)	7-sampling (0)	RL (2)	QRE (1)	stEWA (1)	Nash (0)	REL (2)
NFP (2)	0.397	0.312	0.084	0.030	0.001	0.000	0.000
NRA (0)	0.006	0.194	0.013	0.016	0.000	0.000	0.000
PB0 (0)	0.085	0.545	0.039	0.000	0.000	0.000	0.000
PB1 (1)	0.438	0.945	0.143	0.013	0.000	0.000	0.000
SFP (2)	0.877	0.788	0.462	0.111	0.004	0.000	0.000
IBE (0)	0.000	0.007	0.000	0.001	0.000	0.000	0.000
pNRA (1)	0.009	0.244	0.014	0.017	0.000	0.000	0.000
Payoff-sampling (1)	0.957	0.932	0.400	0.111	0.010	0.000	0.000
NRL (2)		0.297	0.045	0.621	0.057	0.000	0.000
7-sampling (0)	0.297		0.033	0.006	0.004	0.000	0.000
RL (2)	0.045	0.033		0.695	0.064	0.000	0.000
QRE (1)	0.621	0.006	0.695		0.000	0.000	0.000
stEWA (1)	0.057	0.004	0.064	0.000		0.000	0.000
Nash (0)	0.000	0.000	0.000	0.000	0.000		0.053
REL (2)	0.000	0.000	0.000	0.000	0.000	0.053	

Table 11. Significance of model pairwise comparisons cased on prediction scores in the Last 50 Trials. The null hypothesis of no differences in Row and Column Model average scores is tested (Mann-Whitney-Wilcoxon test). Shaded cells indicate comparisons for which we fail to reject the null at the 5% level.

	IBE (0)	pNRA (1)	NFP (2)	7-sampling (0)	SFP (2)	Payoff-sampling (1)	QRE (1)	NRA (0)
IBE (0)		0.253	0.084	0.488	0.004	0.125	0.005	0.000
pNRA (1)	0.253		0.268	0.746	0.045	0.017	0.045	0.000
NFP (2)	0.084	0.268		0.221	0.082	0.169	0.212	0.001
7-sampling (0)	0.488	0.746	0.221		0.582	0.408	0.109	0.000
SFP (2)	0.004	0.045	0.082	0.582		0.800	0.348	0.001
Payoff-sampling (1)	0.125	0.017	0.169	0.408	0.800		0.632	0.008
QRE (1)	0.005	0.045	0.212	0.109	0.348	0.632		0.280
NRA (0)	0.000	0.000	0.001	0.000	0.001	0.008	0.280	
PB1 (1)	0.000	0.000	0.000	0.001	0.000	0.002	0.013	0.011
PB0 (0)	0.000	0.000	0.000	0.000	0.000	0.001	0.004	0.085
Nash (0)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.002
stEWA (1)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
REL (2)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
NRL (2)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
RL (2)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

	PB1 (1)	PB0 (0)	Nash (0)	stEWA (1)	REL (2)	NRL (2)	RL (2)
IBE (0)	0.000	0.000	0.000	0.000	0.000	0.000	0.000
pNRA (1)	0.000	0.000	0.000	0.000	0.000	0.000	0.000
NFP (2)	0.000	0.000	0.000	0.000	0.000	0.000	0.000
7-sampling (0)	0.001	0.000	0.000	0.000	0.000	0.000	0.000
SFP (2)	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Payoff-sampling (1)	0.002	0.001	0.000	0.000	0.000	0.000	0.000
QRE (1)	0.013	0.004	0.000	0.000	0.000	0.000	0.000
NRA (0)	0.011	0.085	0.002	0.000	0.000	0.000	0.000
PB1 (1)		0.054	0.025	0.000	0.000	0.000	0.000
PB0 (0)	0.054		0.050	0.000	0.000	0.000	0.000
Nash (0)	0.025	0.050		0.654	0.000	0.000	0.000
stEWA (1)	0.000	0.000	0.654		0.000	0.000	0.000
REL (2)	0.000	0.000	0.000	0.000		0.124	0.647
NRL (2)	0.000	0.000	0.000	0.000	0.124		0.005
RL (2)	0.000	0.000	0.000	0.000	0.647	0.005	

Table 12. Significance of model pairwise comparisons based on prediction scores in All Trials. The null hypothesis of no differences in Row and Column Model average scores is tested (Mann-Whitney-Wilcoxon test). Shaded cells indicate comparisons for which we fail to reject the null at the 5% level.

	SFP (2)	IBE (0)	NFP (2)	7-sampling (0)	pNRA (1)	Payoff-sampling (1)	QRE (1)	NRA (0)
SFP (2)		0.718	0.088	0.910	0.070	0.015	0.208	0.000
IBE (0)	0.718		0.367	0.429	0.011	0.020	0.007	0.000
NFP (2)	0.088	0.367		0.633	0.322	0.087	0.260	0.000
7-sampling (0)	0.910	0.429	0.633		0.164	0.081	0.048	0.001
pNRA (1)	0.070	0.011	0.322	0.164		0.093	0.193	0.000
Payoff-sampling (1)	0.015	0.020	0.087	0.081	0.093		0.341	0.185
QRE (1)	0.208	0.007	0.260	0.048	0.193	0.341		0.608
NRA (0)	0.000	0.000	0.000	0.001	0.000	0.185	0.608	
PB0 (0)	0.005	0.000	0.009	0.007	0.000	0.067	0.048	0.671
PB1 (1)	0.000	0.000	0.000	0.001	0.000	0.014	0.020	0.001
stEWA (1)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
RL (2)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.019
Nash (0)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
REL (2)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
NRL (2)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

	PB0 (0)	PB1 (1)	stEWA (1)	RL (2)	Nash (0)	REL (2)	NRL (2)
SFP (2)	0.005	0.000	0.000	0.000	0.000	0.000	0.000
IBE (0)	0.000	0.000	0.000	0.000	0.000	0.000	0.000
NFP (2)	0.009	0.000	0.000	0.000	0.000	0.000	0.000
7-sampling (0)	0.007	0.001	0.000	0.000	0.000	0.000	0.000
pNRA (1)	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Payoff-sampling (1)	0.067	0.014	0.000	0.000	0.000	0.000	0.000
QRE (1)	0.048	0.020	0.000	0.000	0.000	0.000	0.000
NRA (0)	0.671	0.001	0.000	0.019	0.000	0.000	0.000
PB0 (0)		0.097	0.000	0.030	0.000	0.000	0.000
PB1 (1)	0.097		0.000	0.112	0.000	0.000	0.000
stEWA (1)	0.000	0.000		0.303	0.225	0.000	0.000
RL (2)	0.030	0.112	0.303		0.203	0.000	0.000
Nash (0)	0.000	0.000	0.225	0.203		0.000	0.000
REL (2)	0.000	0.000	0.000	0.000	0.000		0.382
NRL (2)	0.000	0.000	0.000	0.000	0.000	0.382	

CHAPTER 4

4. LEARNING IN MULTI-GAME EXPERIMENTS

Abstract. I designed and ran multi-game experiments in which subjects played sequences of different two-person, 2x2 games with a unique equilibrium in mixed strategies (MSE). Games in each sequence were obtained by multiplying for a randomly drawn positive constant the payoffs of two completely mixed games. I use these experimental data to test the predictive power of the Perceptron-Based (PB) model and compare it with that of other popular learning and equilibrium models of interactive choice behavior. As a result, the PB model is, by design, the sole model of learning capable to discriminate between the two different classes of games, and it outperforms in accuracy Nash equilibrium and all the other models of learning as well. In addition, experimental results do not provide evidence of learning spillover effects across games, which might provide an explanation for why non-standard stationary models turn out to be the best predictors of observed choice frequencies.

4.1 Economic Models of Generalization and Empirical Evidence

As Stahl and Van Huyck (2002:2) put it, "our ability to understand and predict human behavior would be greatly enhanced by a successful theory of how past experiences with similar situations affect current behavior". The issue of generalization is also economically relevant because most human interactive learning happens in contexts where tasks do not repeat themselves identically over time – as in the typical patterns of interaction that have been empirically studied up to now in the experimental and behavioral economics literature. As seen in the Introduction, generalizing from examples and learning of conditional behavior are among the most fundamental functions of human beings.

Despite the economic relevance of this topic, the experimental and behavioral economics literature on generalization counts only few contributions, which can be classified in three groups based on the approach adopted.

The first is that proposed in Gilboa and Schmeidler's *Case-Based Decision Theory*; this model is designed to describe a decision maker who bases his decisions on the consequences derived from past actions taken in relevant similar cases (*learning by examples*) (Gilboa and Schmeidler, 1995; Rubinstein, 1998). Analogously, Leland (2001) proposes a model of decision-making where agents are assumed to base their decisions on comparisons regarding the similarity or dissimilarity of attributes across alternatives, along the lines suggested by Tversky (1969), although without providing an explicit definition of similarity for games.

The second stream of literature deals with the problem of how past experience fosters the emergence of coordination in a population of players and how subjects anchor to previously learned strategic behaviors. Contributions by Rankin, Van Huyck, and Battalio (2000) and Stahl and Van Huyck (2002) report evidence on the origin of conventions based on payoff dominance in laboratory cohorts playing repeatedly similar but not identical stag-hunt games. Using data from their experiments on repeated play of similar stag hunt games, Stahl and Van Huyck (2002) compare the predictive performance of four models of adaptive beliefs formation, with a number of free parameters ranging from 3 to 8, in which the probability for a player to choose the payoff dominant action is a function of a measure of the distance between the payoffs in the current game and those in the previous one. Their main result is that only the more complex model (allowing for an exogenous belief in the salience of the payoff

dominant action) can explain the data with one set of estimated parameters. In Devetag (2005) the issue of transfer of learning between two different typologies of coordination games (i.e., from *critical mass* to *minimum effort* games) is investigated. Here, the main finding is that subjects use what they have previously learned playing the first game repeatedly as they play the second one; the present study assesses the extent to which efficient achieved precedents can be successfully used as a coordination device in the new situation. In Egidi and Narduzzo (1997), it is shown that past experience could even lead to "strongly routinized behaviors, i.e. groups of player which, after the training phase, adopted one strategy once and for all, and insisted on using it even when hands could not be efficiently played with the strategy adopted" (Egidi and Narduzzo, 1997:1). This above mentioned phenomenon is commonly known as *path-dependence*.

Works by LiCalzi (1992), Sgroi and Zizzo (2002 and 2007) and Sgroi (2003) illustrate the third approach in which the issue of generalization is purely addressed from a modeling point of view. LiCalzi's paper is based on the question raised by Fudenberg and Kreps (1993) – "how does fictitious play extend to situations where players try to extrapolate from past experiences in similar games?" – and provides a new model of fictitious play by "cases". Sgroi and Zizzo present a neural network-based methodology for examining the learning of game playing rules in never previously encountered games. They show how a back-propagation neural network can learn Nash strategies if all other players play Nash equilibrium and the network receives as a feedback target the Nash equilibrium itself. The most important result is that one can teach Nash equilibrium to a neural network with a 60% success rate – similar to the rate experimentally observed on human subjects.

The contribution by Huck, Jehiel, and Rutter (2007) cannot be included in any of the three above-mentioned approaches because it merges experimental and modeling methodologies. However, this contribution adds important insights on the issue. The authors investigate under what conditions learning spillovers arise in a context of multiple interaction tasks i.e., when long run behaviors in one game are affected by behaviors in another one. They find that learning spillovers are a function of the structure of the feedback received by agents. Indeed, when playing two different dominance solvable games, if the information that subjects receive about different interactions is easily separable, then spillovers are minimal; on the contrary, if information is not clearly separated for each game or if it is less accessible, then

learning spillovers do matter and lead subjects' behavior away from that predicted by standard theory. Moreover, Jehiel (2005) provides the concept of analogy-based expectation equilibrium, suggesting a way to broaden the notion of equilibrium in the presence of learning spillovers.

4.2 The Generalizing PB Model

As explained in the second chapter, the PB model of learning is technically a one-layer neural network with continuous output units; neural networks of this kind are called *simple perceptrons* (first proposed by Rosenblatt, 1958). Now, simple perceptrons suffer some severe theoretical limitations in the discrimination tasks they can perform (Hertz, Krogh, and Palmer, 1991; Minsky and Papert, 1969; this latter contribution caused a significant decline in interest and funding of neural network research). However, in spite of these limitations, simulation results discussed in the following Section 6 show that perceptrons are able to discriminate between two different strategic situations and to replicate human choice behavior.

The PB model presents some architectural analogies with established models of learning in games, but it has also some peculiar features that differentiate it from its competitors. Established learning models in economics have two main, cyclically intertwined, component processes:

1. Behavior is generated by some *stochastic choice rule* that maps propensities into probabilities of play.
2. *Learning* employs feedback to modify propensities, which in turn affect subsequent choices.

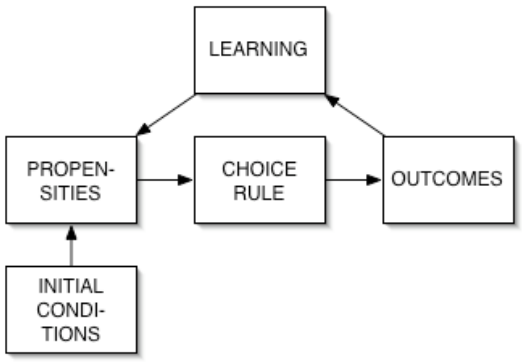


Figure 1. General architecture of a "propensities and stochastic choice rule" learning model.

The PB model's architecture resembles that of other learning models only partially: one can easily interpret network outputs as propensities, whose normalization plays the role of the *stochastic choice rule*. What makes our model different is that choice behavior depends also directly upon game payoffs (represented in the "input layer"). In other words, while in a typical economic learning model choice is a function of propensities only, here it is function of both propensities *and* the payoffs of the game. Furthermore, the learning rule itself depends upon the input payoffs.

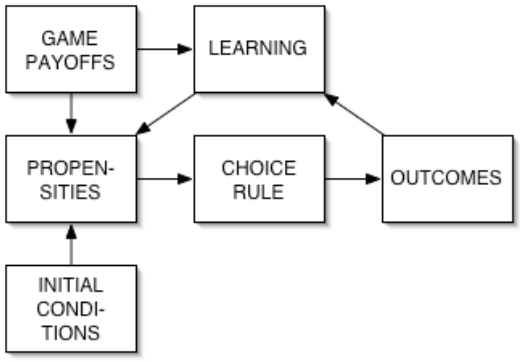


Figure 2. General architecture of the PB model.

This architecture provides the PB model with a peculiar capability to discriminate among different games. Conventional learning models in economics are designed for repeated games. There is learning, but no discrimination or generalization: the simulated learning agent is unable to discriminate between different games at a certain moment. If given abruptly two different games, it would respond in the same way (or just throw away what has been learned). On the other hand, discrimination is something Perceptrons do pretty well: since output is also directly affected by perceived inputs (the activation states of input units), a network, after learning, will respond differently to different games.

4.3 Experimental Design

I designed and ran a multi-game experiment in order to investigate how acquired strategic skills affect subjects' decisions when facing new strategic situations. The experiment consists of two treatments.

In Treatment 1, four cohorts of 8 players each (corresponding to four independent observations) played a sequence of 120 games. This sequence was obtained perturbing the payoffs of two 2x2, two-person constant-sum games (henceforth Game A and

Game B) with a unique MSE. Specifically, perturbations were obtained by multiplying payoffs for a randomly drawn positive constant (from the normal distribution with mean 10 and standard deviation 4). Thus, I obtained two sets of *Type A* and *Type B* games. It is worth noting that Type A and B games are characterized by the same equilibrium probabilities of Game A and B, respectively. The sequence of games was constructed so that in each block of 10 trials there were 5 Type A and 5 Type B games in random order; thus, in each block subjects could play the same number of times Type A and Type B games. Each cohort played a different sequence obtained according to the procedure described above. Due to the structural characteristics of game sequences, I decided to average observed choice frequencies for Type A and Type B games within blocks of 20 trials each. Treatment 2 is equal to the first, except for the pair of games I used to build the sequences (Games A and C). The experimental design is summarized in Table 1. Games A, B, and C were designed in such a way that, in all three games, the sum of all payoffs for row and column players is constantly equal to the same number.

At the beginning of the experiment, subjects in each cohort were randomly assigned the role of either row or column player. At each round and within each cohort, subjects assigned to different roles were randomly and anonymously paired (*random matching protocol*). Random matching was adopted in order to discourage coordination and reciprocating behaviors (Erev and Haruvy, 2005). At the end of each round, subjects were provided with feedback about their and their opponents' actions and outcomes in that round (complete information protocol).

In order to avoid income effects and induce incentives based on performance, subjects were paid on the basis of the outcomes in 12, randomly drawn, rounds. All treatments were run at the experimental laboratory of CEEL (University of Trento).

Subjects could not participate to more than one treatment.

Table 1. Experimental design.

<table>
<tr><td rowspan="2">Treatment 1</td><td>Game A</td><td>Game B</td></tr>
<tr><td>
<table>
<tr><td>Player 2 / Player 1</td><td>L</td><td>R</td></tr>
<tr><td>U</td><td>17,5</td><td>16,6</td></tr>
<tr><td>D</td><td>8,14</td><td>17,5</td></tr>
</table>
NE: P(U) = 0.9, P(L) = 0.1
</td><td>
<table>
<tr><td>Player 2 / Player 1</td><td>L</td><td>R</td></tr>
<tr><td>U</td><td>5,17</td><td>2,20</td></tr>
<tr><td>D</td><td>4,18</td><td>11,11</td></tr>
</table>
NE: P(U) = 0.7, P(L) = 0.9
</td></tr>
<tr><td rowspan="2">Treatment 2</td><td>Game A</td><td>Game C</td></tr>
<tr><td>
<table>
<tr><td>Player 2 / Player 1</td><td>L</td><td>R</td></tr>
<tr><td>U</td><td>17,5</td><td>16,6</td></tr>
<tr><td>D</td><td>8,14</td><td>17,5</td></tr>
</table>
NE: P(U) = 0.9, P(L) = 0.1
</td><td>
<table>
<tr><td>Player 2 / Player 1</td><td>L</td><td>R</td></tr>
<tr><td>U</td><td>17,5</td><td>15,7</td></tr>
<tr><td>D</td><td>15,7</td><td>18,4</td></tr>
</table>
NE: P(U) = 0.6, P(L) = 0.6
</td></tr>
</table>

4.4 Experimental Results

Figures 4 and 5 report my experimental results. Reported choice frequencies correspond, for each treatment, to average choice behavior over four independent observations (four cohorts of subjects). In both figures, I report separately choice frequency trajectories for the two types of games in blocks of 10 trials, although within each block subjects played also 10 games of the other kind. Tables 2 and 3 report empirical choice frequencies averaged over all trials for each independent observation and for each treatment.

I used Game A to build the sequences in both treatments in order to assess to what extent choice behavior is affected by the simultaneous play of another game. If we consider choice behavior in type A games averaged over all periods of play, data do not show any significant difference between average choice behavior in Treatments 1 and 2, for both row (Mann-Whitney-Wilcoxon paired test, p-value = 0.4227) and column players (Mann-Whitney-Wilcoxon paired test, p-value = 0.625). This suggests that learning spillover effects across different games are negligible, at least in this simple case in which the games are of just two types, and that subjects are then able to recognize the structure of the two types of games and play accordingly. This might also explain why equilibrium models outperform dynamic models of learning in predicting data from my experiment.

Empirical results show that in Treatment 1 observed behavior in Type A games is not well approximated by Nash equilibrium. Row players play Nash mixture only in the first two blocks (as Table 4 illustrates), but then observed choice frequencies depart from Nash's mixture. As for column players, play starts from random behavior in the first block, and then converges to values systematically higher than predicted frequencies (observed value of 0.328 versus estimated value of 0.1).

In Treatment 1, Nash descriptive power of choice behavior in type B games is very poor. Column players are supposed to choose action U with probability 0.7, whereas observed play converges to the relative frequency of 0.9. Column players are supposed to choose action L with probability 0.9, but observed behavior converges, after two blocks, to the value of 0.4.

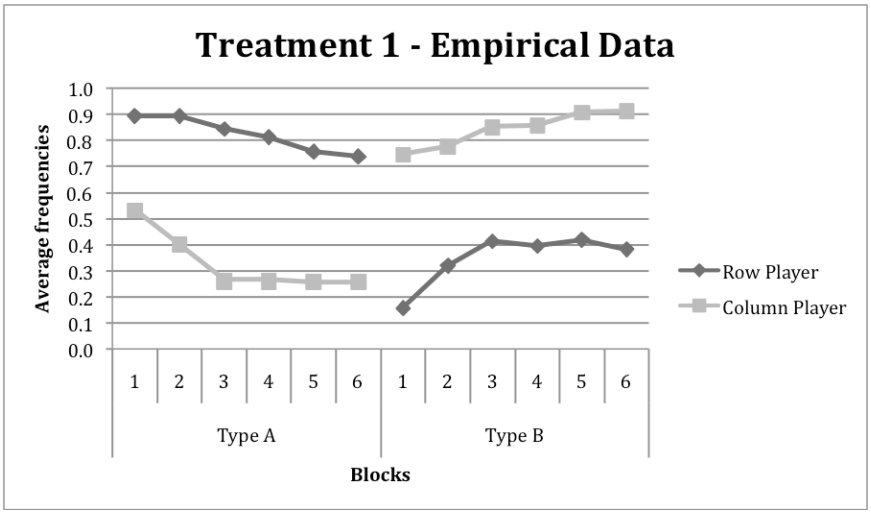


Figure 4. Observed choice frequencies averaged over all independent observations in Treatment 1. Nash predicted behavior for Type A games is P(U) = 0.9 and P(L) = 0.1, whereas for Type B games P(U) = 0.7 and P(L) = 0.9.

In Treatment 2, the relative frequency with which row players choose action U in Type A games is systematically higher than that predicted by standard theory, similarly to what happened in Treatment 1. It is interesting to note here that in Type C games empirical behavior of both row and column players converges to Nash probabilities (P(U) = P(L) = 0.6) in the last block of trials.

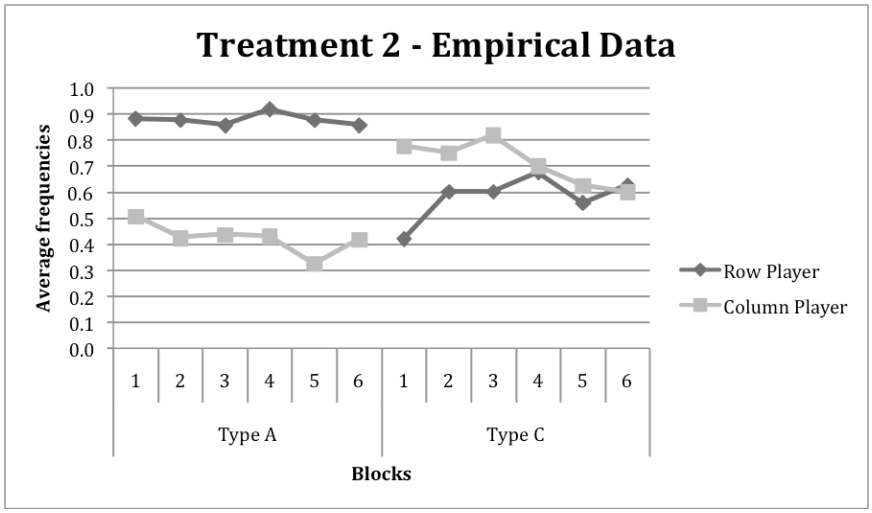


Figure 5. Observed choice frequencies averaged over all independent observations in Treatment 2. Nash predicted behavior for Type A games is P(U) = 0.9 and P(L) = 0.1, whereas for Type C games P(U) = P(L) = 0.6.

Table 2. Observed choice frequencies averaged over all periods in Treatment 1, for Type A and B games.

	Type A					
	Observation 1	**Observation 2**	**Observation 3**	**Observation 4**	**Mean**	**sd**
Row Player	0.829	0.708	0.900	0.854	0.823	0.082
Column Player	0.254	0.292	0.450	0.317	0.328	0.085
	Type B					
	Observation 1	**Observation 2**	**Observation 3**	**Observation 4**	**Mean**	**sd**
Row Player	0.221	0.621	0.454	0.092	0.347	0.236
Column Player	0.917	0.838	0.829	0.779	0.841	0.057

Table 3. Observed choice frequencies averaged over all periods in Treatment 2, for Type A and C games.

	Type A					
	Observation 1	**Observation 2**	**Observation 3**	**Observation 4**	**Mean**	**sd**
Row Player	0.829	0.804	0.892	0.983	0.877	0.080
Column Player	0.404	0.638	0.417	0.238	0.424	0.164
	Type C					
	Observation 1	**Observation 2**	**Observation 3**	**Observation 4**	**Mean**	**sd**
Row Player	0.717	0.517	0.500	0.583	0.579	0.098
Column Player	0.504	0.854	0.704	0.783	0.711	0.151

4.5 Methods

For each model, Mean Squared Deviation (MSD) scores are computed. In order to get the estimated choice frequencies, I ran each model with parameters set to the values that minimize MSD across all datasets considered in the third chapter of my thesis, and computed the MSD with respect to the experimental data. In some sense this procedure corresponds to the *leave-one-out* one (described in the second chapter) that, although indirectly, penalizes models with higher degree of (statistical) complexity. For this reason, in the remainder I will refer to these scores as Prediction scores. I will not consider in this analysis the significance of model pairwise comparisons (as I do in the previous two chapters), since the aim of this study is to test some qualitative aspects of the models.

As always, larger values of Prediction scores correspond to less accurate predictions.

Since learning models are stochastic, the estimated frequency of choice was obtained as the average over 150 simulations, which were run for each experiment and for each parameter configuration. Moreover, in order to make simulation results comparable, the initialization of all dynamic models was set to assure equal probabilities of choosing each action at the first round.

I compare the performances of eight different models of learning and five stationary concepts (for a comprehensive description of some of the most popular stationary and dynamic models see the Introduction). I consider in my analysis the following models of learning: Normalized Fictitious Play (NFP) (Erev et al., 2007); Normalized Reinforcement Learning (NRL) (Erev et al., 2007); Perceptron-Based (PB0 and PB1

with, respectively, zero and one free parameters) (Marchiori and Warglien, 2008); Reiforcement Learning (REL) (Erev and Roth, 1998); Reinforcement Learning (RL) (Erev et al., 2007); Stochastic Fictitious Play (Erev at al., 2007); and Self-tuning Experience Weighted Attraction (stEWA) (Ho, Camerer, and Chong, 2007). The equilibrium concepts I consider are: Nash Equilibrium; Impulse Balance Equilibrium (IBE) (Ockenfels and Selten, 2005); Action-Sampling Equilibrium (Sample-7) (Selten, 2000); Payoff-Sampling Equilibrium (Osborne and Rubinstein, 1998); Net Reward Attractions (NRA) Equilibrium and its parametric version pNRA (proposed and described in Chapter 3).

Logit Quantal Response Equilibrium (McKelvey and Palfrey, 1995) is not invariant to the multiplication of all payoffs for a constant (for a fixed value of the parameter λ), and for that reason it is not included it in my analysis.

4.6 Simulation Results

Experimental results show that subjects can distinguish between the two situations, behaving differently in the two different strategic situations. However, there is no valid reason to expect that this result holds true also in the presence of sequences with more complex games, or with many, structurally different games.

The first important result is that learning models, with the exception of the PB model, are not able to discriminate between the two different strategic situations, providing a poor "average" behavior for both strategic situations, and are always outperformed by Nash equilibrium.

On the contrary, the PB model is able to replicate subjects' conditional behavior, due to its direct dependence of response on game payoffs and performs better than standard theory of equilibrium. Moreover, simulation results show that there is a qualitative parallelism between the behavior produced by the PB model and the observed one; if we consider estimated frequencies for Type A games averaged over all periods in Treatment 1 and 2, the difference is statistically significant for both row (Mann-Whitney-Wilcoxon paired test, p-value < 2.2e-16) and column player (Mann-Whitney-Wilcoxon paired test, p-value < 2.2e-16), but the estimated differences are very small (0.0561 for row player and -0.0875 for column player). This means that not only that the PB model is able to replicate subjects' ability to recognize different strategic situations and act accordingly, but also that its structure is complex enough to avoid spillover effects across games.

Simulation results show that the PB model is the most accurate model of learning. However, non-standard equilibrium models are by far the best predictors of the data, perhaps due to the fact that spillover effects are negligible in my experiment. As a confirmation of this, models of equilibrium give good predictions of behavior in Type A games in both Treatment 1 and 2: this would not be possible if choice behavior were conditioned by the simultaneous play of another game.

Reading from Tables A1 and A2 in Appendix A, we can see that in these games equilibrium models alternative to standard game theory (NRA, IBE, Payoff-sampling, and Action sampling models) provide quite similar predictions. This is the reason for which these stationary models are almost equivalent in predicting data. This is not a shortcoming of my experimental design, since here the aim is to test the predictive power of learning models on data from experiments with a radically new design with respect to the established pattern of analysis proposed up to now in the literature. Moreover, a thorough comparative analysis of equilibrium models has already been proposed in previous chapters. The role of equilibrium models here is rather that of a benchmark, allowing for a better evaluation of the performances of dynamic models.

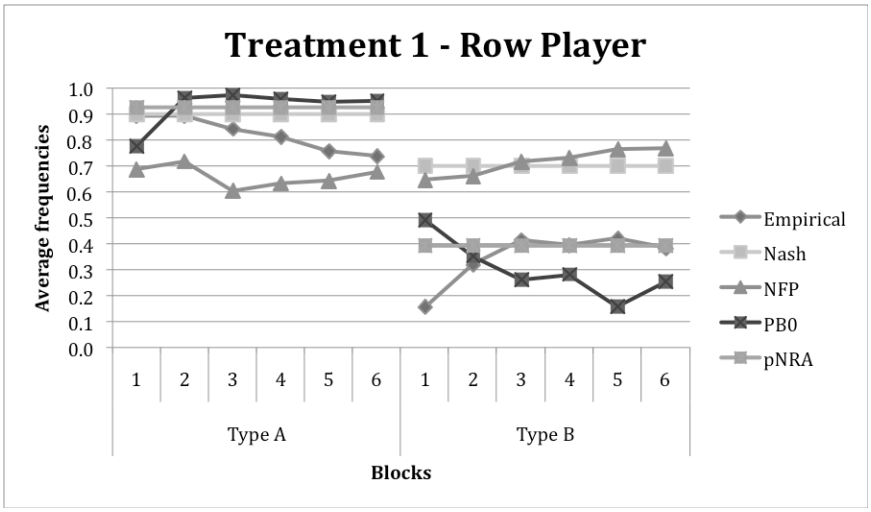


Figure 6. Row Player's predicted and observed choice frequencies in Type A and B games.

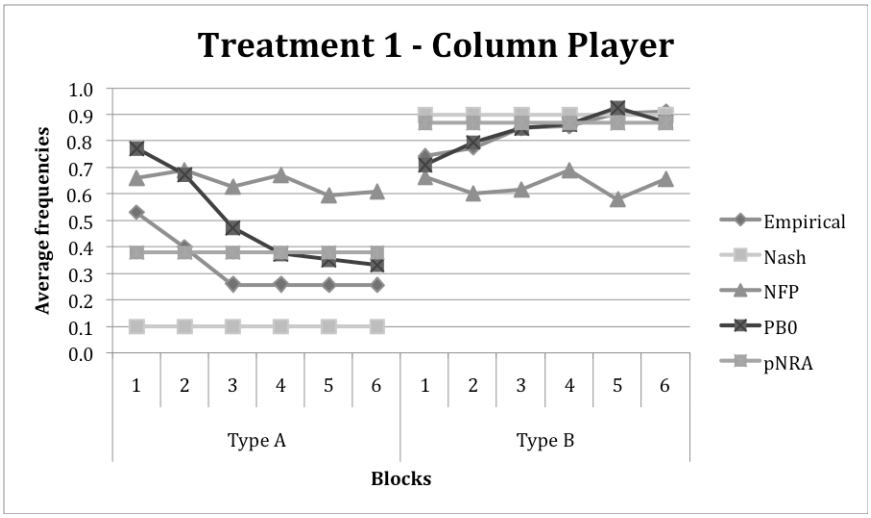


Figure 7. Column Player's predicted and observed choice frequencies in Type A and B games.

Table 4. Average model Prediction scores in Type A and B games (Treatment 1).

Models	MSD Scores	Player Type	Type A Games Average MSD	Type B Games Average MSD
Nash	**0.053**	**Row**	0.010	0.133
		Column	0.063	0.007
NFP	**0.081**	**Row**	0.030	0.139
		Column	0.107	0.048
NRL	**0.094**	**Row**	0.038	0.138
		Column	0.189	0.011
PB0	**0.024**	**Row**	0.023	0.039
		Column	0.035	0.001
PB1	**0.028**	**Row**	0.012	0.019
		Column	0.075	0.005
REL	**0.076**	**Row**	0.107	0.033
		Column	0.041	0.122
RL	**0.097**	**Row**	0.043	0.139
		Column	0.194	0.010
SFP	**0.094**	**Row**	0.037	0.153
		Column	0.126	0.061
stEWA	**0.086**	**Row**	0.051	0.081
		Column	0.201	0.009
IBE	**0.012**	**Row**	0.014	0.010
		Column	0.020	0.005
Payoff-sampling	**0.016**	**Row**	0.015	0.029
		Column	0.012	0.007
NRA	**0.016**	**Row**	0.010	0.008
		Column	0.040	0.004
pNRA	**0.011**	**Row**	0.014	0.010
		Column	0.014	0.005
7-Sampling	**0.012**	**Row**	0.018	0.014
		Column	0.011	0.006

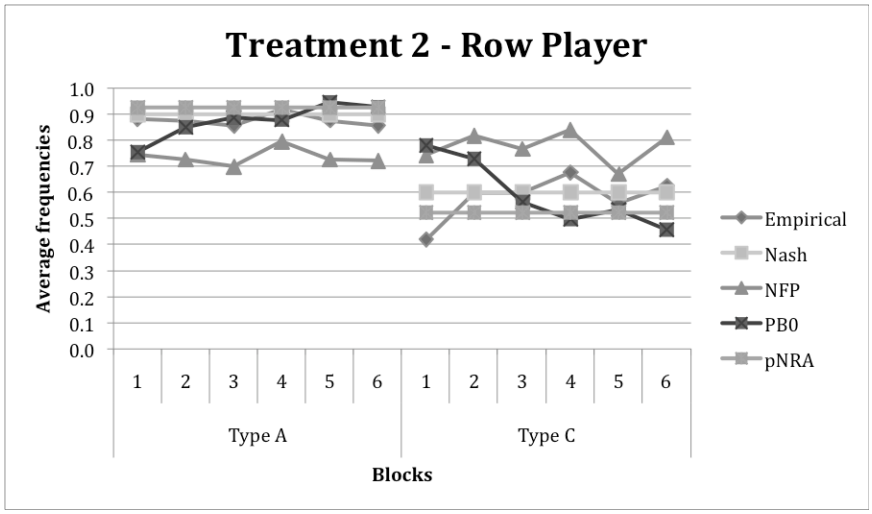


Figure 8. Row Player's predicted and observed choice frequencies in of type A and C games.

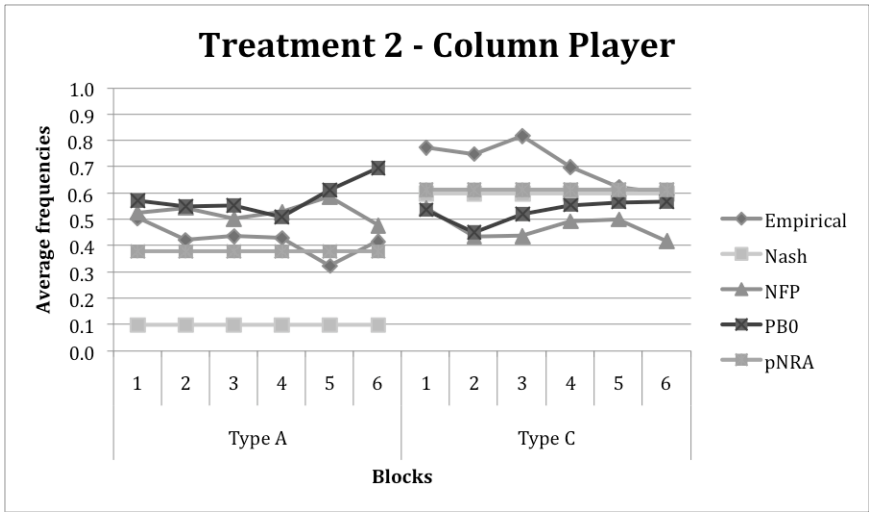


Figure 9. Column Player's predicted and observed choice frequencies in Type A and C games.

Table 5. Average model Prediction scores in Type A and C games (Treatment 2).

Models	MSD Scores	Player Type	Type A Games Average MSD	Type C Games Average MSD
Nash	**0.034**	**Row**	0.001	0.007
		Column	0.108	0.019
NFP	**0.036**	**Row**	0.020	0.043
		Column	0.017	0.065
NRL	**0.047**	**Row**	0.025	0.071
		Column	0.034	0.060
PB0	**0.029**	**Row**	0.005	0.035
		Column	0.033	0.043
PB1	**0.026**	**Row**	0.009	0.020
		Column	0.044	0.032
REL	**0.055**	**Row**	0.144	0.014
		Column	0.009	0.052
RL	**0.053**	**Row**	0.027	0.070
		Column	0.043	0.072
SFP	**0.036**	**Row**	0.021	0.043
		Column	0.011	0.068
stEWA	**0.058**	**Row**	0.141	0.012
		Column	0.071	0.007
IBE	**0.007**	**Row**	0.003	0.008
		Column	0.003	0.012
NRA	**0.010**	**Row**	0.001	0.013
		Column	0.009	0.019
Payoff-sampling	**0.008**	**Row**	0.003	0.007
		Column	0.007	0.015
pNRA	**0.008**	**Row**	0.003	0.010
		Column	0.005	0.016
7-Sampling	**0.008**	**Row**	0.005	0.007
		Column	0.011	0.009

4.7 What Do Subjects Learn?

An important question that arises from the analysis of my experimental results is that of what do subject learn to play. In order to answer this question, a good idea is that of looking at how Prediction scores of a model vary across blocks of trials; this should make clear whether or not subjects' behavior is converging to the behavior predicted by that particular model. Figures in this section report this information, first for data from Treatment 1and then for data from Treatment 2.

Figures 10-13 report the Prediction scores corresponding to each model in each block of trials, for row and column players, in Type A and B games (Treatment 1). These figures help visualize toward which prediction observed behavior is directed.

For what concerns equilibrium models, we can notice the poor predictive power of Nash equilibrium of row players' behavior in type B games (figure 10) and of column

players' behavior in type A games (figure 11). As already noted, equilibrium concepts alternative to standard theory provide very similar predictions and observed behavior converges to the predicted one in Type B games. However, this is not true for Type A games, in which Prediction scores does not seem to be decreasing over time for both row and column players.

As for learning models, we can first note that they are in general less accurate than stationary concepts and, second, that in most of the cases, the produced dynamics diverge greatly from the observed ones (figures 12 and 13). On the other hand, only observed behavior of row player in Type A games does not converge to the PB model prediction (figure 12). In general it can be easily seen that the PB model provides the best approximation of empirical data.

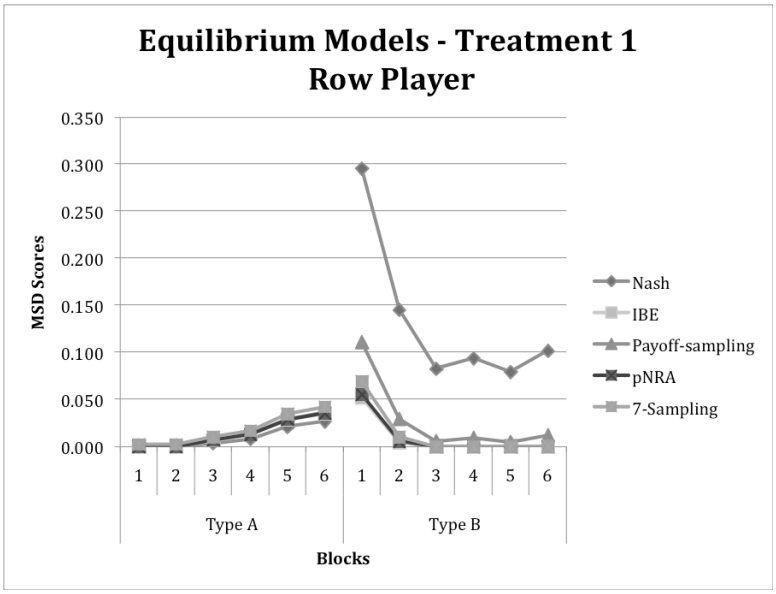


Figure 10. Models of equilibrium. Plot of MSD Scores against blocks of trials for row player in Treatment 1.

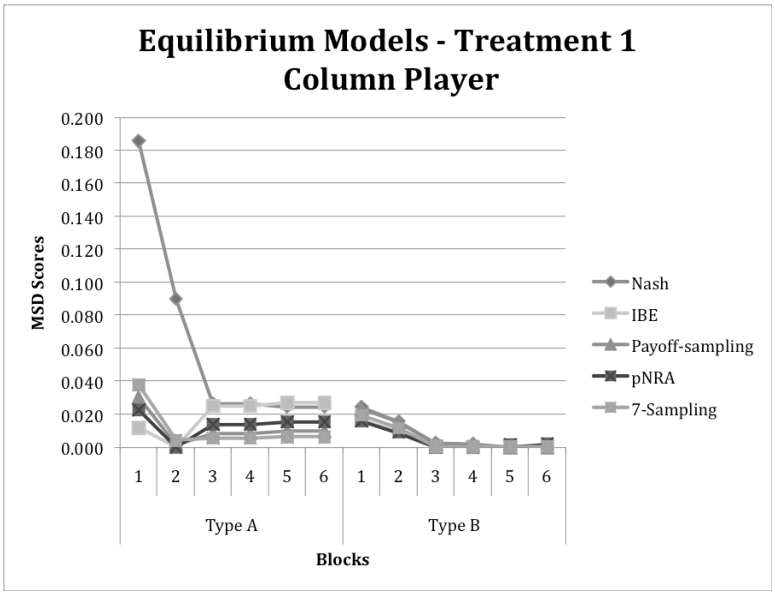


Figure 11. Models of equilibrium. Plot of MSD Scores against blocks of trials for column player in Treatment 1.

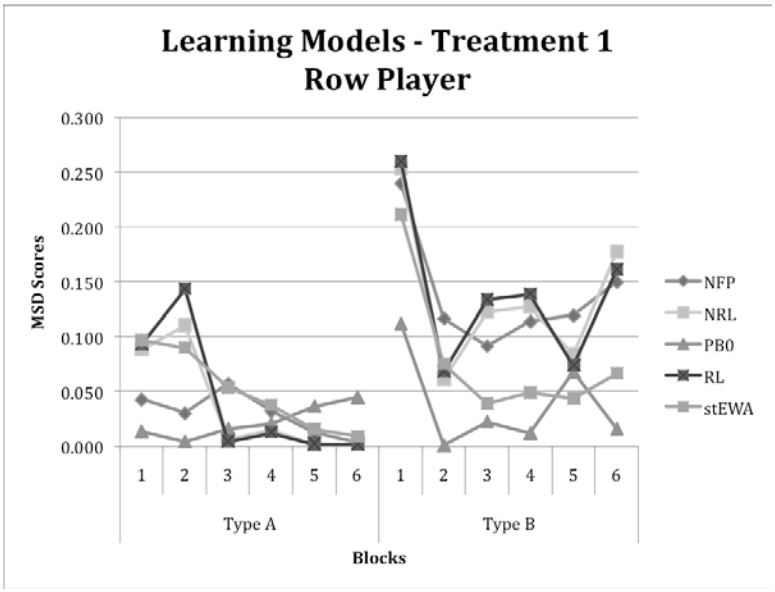


Figure 12. Models of learning. Plot of MSD Scores against blocks of trials for row player in Treatment 1.

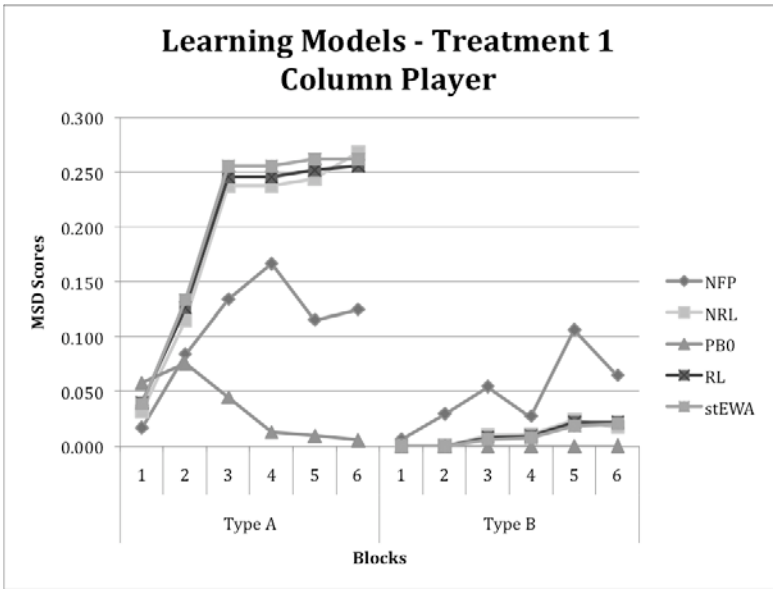


Figure 13. Models of learning. Plot of MSD Scores against blocks of trials for column player in Treatment 1.

In Figures 14-17, choice frequencies on type A and C games are considered (Treatment 2).

For what concerns equilibrium models, in general convergence of observed choice behavior to the estimated frequencies is not monotone as in Treatment 1. It turns out that Nash equilibrium is a very good predictor of row player's behavior in Type C games (figure 14), but a very bad predictor of column player's observed choice frequencies in Type A games (for a quantitative reference see Table 5). Also in this case we can observe that the performances of equilibrium models are quite similar.

On the contrary, learning models provide predictions that are dramatically different and, in general, there is no convergence to empirical frequencies of choice (in particular for column players, see figure 17). However, observed behavior seem to converge toward the PB model predictions, with the exception of column player in Type A games, as in Treatment 1; this is, of course, consequence of the fact that empirical behavior in Type A games is on average the same in the two treatments.

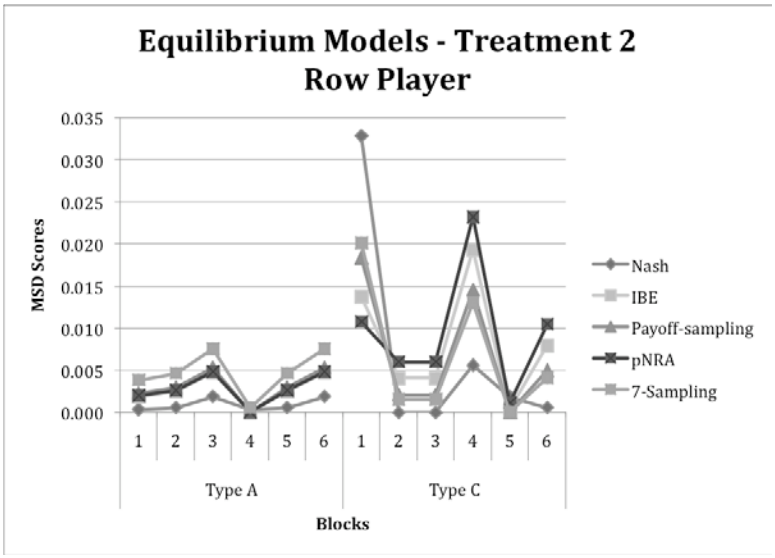


Figure 14. Models of equilibrium. Plot of MSD Scores against blocks of trials for row player in Treatment 2.

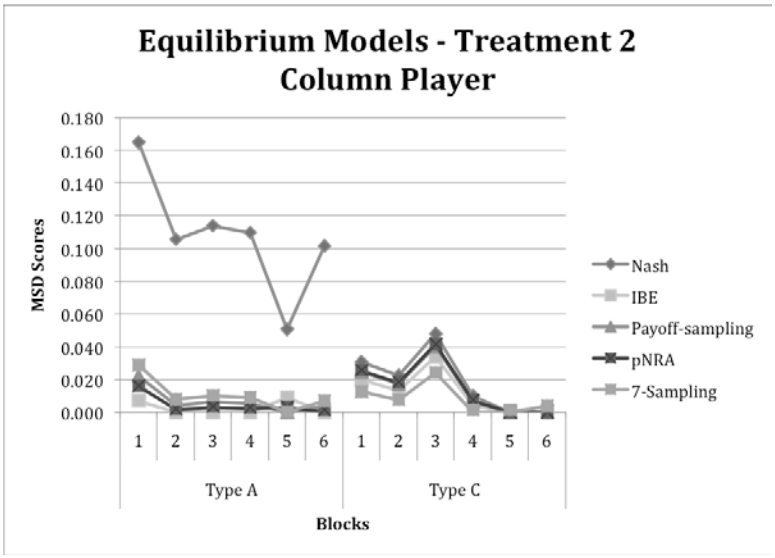


Figure 15. Models of equilibrium. Plot of MSD Scores against blocks of trials for column player in Treatment 2.

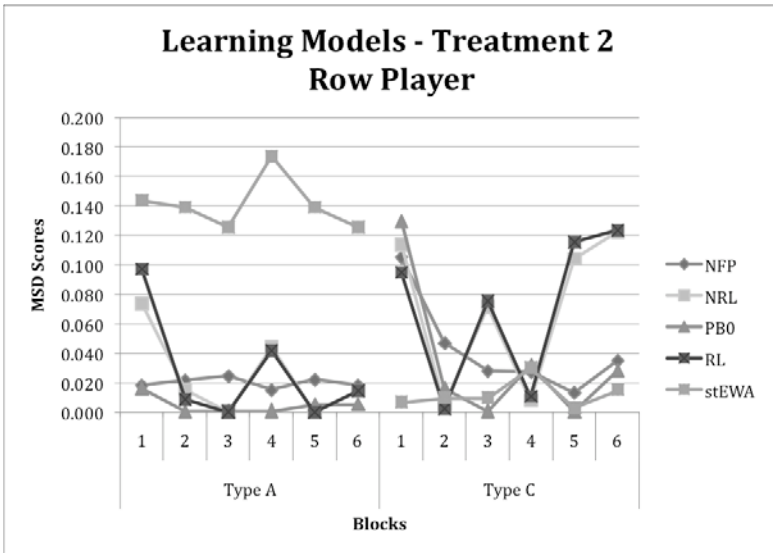


Figure 16. Models of learning. Plot of MSD Scores against blocks of trials for row player in Treatment 2.

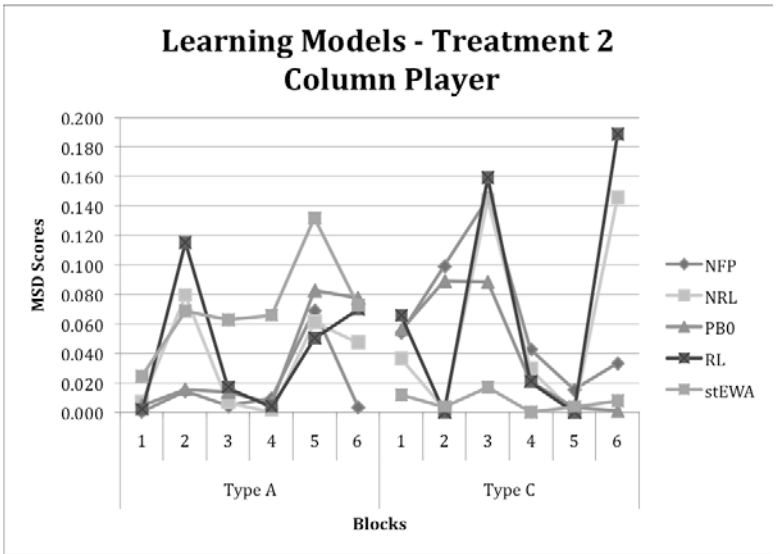


Figure 17. Models of learning. Plot of MSD Scores against blocks of trials for column player in Treatment 2.

4.8 Conclusions and Further Research

I designed and ran multigame experiments in order to assess to what extent past experience affects current choice behavior and to test the capability of some of the most popular dynamic models of choice behavior in providing different responses to different strategic situations.

Empirical results show that in the simple experimental settings considered here, players are able to recognize the two different game structures in each sequence and

play accordingly to this classification. This is particularly interesting because even though the two original games, from which each sequence was obtained, differ only in the predicted probabilities of play, subjects can recognize them, but do not use Nash equilibrium to form their strategies (neither in the early periods, nor in the long run).

Simulation results show that traditional "attraction and stochastic choice rule" learning models are not able to discriminate between the different strategic situations, providing a poor "average" behavior for both strategic situations, and are always outperformed by Nash equilibrium. On the contrary, the PB model is able to replicate subjects' conditional behavior, due to its direct dependence of response on game payoffs and performs better than standard theory of equilibrium. In addition, not only the PB model is able to replicate subjects' ability to recognize different strategic situations and act accordingly, but its structure is also complex enough to avoid spillover effects across games, establishing a qualitative parallelism between predicted and observed choice behavior.

Another important result of my analysis is that non-standard equilibrium models are by far the best predictors of empirical data. I conjecture that this is a consequence of the fact that in my experiments spillover effects are negligible. As a confirmation of this, models of equilibrium give good predictions of behavior in Type A games in both Treatment 1 and 2: this would not be possible if choice behavior were conditioned by the simultaneous play of another game.

If we ask ourselves what do subject learn, then the answer is not unique and straightforward. Indeed, if non-standard equilibrium models and the PB model provide the best approximations of observed behavior, it is also true that empirical choice frequencies do not converge in all cases to the predictions of one of these concepts, thus without giving support to this or that model.

These results leave room for further research.

First of all, further investigations could focus on the effects of some factors on learning spillovers, e.g. different degrees or distributions of payoff perturbations and an increase in the number of different types of games in a sequence. My conjecture is that an increase in the magnitude of payoff perturbations and in the number of different kinds of games in a sequence would lead subjects to confuse games toward an "average" play. However, this conjecture itself and the extent to which it is true have not yet been investigated experimentally. Further research could also include the design of multigame experiments in which the sequences of games are built starting from

more general patterns of strategic interaction other than two-person, 2x2 completely mixed games.

In addition, other potential properties of the PB model are worthy to be investigated. Above mentioned results show that both humans and neural networks are able to categorize games in a sequence, obtained perturbing the payoffs of two games. This parallelism does not hold for any other dynamic model I consider: standard models of economic learning (including the recently proposed NFP and SFP), by design, cannot capture such features of human behavior, because there is no way they can model dependence of behavior from the perception of different game structures. However, it would be also interesting to see under what conditions learning spillovers arise, and test whether neural networks are able to produce equivalent dynamics to those observed under these conditions. It would also be interesting to test the accuracy of equilibrium models in predicting data from multigame experiments in which learning spillovers are present.

4.9 Appendix A. Supporting Materials and Tables

Table A1. Predicted and observed choice frequencies for each game and block in Treatment 1.

Choice Frequencies		Type A						Type B					
		Block 1	Block 2	Block 3	Block 4	Block 5	Block 6	Block 1	Block 2	Block 3	Block 4	Block 5	Block 6
Empirical	P(A1):	0.894	0.894	0.844	0.813	0.756	0.738	0.156	0.319	0.413	0.394	0.419	0.381
	P(A2):	0.531	0.400	0.263	0.263	0.256	0.256	0.744	0.775	0.850	0.856	0.906	0.913
Nash	P(A1):	0.900	0.900	0.900	0.900	0.900	0.900	0.700	0.700	0.700	0.700	0.700	0.700
	P(A2):	0.100	0.100	0.100	0.100	0.100	0.100	0.900	0.900	0.900	0.900	0.900	0.900
NFP	P(A1):	0.686	0.719	0.604	0.633	0.642	0.677	0.646	0.661	0.716	0.732	0.765	0.769
	P(A2):	0.661	0.690	0.629	0.671	0.596	0.610	0.664	0.603	0.617	0.691	0.580	0.658
NRL	P(A1):	0.596	0.561	0.763	0.688	0.715	0.798	0.660	0.566	0.763	0.751	0.709	0.802
	P(A2):	0.710	0.740	0.750	0.750	0.750	0.774	0.735	0.740	0.750	0.750	0.750	0.780
PB0	P(A1):	0.775	0.964	0.973	0.959	0.949	0.949	0.491	0.353	0.262	0.283	0.157	0.255
	P(A2):	0.772	0.676	0.475	0.377	0.355	0.332	0.712	0.796	0.849	0.863	0.927	0.873
PB1	P(A1):	0.730	0.916	0.865	0.868	0.879	0.900	0.451	0.313	0.326	0.338	0.306	0.317
	P(A2):	0.701	0.640	0.477	0.611	0.588	0.554	0.669	0.769	0.792	0.794	0.847	0.784
REL	P(A1):	0.494	0.507	0.516	0.495	0.503	0.497	0.511	0.502	0.493	0.502	0.497	0.500
	P(A2):	0.515	0.490	0.492	0.510	0.504	0.507	0.498	0.502	0.503	0.485	0.497	0.498
RL	P(A1):	0.588	0.515	0.776	0.700	0.701	0.781	0.667	0.580	0.778	0.767	0.691	0.783
	P(A2):	0.731	0.756	0.758	0.758	0.758	0.762	0.755	0.756	0.758	0.758	0.758	0.763
SFP	P(A1):	0.724	0.710	0.520	0.601	0.675	0.686	0.667	0.642	0.779	0.798	0.738	0.774
	P(A2):	0.690	0.696	0.690	0.721	0.591	0.630	0.635	0.589	0.522	0.641	0.579	0.670
stEWA	P(A1):	0.582	0.593	0.611	0.617	0.628	0.640	0.617	0.594	0.612	0.617	0.629	0.640
	P(A2):	0.728	0.766	0.768	0.768	0.768	0.768	0.757	0.766	0.768	0.768	0.768	0.768
IBE	P(A1):	0.925	0.925	0.925	0.925	0.925	0.925	0.386	0.386	0.386	0.386	0.386	0.386
	P(A2):	0.421	0.421	0.421	0.421	0.421	0.421	0.873	0.873	0.873	0.873	0.873	0.873
NRA	P(A1):	0.900	0.900	0.900	0.900	0.900	0.900	0.337	0.337	0.337	0.337	0.337	0.337
	P(A2):	0.500	0.500	0.500	0.500	0.500	0.500	0.821	0.821	0.821	0.821	0.821	0.821
Payoff-sampling	P(A1):	0.929	0.929	0.929	0.929	0.929	0.929	0.490	0.490	0.490	0.490	0.490	0.490
	P(A2):	0.357	0.357	0.357	0.357	0.357	0.357	0.898	0.898	0.898	0.898	0.898	0.898
pNRA	P(A1):	0.926	0.926	0.926	0.926	0.926	0.926	0.392	0.392	0.392	0.392	0.392	0.392
	P(A2):	0.380	0.380	0.380	0.380	0.380	0.380	0.870	0.870	0.870	0.870	0.870	0.870
7-Sampling	P(A1):	0.943	0.943	0.943	0.943	0.943	0.943	0.420	0.420	0.420	0.420	0.420	0.420
	P(A2):	0.336	0.336	0.336	0.336	0.336	0.336	0.883	0.883	0.883	0.883	0.883	0.883

Table A2. Predicted and observed choice frequencies for each game and block in Treatment 2.

Choice Frequencies		Type A						Type C					
		Block 1	**Block 2**	**Block 3**	**Block 4**	**Block 5**	**Block 6**	**Block 1**	**Block 2**	**Block 3**	**Block 4**	**Block 5**	**Block 6**
Empirical	P(A1):	0.881	0.875	0.856	0.919	0.875	0.856	0.419	0.600	0.600	0.675	0.556	0.625
	P(A2):	0.506	0.425	0.438	0.431	0.325	0.419	0.775	0.750	0.819	0.700	0.625	0.600
Nash	P(A1):	0.900	0.900	0.900	0.900	0.900	0.900	0.600	0.600	0.600	0.600	0.600	0.600
	P(A2):	0.100	0.100	0.100	0.100	0.100	0.100	0.600	0.600	0.600	0.600	0.600	0.600
NFP	P(A1):	0.746	0.728	0.700	0.796	0.727	0.721	0.743	0.817	0.767	0.840	0.671	0.812
	P(A2):	0.527	0.544	0.502	0.528	0.588	0.477	0.543	0.435	0.438	0.493	0.501	0.418
NRL	P(A1):	0.609	0.751	0.883	0.708	0.864	0.977	0.756	0.673	0.867	0.764	0.879	0.975
	P(A2):	0.590	0.707	0.355	0.468	0.572	0.201	0.584	0.699	0.439	0.529	0.633	0.218
PB0	P(A1):	0.754	0.849	0.887	0.877	0.946	0.929	0.779	0.728	0.562	0.496	0.534	0.457
	P(A2):	0.573	0.550	0.554	0.509	0.613	0.697	0.538	0.451	0.521	0.556	0.566	0.569
PB1	P(A1):	0.693	0.784	0.812	0.826	0.887	0.864	0.713	0.688	0.597	0.599	0.616	0.509
	P(A2):	0.596	0.624	0.580	0.552	0.629	0.719	0.569	0.515	0.540	0.596	0.556	0.556
REL	P(A1):	0.506	0.496	0.502	0.490	0.497	0.497	0.499	0.494	0.508	0.494	0.492	0.487
	P(A2):	0.503	0.498	0.504	0.496	0.507	0.501	0.491	0.496	0.505	0.497	0.500	0.494
RL	P(A1):	0.569	0.781	0.882	0.714	0.879	0.977	0.727	0.649	0.875	0.780	0.896	0.977
	P(A2):	0.548	0.764	0.308	0.492	0.549	0.154	0.519	0.742	0.420	0.556	0.619	0.165
SFP	P(A1):	0.736	0.752	0.650	0.808	0.731	0.748	0.758	0.816	0.779	0.846	0.665	0.781
	P(A2):	0.521	0.538	0.465	0.529	0.510	0.493	0.523	0.415	0.473	0.416	0.575	0.429
stEWA	P(A1):	0.502	0.502	0.502	0.502	0.502	0.502	0.502	0.502	0.502	0.502	0.502	0.502
	P(A2):	0.663	0.687	0.688	0.688	0.688	0.688	0.666	0.687	0.688	0.688	0.688	0.688
IBE	P(A1):	0.925	0.925	0.925	0.925	0.925	0.925	0.536	0.536	0.536	0.536	0.536	0.536
	P(A2):	0.421	0.421	0.421	0.421	0.421	0.421	0.634	0.634	0.634	0.634	0.634	0.634
NRA	P(A1):	0.900	0.900	0.900	0.900	0.900	0.900	0.500	0.500	0.500	0.500	0.500	0.500
	P(A2):	0.500	0.500	0.500	0.500	0.500	0.500	0.600	0.600	0.600	0.600	0.600	0.600
Payoff-sampling	P(A1):	0.929	0.929	0.929	0.929	0.929	0.929	0.554	0.554	0.554	0.554	0.554	0.554
	P(A2):	0.357	0.357	0.357	0.357	0.357	0.357	0.618	0.618	0.618	0.618	0.618	0.618
pNRA	P(A1):	0.926	0.926	0.926	0.926	0.926	0.926	0.523	0.523	0.523	0.523	0.523	0.523
	P(A2):	0.380	0.380	0.380	0.380	0.380	0.380	0.615	0.615	0.615	0.615	0.615	0.615
7-Sampling	P(A1):	0.943	0.943	0.943	0.943	0.943	0.943	0.561	0.561	0.561	0.561	0.561	0.561
	P(A2):	0.336	0.336	0.336	0.336	0.336	0.336	0.662	0.662	0.662	0.662	0.662	0.662

4.10 Appendix B. Experimental Instructions

INSTRUCTIONS

(Translated from Italian)
You are participating to an experiment on interactive decision-making funded by the Italian Ministry of University and Research (MIUR). This experiment is not aimed at evaluating you neither academically nor personally. We have a policy of strict anonymity and we will never correlate data in such a way that it would allow us or others to identify your responses.
You will be paid on the basis of your performance, privately and in cash, according to the rules described below.

During the experiment, you will not be allowed to communicate with the other participants, neither verbally nor in any other way. If you have any problems or questions, raise your hand and a member of the staff will immediately contact you.

The experiment will consist of 120 round, and in each round you will face an interactive decision task. Specifically, in each round you will be randomly matched with another participant and your payoff will depend on both your decision and that of the other participant. The structure of each decision task will be represented by a *payoff matrix*, as shown in the following figure:

		The Other Player (Column Player)	
		Action 1	Action 2
YOU (Row Player)	Action 1	(6,4)	(4,7)
	Action 2	(3,4)	(5,6)

You have been assigned the role of "row player": therefore, the other player will *always* play the role of "column player".
For each player, two actions are available (labeled "action 1" and "action 2"). For every possible combination of actions by row and column players, there corresponds a cell in the payoff matrix. In every cell there are two numbers between parentheses: the first number corresponds to YOUR payoff (in experimental currency units) and the second corresponds to the payoff of the other player (again expressed in experimental currency units).

As an example, referring to the matrix reported below, if YOU choose to play "action 1" and the other player chooses to play "action 2", then the payoffs will be 4 for YOU (row player) and 7 for the other player (column player).

		The Other Player (Column Player)	
		Action 1	Action 2
YOU (Row Player)	Action 1	(6,4)	(4,7)
	Action 2	(3,4)	(5,6)

Please, remember that the experiment will consist of 120 rounds. In each round, you will be shown a sequence of two screenshots.

The first screenshot will show you the current payoff matrix and you will be invited to make a decision. In order to make a decision, you must type either "1" or "2" in the box labeled "your decision", and then click on the button "confirm". Once you have clicked the confirmation button, you cannot change your decision. You will have a maximum of 30 seconds to choose: after these 30 seconds a blinking red message will appear on the right-up corner of the screen and spur you on to take a decision. Delaying your decision will cause the other participants to wait for you.

Once all players have made their decision, the second screenshot will appear on your monitor. In this second screenshot there will be reported the action you chose, the action chosen by the other player, your respective payoffs, and the payoff matrix you saw in the first screenshot.
The second screenshot will be visible on your monitor for 10 seconds and then another round will start.

This process will be repeated for 120 times. After all rounds have been played, the experiment will be over and the procedure of payment will start. In order to determine your payment, 12 integers between 1 and 120 will be randomly drawn without replacement. In this way, 12 out of the 120 rounds will be randomly selected and you will be paid on the basis of their outcome. One experimental currency unit is equivalent to 10 eurocents (10 experimental units = 1 euro). Moreover, independently from your performance, you will be paid an additional show-up fee of 5 euros.

Before the beginning of the experiment, you will be asked to fill a questionnaire to verify whether the instructions have been understood. Then the experiment will start.
At the end of the experiment, you will be asked to fill a questionnaire for your payment.

Thank you for your kind cooperation!

CHAPTER 5

5. OVERALL CONCLUSIONS AND FURTHER RESEARCH

The Chapters 2 and 3 of my thesis are devoted to the introduction of a new model of learning (the Perceptron-Based model) and of a new model of equilibrium for normal form games (Net Reward Attractions Equilibrium), respectively. I investigate the formal properties of these two models, test their predictive power on data from experiments on two-person, 2x2 completely mixed games, and compare the accuracy of their predictions with those of other stationary and dynamic models representing cutting-edge research in the field of interactive choice behavior modeling.

In the third part of my thesis, I address issues of generalization and conditional behavior in repeated strategic interactions, using both experimental and computational methodologies. Of all parts of my thesis, the last one is that that most deserves further investigation. The reason for that is twofold. First, behavioral and experimental economics literature has paid, up to now, little attention to issues of generalization and conditional behavior in games, in spite of their pervasiveness in everyday life situations and their economic relevance. As a consequence, there is no established methodology to empirically investigate these topics. Second, due to the small number of combinations of strategic situations I study experimentally and to their specific nature, the fourth chapter of my thesis has a rather explorative nature, and a systematic investigation of how human beings generalize and apply their acquired strategic skills to new strategic situations is on the top list of my future research agenda.

All three parts of my thesis share the same methodological approach of model comparison based on *new-game* prediction tasks, opposed to the approach focused on *within-game* predictions (see Introduction and Erev and Haruvy, 2005). The fundamental assumption characterizing the former approach is the use of general parameter values to describe choice behavior over different conditions (games), therefore not allowing for individual or role-related agents' heterogeneity.

The detailed conclusions can be grouped into three distinct sections, one for each part of my thesis, without compromising the unitarity of this work. A section on some of the possible further steps of my research concludes.

5.1 Part One (Chapter 2)

The first important result of my thesis is that regret-based models are always more accurate predictors of empirical data than other models of interactive choice behavior I consider, thus showing that regret for foregone payoffs must play a central role in shaping human choice behavior. The important role of regret in repeated decision tasks has been confirmed also by recent research in the field of neuroscience.

However, the principal aim of the second chapter of my thesis is that of testing the viability of the PB model as predictor of empirical data. As a result, the PB model turns out to be the best predictor of observed data with respect to all other models of learning I consider in my analysis, with the exception of a model (Normalized Fictitious Play proposed by Ert and Erev, 2007) similarly based on regret.

The third important result is the poor performance of reinforcement based models: in some cases they provide less accurate predictions than those of the model of random choice behavior. This result is in contradiction with some recent contributions in the learning literature (Erev and Roth, 1998; Sarin and Vahid, 2001; Erev et al., 2007) and the motivations might be of two different natures. First, testing models on Selten and Chmura's (2008) games seems to particularly penalize reinforcement based models. In particular, the last six games, even though completely mixed, are not constant-sum and, for that reason, might have provided some incentive for cooperative and reciprocating behaviors. Reinforcement learning models do not take into account these cooperative features of human behavior, even indirectly, and are not able to predict behavior in such richer interactive situations. Another reason for the failure of reinforcement models might be that testing models on a large dataset would require the exploration of broader regions of the parameter spaces than those suggested by the authors of the models in previous works, where smaller datasets were considered.

In this chapter I further analyze the predictive power of models fed with game payoffs rescaled according to Kahnemann and Tversky's (1979 and 1992) *prospect theory*. Results show two facts as we pass from actual to rescaled payoffs: first, the ranking of the models remains unaltered; second, the increase in accuracy is significant for regret-based models (NFP, SFP, and PB) and marginal for all the others (stEWA and reinforcement learning models).

Unlike other models of learning, the free parameter PB0 model allows for individual and role-related agents' heterogeneity, as simulations have shown that

connection weights (which directly determine agents' choice behavior) are different for each artificial agent, and that connection weights associated to row players and column players are, on average, different.

5.2 Part Two (Chapter 3)

In Chapter 3, I propose a new behavioral equilibrium concept I call Net Reward Attractions (NRA) equilibrium, and compare its predictive accuracy with that of other five equilibrium concepts and eight models of learning, among the most popular in the literature on interactive decision making modeling. I provide also a parameterized version of NRA I call Parametric NRA (pNRA). It is obtained introducing a parameter $\lambda > 0$ that tunes players' sensitivity to net rewards. According to NRA, it is assumed that, in equilibrium, agents do not maximize their expected utility function, but that, for a player, the propensity of choosing an action is proportional to its corresponding expected net reward – net reward being defined as the difference between the actual payoff and the minimum obtainable one, given other players' moves.

For the comparison, I use here a dataset of experiments on 26 different games, smaller than that I use in the second chapter (which counts 35 games). Indeed, I consider here only datasets for which experimenters made available data for each independent observation (either at the individual or group level, depending on whether fix-pairing or random-matching protocol was used). This allows me to gather a large number of independent conditions on which to test each model and gives the Mann-Whitney-Wilcoxon test more chances to compare models more precisely.

The concept of net reward, as I use it, is very similar to Loomes and Sugden's (1982) concept of *rejoicing* i.e., a measure of the additional pleasure associated to the awareness of having chosen the best action. In this vein, the approach based on net rewards, which I adopt to model choice behavior in the long run, is complementary, although not equivalent, to that based on regret. In Loomes and Sugden's (1982) *regret theory*, these two complementary aspects are fused together in the *Rejoice/Regret* function (see the Introduction), and I show in Chapters 2 and 3 of my thesis that these two components can be separately used to successfully design models of choice behavior.

I tested all models on three prediction tasks, measuring their accuracy in predicting observed choice behavior averaged over the first 50 trials, the last 50 trials, and all

trials. In each prediction task, we can define a set of best performing models i.e., models whose performance is statistically equivalent to that of the model with the smallest Prediction score.

The first important result of my analysis is that in all three prediction tasks, pNRA is in the set of the most accurate models in predicting choice behavior. NRA is outperformed in predicting behavior in the long run by pNRA, showing that the introduction of a parameter tuning sensitivity to net rewards leads to an increase in accuracy. NRA and pNRA are then very accurate predictors of empirical data, performing always significantly better than Nash equilibrium, stEWA, and reinforcement-based models.

In the three prediction tasks, the model that provides the smallest Prediction score is not always the same: NFP in the short run, IBE in the long run, and SFP in all trials. These results not only confirm the robustness and reliability of regret-based learning models (in particular those of SFP and NFP), but also show that some stationary models are very good predictors of behavior in the early periods of play as well as in the long run.

If it is clear from the results that on average regret based models outperform reinforcement-based ones (confirming the results reported in the second chapter of my thesis), the analysis concerning equilibrium models is less straightforward, and the question of why models based on very different assumptions provide equivalently accurate predictions remains unanswered (as also pointed out in Selten and Chmura, 2008). What can be said is that behavioral stationary concepts (IBE and NRA) are never outperformed by QRE, Action-sampling, and Payoff-sampling (i.e., best-response models), although in some cases the two classes of solution concepts are equivalent in predicting data. For this reason, I think that it would be important to include in the set of criteria for model selection the plausibility of the assumptions on which models are based, at least as a tie breaking rule, since the causal relationship between assumptions and model accuracy is, in this context, of particular interest (Burnham and Anderson, 2003). We do not have to forget that best-response models are to be interpreted as "*as if*" models: they do not aim at replicating the mechanisms at the basis of the decision-making process, but merely its effects. In other words, from this point of view, what matters is whether or not models are able to predict data, *as if* agents would act according to them. Obviously, the fact that none of us is able to think rationally (i.e., as prescribed by standard theory of choice) and act accordingly is not

new, and any argument against standard theory based on this objection would be rather poor. The point here is that if we have to choose between two models which perform almost equivalently, why should not we privilege the use of that one that embeds principles about the *real* mechanisms of choice behavior? This approach I suggest would be much more informative, as it would allow us to infer the *real* bases of choice behavior. Of course, the judgment about plausibility of assumptions must be cautiously done because there are no principles that can guide us in this kind of task, and caution is primarily in order in those cases in which we are interested to judge whether certain assumptions are more plausible than others.

The NRA and pNRA models are analytically tractable, straightforwardly generalizable to *n*-person games, and based on assumptions validated by recent research on neural mechanism at the basis of human choice behavior. These features make the NRA and pNRA models particularly appealing.

My analysis confirms the poor predictive power of Nash equilibrium, as reported in many other contributions. Compared to the most accurate model, standard theory provides predictions that are worse of the 106% in the first 50 trials, of the 64% in the last 50 trials, and of the 99% over all periods.

In the long run, reinforcement models provide significantly less accurate predictions than does Nash equilibrium (with the exception of RL in average prediction task). I also find confirmation of the result shown in Chapter 2, according to which regret-based learning models are better than reinforcement-based ones; indeed, NFP, SFP and PB0 always perform significantly better than stEWA, NRL, REL, and RL.

Among models of learning, NFP and SFP are the best predictors: their predictive accuracy is statistically equivalent to that of PB0 and PB1 only in the short run, and predict always significantly better than stEWA and reinforcement models. It is worth noting that out of the eight models of equilibrium I consider, only four (NFP, SFP, PB0, and PB1) perform always significantly better than Nash equilibrium, whereas all equilibrium models give more accurate predictions than does standard theory.

If compared to learning models, stationary concepts are, in general, less complex (statistically, analytically, and computationally). Nonetheless, with the exception of QRE, their predictions of short run behavior are as accurate as those of the best performing learning model, which strongly favors the use of equilibrium models, according to the Occam's razor argument.

5.3 Part Three (Chapter 4)

An important source of advantage for the PB model comes from the nature of the learning tasks that can be modeled. Most human interactive learning happens in contexts where tasks do not repeat themselves identically over time as in the experiments considered here. Generalizing from examples and the learning of conditional behavior (different responses to different inputs) are natural features of human behavior. Standard models of economic learning cannot capture, by design, such features because there is no way they can model dependence of behavior from the perception of different game structures. On the contrary, even simple neural networks, as those investigated here, can easily model generalization and conditional behavior, thus making them a natural tool for describing and predicting learning dynamics in the realistic context of mutating strategic settings.

I designed and ran multigame experiments in order to assess to what extent past experience affects current choice behavior and to test the capability of some of the most popular dynamic models of choice behavior in providing different responses to different strategic situations.

The first important point is that empirical results show that in these simple experimental settings, players are able to recognize the structures of the two games in each sequence and play accordingly to this classification. This is particularly interesting because even though the two original games, from which each sequence was obtained, differ only for the predicted probabilities of play, subjects are able to recognize them, but nonetheless do not use Nash equilibrium to form their strategies (also in the long run).

From a computational point of view, simulation results show that traditional "attraction and stochastic choice rule" learning models are not able to discriminate between the different strategic situations, providing a poor "average" behavior for both situations, and are always outperformed by Nash equilibrium. On the contrary, the PB model is able to replicate subjects' conditional behavior, due to its direct dependence of response on game payoffs and performs better than standard theory of equilibrium. In addition, not only the PB model is able to replicate subjects' ability to recognize different strategic situations and act accordingly, but its structure is also complex enough to avoid spillover effects across games, establishing a qualitative parallelism between predicted and observed choice behavior.

Another important result of my analysis is that non-standard equilibrium models are by far the best predictors of empirical data. I conjecture that this is a consequence of the fact that in my experiments learning spillover effects are negligible. As a confirmation of this, models of equilibrium give good predictions of behavior in Type A games that are played simultaneously with two other different kinds of games in two separate treatments: this would not be possible if choice behavior were conditioned by the simultaneous play of the other games.

If we ask ourselves what do subject learn, then the answer is not unique and straightforward. Indeed, if non-standard equilibrium and PB models provide the best approximations of observed behavior, it is also true that empirical choice frequencies do not converge in all cases to the predictions of one of these concepts, thus not providing unambiguous support for this or that theory.

5.4 Further Research

A further test of the PB and NRA models, and in general of all models of interactive choice behavior proposed until now, on data from experiments on a broader class of strategic interactions (e.g., games with more than two players, with more than two actions available to players, and not necessarily constant sum) is in order. As stated in Selten and Chmura (2008), two-person 2x2 completely mixed games constitute a small class of games where testing models of interactive choice behavior and it would be important to gather data from more general patterns of strategic interaction.

The PB and the NRA models were designed to capture behavior in strategic situations of conflict (constant sum games) in which players' interests are opposed i.e., in which players cannot help their opponents without being damaged. Although this pattern of empirical investigation is necessary if we want to disentangle the effects of adaptation and reciprocation, in many economically interesting situations reciprocating and cooperative behaviors do matter (also recent neuroscientific research supports this claim, see Fliessbach et al., 2007); thus, it would be interesting to generalize these models by, for example, introducing a social preference or inequality aversion component.

In addition, potential properties of the PB model are worthy to be investigated. Above mentioned results show that both humans and neural networks are able to categorize games in a sequence, obtained perturbing the payoffs of two games. This

parallelism does not hold for any other dynamic models that I consider: standard models of economic learning (including the recently proposed NFP and SFP), by design, cannot capture such features of human behavior, because there is no way they can model dependence of behavior from the perception of different game structures. However, it would be also interesting to see under what conditions learning spillovers arise, and test whether neural networks are able to produce dynamics of behavior equivalent to those observed under these conditions. It would also be interesting to test the accuracy of equilibrium models in predicting data from multigame experiments in presence of learning spillovers.

Issue of generalization deserves further investigation also from an experimental point of view. First of all, further research could focus on the effects of some factors on learning spillovers, e.g. different degrees and distributions of payoff perturbations and an increase in the number of different types of games in a sequence. My conjecture is that an increase in the magnitude of payoff perturbations and in the number of different kinds of games used to build a sequence would lead subjects to confuse games toward an "average" play. However, this conjecture itself and the extent to which it is true have not yet been investigated experimentally. Further research could also include the design of multigame experiments in which the sequences of games are built starting from patterns of strategic interaction more general than two-person, 2x2 completely mixed games.

6. BIBLIOGRAPHY

Akaike, Hirotogu. 1973. "Information Theory and an Extension of the Maximum Likelihood Principle." Pages 267-281 *in* B. N. Petrov, and F. Csaki, (eds.) *Second International Symposium on Information Theory*. Akademiai Kiado, Budapest.

Arthur, Brian. 1994. "On Designing Economic Agents that Behave Like Human Agents." *Journal of Evolutionary Economics*, 3 (1), 1-22.

Atkinson, Richard C., and Patrick Suppes. 1958. "An Analysis of Two-Person Game Situations in Terms of Statistical Learning Theory." *Journal of Experimental Psychology*, 55(4), 369-78.

Aumann, Robert J. 1974. "Subjectivity and Correlation in Randomized Strategies." *Journal of Mathematical Economics*, 1 (1), 67–96.

Avrahami, Judith, Werner Güth, and Yaakov Kareev. 2005. "Games of Competition in a Stochastic Environment." *Theory and Decision*, 59 (4), 225-294.

Bar-Hillel, Maya, and Efrat Neter. 1996. "Why Are People Reluctant to Exchange Lottery Tickets?" *Journal of Personality and Social Psychology*, 70 (1), 17-27.

Baum, Eric B., and Frank Wilczek. 1988. "Supervised Learning of Probability Distributions by Neural Networks." In *Neural Information Processing Systems*, ed. D. Z. Anderson, 52-61. New York: American Institute of Pshysics.

Bell, David E. 1982. "Regret in Decision Making Under Uncertainty." *Operations Research*, 30 (5), 961-981.

Binmore, Ken, Joe Swierzbinski, and Chris Proulx. 2001. "Does Minimax Work? An Experimental Study." *Economic Journal*, 111 (473), 445-464.

Bishop, Christopher M. 1995. "Neural Networks for Pattern Recognition." *Oxford University Press*, Oxford, NY.

Blackburn, James M. 1936. "Acquisition of Skill: An Analysis of Learning Curves." *IHRB Report No. 73*.

Brandenburger, Adam. 1992. "Knowledge and Equilibrium in Games." *Journal of Economic Perspectives*, 6 (4), 83-101.

Browne, Michael W. 2000. "Cross-Validation Methods." *Journal of Mathematical Psychology*, 44 (1), 108-132.

Burnham, Kenneth P., and David Anderson. 2003. "Model Selection and Multimodel Inference. A Practical Information-Theoretic Approach." *Springer-Verlag,* New York, NY.

Busemeyer, Jerome R., and Yin-Min Wang. 2000. "Model Comparisons and Model Selections Based on Generalization Criterion Methodology." *Journal of Mathematical Psychology,* 44 (1), 171-189.

Busemeyer, Jerome R., and Julie C. Stout. 2002. "A Contribution of Cognitive Decision Models to Clinical Assessments: Decomposing Performance on the Bechara Gambling Task." *Psychological Assessment*, 14 (3), 253-262.

Camerer, Colin F. 2003. "Behavioral Game Theory: Experiments in Strategic Interaction." *Princeton University Press*, Princeton, NJ.

Camerer, Colin F. 2003. "Psychology and Economics. Strategizing in the brain." *Science,* 300 (5626), 1673-1675.

Camerer, Colin F., and Teck-Hua Ho. 1999. "Experience-Weighted Attraction Learning in Normal Form Games." *Econometrica*, 67 (4), 837-874.

Camille, Nathalie, Giorgio Coricelli, Jerome Sallet, Pascale Pradat-Diehl, Jean-René Duhamel, and Angela Sirigu. 2004. "The Involvement of Orbitofrontal Cortex in the Experience of Regret." *Science*, 304 (5674), 1167-1170.

Carroll, Douglas J., and Myron Wish .1974. "Models and Methods for Three-Way Multidimensional Scaling." in D.H. Krantz, R.C. Atkinson, R.D. Luce & P. Suppes (Eds.), *Contemporary developments in mathematical psychology: Vol. 2 Measurement, psychophysics, and neural information processing*, 283-319, New York, Academic Press.

Cheung, Yin-Wong, and Daniel Friedman. 1997. "Individual Learning in Normal Form Games: Some Laboratory Results." *Games and Economic Behavior*, 19 (1), 46-76.

Cheung, Yin-Wong, and Daniel Friedman. 1998. "Comparison of Learning and Replicator Dynamics Using Experimental Data." *Journal of Economic Behavior and Organization*, 35 (3), 263-280.

Chiappori, Pierre A., Steven D. Levitt, and Timothy Groseclose. 2002. "Testing Mixed Strategy Equilibrium When Players Are Heterogeneous: The Case of Penalty Kicks." *American Economic Review,* 92 (4), 1138-1151.

Connolly, Terry, and Marcel Zeelenberg. 2002. "Regret in Decision Making." *Current Directions in Psychological Science*, 11 (6), 212-216.

Coricelli, Giorgio, Hugo D. Critchley, Mateus Joffily, John P. O'Doherty, Angela Sirigu, and Raymond J. Dolan. 2005. "Regret and its Avoidance: a Neuroimaging Study of Choice Behavior." *Nature Neuroscience,* 8, 1255-1262.

Crawford, Vincent P. 1985. “Learning Behavior and Mixed-Strategy Nash Equilibria.” *Journal of Economic Behavior & Organization*, 6 (1), 69-78.

Daw, Nathaniel D., John P. O'Doherty, Peter Dayan, Ben Seymour, and Raymond J. Dolan. 2006. “Cortical Substrates for Exploratory Decisions in Humans.” *Nature,* 441, 876-879.

Denker, John S. 1986. “Neural Network Refinements and Extensions.” In Neural Networks for Computing , ed. John S. Denker, 121-128. *American Institute of Physics Inc.*, Woodbury, NY.

Devetag, Giovanna. 2005. “Precedent Transfer in Coordination Games: An Experiment.” *Economic Letters*, 89 (2), 227-232.

Egidi, Massimo, and Alessandro Narduzzo. 1997. “The Emergence of Path-Dependent Behaviors in Cooperative Contexts.” *International Journal of Industrial Organization*, 15 (6), 677-709.

Erev, Ido, and Alvin E. Roth. 1998. “Predicting How People Play Games: Reinforcement Learning in Experimental Games with Unique, Mixed-Strategy Equilibria.” *American Economic Review*, 88 (4), 848-81.

Erev, Ido, Yoella Bereby-Meyer, and Alvin E. Roth. 1999. “The Effect of Adding a Constant to All Payoffs: Experimental Investigation, and Implications for Reinforcement Learning Models.” *Journal of Economic Behavior & Organization*, 39 (1), 111-128.

Erev, Ido, and Ernan Haruvy. 2001. “On the Potential Uses and Current Limitations of Data Driven Learning Models.” *Working paper*.

Erev, Ido, Alvin E. Roth, Robert L. Slonim, and Greg Barron. 2002. “Predictive Value and the Usefulness of Game Theoretic Models.” *International Journal of Forecasting,* 18 (3), 359-368.

Erev, Ido, and Ernan Haruvy. 2005. “Generality, Repetition, and the Role of Descriptive Learning Models.” *Journal of Mathematical Psychology*, 49 (5), 357-371.

Erev, Ido, Alvin E. Roth, Robert L. Slonim, and Greg Barron. 2007. “Learning and Equilibrium as Useful Approximations: Accuracy of Prediction on Randomly Selected Constant Sum Games.” *Journal of Economic Theory*, 33 (1), 29-51.

Ert, Eyal, and Ido Erev. 2007. “Replicated Alternatives and the Role of Confusion, Chasing, and Regret in Decisions from Experience.” *Journal of Behavioral Decision Making*, 20 (3), 305-322.

Fehr, Ernst, and Klaus M. Schmidt. 1999. "A Theory of Fairness, Competition, and Cooperation." *The Quarterly Journal of Economics*, 114 (3), 817-868.

Fishburn, Peter C. 1982. "Non Transitive Measurable Utility." *Journal of Mathematical Psychology*, 26 (1), 31-67.

Fishburn, Peter C. 1983. "Transitive Measurable Utility." *Journal of Economic Theory*, 31 (2), 293-317.

Fliessbach, K., B. Weber, P. Trautner, T. Dohmen, U. Sunde, C. E. Elger, and A. Falk. 2007. "Social Comparison Affects Reward-Related Brain Activity in the Human Ventral Striatum." *Science*, 318 (5854), 1305-1308.

Friedman, Daniel. 1983. "Effective Scoring Rule for Probabilistic Forecasts." *Management Science*, 29 (4), 447-454.

Fudenberg, Drew, and David M. Kreps. 1993. "Learning Mixed Equilibria." *Games and Economic Behavior*, 5 (3), 320-367.

Gilboa, Itzhak, and David Schmeidler. 1995. "Case-Based Decision Theory." *Quarterly Journal of Economics*, 110 (3), 605-639.

Harsanyi, John C. 1973. "Games with Randomly Disturbed Payoffs: A New Rationale for Mixed-Strategy Equilibrium Points." *International Journal of Game Theory*, 2 (1), 1-23.

Hart, Sergiu, and Andreu Mas-Collel. 2000. "A Simple Adaptive Procedure Leading to Correlated Equilibrium." *Econometrica,* 68 (5), 1127-1150.

Hart, Sergiu, and Andreu Mas-Collel. 2003. "Regret-Based Continuous-Time Dynamics." *Games and Economic Behavior*, 45 (2), 375-394.

Hart, Sergiu. 2005. "Adaptive Heuristics." *Econometrica*, 73 (5), 1401-1430.

Haruvy, Ernan, and Dale O. Stahl. 2000. "Robust Initial Conditions for Learning Dynamics." *Working paper, University of Texas at Austin*.

Hertz, John A., Anders S. Krogh, and Richard G. Palmer. 1991. "Introduction to the Theory of Neural Computation." *Addison-Wesley Publishing Company*, Redwood City, CA.

Hetts, John J., David S. Boninger, David A. Armor, Faith Gleicher, Ariel Nathanson. 2000. "The Influence of Anticipated Counterfactual Regret on Behavior." *Psychology and Marketing,* 17 (4), 345-368.

Ho, Teck-Hua, Colin F. Camerer, and Juin-Kuan Chong. 2007. "Self-tuning Experience-Weighted Attraction Learning in Games." *Journal of Economic Theory*, 133 (1), 177-198.

Holland, John H., Keith J. Holyoak, Richard E. Nisbett, and Paul R. Thagard. 1986. "Induction: Processes of Inference, Learning, and Discovery." *The MIT Press*, Cambridge, MA.

Hopfield, John J. 1982. "Neural Networks and Physical Systems with Emergent Collective Computational Abilities." *Proceedings of the National Academy of Sciences of the USA*, 79 (8), 2554-2558.

Hopfield, John J. 1987. "Learning Algorithms and Probability Distributions in Feed-Forward and Feed-Back Networks." *Proceedings of the National Academy of Sciences of the USA,* 84 (23), 8429-8433.

Huck, Steffen, Philippe Jehiel, and Tom Rutter. 2007. "Learning Spillover and Analogy-Based Expectations: A Multi-Game Experiment." *Working Paper*.

Hutchinson, J. Wesley, and Gregory R. Lockhead. 1977. "Similarity as Distance: A Structural Principle for Semantic Memory." *Journal of Experimental Psychology: Human Learning and Memory*, 3 (6), 660-678.

Jehiel, Philippe. 2005. "Analogy-Based Expectations Equilibrium." *Journal of Economic Theory*, 123 (2), 81-104.

Kahneman, Daniel, and Amos Tversky. 1979. "Prospect Theory: An Analysis of Decision Under Risk." *Econometrica*, 47 (2), 263-291.

Kahnemann, Daniel, and Amos Tversky. 1982. "Judgment Under Uncertainty: Heuristics and Biases." *Cambridge University Press*, Cambridge, NY.

Kahneman, Daniel, and Dale T. Miller. 1986. "Norm Theory: Comparing Reality to its Alternatives." *Psychological Review*, 93 (2), 136-153.

Kruskal, Joseph B. 1964. "Nonmetric Multidimensional Scaling: a Numerical Method." *Psychometrika*, 29 (2), 115-129.

Kullback, Solomon. 1959. "Information Theory and Statistics." *John Wiley and Sons*, New York, NY.

Larrick Richard P., and Terry L. Boles. 1995. "Avoiding Regret in Decisions with Feedback: A Negotiation Example." *Organizational Behavior and Human Decision Processes,* 63 (1), 87-97.

Lee, Daeyeol. 2006. "Neuroeconomics: Best to Go with What You Know?" *Nature,* 441, 822-823.

Leland, Jonathan W. 2002. "Similarity, Uncertainty and Time – Tversky (1969) Revisited." (under revision, *Organizational Behavior and Human Decision Processes*).

LiCalzi, Marco. 1992. “Fictitious Play by Cases.” *Games and Economic Behavior*, 11 (1), 64-89.

Loomes, Graham, and Robert Sugden. 1982. “Regret Theory: An Alternative Theory of Rational Choice Under Uncertainty.” *Economic Journal*, 92 (368), 805-824.

Loomes, Graham, and Robert Sugden. 1983. “Regret Theory and Measurable Utility.” *Economics Letters*, 12 (1), 19-21.

Loomes, Graham, and Robert Sugden. 1987a. “Some Implications of a More General Form of Regret Theory.” *Journal of Economic Theory*, 41 (2), 270-287.

Loomes, Graham, and Robert Sugden. 1987b. “Testing for Regret and Disappointment in Choice Under Uncertainty.” *Economic Journal*, 97 (388a), 118-129.

Marchiori, Davide, and Massimo Warglien. 2008. “Predicting Human Behavior by Regret Driven Neural Networks.” *Science*, 319 (5866), 1111-1113.

McCabe, Kevin A., Arijit Mukherji, and David E. Runkle. 2000. “An Experimental Study of Information and Mixed-Strategy Play in Three-Person Matching-Pennies Game.” *Economic Theory*, 15 (2), 421-462.

McClelland, James L., David E. Rumelhart, and the PDP Research Group. 1986. “Parallel Distributed Processing. Explorations in the Microstructure of Cognition. Volume 2: Psychological and Biological Models.” *The MIT Press*, Cambridge, MA.

McCulloch, Warren S., and Walter Pitts. 1943. “A Logical Calculus of the Ideas Immanent in Nervous Activity.” *Bulletin of Mathematical Biophysics*, 5, 115-133.

McFadden, Daniel L. 1976. “Quantal Choice Analysis: A Survey.” *Annals of Economics and Social Measurement*, 5 (4), 363-390.

McKelvey, Richard D., and Thomas R. Palfrey. 1995. “Quantal Response Equilibria for Normal Form Games.” *Games and Economic Behavior*, 10 (1), 6-38.

Mellers, Barbara, Alan Schwartz, Katty Ho, and Ilana Ritov. 1997. “Decision Affect Theory: Emotional Reactions to the Outcomes of Risky Options.” *Psychological Science,* 8 (6), 423-429.

Mellers, Barbara, Alan Schwartz, and Ilana Ritov. 1999. “Emotion-Based Choice.” *Journal of Experimental Psychology: General*, 128 (3), 332-345.

Mervis, Carolyn B., and Eleanor H. Rosch. 1981. “Categorization of Natural Objects.” *Annual Review of Psychology*, 32, 89-115.

Minsky, Marvin L., and Seymour Papert. 1969. “Perceptrons.” The MIT Press, Cambridge, MA.

Mookherjee, Dilip, and Barry Sopher. 1994. “Learning Behavior in an Experimental

Matching Pennies Game." *Games and Economic Behavior*, 7 (1), 62-91.

Mookherjee, Dilip, and Barry Sopher. 1997. "Learning and Decision Costs in Experimental Constant Sum Games." *Games and Economic Behavior*, 19 (1), 97-132.

Mosier, Charles I. 1951. "The Need and Means of Cross Validation. I. Problems and Designs of Cross-Validation." *Educational and Psychological Measurement*, 11, 5-11.

Nash, John F. 1950. "Equilibrium Points in n-Person Games." *Proceedings of the National Academy of Sciences,* 36 (1), 48-49.

Neugebauer, Tibor, and Reinhard Selten. 2006. "Individual Behavior of First-Price Auctions: The Importance of Information Feedback in Computerized Experimental Markets." Games and Economic Behavior, 54 (1), 183-204.

Neuringer, Allen. 1986. "Can People Behave 'Randomly'? The Role of Feedback." *Journal of Experimental Psychology: General*, 115 (1), 62-75.

Nosofsky, Robert M. 1990. "Similarity Scaling and Cognitive Process Models." *Annual Review of Psychology*, 43, 25-53.

Ockenfels, Axel, and Reinhard Selten. 2005. "Impulse Balance Equilibrium and Feedback in First Price Auctions." *Games and Economic Behavior*, 51 (1), 155-170.

Osborne, Martin J., and Ariel Rubinstein. 1994. "A Course in Game Theory." *The Mit Press*, Cambridge, MA.

Osborne, Martin J., and Ariel Rubinstein. 1998. "Games with Procedurally Rational Players." *American Economic Review*, 88 (4), 834–47.

Palacios-Huerta, Ignacio. 2003. "Professionals Play Minimax." *Review of Economic Studies*, 70 (2), 395-415.

Palacios-Huerta, Ignacio, and Oscar Volij. 2006a. "Appendix to Experientia Docet: professionals Play minimax in Laboratory Experiments." *Working paper, Brown University*.

Palacios-Huerta, Ignacio, and Oscar Volij. 2006b. "Field Centipedes." *Working paper, Brown University*.

R Development Core Team. 2009. "R: A Language and Environment for Statistical Computing." R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

Rankin, Frederick W., John B. Van Huyck, and Raymond C. Battalio. 2000. "Strategic Similarity and Emergent Conventions: Evidence from Similar Stag Hunt Games." *Games and Economic Bahavior*, 32 (2), 315-337.

Ripley, Brian D. 1996. "Pattern Recognition and Neural Networks." *Cambridge University Press*, Cambridge, MA.

Ripley, Brian D., and William N. Venables. 2002. "Modern Applied Statistics with S." *Springer-Verlag*, New York, NY.

Ritov, Ilana. 1996. "Probability of Regret: Anticipation of Uncertainty Resolution in Choice." *Organizational Behavior and Human Decision Processes,* 66 (2), 228-236.

Roese, Neal J. 1994. "The Functional Basis of Counterfactual Thinking." *Journal of Personality and Social psychology*, 66 (5), 805-818.

Roese, Neal J. 1997. "Counterfactual thinking." *Psychological Bulletin*, 121 (1), 133-148.

Rosch, Eleanor H. 1973. "Natural Categories." *Cognitive Psychology*, 4 (3), 328-350.

Rosch, Eleanor H. 1975. "Cognitive Representations of Semantic Categories." *Journal of Experimental Psychology: General*, 104 (3), 192-233.

Rosch, Eleanor H., and Carolyn B. Mervis. 1975. "Family Resemblances: Studies in the Internal Structure of Categories". *Cognitive Psychology*, 7 (4), 573-605.

Rosch, Eleanor H., Carolyn B. Mervis, Wayne D. Gray, David M. Johnson, and Penny Boyes-Braem. 1976. "Basic Objects in Natural Categories." *Cognitive Psychology*, 8 (3), 382-439.

Rosenblatt, Frank. 1958. "The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain." *Psychological Review*, 65 (6), 386-408.

Rosenblatt, Frank. 1962. "Principles of Neurodynamics: Perceptrons and the Theory of Brain Mechanisms." *Spartan Books*, Washington, DC.

Rosenthal, Robert W., Jason Shachat, and Mark Walker. 2003. "Hide and Seek in Arizona." *International Journal of Game Theory*, 32 (2), 273-293.

Roth, Alvin E., and Ido Erev. 1995. "Learning in Extensive-Form Games: Experimental Data and Simple Dynamic Models in the Intermediate Term." *Games and Economic Behavior,* 8 (1), 164-212.

Rubinstein, Ariel. 1998. "Modeling Bounded Rationality." *The MIT Press,* Cambridge, MA.

Rumelhart, David E., Geoffrey E. Hinton, and Ronald J. Williams. 1986a. "Learning Representations by Back-Propagating Errors." *Nature*, 323, 533-536.

Rumelhart, David E., Geoffrey E. Hinton, and Ronald J. Williams. 1986a. "Learning Internal Representations by Error Propagation." In *Parallel Distributed Processing. Explorations in the Microstructure of Cognition*, vol. 1, 318-362. *MIT Press*, Cambridge, MA.

Sarin, Rajiv, and Farshid Vahid. 2001. "Predicting How People Play Games: A Simple Dynamic Model of Choice." *Games and Economic Behavior*, 34 (1), 104-122.

Savage, Leonard J. 1951. "The Theory of Statistical Decision." *Journal of the American Statistical Association*, 46 (253), 55-67.

Savage, Leonard J. 1954. "The Foundations of Statistics." *John Wiley and Sons*, New York, NY.

Selten, Reinhard, and Rolf Stöcker. 1986. "End Behavior in Sequences of Finite Prisoner's Dilemma Supergames: A Learning Theory Approach." *Journal of Economic Behavior and Organization*, 7 (1), 47-70.

Selten, Reinhard. 1998. "Axiomatic Characterization of the Quadratic Scoring Rule." *Experimental Economics,* 1 (1), 43-62.

Selten, Reinhard, and Joachim Buchta. 1999. "Experimental Sealed Bid First Price Auctions with Directly Observed Bid Functions." In: D. Budescu, I. Erev., and R. Zwick (eds.), *Games and Human Behavior: Essays in the Honor of Amnon Rapoport*. Lawrenz Associates Mahwah, NJ.

Selten, Reinhard, Klaus Abbink, and Ricarda Cox. 2005. "Learning Direction Theory and the Winner's Curse." *Experimental Economics*, 8 (1), 5-20.

Selten, Reinhard, and Thorsten Chmura. 2008. "Stationary Concepts for Experimental 2x2-Games." *American Economic Review*, 98 (3), 938-66.

Seta, Catherine E., John J. Seta, Todd G. McElroy, and Jessica Hatz. 2008. "Regret: the Role of Consistency-Fit and Counterfactual Salience." *Social Cognition*, 26 (6), 700-719.

Sgroi, Daniel, and Daniel J. Zizzo. 2002. "Strategy Learning in 3x3 Games by Neural Networks." *Working paper.*

Sgroi, Daniel. 2003. "Using Neural Network to Model Bounded Rationality in Interactive Decision-Making." *Greek Economic Review*, 22, 113-132.

Sgroi, Daniel, and Daniel J. Zizzo. 2007. "Neural Networks and Bounded Rationality." *Physica A,* 375 (2), 717-725.

Sgroi, Daniel, and Daniel J. Zizzo. 2009. "Learning to Play 3x3 Games: Neural Networks as Bounded-Rational Players." *Journal of Economic Behavior and Organization*, 69 (1), 27-38.

Shachat, Jason M. 2002. "Mixed Strategy Play and the Minimax Hypothesis." *Journal of Economic Theory*, 104 (1), 189-226.

Shepard, Roger N. 1958. "Stimulus and Response Generalization: Deduction of the Generalization Gradient From a Trace Model." *Psychological Review*, 65 (4), 242-256.

Shepard, Roger N. 1962. "The Analysis of Proximities: Multidimensional Scaling with an Unknown Distance Function." Psychometrika, 27 (2), 125-140.

Shepard, Roger N. 1974. "Representation of Structure in Similarity Data: Problems and Prospects." Psychometrika, 39 (4), 373-421.

Stahl, Dale O. 1999. "A Horse Race Among Action Reinforcement Learning Models." University of Texas working paper.

Stahl, Dale O., and John B. Van Huyck. 2002. "Learning Conditional Behavior in Similar Stag Hunt Games." *Working Paper*.

Tang, Fang-Fang. 2001. "Anticipatory Learning in Two-Person Games: Some Experimental Results." *Journal of Economic Behavior and Organization*, 44 (2), 221-232.

Thorndike, Edward L. 1898. "Animal Intelligence: An Experimental Study of the Associative Processes in Animals." *Psychological Monographs*, 2(8).

Tobler, Philippe N., Christopher D. Fiorillo, and Wolfram Schultz. 2005. "Adaptive Coding of Reward Value by Dopamine Neurons." *Science*, 307 (5715), 1642-1645.

Tremblay, Leon, and Wolfram Schultz. 1999. "Relative Reward Preference in Primate Orbitofrontal Cortex." *Nature*, 398, 704-708.

Tversky, Amos. 1969. "Intransitivity of Preferences." *Psychological Review*, 76 (1), 31-48.

Tversky, Amos. 1977. "Features of Similarity." *Psychological Review*, 84 (4), 327-352.

Tversky, Amos, and Itamar Gati. 1982. "Similarity, Separability, and the Triangle Inequality." *Psychological Review*, 89 (2), 123-154.

Tversky, Amos, and Daniel Kahneman. (1992). "Advances in Prospect Theory: Cumulative Representation of Uncertainty." *Journal of Risk and Uncertainty*, 5 (4), 297-323.

Von Neumann, John, and Oskar Morgenstern. 1947. “Theory of Games and Economic Behavior.” *Princeton University Press*, Princeton, NJ.

Walker, Mark, and John Wooders. 2001. “Minimax Play at Wimbledon.” *American Economic Review*, 91 (5), 1521-1538.

Walker, Mark and John Wooders. “Mixed Strategy Equilibrium.” The New Palgrave Dictionary of Economics. Second Edition. Eds. Steven N. Durlauf and Lawrence E. Blume. Palgrave Macmillan, 2008. The New Palgrave Dictionary of Economics Online. Palgrave Macmillan. 05 January 2010.

Yechiam, Eldad, and Jerome R. Busemeyer. 2005. “Comparison of Basic Assumptions Embedded in Learning Models for Experience-Based Decision Making.” *Psychonomic Bulletin & Review*, 12 (3), 387-402.

Young, Peyton H. 2004. “Strategic Learning and its Limits.” *Oxford University Press*, Oxford, NY.

Zeelenberg, Marcel, Jane Beattie, Joop van der Pligth, and Nanne K. de Vries. 1996. “Consequences of Regret Aversion: Effects of Expected Feedback on Risky Decision-Making.” *Organizational Behavior and Human Decision Processes,* 65 (2), 148-158.

Zeelenberg, Marcel, and Jane Beattie. 1997. “Consequences of Regret Aversion 2: Effects of Expected Feedback on Risky Decision-Making.” *Organizational Behavior and Human Decision Processes,* 72 (1), 63-78.

Zeelenberg, Marcel, Wilco W. van Dijk, and Antony S. R. Manstead. 1998. “Reconsidering the Relation between Regret and Responsibility.” *Organizational Behavior and Human Decision Processes*, 74 (3), 254-272.

Zeelenberg, Marcel, Wilco W. van Dijk, Joop van der Pligt, Antony S. R. Manstead, Pepijn van Empelen, and Dimitri Reinderman. 1998. “Emotional Reactions to the Outcomes of Decisions: The Role of Counterfactual Thought in the Experience of Regret and Disappointment.” *Organizational Behavior and Human Decision Processes*, 75 (2), 117-141.

Zeelenberg, Marcel. 1999. “Anticipated Regret, Expected Feedback and Behavioral Decision Making.” *Journal of Behavioral Decision Making*, 12 (2), 93-106.