

**DO PLAYERS LEARN HOW TO LEARN?  
EVIDENCE FROM CONSTANT SUM  
GAMES WITH VARYING NUMBER OF  
ACTIONS**

A Master's Thesis

by  
SARAÇGİL, İHSAN ERMAN

Department of  
Economics  
Bilkent University  
Ankara  
June 2009

**DO PLAYERS LEARN HOW TO LEARN?  
EVIDENCE FROM CONSTANT SUM  
GAMES WITH VARYING NUMBER OF  
ACTIONS**

The Institute of Economics and Social Sciences  
of  
Bilkent University

by

SARAÇGİL, İHSAN ERMAN

In Partial Fulfillment of the Requirements For the Degree  
of  
MASTER OF ARTS

in

THE DEPARTMENT OF  
ECONOMICS  
BILKENT UNIVERSITY  
ANKARA

June 2009

I certify that I have read this thesis and have found that it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Arts in Economics.

---

Assist. Prof. Dr. Kevin Hasker

Supervisor

I certify that I have read this thesis and have found that it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Arts in Economics.

---

Prof. Dr. Semih Koray

Examining Committee Member

I certify that I have read this thesis and have found that it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Arts in Economics.

---

Assist. Prof. Dr. Zafer Akin

Examining Committee Member

Approval of the Institute of Economics and Social Sciences

---

Prof. Dr. Erdal Erel

Director

# ABSTRACT

## DO PLAYERS LEARN HOW TO LEARN? EVIDENCE FROM CONSTANT SUM GAMES WITH VARYING NUMBER OF ACTIONS

SARAÇGİL, İHSAN ERMAN

M.A., Department of Economics

Supervisor: Assist. Prof. Dr. Kevin Hasker

June 2009

This thesis investigates the learning behaviour of individuals in strategic environments that have different complexity levels. A new experiment is conducted in which ascending or descending series of constant sum games are played by subjects and the experimental data including both stated beliefs and actual plays are used to estimate which learning model explains the subjects' behaviour best within and across these games. Taking learning rules that model the opponent as a learning agent and heterogeneity of the population into consideration, the estimation results support that people switch learning rules across games and use different models in different games. This game-dependency is confirmed by both action, beliefs and the joint estimations. Although their likelihoods vary from game to game, best response to uniform beliefs and reinforcement learning are the most commonly used learning rules in the four games considered in the experiment, while fictitious play and iterations on that are rare instances observed only in estimation by stated beliefs. Despite the change across games, there is no significant link between complexity of the game and the cognitive hierarchy of learning models. Belief statements and best response behaviour also differ across games as we observe

people making smoother guesses in large action games and more dispersed beliefs statements in small action games. Inconsistency between actions and stated beliefs is stronger in large action games. The evidence strongly supports that learning and belief formation are both game-dependent.

*Keywords:* Reinforcement learning, fictitious play, iterated best response, elicited beliefs, constant sum games, experimental economics.

## ÖZET

# OYUNCULAR ÖGRENMEYİ ÖGRENİYORLAR MI? DEĞİŞEN SAYIDA HAREKETLİ SABİT TOPLAM OYUNLARINDAN ÖRNEKLER

SARAÇGİL, İHSAN ERMAN

Yüksek Lisans, Ekonomi Bölümü

Tez Yöneticisi: Yard. Doç. Dr. Kevin Hasker

Haziran 2009

Bu tez çalışmamızda insanların farklı karmaşıklık düzeyine sahip stratejik ortamlarda öğrenme davranışı incelenmektedir. Artan ve azalan olmak üzere her biri dört oyundan oluşan iki farklı sabit toplam oyun dizisi kullanılarak yapılan deneyler süresince toplanan tahmin bilgileri ve oyundaki seçimler, hangi öğrenme modeli ya da modellerinin tek tek oyunlarda ve de farklı oyunlar arasında sergilenen davranışı açıklamada kullanılmıştır. Karşısındakini öğrenen bir oyuncu olarak gören modeller ve toplum içinde heterojenliği de göz önüne alınarak yapılan ekonometrik analizler sonucunda insanların öğrenme davranışını oyundan oyuna değiştirdiği, farklı oyunlara farklı şekilde yaklaştığı gözlemlenmiştir. Bu oyuna-bağlılık yalnız seçimler, yalnız tahminler ve hem seçim hem tahminleri kullanılan analizlerce desteklenmiştir. Kullanılma sıklıkları oyundan oyuna ve analiz edilen veriye göre değişmekle beraber, üniform tahminlere en iyi cevap ve Destek-temelli öğrenme her dört oyunda da tutarlı bir şekilde kullanılan öğrenme kuralı olsa da hayâli oyun ve ona en iyi cevap vermek suretiyle oluşturulabilecek modeller yalnızca tahmin verisine dayalı analizde ve tek bir tür oyunda destek bulmuştur. Oyundan oyuna öğrenme davranışındaki değişiklikler ile oyunların karmaşıklık seviyesi

arasında net bir ilişki bulunmamaktadır. Tahmin belirtme ve tahminlere en iyi cevap davranışının da oyuna göre değiştiği görülmüştür. Az aksiyonlu oyunlarda daha dağınık tahminler verilmişken çok aksiyonlu oyunlarda tahminler üniforma yaklaşmaktadır. Seçimler ile tahminler arasındaki tutarsızlık ise tüm oyunlarda görülmekle beraber tahminler seçimleri yalnız az aksiyonlu oyunlarda kısmen açıklayabilmektedir. Sonuç olarak öğrenme ve tahmin üretme davranışı net bir şekilde oyuna bağlıdır.

*Anahtar Kelimeler:* Destek-temelli öğrenme, hayali oyun, sabit toplam oyunlar, deneysel iktisat

# ACKNOWLEDGMENTS

I would like to express my deepest gratitude to;

Assist. Prof. Kevin Hasker and Zafer Akin for excellent supervision they provided during my graduate studies. Without their guidance and support, this thesis would never come to this stage.

TUBITAK for believing and investing in this research by providing a generous grant and for their financial support during my graduate studies in Bilkent.

All my professors who taught me Economics and particularly Örsan Öрге who, on the contrary, encouraged me to always remain critical to it.

Finally, my parents for patience, care and encouragement and friends for amity.

# TABLE OF CONTENTS

<b>ABSTRACT</b> . . . . .	<b>iii</b>
<b>ÖZET</b> . . . . .	<b>v</b>
<b>TABLE OF CONTENTS</b> . . . . .	<b>viii</b>
<b>LIST OF TABLES</b> . . . . .	<b>x</b>
<b>LIST OF FIGURES</b> . . . . .	<b>xi</b>
<b>CHAPTER 1: INTRODUCTION</b> . . . . .	<b>1</b>
<b>CHAPTER 2: LITERATURE REVIEW</b> . . . . .	<b>6</b>
<b>CHAPTER 3: EXPERIMENTAL DESIGN AND PROCES-</b> <b>DURES</b> . . . . .	<b>12</b>
3.1 The Experiment . . . . .	12
3.2 Games . . . . .	14
3.3 Subjects . . . . .	17
<b>CHAPTER 4: DESCRIPTIVE ANALYSIS OF ACTIONS</b> <b>AND PREDICTIONS</b> . . . . .	<b>20</b>
4.1 Predictions . . . . .	20
4.2 Best Response to Predictions . . . . .	26
<b>CHAPTER 5: LEARNING MODELS AND ESTIMATION</b> .	<b>34</b>
5.1 Learning Models . . . . .	34

5.2	Econometric Methods . . . . .	40
5.3	Estimation Results . . . . .	43
	<b>CHAPTER 6: CONCLUSION . . . . .</b>	<b>49</b>
	<b>BIBLIOGRAPHY . . . . .</b>	<b>51</b>
	<b>APPENDICES . . . . .</b>	<b>55</b>
	APPENDIX A . . . . .	56
	APPENDIX B . . . . .	60

# LIST OF TABLES

3.1	Games used in the experiments . . . . .	16
3.2	Subject Pool . . . . .	17
3.3	Heterogeneity in the Subject Pool . . . . .	18
3.4	Analysis of Subject Pool . . . . .	19
4.1	MSDU and MSDB Scores . . . . .	21
4.2	APS and MSDO Scores . . . . .	24
4.3	Best Response Rates . . . . .	28
4.4	Approximate Best Response Rates . . . . .	32
5.1	Cognitive Hierarchy of Learning Rules . . . . .	40
5.2	Estimation by Actions . . . . .	45
5.3	Estimation by Stated Beliefs . . . . .	46
5.4	Joint Estimation . . . . .	47

## LIST OF FIGURES

6.1	Distribution by Faculty . . . . .	56
6.2	MSDU and MSDB Scores . . . . .	56
6.3	Subjects in the First Quantile . . . . .	57
6.4	Subjects in Highest Quantile . . . . .	57
6.5	BR% over all games per subject . . . . .	57
6.6	Best Response Rates per Game . . . . .	58
6.7	Percentile Values of BR% . . . . .	58
6.8	Distribution of Population in Quantiles . . . . .	58
6.9	ABR% over games . . . . .	59

# CHAPTER 1

## INTRODUCTION

In an experiment how do subjects learn which actions to take? Do they use a simple reinforcement learning model (if an action does well, use it again), a more complicated model of belief formation like fictitious play, or some variation or combination of these models? Erev and Roth (1988) find that in a wide range of experiments the predominant model was a variation of reinforcement learning—but show in some instances that other models do better. Feltovich (2000) and Nyarko and Schotter (2002) find evidence for belief based models like fictitious play, Camerer and Ho (1999) develop a general model that is a convex combination of reinforcement learning and fictitious play. A motivation for this literature is to find a fundamental model of learning that subjects use in a wide variety of environments. But the question we ask in this thesis is does such a universal rule exist?

Consider two simple games that have been extensively analyzed in this literature, coordination games and constant sum games. The former generally has several pure strategy equilibria, the latter generally has a unique equilibrium in mixed strategies. We would like the analyst to ask himself how he would learn in these two environments. Most probably in the coordination game an experienced analyst would experiment over the Pareto efficient and fair equilibria, possibly sticking to one particular strategy in the

hope that whomever they are playing with will eventually best respond. In the latter environment learning is more complicated, and indeed this is the class of games most frequently analyzed in this literature. In this case one might try to develop a model of your opponent's play and then best respond to that model. As the literature has shown which particular model players are using may vary with game, but if this strategy is successful then one is assured of a high payoff. From this simple reflection one will realize there probably is not one, unique, optimal learning strategy, thus perhaps this research agenda can at best find an optimal learning rule for a given class of games. But considering the latter problem in more depth one should realize that it may not even be possible to find the optimal learning rule in constant sum games. Simply stated if one decides to use a particular learning rule and your opponent learns this then they should best respond to that model, thus your model is no longer correct.

In this thesis we report our findings from a simple variation of the standard constant sum game experiment designed to see whether there is or is not a universal learning rule used in constant sum games. The critical difference between our experiments and previous experiments is that we have the same subjects playing a series of games, games which are very similar except that they have a different number of actions. The null hypothesis is that subjects will use the same learning model in all games. We conclusively reject this null, instead we find that the learning model depends on the game—even though the games are similar. We also follow Nyarko and Schotter (2002) by eliciting players' beliefs about what strategy their opponent will use and matching players against a fixed opponent to enable structural learning. A novelty of our analysis is that we use this information directly in our estimations, we find a simple and easily generalizable method to estimate the likelihood of the guesses given the assessments that are appropriate for a given model. Costa-Gomes and Weiszäcker (2008) also use stated beliefs in estimation, but

they do not use this to estimate the probability of models and their method is not easily generalizable. Another difference between our experiments and a majority of the literature is that we do not analyze 2x2 games. In the words of Nyarko and Schotter (2002), most of the models almost become strategically equivalent in 2x2 games. Salmon (2001) also shows that when he simulates experiments using many games in the literature that if there are only two actions then most models are accepted. In response to these arguments we look at 3x3, 4x4, 5x5, and 6x6 action games, with subjects playing these games either in an ascending or descending sequence. Salmon (2001) also shows that empirically it is hard to select a model using the number of subjects used most experiments, in response to this we have collected one of the largest data sets in the literature. We analyze 456 subjects in our final estimation. As well Hopkins (2002) has shown that reinforcement learning and fictitious play are behaviourally equivalent in the long run, thus we have our subjects play each game for only twenty five periods. Since we hypothesize that players are using different learning rules in different games (and indeed based on whom they are playing with) our analysis is done assuming a heterogeneous population like Costa-Gomes et al. (2001). We also allow for players to best respond to the simpler models in our estimation, since as we argued above this is the likely methodology for sophisticated learners.

Our baseline models in estimation are reinforcement learning (RL), fictitious play (FP), and uniform randomization (Rnd). We allow consider two simple variations on reinforcement learning and fictitious play that ignore the entire history except for last period. With fictitious play this is the simple Cournot Best Response model (Co), with reinforcement learning the model is less straightforward. We have chosen a satisficing model (Sf). Players have a goal for their payoff, and if the given action meets that goal then it is played with higher probability in the next period, with the reference probability for playing an action given by the priors. We consider best responses to all of

these models (the player believes his opponent is using this model) and for reinforcement learning, satisficing, and uniform randomization we consider the best response to this best response (the player believes his opponent believes he is using this model.) In the terminology of the literature on cognitive hierarchy (Camerer et al. 2004) this means that the highest level model we are estimating has a level of two. Costa-Gomes et al. (2001) and Camerer et al. (2004) find little empirical evidence for players using more complex models and we found similar results in preliminary estimates, thus to limit the space of models under consideration we have imposed this ceiling on complexity.

Our baseline finding is that people use different models in different games. We estimate results based on assessments, actions, and jointly. While the results depend on the estimation methodology we always find that the proportion of people using a given model changes significantly. In preliminary estimates using both assessments and actions most people use standard reinforcement learning in 4 and 5 action games, but the percentage using this model in 3 and 6 action games is much lower, with either uniform randomization or the best response to that model (the "naive" model) being the most common model in those games. When our estimates are based solely on action in small action games the naive model or reinforcement learning are selected, in larger games the best response to reinforcement learning and a satisficing model joins these two. Looking at stated beliefs and consistency of actions with stated beliefs, we find that stated beliefs perform poorly since they are statistically inaccurate in prediction per period and per game play of opponent. In contrast to Nyarko and Schotter (2002) but like Costa-Gomes and Weiszäcker (2008) we find that people only rarely best respond to their stated beliefs, yet in our experiment the percentage who were best responding is dramatically below theirs even in three action games. We finally observe that the dispersion in stating beliefs has almost a monotonic pattern: in three action games the beliefs tend to be more dispersed, while as the num-

ber of actions increases, they tend to be close to believing the opponent will use a uniform randomization. In a nutshell, our evidence supports the idea that both learning and belief formation are game-dependent, hence neither behaviours are generalizable over games - strictly speaking even over games with the same structure.

The rest of the thesis is organized as follows. In Section 2 we review the experimental analysis learning literature in the scope of our experiment. In Section 3, the experimental procedure and the games are described in detail and section 4 gives descriptive analysis of action and stated beliefs data. We discuss the model space, construction of the likelihood for the action, beliefs and joint data and the results in Section 5. Section 6 concludes.

## CHAPTER 2

### LITERATURE REVIEW

Theories of learning and equilibrium mechanisms in repeated game-theoretic environments have been a driving interest in research agenda of experimental economics. The yardstick for understanding learning models and searching for evidence of their use is that learning models' predictions are more accurate explanations of game-play history and they outperform mainstream equilibrium concepts in terms of predicting long run outcomes. There are numerous approaches to model learning in games: evolutionary and replicator dynamics, reinforcement learning, adaptive belief-formation models like fictitious play, experience-weighted attractions due to Camerer and Ho (1999), imitation and population matching models, learning direction and sophistication due to Selten and Stocker (1986) and rule learning due to Stahl (1999) and (2000). We will not be interested in theoretical and asymptotic properties of learning models, leaving a large class of evolutionary models whose main interest is to derive limiting behaviours that govern long run equilibrium. The second large class of models we left out of the discussion is population-based social learning models since our experiment focus on individual learning while subjects are matched with the exact same opponent for the entire experiment.

The first important model we will consider is reinforcement learning. Originated in psychology literature, reinforcement learning is based on two princi-

ples due to Erev and Roth (1998): the law of effect and the power of practice. The former means actions that proved its success before are more likely to be taken in the future and the latter implies that with more experience, behaviour becomes smoother and converges. Hopkins (2002) note that this is the type of learning model which processes the minimal level of information. The decisions generated by this model only bears the player's own history of play and do not form any beliefs about the opponent. Erev and Roth (1998) stands as the major paper who supports this learning model in a large class of games in the experimental literature, despite the fact that the majority of these games are 2x2 games with unique mixed strategy equilibrium, in terms of its descriptive and predictive power over game theoretic solution concepts and belief-based models. Camerer (2003) gives numerous other papers who estimated or applied reinforcement learning to laboratory or field experiments in his excellent survey. With its several variations like discounted reinforcement, reinforcement spillovers where unplayed actions are reinforced by some amount, and averaged reinforcement learning where people normalize the reinforcement to what they would obtain at most or at least, this model is among the most popular learning models in the experimental literature.

A second large class of learning models with experimental support is fictitious play and its immediate variants like Cournot Best-Reply process where subjects adaptively form beliefs and best respond to them. More precisely, Fudenberg and Levine (1998)'s classical formulation is based on estimating the distribution of opponent's history of play, deducing opponent's likelihood of playing each strategy assuming that his strategies come from the estimated distribution and finally best responding to that probability vector. Cournot Best-Reply, which can be parametrized as a special case of fictitious play where only the immediate past actions of the opponent matters, simply takes the last action of the opponent and best responds to that assuming that it will be repeated. The information processing feature of these two models

differ significantly from reinforcement learning. As Hopkins (2002) argues, in fictitious play subject's play does not depend on his own history but totally depends on those of the opponent, which is the opposite of reinforcement learning. In experimental analysis, Cheung and Friedman (1998) found support for these types of models in various 2x2 games, Feltovich (2000) in two-stage 2x2 games. However as Camerer (2003) points out, in the class of games with unique mixed strategy equilibrium (the category into which our experiment also falls despite the important distinction that we do not look at 2x2 games), fictitious play does not fit exclusively better than reinforcement learning.

Camerer and Ho (1999) offered a general model which encompasses reinforcement learning and weighted fictitious play as special cases and allows to estimate intermediate cases. In experience-weighted attraction (EWA) model, people can choose which information to use as opposed to the other two whose information process is pre-given. They find significant support for intermediate case except for 2x2 games with unique mixed strategy equilibrium - there a parametrization close to reinforcement learning outperforms all other models. Their findings motivate using more general and comprehensive learning models instead of horse-racing strictly defined algorithms which was a mainstream approach in the early literature.

Claiming that action data alone does not suffice to correctly reveal belief-based models in experiments, Nyarko and Schotter (2000) and (2002) elicited beliefs by a proper quadratic scoring rule whose validity is justified later by Palfrey and Wang (2008). In Nyarko and Schotter (2002) it is shown that belief statements are not approximated by fictitious play variants but without underpinning the algorithm behind these statements, best response to them is the best approximation of game play in the 2x2 game they considered, in (2000) which is under revision, they illustrate that in the same experiment best response to stated beliefs outperforms reinforcement learning and EWA

model as well. However, the criticism directed to this approach is two folds: first of all best-response to a mysterious belief statement is not a learning or a decision-making model unless we pin down how those beliefs are formed. Secondly as the later studies (including ours) have shown, their analysis pertains to 2x2 games since there is no large space over which subjects can form beliefs, putting them in a situation where they are asked to justify their actions. Finally, there is a practical concern with elicited beliefs that they are unavailable outside the laboratory.

Despite the fact they refrain from analyzing learning models per se, behavioural models of decision making and belief formation have important shared parts with the learning literature. Costa-Gomes et al. (2001) examine the correlation between information search and decision-making and find significant link between the two. They looked at a large class of games (including dominance solvable games) and found that population is too heterogeneous to be described by a single decision rule. Best response to uniform randomization (naïve player in their terms) and best response to a naïve player are most commonly observed types in the population. Costa-Gomes and Weiszäcker (2008) looked at a sequence of one-shot 3x3 games while eliciting beliefs at the same time and illustrated the existence of the same types in the population. More interestingly, they show that stated beliefs are not accurate and not representative of subjects' true beliefs in general, contradicting Nyarko and Schotter (2002) result that best response to stated beliefs best predicts and calibrates the game-play in a 2x2 game. The common theme between two papers discussed here is that they look at decision-making models where subjects model his opponent. The same idea can be analyzed in a paper on learning. Selten and Stocker (1986) is a prominent paper which introduces learning directions or anticipatory learning where subjects form more sophisticated beliefs by examining his opponent's previous choices along with his previous payoffs and predicting the changes he is going to make. In the

original paper, the model resembles a Cournot-style behaviour where subjects also consider iterations on Cournot Best-Reply as well. Stahl (1999) and (2000) more neatly generalized this idea by introducing rule learning. In rule learning subjects consider their opponents as learning agents and try to accommodate their decision rules for a learning opponent. These papers find support for level-by-level iterations on rules or step-by-step reasoning. Camerer et al (2004) put all these approaches into one comprehensive perspective by introducing cognitive hierarchy of models. At level 0, people do not form beliefs or make computations and play according to some prior they have in mind. One level of iteration is thinking what his opponent can do, this corresponds to one period elimination of a dominated strategy of the opponent or best responding to an assessment of the opponent. Proceeding like this, one can achieve a clear separation of models by their level of computation or iteration of best response. Camerer et al (2004) looked at very large class of previously studied games and estimated the level of sophistication to be between 1 to 2, depending on the competition level, number of opponents and complexity. This paper gives an important layout to classify learning models by their cognitive demand and information processing requirements, providing a unified approach to learning-to-learn.

Finally we will discuss three papers that deal exclusively with econometrics of experimental data and identification of learning models. Cabrales and Garcia-Fontes (2000) is a theoretical exploration of econometric tools for estimating learning models from experimental data. It bases its analysis on Camerer and Ho (1999), which is also the baseline model from which we construct or model space, and give conditions for consistency and identification in maximum likelihood estimations with unobserved heterogeneity. Salmon (2001) offers a very intuitive method to check for robustness of econometric identification of learning models. It simulates action data using each model and then estimates all models based on this known underlying model. The

results are pessimistic in the sense that a correct identification is hardly likely to be achieved with 2x2 games and with small samples. The paper suggests using large action games with larger sample size, which we use as guidelines in designing our experiment. Wilcox (2006) shows that ignoring heterogeneity in the population has a cost, such pooled-estimations favours reinforcement learning over belief-based models. As the main source of bias appears to be precision parameter, namely  $\lambda$ , the paper strongly suggests using different precision parameters for different models and allow individual heterogeneity, if possible, to reduce this bias against belief-based models. Yet the paper, along with the previous two, concludes that there is no perfect econometric method that has no identification problems.

## CHAPTER 3

# EXPERIMENTAL DESIGN AND PROCEDURES

### 3.1 The Experiment

In the light of our criticisms to the state of art in experimental analysis of learning, we look at learning in constant sum games where we present subjects with a sequence of games of varying complexity - where complexity is characterized by the number of available strategies . Feltovich (2000) justifies the use of constant sum games by arguing that it works best to closely tie subjects' preferences to the payoffs of the game and to limit the influence of subjects' tastes for non-pecuniary aspects of outcome (i.e. fairness etc.). This is important because it allows us saying that any lack of equilibrium play is not due to subjects' unwillingness to play their theoretically optimal strategy but due to subjects' approach in deciding what the optimal strategy is. We use a belief elicitation process as in Nyarko and Schotter (2001) where subjects earn additional profit by correctly estimating what his opponent will do.

We present our subjects at a sequence of constant sum games with unique stable mixed strategy Nash equilibrium. In each sequence there exists four square matrix game of this sort and they differ in number of actions which

are 3, 4, 5 and 6. We did not consider 2x2 games interesting enough to be in the sequence because as Nyarko and Schotter (2001) points out, most of the learning models with 2x2 game of unique equilibrium tend to be strategically-equivalent; i.e. different learning models yield the same strategy most of the time. This was indeed illustrated in Salmon (2001) as he explicitly showed that there exist a severe identification problem in constant sum 2x2 games even if data is generated via simulation by the selected model. Thus we drop 2x2 games and look at larger games where learning models are more strategically-separable, hence identifiable. Our experiments are run for 100 periods and the duration of each game in the sequel is equal, that is 25 periods per game. Subjects are randomly matched with an unknown subject in the same lab but they play against the same opponent for the entire 100 periods. All subjects are told explicitly that their opponent will not change during the experiment. The server randomly assigns row player role to one subject in each randomly matched pair and the other subject automatically becomes the column player - but for the sake of making the experiment easy to understand for the subjects, column players are presented the transpose of the relevant payoff matrix, therefore everyone plays as if they are the row player. Each period subjects are shown three screens: in the first screen, subjects see the payoff matrix of the game (or the transpose depending on their role) and they are asked to report their predictions of the opponents strategies in that period. They are given 30 seconds to produce and confirm their estimates. The subjects are asked to give predictions in percentages, but if the predictions did not sum up to 100 the program normalizes the elicited predictions . They move to the next screen after every subject in the lab confirms their estimates or after 30 seconds (those who do not report predictions receive no payoff.) The payoff of the subject is given by the following scoring rule:

$$\text{Prediction Payoff} = \frac{1}{2} \times (1 + 2g_i - \sum_{j=1}^I g_j^2) \quad (3.1)$$

$g_i$  denotes the estimated probability of the action that is committed by the opponent,  $I$  denotes the total number of available actions in that game. Note that the above formula gives maximum payoff 1 if the player predicts 100% on one action and is correct, and a minimum payoff 0 if he assigns probability 0 to action played by his opponent. The use of proper quadratic scoring rules were justified in detail in Palfrey and Wang (2008). In the second screen they again see the same payoff matrix but this time they are asked to click on the buttons that correspond to their strategy in the game. Once more experiment proceeds when everyone submits a decision. At the third and the last screen, subjects observe the summary of the period which includes useful what his opponent did that period, how much he and his opponent earn from the game, how much he gained from his predictions and monetary value of total payoff. This procedure is repeated for 100 periods each session. Monetary payoffs are calculated according to the formula:

$$\text{Payoff} = 0.1TL \times \text{GamePayoff} + 0.06TL \times \text{PredictionPayoff} + 5TL \quad (3.2)$$

## 3.2 Games

We want to look at  $I$  action constant sum games with a unique mixed strategy Nash equilibrium. To make the games easy to understand  $u_1(i, j) \in \{0, 1, 2, 3, 4\}$  and  $u_2(i, j) = 4 - u_1(i, j)$ . We also want the following characteristics:

1. *Uniqueness of best response to pure strategies*: This is giving players simplicity in figuring out what to do to the first order. Of course when they recognize the other player is using a mixed strategy this simplic-

ity will lessen, but we should be sure of at least first order simplicity. Uniqueness guarantees that the players will always be learning the same equilibrium in the game, it is not possible for some players to be learning equilibrium  $\alpha$  and others learning equilibrium  $\beta$ . We should note that theoretic studies of learning algorithms always analyze behavior “close” to an equilibrium. This restriction guarantees that players are always “close” to the given equilibrium.

2. *iid fair*: This restriction means that if both players choose an action using iid randomization then their average payoff should be the same. If this was not true one party could feel like they are disadvantaged.
3. *Minimal equitable payoffs*: This restriction is meant to reduce the opportunity for players to use an “equitable” strategy. If too many cells in a given row give both players the same payoff then a player could select this strategy because it is equitable. If  $I$  is even this means that no payoffs are 2, if  $I$  is odd we need one payoff of 2.

To make sure they actually have to use a learning algorithm we want:

1. *Full support of equilibrium*: This ensures that players can not discard strategies, effectively making learning less complicated than the number of actions in the game.
2. *Minimize the number of symmetric equilibrium* : We want to minimize this in two ways, first we do not want them to be using the same strategy, and second that strategy should not be equally likely. These criteria can be met if  $I \geq 3$ .

Let  $p_{niI}^*$  be the equilibrium probability for player  $n \in \{1, 2\}$  of action  $i \in \{1, 2, 3, \dots, I\}$  given there are  $I$  actions.

3. *Minimize the number of occurrences of  $p_{niI}^*$* . This restriction is due to  $p_{niI}^*$  being an attractor for our learning rules. If too many actions in

too many games (and possibly for too many players) have the same attractor then players will be more likely to rely on rules of thumb, or classify several different strategies and use the same ad-hoc learning rule for all of them.

We note that by permuting the rows and columns of a matrix we can make a different game (from the subjects point of view) thus we have a total of  $(I!)^2$  games for each game we present below. Such permutations also prevent top row bias or other rules of thumb from biasing our results. We will always present the payoff matrix for the row player (player 1) and payoffs for the column player (2) are  $u_2(i, j) = 4 - u_1(i, j)$ , and  $p$  will be player 1's strategy,  $q$  will be player 2's strategy. We should note that there is only one game which satisfies all of our criteria in  $3 \times 3$  games, in all other cases we selected a game that best fit our extra desiderata. Table 3.1 displays the full set of games used in the experiment.

**Table 3.1: Games used in the experiments**

$3 \times 3$ <table border="1" style="margin: auto;"> <tr><td>4</td><td>0</td><td>2</td></tr> <tr><td>3</td><td>4</td><td>0</td></tr> <tr><td>0</td><td>1</td><td>4</td></tr> </table>	4	0	2	3	4	0	0	1	4	$p^*$ $\frac{1}{5}$ $\frac{2}{5}$ $\frac{2}{5}$	$4 \times 4$ <table border="1" style="margin: auto;"> <tr><td>4</td><td>0</td><td>3</td><td>3</td></tr> <tr><td>1</td><td>4</td><td>0</td><td>1</td></tr> <tr><td>1</td><td>3</td><td>4</td><td>0</td></tr> <tr><td>0</td><td>1</td><td>3</td><td>4</td></tr> </table>	4	0	3	3	1	4	0	1	1	3	4	0	0	1	3	4	$p^*$ $\frac{3}{8}$ $\frac{3}{8}$ $\frac{1}{8}$ $\frac{1}{8}$											
4	0	2																																					
3	4	0																																					
0	1	4																																					
4	0	3	3																																				
1	4	0	1																																				
1	3	4	0																																				
0	1	3	4																																				
$5 \times 5$ <table border="1" style="margin: auto;"> <tr><td>4</td><td>2</td><td>3</td><td>0</td><td>1</td></tr> <tr><td>1</td><td>4</td><td>0</td><td>1</td><td>3</td></tr> <tr><td>0</td><td>1</td><td>4</td><td>3</td><td>3</td></tr> <tr><td>1</td><td>3</td><td>1</td><td>4</td><td>0</td></tr> <tr><td>3</td><td>0</td><td>1</td><td>3</td><td>4</td></tr> </table>	4	2	3	0	1	1	4	0	1	3	0	1	4	3	3	1	3	1	4	0	3	0	1	3	4	$q^*$ $\frac{2}{7}$ $\frac{2}{7}$ $\frac{3}{7}$	$q^*$ $\frac{1}{6}$ $\frac{7}{18}$ $\frac{1}{6}$ $\frac{5}{18}$	$p^*$ $\frac{1}{5}$ $\frac{7}{25}$ $\frac{4}{25}$ $\frac{4}{25}$ $\frac{1}{5}$ $\frac{4}{25}$											
4	2	3	0	1																																			
1	4	0	1	3																																			
0	1	4	3	3																																			
1	3	1	4	0																																			
3	0	1	3	4																																			
$6 \times 6$ <table border="1" style="margin: auto;"> <tr><td>4</td><td>3</td><td>1</td><td>0</td><td>3</td><td>3</td></tr> <tr><td>0</td><td>4</td><td>1</td><td>3</td><td>1</td><td>3</td></tr> <tr><td>1</td><td>1</td><td>4</td><td>3</td><td>0</td><td>1</td></tr> <tr><td>1</td><td>3</td><td>0</td><td>4</td><td>1</td><td>3</td></tr> <tr><td>3</td><td>1</td><td>3</td><td>1</td><td>4</td><td>0</td></tr> <tr><td>3</td><td>0</td><td>3</td><td>1</td><td>1</td><td>4</td></tr> </table>	4	3	1	0	3	3	0	4	1	3	1	3	1	1	4	3	0	1	1	3	0	4	1	3	3	1	3	1	4	0	3	0	3	1	1	4	$q^*$ $\frac{9}{31}$ $\frac{5}{31}$ $\frac{6}{31}$ $\frac{6}{31}$ $\frac{5}{31}$	$q^*$ $\frac{3}{22}$ $\frac{3}{22}$ $\frac{5}{22}$ $\frac{5}{22}$ $\frac{3}{22}$ $\frac{3}{22}$	$p^*$ $\frac{2}{17}$ $\frac{3}{17}$ $\frac{2}{17}$ $\frac{2}{17}$ $\frac{3}{17}$ $\frac{5}{17}$ $\frac{2}{17}$
4	3	1	0	3	3																																		
0	4	1	3	1	3																																		
1	1	4	3	0	1																																		
1	3	0	4	1	3																																		
3	1	3	1	4	0																																		
3	0	3	1	1	4																																		

### 3.3 Subjects

The experiments were conducted in Bilkent University in Ankara, Turkey and the entire subject body consists of students most of which are undergraduates. Excluding preliminary sessions, 456 subjects participated in the experiments that took place in between March and May 2008. To increase heterogeneity of the subject pool, we invited 35% of our subjects from the nearby campus of Middle East Technical University, Ankara. The experiments are announced to the whole student body thus our subjects come from numerous backgrounds most of which did not take any, or if any limited number of, economics and game theory classes. 5 TL was given as a show-up-fee to each subject who arrived the experiment lab on prescribed time. Looking at the population proportions of our subjects in Figure 6.1, 45% of our subjects are from various engineering departments and adding 12% Natural Sciences to this, more than half of our subjects have high level of mathematical and computational exposure. Social Sciences, including economics is the second largest group in the subject pool.

**Table 3.2: Subject Pool**

	Profit	Year	Income
Average	27.73	2.26	2533.21
Max	43.90	5.00	39934.00
Min	15.06	1.00	300.00
StDev	2.15	1.17	3614.12
	METU	Male	Game Theory
% Percent	35.96	71.71	13.60

Table 3.2 presents descriptive analysis of the subject pool. The average payoff for a subject was 28 TL which corresponds approximately to \$20 at that time. The maximum and the minimum are respectively 43 TL and 15 TL while the maximum possible payoff was 50 TL. Note that 43 TL was an obvious outlier; the realized payoffs are clustered between 25-30 TL in general. The majority of our subjects were 2nd or 3rd year students and 72%

of them are male. Only 13,5% reported that they have taken a class in game theory, and the average income is comparable to a general Bilkent student body

**Table 3.3: Heterogeneity in the Subject Pool**

	Mean Earnings per Session		Variation per Session	
	Constant	%METU	Constant	%METU
Coef.	27.71 (70.41)	0.23 (0.31)	0.89 (1.21)	2.41 (1.73)

To test for differences in payoffs generated by population characteristics, we first check whether coming from either one of the two universities make any difference to session earnings or variation of that. The results in Table 3.3 indicate that neither per session mean earnings or variation of that can be explained by the density of students coming from a particular university. Then we regress payoffs on all the variables in the summary table. The results in table 3.4 indicate that none of the characteristics above has a significant effect on the performance of individuals in the experiment. The next estimation displayed at the bottom half of table 3.4 tests for differences generated by the training and it clearly shows that only Fine Arts students, who account for only 2% of the subject pool, perform better than rest of the student body. The remaining backgrounds do not have a significant impact on performance (these backgrounds are EN-engineering, NS-natural sciences, SS-social sciences, ART-fine arts, VOC-vocational training, EDU-education, HUM-humanities, BUS-business, LAW-law). Finally we test whether percentage of METU students in a given session affect the mean earnings or standard deviation of earnings in that session. This is done to be certain that there are no "buddy effects." In these games one could take the attitude that since one is playing with one's friends win or loose the total payoff will be the same, or alternatively act especially cut throat for the potential bragging rights.

**Table 3.4: Analysis of Subject Pool**

	Coef.		Coef.
Const	28.00 (71.46)	Const. (EN)	27.85 (197.04)
Uni	0.05 (0.17)	NS	-0.18 (-0.59)
Year	0.04 (0.37)	SS	-0.51 (-1.90)
Inc	0.00 (-1.48)	ART	2.08 (3.01)
Sex	-0.26 (-0.84)	VOC	-0.02 (-0.04)
GT	0.03 (0.07)	EDU	0.45 (0.77)
		HUM	-0.59 (-0.89)
		BUS	0.01 (0.02)
		LAW	-1.08 (-1.65)

## CHAPTER 4

# DESCRIPTIVE ANALYSIS OF ACTIONS AND PREDICTIONS

### 4.1 Predictions

One of the conclusions in Salmon (2001) and Nyarko and Schotter (2000) concerning the identification problem of belief-based models in experiments was that action data alone is biasing the estimates. It is an intuitive argument since estimating by actions only does not give weight to the beliefs that underscore these actions. Eliciting beliefs by a proper scoring become popular in experimental analysis as a consequence of this observation. While we leave the close relationship between actions and stated beliefs to the other subsection, we will first analyze the predictions against several benchmark models.

All the important papers who elicited beliefs to study learning or behavioural models on game-play looked at either 2x2 or 3x3 games which allow them to graphically illustrate predictions, but with our larger games we can not do this in a comprehensible fashion. Instead, we will analyze the dispersion and statistical accuracy. First we will look at how far the predictions are from the (infinitely risk averse) prediction of uniform probabilities and with respect to (infinitely risk loving) prediction of predicting one pure ac-

**Table 4.1: MSDU and MSDB Scores**

	Total		Descending		Ascending	
	MSDU	MSDB	MSDU	MSDB	MSDU	MSDB
3x3	0.30	0.55	0.33	0.52	0.26	0.52
4x4	0.25	0.60	0.26	0.60	0.23	0.61
5x5	0.22	0.64	0.23	0.62	0.20	0.67
6x6	0.22	0.63	0.23	0.60	0.20	0.68
Overall	0.25	0.60	0.27	0.59	0.22	0.62

tion with probability 1. For the distance from the uniform we will normalize our measure by the number of actions. Such a normalization for number of actions is motivated by the following example: suppose player states a vector (1,0,0) in a 3 action game and MSDU score is 0,667. At a 4x4 game if he states (1,0,0,0) then his score becomes 0,75 although his statement is same in characteristics. Therefore the normalized MSDU score we employ is:

$$MSDU = \left( \sum_{i=1}^I (g_i - \frac{1}{I})^2 \right) * \frac{I}{I-1} \quad (4.1)$$

where  $g_i$  is the stated probability for action i. This gives a score between [0,1] regardless of the game, 0 for a uniform statement and 1 for a pure action prediction. Analogously, one can measure how distant a statement is from a boundary prediction. Note that this time there are I possible vectors to check against hence:

$$MSDB = \min_{j \in I} \left( \sum_{i=1}^I (g_i - \Theta_j)^2 \right) * \frac{I}{I-1} \quad (4.2)$$

where  $\Theta_j$  denotes the vector that gives probability 1 to j'th action.

We report the computed scores per game averaged by individuals in Table 4.1. These results indicate that the predictions do change with the game. If we do not differentiate the order of play, we observe a monotonic move towards uniform statements as the game gets larger and in line with this observation, the statements move away from boundaries in large action games. In other

words, more smooth and risk averse statements are formed in more complex games. Before digging this result further, note that subjects in two treatments predictions with different levels of dispersion. The people playing in the games in a descending order of complexity on average predictions that are slightly yet significantly more distant from the uniform across all the games. However, this does not change the observation that subjects' predictions become close to uniform when they are playing a large number actions, while they form relatively bolder predictions in small action games. A parallel finding was also available in Costa-Gomes and Weiszäcker (2008) where they observe people stating close to uniform predictions when the game only has undominated action hence relatively more complex than games with dominated actions. But compared to Nyarko and Schotter (2002) predictions are much more smooth, that paper observed that the predictions exhibit very frequent jumps over two possible actions. This is confirmed in whiskers plot of MSDU and MSDB in Figure 6.2 as well, belief statements are significantly distant from boundaries even at the 3x3 game.

Given this change in behaviour, it is legitimate to ask whether it is the entire population that exhibits a change. The pie charts in Figures 6.3 and 6.4 show that this is not the case. While the average of MSDU decreases almost smoothly as the games get larger, there are always people stuck at very higher scores (people stating boundary predictions ) and at very low values (people stating uniform predictions ) regardless of the game. Note that the quartiles in this chart are over the entire sample, thus if someone is always in the top quartile it means their MSDU scores are in the top 25% of all MSDU scores for all games. The relative population of these two groups who exhibit constant behaviour in stated beliefs is around 20%. If we compute the relative frequency of subjects that are in the lowest or highest quartile with respect to MSDU in the population, we observe that approximately 10% of the population are in the lowest quartile and 10% in the highest quartile in all of

the four games while these figures double if we allow being in the same lowest or highest quartile in at least 3 of the 4 games in the experiment. It appears that 20% of the population can be seen as marginal subjects since they are either consistently at the lowest quartile (always close to uniform statements) or highest quartile (always close to boundary statements). Likewise 20% of the population never falls into any marginal category in all of the four games. There is almost an uniformly distributed population in terms of number of times they fall into a marginal category.

We now turn our attention to statistical accuracy of stated beliefs in explaining the opponent's play. Here the analysis will depend on two measures commonly used in the literature: average probability score (APS) which is mean squared deviation of stated beliefs from opponent's actual play at each period averaged per game and the mean squared deviation from opponent's frequency of play (MSDO). Formally defined:

$$APS = \frac{1}{25} \sum_{t=1}^{25} \sum_{i=1}^I (g_{it} - \Theta_{opp})^2 \quad (4.3)$$

where  $\Theta_{opp}$  is the action vector of the opponent, 25 is number of times I-action game is played

$$MSDO = \frac{1}{25} \sum_{t=1}^{25} \sum_{i=1}^I (g_{it} - \rho_{opp})^2 \quad (4.4)$$

where  $\rho_{opp}$  is the frequency vector of the opponent in I-action game

One intuitive explanation of what these two popular metrics measure is that APS score is an indicator of average accuracy of stated beliefs in prediction opponent's behaviour at each period of a given game. Stating purely uniform beliefs each period will guarantee a score of  $\frac{I-1}{I}$  so any score above that we consider as inaccuracy on behalf of stated beliefs. MSDO however measures how close the stated beliefs approximate the average frequency of opponent's actions in a given game. In a sense it is an accuracy measure

**Table 4.2: APS and MSDO Scores**

	APS				MSDO			
	3x3	4x4	5x5	6x6	3x3	4x4	5x5	6x6
Total	0.86	0.90	0.95	0.99	0.25	0.23	0.21	0.22
Descending	0.87	0.92	0.97	1.01	0.25	0.24	0.23	0.24
Ascending	0.84	0.88	0.92	0.96	0.25	0.21	0.19	0.20

of the aggregate play in a game. Although look similar in nature, achieving a low APS score is harder than achieving a low MSDO score, which implies that subjects who accurately predict opponent’s action each period at each period of a given game will also have accurately estimate the average frequency of the opponent while the converse is not necessarily true. It is possible, and observed in our analysis and Costa-Gomes and Weiszäcker (2008) that the stated beliefs fail to account for opponent’s play at each period but approximate the average frequency of game-play remarkably good. The table 4.2 summarizes computed APS and MSDO scores per game with ascending/descending treatments and aggregated data.

We observe several interesting patterns in APS and MSDO scores. For APS, accuracy tends to decrease in large games as expected. In more complex games, people have harder time predicting the game-play at each period. The difference in two treatments are significant and noteworthy since appears that people starting the experiment with an easier game form 4% more accurate predictions than people playing in reverse order. Yet if we compare the APS accuracies with uniformly random predictions, the stated beliefs perform poorly since all the scores regardless of the game is significantly higher than  $\frac{I-1}{I}$ . Put differently, stating equal probability to each action at each game will be more accurate description of the opponent than the subject’s proper statements. Moreover, this conclusion does not pertain to population averages only, there is one single individual who scored lower than  $\frac{I-1}{I}$  across all games but that agent was an outlier since his opponent played the same action in almost the entire experiment. Excluding him, there exist no sub-

ject whose stated beliefs scored lower than uniform statement benchmark in more than 2 games out of 4 and those with lower scores at least one game is approximately 2% of the population. Hence stated beliefs' are statistically accurate. The conclusions slightly change when we consider MSDO scores. Here we observe that people actually perform considerably well in predicting the average frequency of opponent's play in a given game. For instance looking at 3x3 games only Costa-Gomes and Weiszäcker (2008) found this rate around 0.20-0.25 slightly lower than ours. Moreover, the accuracy in MSDO sense is even increasing as the games get larger. However, if we use uniformly random statements as a benchmark for accuracy as we did in APS case, predictions are comparatively less accurate than predictions with equal probability at each action. Uniform statements would generate a MSDO score of 0.06 at each game we have, which is significantly lower than stated beliefs scores. Interestingly, uniform statements would be lower for descending order participants while the stated beliefs of ascending order participants are more accurate than descending order participants.

To summarize the descriptive analysis of predictions in terms of dispersion and accuracy, subjects formed beliefs that cluster around uniform statements as the games get larger and more dispersed in small action games. Although there significant differences across subjects in two treatments, the general trends are not affected by the order of play. As the statements become smoother in complex games, their per period accuracy decreases while per game accuracy increases. Nonetheless, this observation do not change the fact that accuracy of stated beliefs are low with respect to both per period and per game metrics compared to putting equal probability to each action.

## 4.2 Best Response to Predictions

We begin our analysis with a broad look at the actions data and check whether actions are best response to the predictions. We would like to mention that our data fails the standard test of whether actions are generated by a uniform distribution. The reason for this is that we have full support equilibria in the neighborhood of  $\frac{1}{7}$ , thus it is hard to differentiate between playing equilibrium every period and the uniform distribution with an exact  $\chi^2$  test.

Assuming that actions are not random, we will look at the consistency between stated beliefs and actions in the spirit of Nyarko and Schotter (2002). The figure 6.5 gives the best response percentages calculated per individual for 100 periods aggregated over four different games (i.e. throughout the experiment). The figure is not meant to account for differences across best response behaviour in different games, but it conveys an important message regarding how much the subjects stick to their own belief statements. It is evident that the best response to stated beliefs for an average individual is between 20-30% which is strikingly low compared to Nyarko and Schotter (2002) which estimates this rate as high as 70% in 2x2 game they had in the experiment. Compared to Costa-Gomes and Weiszäcker (2008) which looks at 3x3 games only the same rate in around 50% on average. In contrast to Nyarko and Schotter (2000) and (2002) we find that predictions are poor at explaining the actions of players in the game. This is despite the fact that the ratio of game payoffs to payoffs from predictions is essentially the same. To be precise when aggregated over games per individual, the mean rate of best response to stated beliefs is 26% while the maximum and the minimum rates were 48% and 15% respectively. This means that the person who stuck to his belief statements most did it less than half of the experiment and on average subjects best responded to their predictions in one-fourth of the entire experiment. Note moreover that, if subjects were purely randomizing over actions available in each game, the probability that their actions correspond to best

response to stated beliefs equal 23% (i.e.  $\frac{1}{4} \sum_{I=3}^6 \frac{1}{I}$ ) although the difference is statistically significant, it is very small. Looking at game-by-game results in Figure 6.6, we observe that best response percentages are slightly higher than pure chance at each game and given the large standard deviation, we cannot reject the hypothesis that for an average subject, level of best response to stated beliefs is pure chance. This is by itself an essential result given the recent studies that favor using elicited belief proxies to explain game-play. Like Costa-Gomes and Weiszäcker (2008) observed a decrease in best response to stated beliefs as we move to 3x3 games (although not as dramatic as ours but this may be due to structure of the experiment; in their study 14 different 3x3 games with varying equilibrium predictions are played sequentially while in our case the same 3x3 game is played for 25 consecutively), our finding indicates the high success of stated beliefs in explaining the game-play pertains to simple 2x2 games.

While on average we cannot reject that best-response to predictions might have occurred by chance, the distribution of subjects with regards to their best response behaviour is of interest as it is the main problematic of this paper, we now look at best response to stated beliefs in a given game while we separate ascending and descending treatments. The table 4.3 summarizes the main characteristics of two treatments. Although the best response to predictions rate differ between two treatments significantly, the most noteworthy observation derived from these differences is that consistency of actions and elicited beliefs increase relatively as the experiment proceeds. Descending treatment participants started with large action games, and their best response rates are lower than those of ascending order players to played the same games at later stages of the game. Likewise descending order participants played more consistently with their predictions in small action games that they played at later stages. Nonetheless, given such high standard deviations, we fail to reject that these best responses may occur by chance.

**Table 4.3: Best Response Rates**

	3x3	4x4	5x5	6x6
Mean	33.43%	27.71%	24.29%	18.91%
Descending	35.51%	27.92%	23.30%	18.97%
Ascending	30.55%	27.34%	25.60%	19.24%
Random	33.33%	25.00%	20.00%	16.67%
Max Desc	80.00%	76.00%	62.50%	68.00%
Max Asc	68.00%	68.00%	84.00%	50.00%
Min Desc	4.00%	0.00%	0.00%	0.00%
Min Asc	0.00%	0.00%	0.00%	0.00%
StDev Desc	12.52%	13.13%	10.93%	11.10%
StDev Asc	12.47%	13.28%	12.68%	10.18%

Although we reject the consistency of actions and elicited beliefs on averages, we suspect that there are subjects with seemingly high best response percentages. As the percentile cutoff values in Figure 6.7 illustrates, even the fourth quartile best response rates are less than 10% above from the random, staying within one standard deviation distance from it. Looking at the distribution of individuals with respect to these cutoff values in Figure 6.8, we observe that the subjects that lie at the fourth quartile constitute approximately  $\frac{1}{4}$  of the entire population at each game. However, this 25% population at each game does not consist of the same people over different games. For example, there is one single individual who is ranked in the lowest quartile in all of the four games. Likewise there are only 3 subjects over 456 who are in the fourth quartile across four games, and 28 subjects are in the same quartile in at least 3 of the 4 games. This finding support the hypothesis that for the vast majority of the population, the conclusion that best response to stated beliefs is not significantly different from random prevails.

Given this high level of inconsistency between elicited beliefs and actions motivates us think over possible reasons that might have stimulated it. Remember from the previous section that stated beliefs perform poorly to predict opponent's action per period while they are relatively better in predicting the per game frequencies. If subjects have known that their predictions were

poor proxies of opponent's actions, that would make sense to observe that these guesses are not in use but of course, subjects were not aware of this fact during the experiment. However it appears that they are behaving as if they know that what they conjecture about the opponent is not correct yet, they do not seem to do anything significant to correct their belief statements. One plausible argument that envisages such an inconsistency is that due to our presentation of belief elicitation and decision giving in experimental design as two separate processes, subjects might be just reacting this design by separately considering belief-elicitation problem and decision problem but not conjointly as the belief-learning models conjecture. When these two problems are handled separately, owing to the fact that monetary incentives for the game-play are much higher as opposed to incentives for correct prediction, subjects may be behaving in risk averse manner by stating something that guarantee a satisficing amount from predictions and focus their attention to game-play without much pondering about their statements. This risk aversion argument was not supported in Nyarko and Schotter (2002) and Costa-Gomes and Weiszäcker (2008) however it is to a certain extent in our case. In the previous section, we reported that there is an evident tendency to submit predictions around equal probability vector and this tendency increases almost monotonically as the games get larger and consequently, their accuracy decreases even more. The rationale for many subjects in doing this may be securing a payoff around 0.5-0.6 by stating a vector close to equal probabilities and moving on to decision-making stage without considering their statements they have made few seconds ago. After all, they earn something positive (or at least will not lose anything) by submitting something and even if a subject is reinforcement learner who does not pay any attention to opponent's play, there is a positive incentive to submit something. However like we have discussed before, consistent risk averse statements are formed by approximately 20% of the population across at least 3 of the 4 games so

risk aversion argument does not account for the inconsistency observed in the entire population.

The second plausible argument to account for this inconsistency is the noise apparent in decision-making. In a sense we are assuming that predictions are representative for true beliefs yet due to noise, people fail to best reply to them during the experiment. Although Costa-Gomes and Weiszäcker (2008) find that there is no significant change in best response behaviour when the subjects are endowed with calculators, a further justification for the existence of noise in decision-making might be the time constraints in the experiments. People feel forced to submit a decision as soon as possible for the experiment to proceed, therefore do not perform rigorous expected value calculations to pick up the *best* reply but heuristically figure out a *good* reply to their predictions. Hence if we develop a metric to measure the level of noise in picking the best response, it will prove useful to judge whether such noisy processes affect the inconsistency between actions and elicited beliefs. In the standard best response analysis, we were punishing all non-best replies by assigning 0 if they were chosen, therefore the new metric will be based on relaxing this punishment as expected. Formally defined:

$$\textit{Approximate Best Response}(ABR) = \frac{\textit{Actual Reply} - \textit{Worst Reply}}{\textit{Best Reply} - \textit{Worst Reply}} \quad (4.5)$$

ABR gives a best response score of 1 if best response is chosen and 0 to worst possible action, and a score in between depending on how close the actual reply is to the best reply. Computing this scores for each individuals and averaging game-by-game, we obtain the figures in Table 4.4. It is expected that the numbers will be inflated since we assign a positive value to chosen actions as long as they are not the worst response in ABR. However, the trend across games turned upside down as opposed to pure best response analysis:

while we used to observe monotonic decrease in consistency between actions and stated beliefs as the games get larger, in approximate sense we see an monotonic increase in consistency towards large action games. Note moreover that allowing approximate best responses did not much affect the computed score for 3x3 game, however in large actions people seem to stick more to their predictions given the noise in decision-making. Yet this trend is more tricky than it seems, as we have the uniform randomization as a benchmark for pure best response to predictions, the computed ABR scores are labeled *Random* in Table 4.4. Checking with respect to this benchmark, we observe that ABR in small action games are 10% higher than pure randomization and the difference is statistically significant although the standard deviations are again quite high as it was the case in pure best response analysis. The gap between ABR to predictions and uniform randomization closes as we move to large action games and the differences become statistically insignificant - even in 6x6 games random and ABR to stated beliefs perform equally well. Consequently, we fail to reject the hypothesis that people approximately best respond to their stated beliefs in 5 and 6 actions game, although we can in small action games. This new result gives a quantifier of noise in decision-making assuming that the stated-beliefs represent the true beliefs. It appears that subjects are consistent with their predictions in small action games with a noteworthy level of noise but inconsistency is undeniable in large action games even we allow for noisy decision-making process. The figure 6.9 gives graphical comparison of aggregated, descending and ascending play averages game-by-game with uniform randomization over actions. We argue that the treatment effects are insignificant and the conclusions we derived with the aggregated data holds for both treatments. Notice another observation in figure 6.9 that the standard deviation of ABR over games tends to decrease in large action games.

To summarize our descriptive analysis for consistency in between pre-

**Table 4.4: Approximate Best Response Rates**

	3x3	4x4	5x5	6x6	Overall
Total	35.67%	43.68%	46.89%	50.05%	44.07%
Descending	36.28%	43.32%	47.65%	50.35%	44.40%
Ascending	35.01%	44.42%	46.10%	49.88%	43.85%
Random	25.02%	33.41%	41.60%	50.17%	37.55%
Max Desc	67.59%	81.09%	76.50%	73.85%	58.87%
Max Asc	72.25%	77.55%	83.09%	74.93%	61.97%
Min Desc	6.00%	13.37%	23.04%	26.60%	30.91%
Min Asc	5.64%	18.00%	17.00%	23.57%	29.94%
StDev	12.32%	11.81%	10.98%	9.82%	6.43%
StDev Desc	12.22%	12.11%	11.03%	9.88%	6.99%
StDev Asc	12.46%	11.37%	10.88%	9.75%	5.56%

dictions and actions, we observe an undeniably strong level of inconsistency between actions and belief statements across all of the four games we have considered. We cannot reject the hypothesis that the observed levels of consistency are of pure chance hence the evidence strongly suggest that using elicited belief is a poor proxy to predict game-play in all four games which implies that Nyarko and Schotter (2000) and (2002) results are due to simplicity of 2x2 game they used in the experiment. We present two alternative arguments that may underpin this inconsistency, the first hypothesis due to Costa-Gomes and Weiszäcker (2008) which relies upon the separation of belief statement and decision-making tasks due to their presentation in experimental design and inequality of incentives that undermine belief elicitation which incites subjects to submit risk averse statements and not use them in decision-making process. Despite the fact that this hypothesis is not supported by the original founders, there is enough evidence in our case especially in large action games. The second hypothesis assumes that stated beliefs represent true beliefs but the inconsistency is due to noise, we show that this argument is plausible to account for the inconsistency in small action games where people actually are best responding to their stated beliefs but with a loss of precision. However, the noisy decision scenario cannot still explain the inconsistency in

large action games where it is evident that even when approximate best responses are allowed in decision-making.

## CHAPTER 5

### LEARNING MODELS AND ESTIMATION

#### 5.1 Learning Models

In this section we will formulate the space of learning models that we will consider in estimating by actions and elicited belief data that we generate during the experiment. Current experimental analysis of learning literature is concentrated around variations and generalizations of two learning models: reinforcement learning and fictitious play. For the sake of consistency, we will place these two models at the center of our learning models as well. No matter the principles and assumptions behind a learning model in a game-theoretic environment, it can be expressed by two equations; if we consider the propensity vector of actions as the state, each learning model has a decision rule where the agent chooses an action given the propensities of actions as the state of that period. Like a state evolution equation in dynamic programming problems, the propensities need to be updated by some rule which is the second and the last equation of a learning model. Despite the analogy between learning models and optimal control problems, learning models have a comparative advantage since the objective criterion, decision rule, can be written in a convenient common form no matter how the update rule works. Formally defined, each individual will be assumed to have a propensity they

attach to each action which will be denoted  $q \in [0, 4]^I$  where  $I \in \{3, 4, 5, 6\}$  is the number of actions. Given this, they will also have a precision parameter  $\lambda(t)$  and the probability they take action  $i$  is:

$$p_i(q) = \frac{e^{\lambda(t)q_i}}{\sum_{j=1}^I e^{\lambda(t)q_j}} \quad (5.1)$$

$p(q) \in \Delta^I$  can be interpreted as the decision produced out of given propensities and  $\lambda(t)$  which can be interpreted as how much they stick to their propensities when they are giving their decision at period  $t$ . Hence it is also a noise parameter as  $\lambda(t) \rightarrow 0$  implies that propensities are not even considered in giving the decision. At the other extreme,  $\lambda(t) \rightarrow \infty$  implies decisions are perfect best responses to the propensity vector of actions. Despite a sort of weighted average representation of the decision rule which seems as if there is no optimization in decision making, Hopkins (2002) shows that belief-based models with optimization can be written in this form too, therefore equation 5.1 will be the common decision rule we will consider for all learning models whether there is optimization or not in decision-making procedure.

Given the common decision rule, the learning models differ in how  $q$  is determined. We first discuss two classic models, reinforcement learning and fictitious play. Reinforcement learning is first originated in psychology literature and Erev and Roth (1998) popularized the use of this model of leaning where they show among a large set of experiments, simple formulation of reinforcement learning explains the findings better than belief-based models of any sort and even has a higher predictive power. The intuition behind the model is the conjecture that actions that yielded good outcomes in the past are more likely to be played today. One can note that it is similar to evolutionary approach to repeated games thus it is no surprise that given sufficient time, evolutionary replicator dynamics approximate reinforcement learning dynamics. It is less cognitively demanding as opposed to belief-

based models: what is expected from a reinforcement learner is to update the probability of played strategy by the payoff each by playing it, increasing the likelihood of that action in the future if it is reinforced positively. Assume that a given player chose action  $i$  and that his opponent choose action  $j$ . Let  $\pi \in [0, 4]^{I^2}$  be the player's payoff matrix;  $x_j \in \{0, 1\}^I$  be a vector with a one in the  $j$ 'th row and zero else;  $I(i) \in \{0, 1\}^{I^2}$  be a matrix which has ones in the  $i$ 'th row and zeros else so that it is an indicator function of the player's action, and  $\phi \in (0, 1)$  the discount factor. Then given the old propensities ( $q$ ) reinforcement learning is:

$$q' = \phi q + (1 - \phi) I(i) \pi \cdot x_j \quad (5.2)$$

Fictitious Play, on the other hand, is the most commonly used learning model among the whole class of belief-based models. The intuition behind this model is somewhat similar to the reinforcement learning in terms of updating the propensities but it is a major difference since it includes a best-response mechanism. There are two main formulations of fictitious play depending on how to update the frequencies. Fudenberg and Levine's (1998) classical formulation is based on estimating the cumulative frequency of past plays and best responding to the estimated distribution but we prefer the propensity-based formulation in Hopkins (2002) since it shares the same decision rule (5.1) with the reinforcement learning. The update rule can be expressed as:

$$q' = \phi q + (1 - \phi) \pi \cdot x_j. \quad (5.3)$$

The primary difference between these models is that with reinforcement learning you do not update the propensities of actions that were not taken. With fictitious play you conduct the counter-factual exercise of considering what payoff you would have gotten if you had done something else. In a sense, while only taken actions are reinforced in the former, strategies that

are not used are hypothetically reinforced by the payoff it would have generated when the opponent played his  $j^{\text{th}}$  action. As expected, fictitious play converges faster than reinforcement learning since it picks the successful actions more rapidly than the reinforcement learning. However as Hopkins (2002) shows, if reinforcement learning or fictitious play ever converges, they tend to be qualitatively different than Nash equilibrium, yet there are numerous cases where neither of the two learning models converge therefore when experimental results support either of the models being in use, one should not expect Nash equilibrium as a long-run outcome.

Notice there are three primary differences between our formulation of these models and the standard models in the literature. First in those models, the second term in each of these expressions is not multiplied by  $(1 - \phi)$ ; in other words weighted average of past propensities and payoffs is not common in empirical papers while the impact of payoff in updating the propensity is taken for granted. However as is derived from a slightly extended Camerer and Ho (1999) model in the appendix B, standard formulation is equivalent to our weighted expression except that in the common models  $\lambda(t)$  grows at a geometric rate, which is unreasonably fast to assume a priori in an empirical analysis. This, in fact, is the motivation for the second difference—we allow  $\lambda$  to change over time but we do it at a log rate which is generally more empirically reasonable. Finally in general formulation  $q$  is unconstrained, however this means that it is very hard to compare results from different analysis. If  $q$  is on the order of one thousand in one paper and ten in another then for given  $\phi$  the impact of new observations will be much more significant in the latter. In contrast in our model we can tell precisely what a given  $\lambda(t)$  and  $\phi$  means in simple counterfactual analysis, thus allowing us making comparisons over games and ascending versus descending treatments.

We also consider variations on these models. The simplest is that some people are using the traditional uniform randomization, in other words people

just choose an action at random. If  $1_I = \{1\}^I$  is a vector with 1's in every dimension then:

$$q' = \frac{1}{I} 1_I \quad (5.4)$$

We create two other models by looking at extreme cases of Fictitious Play and Reinforcement Learning above, in both models we essentially set  $\phi$  to zero. With Fictitious Play this results in the well known Cournot Best Response model where the subject best responds to the last action of the opponent:

$$q' = \pi \cdot x_j \quad (5.5)$$

And with reinforcement learning the implied model is not so transparent. Essentially we are looking for a “satisficing” model. If an action does well in a given period then it should be used in the next, if it does badly then we want the person to switch to another action. But to which action? It is known as win-stay-lose-switch model in two action games, since we have more than two actions in all of our games, it is not evident to which actions the subject is expected to switch when the previous strategy is not satisficing for him. We solve this by having them refer to a prior, thus if  $q_0$  is the prior,  $c \in [0, 4]$  is the expectations level of the player, and  $\alpha \in (0, 1)$

$$q' = I(i) (\pi \cdot x_j - c 1_I) + \alpha q_0 \quad (5.6)$$

where  $\alpha$  is chosen so that if  $\pi_{ij} = 4$  the most likely action to choose will be  $i$ , if  $\pi_{ij} = 0$  the least likely is  $i$ . In words, people are more likely to switch to an action which is best response to the opponent's last action if they are not satisfied with their previous strategy. We then re-scale these payoffs so that  $q' \in [0, 4]^I$  in line with the previous models.

Thus our five baseline models (in order of computational complexity) are randomization (Rnd), satisficing (Sf), reinforcement learning (RL), Cournot

Best Response (Co), and Fictitious Play (FP). We will now extend the space of models by allowing that people best respond to a model of opponent. In the spirit of Stahl et al (1999) and (2000), people may model the opponent as a learning agent by Reinforcement Learning, Cournot Best Response or Fictitious Play, or by a behavioural type of *randomizer* or *satisficer*. Depending on the model of the opponent, agent can calculate the likelihood of actions his opponent can play and give a best response to that. We consider several iterations of best responses to these five base models. A model is denoted by  $q_m$ , then if  $\pi_i$  is the  $i$ 'th row of the payoff matrix:

$$p_i(BR(q_m)) = \frac{e^{\lambda(t) \pi_i p(q_m)}}{\sum_{j=1}^I e^{\lambda(t) \pi_j p(q_m)}} \quad (5.7)$$

gives the decision rule of the model. Notice that this type of learn-to-learn does not have an update rule of its own but uses the update rule of the underlining model of the opponent. The mechanism works iteratively: if my opponent is using model  $m$  then his propensity vector would be  $q_m$  and by the decision rule 5.1 the actions vector would be  $p(q_m)$ . Now we set  $q' = \pi p(q_m)$  we obtain the analogous formulation of the base models for this type of sophisticated learning model. Note that in principle we could consider an unlimited number of best responses to any of the above models then each step of iterations imply thinking more sophisticated models of the opponent and making deeper computations therefore moving up in the cognitive hierarchy of models due to Camerer et al (2004). The categorization of our space of models in terms their cognitive demand is tabulated in table 5.1. The reasoning behind this categorization relies upon the number of iterations of best response to a belief vector. Neither randomization (Rnd), satisficing (Sf), reinforcement learning (RL) form beliefs about the opponent and gives a best response therefore they are considered as level-0 models. Cournot and Fictitious Play form beliefs so they are automatically begin at level-1. However in practice, we do not consider any models that are above level 2 rationality;

**Table 5.1: Cognitive Hierarchy of Learning Rules**

	Learning Model
Level-0	Rnd, Sf, RL
Level-1	Br(Rnd), Br(Sf), Br(RL), Co, FP
Level-2	Br <sup>2</sup> (Rnd), Br <sup>2</sup> (Sf), Br <sup>2</sup> (RL), Br(Co), Br(FP)
....	....
Level-k	Br <sup>k</sup> (Rnd), Br <sup>k</sup> (Sf), Br <sup>k</sup> (RL), Br <sup>k-1</sup> (Co), Br <sup>k-1</sup> (FP)

this means we consider up to the second iteration of best responses to Rnd, Sf, and RL, and only the best responses to Co and FP. This has two underlying motivations behind. The first one is Camerer et al (2004) which does not support more than level-2 models in a large class of experiments but the second and the most important motivation is that our preliminary estimations did not indicate any significant support for more than two iterations, therefore including more iterations in the estimations do not add more than computational overhead.

## 5.2 Econometric Methods

Given our motivation to include belief statements in estimating the learning model, we will extend the mainstream model selection problem in Feltoich(2000), Cabrales and Garcia-Fontes (2000) and Camerer and Ho (1999), which uses only the actions to determine the model. Costa-Gomes and Weiszäcker (2008) appears to be the sole structural attempt to incorporate belief statements in estimation but in their case, the motivation behind extending model selection by actions is to offer a formal econometric solution to understand the inconsistency between predictions and actions; the same problem that we have observed in previous chapter. And they support the hypothesis that the stated beliefs do not represent the true beliefs that would justify the actions but they do not estimate the model which would generate the predictions in the first place.

We will construct the likelihood-by-actions, likelihood-by-stated beliefs and the joint likelihood of actions and stated beliefs one by one. Let  $x_{nt}$  be player  $n$ 's action vector in period  $t$  which will have a one in the row of the action that they took and a zero elsewhere. Let  $g_{nt}$  be their prediction at how agent  $n$  thinks his opponent will play at period  $t$ . With each model ( $q_{mt}$ ) there is an associated probability that they take each action ( $p(q_{mt})$ ) and assessment of how they should guess that their opponent will play ( $a(q_{mt})$ ). Notice that some of our models do not have formal belief formation (i.e. Random, Satisficing and Reinforcement Learning), however we do ask them to predict the probability their opponents will take each action. We will assume that the probabilities that they will state will be the *assessments* that would justify their planned actions if they are using one of the three mentioned models. It is a legitimate concern whether asking for predictions would bias subjects to use belief-based models but as Nyarko and Schotter (2001) and Costa-Gomes and Weiszäcker (2008) shows, the timing of belief elicitation process (before, after or during the play) or the process as a whole do not have significant effects on the uses of learning models.

For a given model the likelihood of  $x_{nt}$  is

$$L^A(q_{mt}|x_{nt}) = x'_{nt} p(q_{mt}) \quad (5.8)$$

Now with the assessments we do not have a well defined standard methodology. The problem is that the guesses and assessments are from the  $I$ -dimensional simplex. From the first order conditions of our objective function we have:

$$g_{nt} = a(q_{mt}) \quad (5.9)$$

that predictions truly represent the assessments. One can easily transform

these first order conditions so that

$$\forall i \in \{2, \dots, I\}, \ln \left( \frac{g_{nt}(i)}{g_{nt}(1)} \right) = \ln \left( \frac{a_i(q_{mt})}{a_1(q_{mt})} \right) \quad (5.10)$$

where the statements are normalized with respect to the first component of the vectors. Thus we assume that

$$\ln \left( \frac{g_{nt}(i)}{g_{nt}(1)} \right) = \ln \left( \frac{a_i(q_{mt})}{a_1(q_{mt})} \right) + \varepsilon \quad (5.11)$$

where  $\varepsilon \sim N(0, \sigma)$ . Let  $L^B(q_{mt}|g_{nt})$  be the associated probability distribution function. As Costa-Gomes and Weizsäcker (2008) argue, the statement of beliefs and the choice of actions are independent questions as it is motivated by the design of the experiment. Thus while we expect that there should be some correlation because they use the same model in both cases the errors are independent, and hence we can write the joint likelihood of the model as:

$$L(q_{mt}|x_{nt}, g_{nt}) = L^A(q_{mt}|x_{nt}) L^B(q_{mt}|g_{nt}) \quad (5.12)$$

We allow for heterogeneity in the population thus we estimate the joint likelihood of the models for person  $n$  at time  $t$  as:

$$L_{nt} = \sum_{m=1}^{13} \frac{e^{\rho_m}}{\sum_{l=1}^{13} e^{\rho_l}} L(q_{mt}|x_{nt}, g_{nt}) \quad (5.13)$$

where  $\rho_m$  denotes the probability of model  $m$  being in use. We then take the log of  $L_{nt}$  and sum it over time and the population which would give the likelihood we will use in the joint estimation.

$$L = \sum_n^N \sum_t^T \ln L_{nt} \quad (5.14)$$

Likewise the likelihoods for estimation by actions or stated beliefs only can be written as

$$L_{nt}^A = \sum_{m=1}^{13} \frac{e^{\rho_m}}{\sum_{l=1}^{13} e^{\rho_l}} L^A(q_{mt}|x_{nt}) \quad \text{and} \quad L^A = \sum_n^N \sum_t^T \ln L_{nt}^A \quad (5.15)$$

$$L_{nt}^B = \sum_{m=1}^{13} \frac{e^{\rho_m}}{\sum_{l=1}^{13} e^{\rho_l}} L^B(q_{mt}|g_{nt}) \quad \text{and} \quad L^B = \sum_n^N \sum_t^T \ln L_{nt}^B \quad (5.16)$$

Notice that we allow  $\rho_m$  to change with the game by estimating for 25 periods each with the people playing each game thereby allowing different models to be optimal in different games. We will also treat ascending and descending experiments separately by altering  $N$ .

### 5.3 Estimation Results

In this section, we will report estimation results based on actions data, elicited beliefs data and joint data separately. The reported results come from come from a sample of 50 subjects in descending treatment but not from the entire subject pool. The justification for this sampling is twofold. The first and the most prominent reason is to obtain converged results in Bayesian estimation, which is interestingly not common due to computational complexity of the model selection problem described in the previous section. Using such a large subject pool is one of the promises of our experiment but then one needs to report unstable and non-robust results since convergence with such a cumbersome problem is hard to come by within a reasonable time frame. Secondly, the amount of data is of comparable size with the rest of the experimental literature which allows us to compare parameter estimates and probability of models.

Table 5.2 summarizes the estimation results based on actions. Before contemplating on the changes in probability of models, there are basic key observations to make from this table. Estimation by Actions strongly rejects fictitious play as a learning model since it is not observed at all in any of

four games together with best response to fictitious play and its Cournot-type refinements, except for 3x3 game where best response to Cournot has positive yet very small probability. Reinforcement learning with estimated  $\phi$  around 0.97 turns out to be the only model that is constantly being used in all of the four games and the unique learning model in 4x4 games. Except for 4x4 case, the population is heterogeneous in the learning models they use and although reinforcement learning can explain almost half of the game-play over all games, it is not the winning model in the remaining three games. Best response to random, *naïve* play in Costa-Gomes et al (2001), is the second top model overall after reinforcement learning but it is identified as the top models in 3x3 and 6x6 games and almost non-existent in the other two games. Best response to reinforcement learning as a level-1 model is frequently used in large action games, even the top model in 5x5 games is not apparent in small action games. A similar argument can be formed for satisficing model as well. 10 to 20% of the population is using a simply satisficing model in large action games while it completely disappears in small action games. Despite the changes in models across games, there is no clear pattern in these changes related to cognitive hierarchy although we observe a slight increase in cognitive effort in large action games. Notice that moving from 6x6 to 5x5 games, we see an increase in the probability of models which are best responses to the top models in the 6x6 game but this evidence is both minor and do not generalize to other games in the sequence. The precision parameter  $\lambda(t)$  estimates for these result is a constant around 1 and increasing at rate 0,52 for reinforcement learning and 0,78 for best response to random model which implies a increase in best responsiveness with time.

The second set of estimation results that we will analyze are belief statements. Table 5.3 reports the probability that predictions are generated by each model. Here we see a much more clustered picture in terms of models used per game, in other words estimation results imply that one can pick a

**Table 5.2: Estimation by Actions**

	3x3	4x4	5x5	6x6	Overall
Rnd	0.0%	0.0%	1.0%	0.2%	0.3%
Br(Rnd)	<b>43.0%</b>	0.0%	0.7%	<b>39.2%</b>	<b>20.7%</b>
Br <sup>2</sup> Rnd	0.0%	0.0%	<b>10.4%</b>	0.1%	2.6%
Sf	0.0%	0.0%	<b>18.9%</b>	<b>10.0%</b>	7.2%
Br(Sf)	9.3%	0.0%	0.0%	0.0%	2.3%
Br <sup>2</sup> Sf	0.0%	0.0%	7.0%	0.0%	1.8%
Rl	<b>41.2%</b>	<b>100.0%</b>	<b>26.2%</b>	<b>28.9%</b>	<b>49.1%</b>
Br(Rl)	0.0%	0.0%	<b>31.4%</b>	<b>17.2%</b>	<b>12.1%</b>
Br <sup>2</sup> Rl	0.0%	0.0%	4.4%	4.4%	2.2%
Co	0.0%	0.0%	0.0%	0.0%	0.0%
Br(Co)	6.4%	0.0%	0.0%	0.0%	1.6%
Fp	0.0%	0.0%	0.0%	0.0%	0.0%
Br(Fp)	0.0%	0.0%	0.0%	0.0%	0.0%

single model to generate those belief statements despite the model changes from game to game. 3x3 and 5x5 games are characterized by a best response to random model, 4x4 by fictitious play with  $\phi = 0.25$  and Cournot with a slight importance and 6x6 games by reinforcement learning assessments with  $\phi = 0.97$ . The precision parameters for three models are of the form  $\lambda(t) = 1 + x \ln(t)$  where  $x$  is estimated around 3 for best response to random, 0.6 for fictitious play and 8 for reinforcement learning. These parameter values imply that people are convinced very rapidly except for fictitious play whose growth rate is comparatively low. Although again the game-dependency is evident, notice the inconsistencies between two set of results. By comparing Table 5.2 and Table 5.3, there is almost no match between the models that generate the predictions and the model that determines the action; the inconsistency we put through in previous sections. The only instances that there is a match to a certain extent is 3x3 and 6x6 games where half of the population is consistent with best response to random in the former and reinforcement learning in the latter. In the other two games played between 25<sup>th</sup>-75<sup>th</sup> periods, what they state is completely different than what people actually play. In 4x4 games people seem to submit fictitious play assessments which

**Table 5.3: Estimation by Stated Beliefs**

	3x3	4x4	5x5	6x6	Overall
Rnd	0.0%	0.0%	0.0%	0.0%	0.0%
Br(Rnd)	<b>99.8%</b>	0.0%	<b>100%</b>	0.7%	<b>50,1%</b>
Br <sup>2</sup> Rnd	0.0%	0.0%	0.0%	0.0%	0.0%
Sf	0.0%	0.0%	0.0%	0,2%	0.0%
Br(Sf)	0.0%	0.0%	0.0%	0.0%	0.0%
Br <sup>2</sup> Sf	0.0%	0.0%	0.0%	0.0%	0.0%
Rl	0.0%	0.0%	0.0%	<b>98.3%</b>	<b>24.6%</b>
Br(Rl)	0.0%	0.0%	0.0%	0.0%	0.0%
Br <sup>2</sup> Rl	0.0%	0.0%	0.0%	0.0%	0.0%
Co	0.1%	<b>12.5%</b>	0.0%	0.6%	3.3%
Br(Co)	0.0%	0.0%	0.0%	0.0%	0.0%
Fp	0.1%	<b>87.5%</b>	0.0%	0.1%	<b>22.0%</b>
Br(Fp)	0.0%	0.0%	0.0%	0.0%	0.0%

does not explain the game-play depending on estimation by action results. Likewise there is an inconsistency between best response to uniform beliefs, which would state  $1/I$  for each of opponent's actions, and actions given by a large class of models which does not include best response to random. This estimation may be considered as a more revealing version of analysis in Costa-Gomes and Weiszäcker (2008). There the authors tested whether true beliefs that undermine the actions are represented by stated beliefs and by their estimation results they rejected this hypothesis. However, their methodology does not prescribe the models that could have generated those belief statements in the first place. Therefore comparison of table 5.2 and table 5.3 is a way of achieving this goal while showing that true beliefs or assessments that justify the actions are not represented by the stated beliefs or predictions.

Finally we will report the joint estimation results where belief statement and decision-making processes are assumed to be independent tasks but both of them are used jointly govern the behaviour throughout the experiment. In other words, by estimating equation 5.14, we are estimating the probability that people use each model jointly for stating the beliefs and determining the action. Table 5.4 summarizes the findings. We see a large concentration on

**Table 5.4: Joint Estimation**

	3x3	4x4	5x5	6x6	Overall
Rnd	0.1%	0.0%	0.0%	<b>43.2%</b>	<b>10.8%</b>
Br(Rnd)	<b>33.1%</b>	0.0%	0.0%	<b>27.6%</b>	<b>15.2%</b>
Br <sup>2</sup> Rnd	0.0%	0.0%	0.0%	0.0%	0.0%
Sf	1.0%	0.0%	0.0%	<b>12.4%</b>	3.4%
Br(Sf)	0.0%	7.4%	0.0%	0.0%	1.9%
Br <sup>2</sup> Sf	0.0%	0.0%	0.0%	0.0%	0.0%
Rl	<b>56.7%</b>	<b>92.6%</b>	<b>100.0%</b>	4.6%	<b>63.5%</b>
Br(Rl)	0.0%	0.0%	0.0%	<b>11.8%</b>	3.0%
Br <sup>2</sup> Rl	0.0%	0.0%	0.0%	0.0%	0.0%
Co	0.0%	0.0%	0.0%	0.0%	0.0%
Br(Co)	9.1%	0.0%	0.0%	0.0%	2.3%
Fp	0.0%	0.0%	0.0%	0.0%	0.0%
Br(Fp)	0.0%	0.0%	0.0%	0.0%	0.0%

reinforcement learning except for 6x6 game which is mostly random or best response to random. Interestingly, best response to reinforcement learning is observed in 6 action games instead of the reinforcement learning itself. These results imply that if we forcefully think of agents as people who use a single model to predict the opponent and choose the actions jointly, then their behaviour is best explained by reinforcement learning in most of the games although its success depends on the game. Notice also that this is the only estimation that chooses random as a widely used model in a particular game. Comparing these results with the previous two, it is tempting to argue that since we seem to support a discrepancy between actions and predictions, it is hardly likely to probabilities in Table 5.4 have any match between those of estimation by actions and stated beliefs separately. However, we observe that there is ambiguous sign of similarity between estimation by actions and joint estimations. Remember that in section 4 approximate best response analysis indicate a noisy use of stated beliefs in determining the actions in 3 and 4 action games and this is where we see a close relationship between two estimation results. Two methods pick the same models with similar probabilities in 3x3 and 4x4 games where we know that predictions have an

explanatory power of actions to a certain extent but in the light of very same results, the relationship is much vague when we move to large action games. In addition to these remarks in any of the three different estimation method used, we do not find a significant relationship between the cognitive hierarchy of learning models with the complexity of the games presented.

## CHAPTER 6

### CONCLUSION

In this thesis we looked at dynamic environments where subjects play a sequence of similar matrix games with unique stable equilibrium. Our questions were (1) whether there exists universal learning rules that are used across all games, (2) whether learning behaviour adapts complexity and (3) to what extent there is consistency between actions and stated beliefs. The answer to the latter question is given by best response and approximate best response analysis in section 4 that the discrepancy between the two is undeniably strong. That is, as opposed to Nyarko and Schotter (2001), the pioneer of belief elicitation to understand observed actions in learning experiments, we observe that best response to stated beliefs is dramatically low, which confirms and strengthens Costa-Gomes and Weiszäcker (2008) who also observed the same decrease in best responses when they look at slightly more complex games than a 2x2 game. These two findings together support Salmon (2001) critique that 2x2 games are too simple game-forms to understand belief formation and learning behaviour of experimental subjects. In order to describe the inconsistency between actions and stated beliefs, we looked for evidence for three potential scenarios. The first one is risk aversion in stating beliefs and as it is the case for the first two major papers in the literature, we do not find strong support for risk averse guesses although there is a tendency

to smoothen belief statements as the games get larger. The separation of the belief elicitation and action tasks and inequality of incentives is a plausible scenario and the inconsistency between estimation results by actions and assessments seem to confirm this hypothesis although it cannot quantify to what degree the separation of task can account for it. The third and the last, decision noise quantified by approximate best reply process can account for the inconsistency in small action games. That is, people best respond to their stated beliefs approximately in small action games but not in large action games. This finding implies that people actually use their stated beliefs in determining the actions in 3x3 and 4x4 games but do not or can not perform the best reply process perfectly. Another support for this scenario is the paralleling results in estimation by only actions and joint data in small action games. Yet our experimental design does not specifically aim to control the decision noise, the reasons that might cause this imperfection in best replying to stated beliefs remain hypothetical. A tempting claim one can make is that since the design forces subjects to report their guesses first and then make a decision, the mapping from guesses to actions is flawed due to computational disability. Costa-Gomes and Weiszäcker (2008) test whether an increase in computational ability (e.g. when subjects are endowed with a calculator) affect the relationship between actions and stated beliefs and find that it is not the case therefore discredit this claim. Despite the given ordering of tasks, we can hypothesize that people think of their action first and then reported their guesses in a justifying manner. The rationale for such a behaviour might very well be the inequality of monetary incentives between making correct guesses and giving a best reply to the opponent. However, this would then motivate the subject to state a uniform belief, secure a reasonable payoff and focus on game-play instead but we have already discredited such a risk averse behaviour. In conclusion, the underlying motivation for this inconsistency remains an open problem and a matter of further research by integrating two

tasks and incorporate quantifiers for the match between guesses and actions in the experiments.

To the question of whether there exists universally successful learning rules across different games. Estimations with a sample of descending subjects show that the population seems heterogeneous in most of the games. Best response to random and reinforcement learning are two learning models that are chosen with high probabilities across all games but their explanatory power depends on which data used for estimation and which game we are looking at. Fictitious play and variants are rejected by actions and joint data estimations, the only instance it finds support is belief statements in 4x4 games. In that regard, one can answer the first core question by rejecting the existence of a universal learning model at least in the class of games we are looking at in this experiment. Note however that if this hypothesis is rejected in a class of extremely similar games where the only changing factor is the number of actions, then it is hard to assume and hypothesize that there exists such universal learning models in larger class of games in general. This argument also disfavours cross-experimental comparisons as a way of aggregating support for a learning model being in use. Finally, despite the changing learning behaviour across games, the change does not follow a significant pattern with the complexity. The models estimated by three different methods are either level-0 or level-1 in the cognitive hierarchy of learning models due to Camerer et al (2004) and level-2 models are rarely observed but in general, learning models do not seem to adapt the complexity of the games in an intuitive manner. We conclude by indicating that the future research should be extending the estimation results to the entire subject pool including the ascending players and test for differences among two treatments. This extension will increase the credibility of our results since it will be largest data set available in the experimental literature.

## BIBLIOGRAPHY

- Cabrales, Antonio and Garcia-Fontes, Walter. 2000. "Estimating Learning Models with Experimental Data," Working Paper. Barcelona, Spain: Universidad de Pompeu Fabra
- Camerer, Colin and Ho, Teck-Hua. 1999. "Experience-weighted Attraction Learning in Normal Form Games," *Econometrica* 67(4): 827-874
- Camerer, Colin. 2003. *Behavioral Game Theory: Experiments in Strategic Interaction*. Princeton: Princeton University Press
- Camerer, Colin, Ho, Tech-Huan and Chong, Juin-Kuan. 2004. "A Cognitive Hierarchy Model of Games," *Quarterly Journal of Economics* 119(3): 861-898
- Cheung, Yin-Wong and Friedman, Daniel. 1998. "A Comparison of Learning and Replicator Dynamics Using Experimental Data," *Journal of Economic Behavior & Organization* 35(3): 263-280
- Costa-Gomes, Miguel, Crawford, Vincent P. and Broseta, Bruno. 2001. "Cognition and Behavior in Normal-Form Games: An Experimental Study," *Econometrica* 69(5): 1193-1235
- Costa-Gomes, Miguel and Weiszäcker, Georg. 2008. "Stated Beliefs and Play in Normal-Form Games," *Review of Economic Studies* 75:729-762

- Erev, Ido and Roth, Alvin E. 1998. "Predicting How People Play Games: Reinforcement Learning in Experimental Games with Unique, Mixed Strategy Equilibria," *The American Economic Review* 88(4): 848-881
- Feltovich, Nick. 2000. "Reinforcement-Based vs. Belief-Based Learning Models in Experimental Asymmetric-Information Games," *Econometrica* 68(3): 605-641
- Fudenberg, Drew, and Levine, David K. 1998. *The Theory of Learning in Games*. Cambridge: MIT Press
- Hopkins, Ed. 2002. "Two Competing Models of How People Learn in Games," *Econometrica* 70(6): 2141-2166
- Nyarko, Yaw and Schotter, Andrew. 2000. "Comparing Learning Models with Ideal Micro-Experimental Data Sets," Under Revision. New York, NY: New York University, New York
- Nyarko, Yaw and Schotter, Andrew. 2002. "An Experimental Study of Belief Learning Using Elicited Beliefs," *Econometrica* 70(3): 971-1005
- Palfrey, Thomas and Wang, Stephanie. 2008. "On Eliciting Beliefs in Strategic Games". Working Paper. Pasadena, CA: California Institute of Technology, Pasadena
- Salmon, Timothy C. 2001. "An Evaluation of Econometric Models of Adaptive Learning". *Econometrica* 69(6): 1597-1628
- Selten, Reinhard and Stoecker, Rolf. 1986. "End Behavior in Sequences of Finite Prisoner's Dilemma Supergames: A Learning Theory Approach," *Journal of Economic Behavior & Organization* 7(1): 47-70
- Stahl, Dale. 1999. "Evidence Based Rule Learning in Symmetric Normal-Form Games," *International Journal of Game Theory* 28: 111-130

Stahl, Dale. 2000. "Rule Learning in Symmetric Normal-Form Games: Theory and Evidence," *Games and Economic Behavior* 32: 105-138.

Wilcox, Nathaniel T. 2006. "Theories of Learning in Games and Heterogeneity Bias," *Econometrica* 74(5): 1271-1292

# APPENDICES

## APPENDIX A

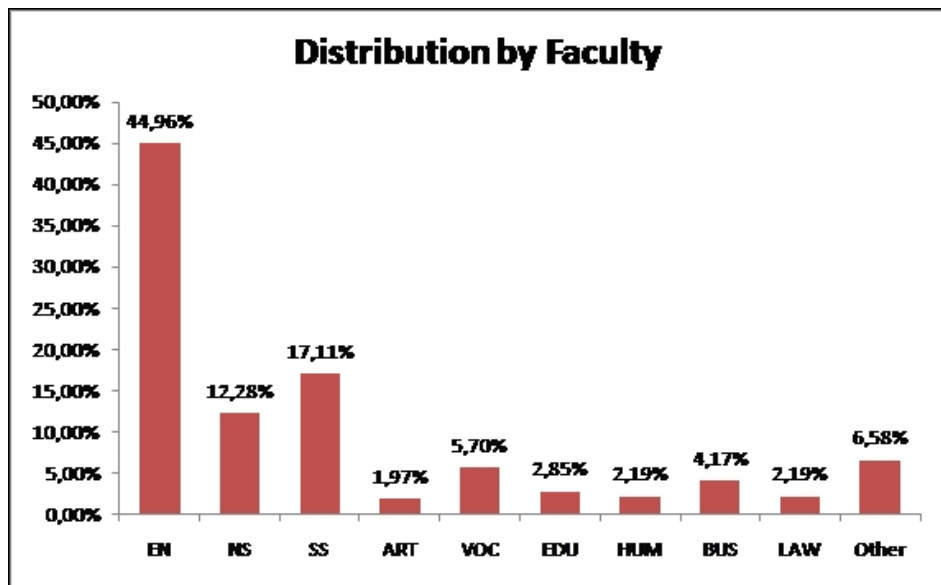


Figure 6.1: Distribution by Faculty

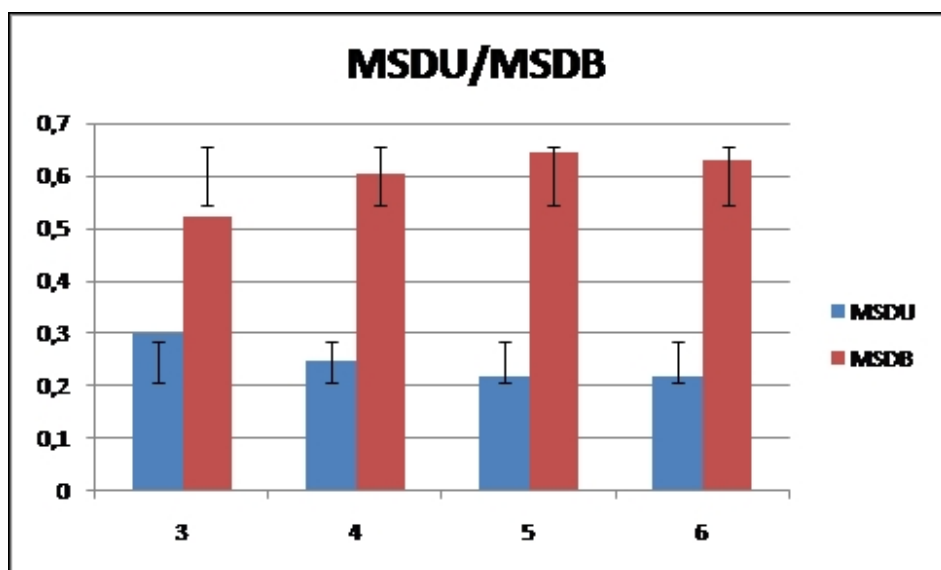


Figure 6.2: MSDU and MSDB Scores

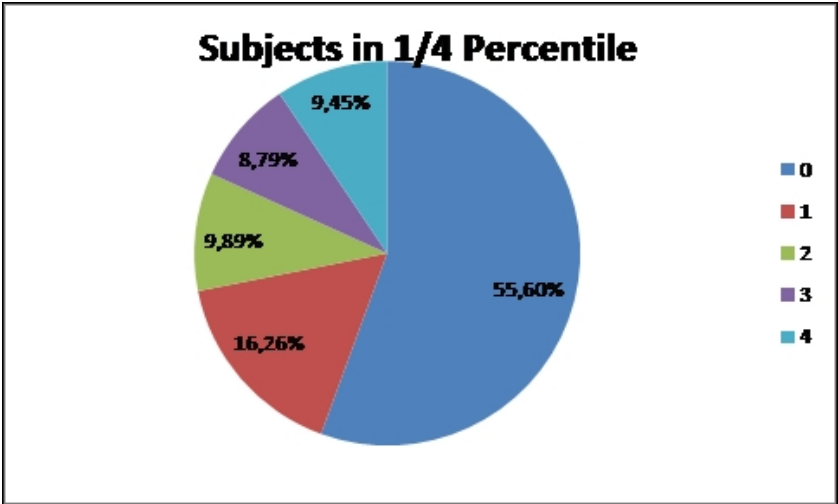


Figure 6.3: Subjects in the First Quantile

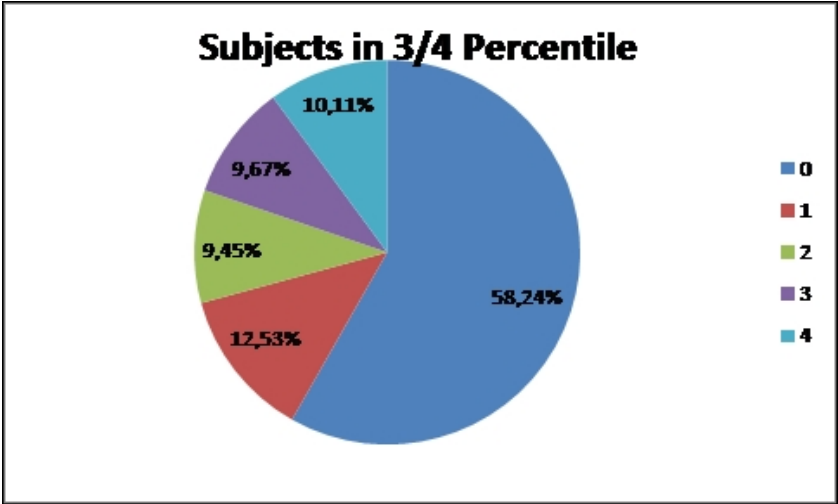


Figure 6.4: Subjects in Highest Quantile

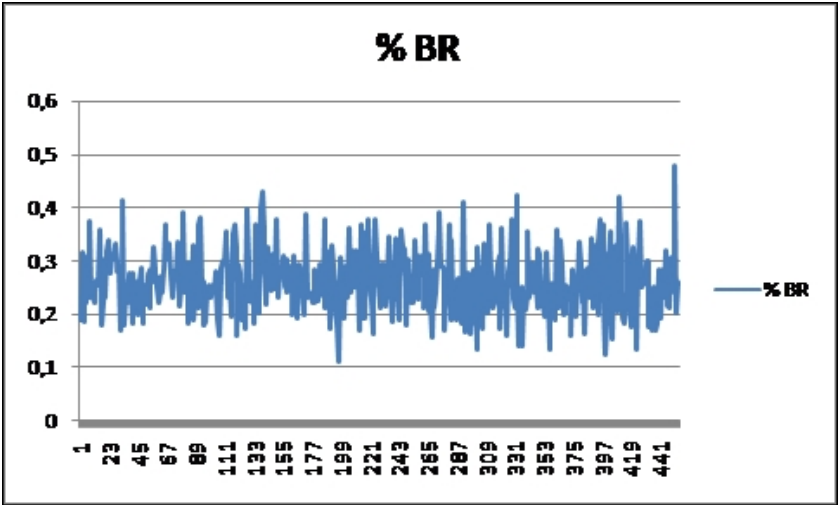


Figure 6.5: BR% over all games per subject

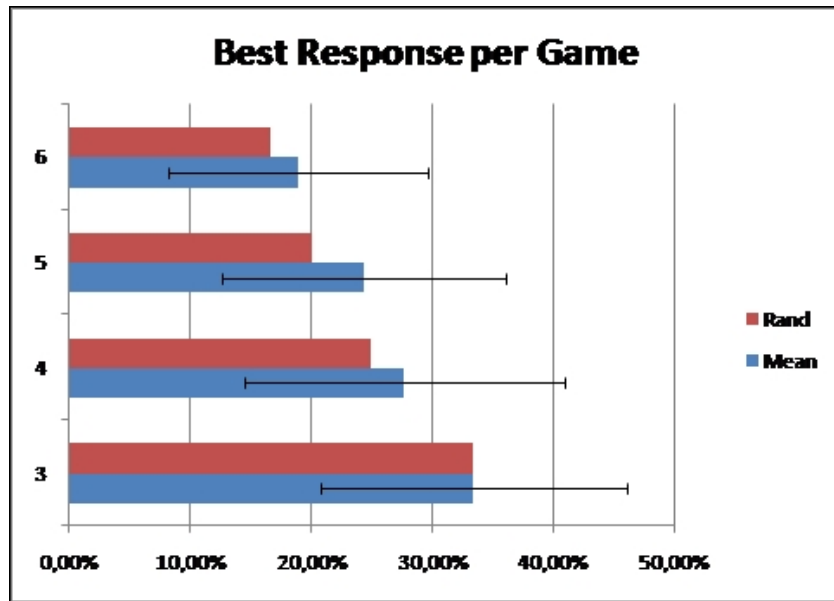


Figure 6.6: Best Response Rates per Game

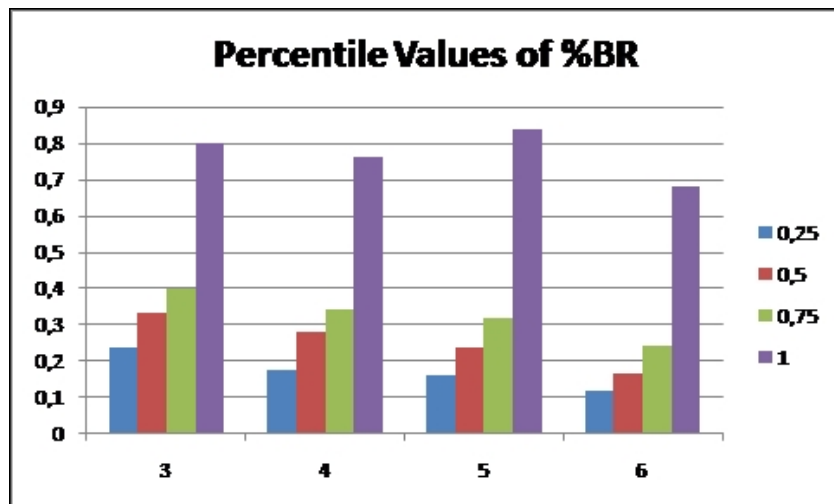


Figure 6.7: Percentile Values of BR%

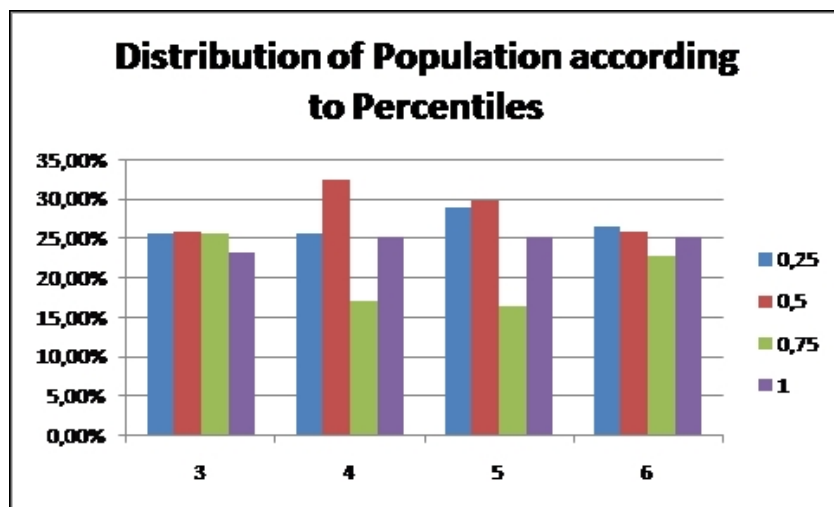


Figure 6.8: Distribution of Population in Quantiles

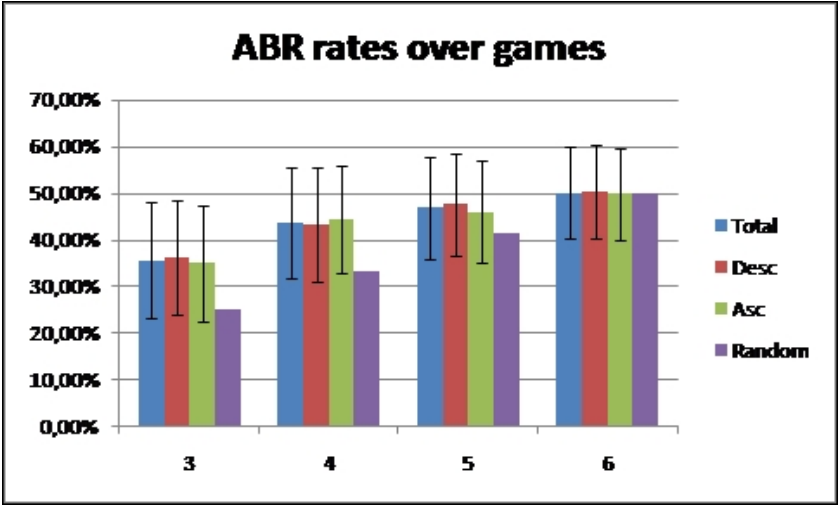


Figure 6.9: ABR% over games

## APPENDIX B

In Camerer and Ho (1999) they define the attractions of the models as:

$$q' = \psi \frac{N(t)}{N(t+1)} q + \frac{1}{N(t+1)} (\delta + (1 - \delta) I_j) \pi_i \quad (6.1)$$

where experience weights evolve according to:

$$N(t+1) = \rho N(t) + 1 \quad (6.2)$$

we first want to normalize the  $q_0$ 's: the initial priors. In general  $\sum_{i=1}^I q_{i0} = \frac{1}{\beta}$ , usually  $\beta$  is not chosen by the econometrician, rather  $q_{10}$  is normalized to zero and this is a free variable. If we normalize the equation:

$$q' = \psi \frac{N(t)}{N(t+1)} q + \frac{\beta}{N(t+1)} (\delta + (1 - \delta) I_j) \pi_i \quad (6.3)$$

where we can interpret  $\beta$  as the "impact factor" of the new observations. We can write this equation as:

$$q' = \frac{1}{\phi(t)} [\phi(t) q + (1 - \phi(t)) (\delta + (1 - \delta) I_j) \pi_i] \quad (6.4)$$

where

$$\phi(t) = \frac{\psi N(t)}{\psi N(t) + \beta} \in (0, 1) \quad (6.5)$$

at this point we impose the normalization that  $N(0) = \frac{1}{1-\rho}$  then

$$\phi(t) = \phi = \frac{\psi}{\psi + \beta(1 - \rho)} \in (0, 1) \quad (6.6)$$

notice that after this normalization we can interpret  $1 - \rho$  as the impact factor, so the initial rescaling was not necessary. However as many analysts assume  $\rho = 0$  we wish to point out that normalizing  $q_0$  has the same effect. Now we note that what matters for decision making is  $\lambda(t+1)q'$ , thus we can rewrite this as:

$$\lambda(t+1)q' = \lambda(t+1)[\phi q + (1 - \phi)(\delta + (1 - \delta)I_j)\pi_i] \quad (6.7)$$

if we assume

$$\lambda(t+1) = \left(\frac{1}{\phi}\right)^t \lambda(0) . \quad (6.8)$$

At this point we note an undesirable parametric restriction in the base model. In the standard model lambda is constrained to grow at a geometric pace, and moreover a geometric pace that is dependent on coefficients that have other interpretations. First note that a geometric pace is a very high rate of growth, second note that the fact that it is parametrically linked to coefficients for which we have other interpretations is not supportable. Thus in this paper we use the model:

$$\lambda(t+1) = \lambda_0 + \lambda_{dt} \ln(t_d) \quad (6.9)$$

where  $t_d$  is the length of time that a player has been playing a particular game. The coefficients are constrained to be positive, thus we are assuming that the certainty can only grow over time. Other parametric specifications could be attempted, but have not here. Indeed the original model could be tested by seeing if it can be accepted as a null hypothesis in this regression. We have not done so because we reject a-priori such a parametric restriction.

In passing we should point out that if  $\lambda(t)$  is increasing and one implicitly imposes equation 6.8 then this will bias the estimates of  $\phi$  (and the standard coefficient  $\psi$ ) towards zero. This could explain why our estimates of  $\phi$  are generally higher than those found in the literature.

Thus in the normalized model,  $\phi$  is the amount past observations are discounted,  $\delta$  is the degree to which counter-factual information is considered, and  $\lambda(t)$  is the rate at which the certainty of the subject is increasing over time. We note that we have not normalized the  $q$ 's to being in the interval  $[0, 1]$  but rather  $[0, 4]$  since this is the space of the payoffs in the stage game. If one wants to conduct this final normalization (of both  $q$ 's and  $\pi$ 's) then one should multiply our  $\lambda$ 's by four.

Due to this normalization and the fact that subjects played a game for 25 periods we can state, for example, that the weight on the priors have at most half weight in the final period if  $\phi \leq 0.97153$ , we have restricted  $\phi$  to be less than .97 to be sure that these models are not essentially putting weight one on the prior in every period—in which case they would be a type of *a-priori* randomization. Furthermore we can construct arguments to find reasonable bounds on  $\lambda_0$ . To bound  $\lambda_0$  let  $q_i$  be the highest attraction and  $q_j$  be the second highest, then the likelihood ratio of these two actions is:

$$\frac{e^{\lambda_0 q_i}}{e^{\lambda_0 q_j}} = LR_{ij} \quad (6.10)$$

thus a bound on  $\lambda_0$  is:

$$\lambda_0 \geq \frac{\ln(LR_{ij})}{q_i - q_j} \quad (6.11)$$

we only need to select one variable instead of the two available, so we let  $q_i - q_j = 1$  and then we need to select a reasonable likelihood ratio if the difference between the best response and the second best response is 25% (after normalizing payoffs). Notice that if the opponent is expected to play a pure strategy this is exactly the difference implied by a best response in a

belief model. On the other hand the likelihood of this high of a difference between the  $q$ 's in most models is relatively low. We arbitrarily choose  $LR_{ij} = e = 2.7183$ , thus this implies that  $\lambda_0 \geq 1$ .

Note that neither of these comparisons would make sense in the standard model. In that model since  $q^0$  and  $\pi$  are not in the same space one can not precisely interpret  $\phi^t q^0$ . As well since these values are changing (and usually increasing)  $q_i^t - q_j^t$  will be changing over time and bounding  $\lambda$  based on the likelihood of these actions is meaningless. On a deeper level notice that this means that if in two models the estimated values of  $\sum_{i=1}^I q_i^0$  are different then the interpretation of  $\psi$  is different and the coefficients can not be compared. As well the estimated value of  $\lambda$  has no clear interpretation since  $q^t$  is growing over time.