THE SIGN-UP GAME, SOPHISTICATED LEARNING

AND LEARNING VARIABLE DEMAND

A Dissertation

by

MEGHA WEERAKOON WATUGALA

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

August 2008

Major Subject: Economics

THE SIGN-UP GAME, SOPHISTICATED LEARNING

AND LEARNING VARIABLE DEMAND

A Dissertation

by

MEGHA WEERAKOON WATUGALA

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Approved by:

Chair of Committee,	John Van Huyck
Committee Members,	Timothy Gronberg
	Guoqiang Tian
	Peter Howard
Head of Department,	Larry Oliver

August 2008

Major Subject: Economics

ABSTRACT

The Sign-Up Game, Sophisticated Learning

and Learning Variable Demand. (August 2008)

Megha Weerakoon Watugala, B.S., California Institute of Technology

Chair of Advisory Committee: Dr. John Van Huyck

This dissertation makes contributions in topics related to mechanism design and learning in game theoretic environments through three essays. The first essay deals with the question of mechanism design in the principal-agent model. The main contribution of this essay is in extending the work by Piketty (1993). It prescribes a mechanism in incomplete informational settings where the principal is able to implement first-best contracts while extracting the entire surplus. Importantly, the mechanism is such that the desired outcome can be uniquely obtained when agents play the action that survives iterative elimination of dominated strategies. Furthermore, given the mechanism, the desired outcome is shown to be a truth-revealing Nash equilibrium which is also Pareto-efficient. It is shown that the proposed mechanism also has the feature that none of the agents prefer any of the other possible Nash Equilibria to the status quo. It thus gives insights into possible mechanisms in finite agent settings that could improve upon the traditional second-best results.

In the second essay, a model of sophisticated learning is developed where it assumes that a fraction of the population is sophisticated while the rest are adaptive

learners. Sophisticated learners in the model try to maximize their cumulative payoff in the entire length of the repeated game and are aware of the way adaptive learners learn. Sophisticated learning contrasts other models of learning which typically tend to maximize the payoff for the next period by extrapolating the history of play. The sophisticated learning model is estimated on data of experiments on repeated coordination games where it provides evidence of such learning behavior.

The third essay deals with the optimal pricing policy for a firm in an oligopoly that is uncertain about the demand it faces. The demand facing the oligopoly, which can be learned through their pricing policy, changes over time in a Markovian fashion. It also deduces the conditions in which learning (experimentation) is not achievable and outlines the different learning policies that are possible in other settings. The model combines the monopoly learning literature with that of the literature on pricing behavior of firms over business cycles. The model has interesting insights on the pricing behavior over business cycles. It predicts that prices jump as the belief of a possible future boom rises over a certain threshold. The model also predicts competition to be quite vigorous following a boom while firms are predicted not to experiment with their (pricing) policies for many periods following a bust.

To my parents

ACKNOWLEDGMENTS

There are many people that I would like to thank for their time and effort during my time at Texas A&M.

First of all, I would like to express my gratitude to my advisor, Prof. Van Huyck, for guiding me through my research. I would also like to thank the members of my committee, Prof. Timothy Gronberg, Prof. Guoqiang Tian and Prof. Peter Howard, and the rest of the Faculty at the Department of Economics whose advise and support were beneficial to my studies and research.

I would also like to thank all my friends. They made my graduate life enjoyable. Everyone I interacted with ensured that my graduate days went smoothly with many fond memories to last a lifetime.

Finally, I would like to thank my parents, sister Sumudu and brother Chula, who have for years supported me, shown confidence in me and have continued to encourage me in many regards throughout my academic career.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

Over the last century, the theory of games has become an important aspect of economic thought. A game is a situation where there is strategic interaction between individuals. That is, the payoff (outcome) to an agent is dependent not only on his own action, but also on the action of the agents interacting with him. Game Theory studies interactions in such settings. Almost all social interactions in our everyday life exhibit strategic interdependence and, hence, have an aspect of a game in them. This explains why Game Theory has had such a large influence on economic reasoning and social science.

The interactions between a taxpayer and the government, an employee and the employer, and a customer and a monopolist, all exhibit strategic interdependence. The actions of agents in these types of situations are influenced by the rules set at the outset. These types of scenarios are studied extensively in mechanism design. Understanding how rules, laws, contracts, etc. influence outcomes is the key to successful institution and policy design.

The second chapter uses a mechanism design approach to investigate the allocation rule a monopolist should enforce among his buyers, given limited information on

This dissertation follows the style and format of the *Journal of Economic Theory.*

his customers' demand characteristics. The classical solution to this sort of principal-agent problem where a single principal is facing a single agent drawn randomly from an underlying characteristics distribution, shows that the best the principal can do is end up implementing second-best outcomes. Piketty (JET 1993) shows that in the principal-agent problem with a finite agent population, if the agent population's underlying characteristics distribution is known, a mechanism can in fact be designed to implement first-best allocations. The research of the second chapter finds and characterizes other incomplete informational settings, where the principal still ends up implementing first-best contracts while extracting the entire surplus.

The third chapter investigates sophisticated learning. Experiments show that sometimes these models cannot explain human behavior. For example, experiments done on behavior on repeated games show that subjects do not initially play an equilibrium and that play evolves with experience (see Van Huyck, Cook & Battalio 1997). Such adaptive behavior is modeled by learning. Most learning models (reinforcement and belief learning) use the history of play to deduce possible future actions even in finitely repeated games. Indeed, backward induction and subgame perfect equilibria, both predict that players would take action in the present only in contemplating the future dynamics of the repeated game. Such players in the learning literature would be sophisticated players.

The third chapter finds evidence for sophisticated learning in repeated coordination games. The learning model used assumes that a fraction of the population is

adaptive, while the rest are sophisticated. Sophisticated learners in the model would be aware of the way adaptive learners learn, and behave with the intention of maximizing their payoff in the entire length of the repeated game. The model explains many behavior such as teaching and signaling regularly found in experiments with repeated games. The data for estimation was obtained from experiments conducted at the Economic Research Laboratory at Texas A&M University (Van Huyck, Cook & Battalio 1997). The analysis reported in the third chapter find some evidence of sophisticated learning.

The fourth chapter develops a model of learning in an duopoly setting. Firms must learn stochastic demand, while interacting with another firm. The duopolist must learn demand through their pricing policy. Demand evolves as a Markovian stochastic process. The chapter finds the pricing policy in a grim-trigger strategy setting and analyzes its properties. The model arbitrages the monopoly demand learning literature with the oligopoly pricing literature. The model has interesting insights on pricing behavior over business cycles. It predicts that prices jump when the belief of a possible future boom crosses a certain threshold. This behavior is similar to that of a monopoly. It also predicts competition to be quite vigorous following a boom, while firms are predicted not to experiment with their pricing policies for many periods following a bust, which contrasts predicted behavior of a monopoly. The need for the duopolists to learn about demand changes the duopolists pricing behavior in the existing literature.

CHAPTER II

IMPLEMENTING FIRST-BEST ALLOCATIONS IN THE PRINCIPAL-AGENTS

MODEL

A. Introduction

This chapter deals with the problem of adverse selection in the principal-agent model. In the principal-agent problem, the principal is assumed only to know the distribution, $F(\theta)$, of skills or tastes, θ, of the population. The problem boils down to the principal constructing a sorting mechanism, given his information about the agents, to extract as much rent as possible. For the simplest case, when the principal is dealing with one agent, incomplete information on the part of the principal results in the well known second-best contracts. Extensive work has been done to deduce conditions under which we could improve on second-best allocations when many agents are involved.

Hammond (1979) and Guesnerie (1981) show with a continuum of agents, that if characteristics of each agent are drawn independently from the same initial distribution, then the principal cannot improve on the classical contract schedules. Dierker and Haller (1990) give the counterpart of this result in large finite economies which is based on the same independence assumption.

In the case where the number of agents is finite, there are two common inter-

pretations of the information (about the distribution $F(\theta)$) available to the principal. They are:

1. Each agent is drawn randomly from the underlying distribution $F(\theta)$.

2. $F(\theta)$ is the true realized distribution of the agent population in question.[1]

The latter is the more traditional interpretation.[2] Piketty (1993) shows that when the number of agents is finite, under the second interpretation, the principal can design a game whose unique Bayesian Nash Equilibrium (through iterative elimination of strictly dominated strategies) yields first-best allocations. Thus, there is no loss in efficiency and the principal is able to extract all the rent. The fact that the principal obtains the same rent as under the complete information setting makes this an interesting result. However, the principal under the second interpretation, in some sense, has more information than the first under interpretation. In the first interpretation, since each agent is randomly drawn from the underlying distribution, the principal has less information on the realized distribution of the agent population's types. Indeed, in such a setting, it is possible for all the agents to be of the same type. Such a setting being too broad, Piketty (1993) went to the other extreme where the principal is completely aware of the realized type distribution of the agents

---

[1]That is, the principal knows how many agents of each type there are in the population with certainty.

[2]See Maskin and Riley (1984).

in question. Here, the principal is aware of the number of agents of each possible type with certainty.

Thus, depending on the information available to the principal, contrasting allocations result. If the agent types are drawn randomly from the distribution the principal is aware of, then second-best allocations can result. On the other hand, if the principal is aware of the number of agents of each type in the finite agent population, he can design a mechanism (a game) which results in first-best allocations. Given such solutions, what will happen to the allocations in other information settings? Are there any other incomplete information settings where the principal can design a mechanism to end up allocating first-best outcomes?

Riordan and Sappington (1988) and Cremer and McLean (1985, 1988) derive conditions on the incomplete information structure in a finite population in which full extraction of surplus is possible. All these work assume that agents are risk neutral. The results of Riordan and Sappington (1988) hinges on a public signal received after signing the contracts that is correlated with the private information of the agents. Riordan and Sappington derive necessary and sufficient conditions of the public ex post signal for first-best outcomes to result. The result of Cremer and McLean (1985, 1988) hinges on the principal's ability construct lotteries which the agents have to accept when participating. Due to the presence of the lottery, agents maximize their expected payoff and it is quite possible for some of them to end up with a lower payoff than if they had not dealt with the principal (status quo) at all.

This chapter deduces other incomplete information settings where the principal is still able to design a game to implement first-best allocations while extracting all the rent. The main feature of the information setting is that by an agent's revelation of his type, the principal is able to better deduce the types of the remaining agents. It also uses the fact that a self-interested agent has no concern for the information of the other agents he reveals through his actions. Thus, manipulating this non-cooperative behavior of the agents, the principal is able to design a game in certain settings where agents would truthfully reveal themselves to obtain the desired results. Also, unlike the results of Cremer and McLean, agents in the mechanisms proposed are guaranteed to, ex post, prefer the truthful revealing outcome to the status quo.

B.   The Model

Adverse selection in the principal-agent model is applicable to many situations[3] and we will illustrate our result through a simple application of a monopoly (the principal) selling a single good to many buyers (the agents).[4] The profit maximizing monopoly is assumed to be able to produce the good at a constant marginal cost of c and will

---

[3]Price discrimination with quantity discounts (Goldman, Leland and Sibley (1984), Roberts (1979), Spence (1977), Maskin and Riley (1984)), monopoly pricing of goods of differing quality (Mussa and Rosen (1978)), optimal income taxation (Mirrlees (1971)), monopoly pricing of insurance (Stiglitz (1997)) and labor contracts (Hart (1983)).

[4]The description of the problem is motivated by the problem in "Monopoly with incomplete information" of Maskin and Riley (1984).

discriminate among the agents by offering them non-linearly priced bundles of the good.

1. The Agents

The economy consists of the principal (monopoly) selling a good to a set of agents (buyers)[5] $I = \{1, ..., n\}$. The finite set $\Theta = \{\theta_1, ..., \theta_r\}$ capture the characteristics (types) of the agents (w.l.o.g. let us assume that $0 < \theta_1 < ... < \theta_r$).

The utility of agent $i \in I$, of type $\theta(i)$, is given by

$$U\left((q, T), \theta(i)\right) = \int_0^q p\left(x, \theta(i)\right) dx - T \tag{2.1}$$

where q is the quantity of the good purchased while T is the payment (transfer paid) to the monopolist. Viewing $p(q, \theta)$ to be an agent's inverse demand for the good, the utility function thus takes the form of consumer surplus. Throughout we will assume the standard assumptions on $U(.)$ and $p(.)$ as listed below so that the functions are well-behaved for the problem at hand.

---

[5]The words principal and monopoly and the words agents and buyers and the words contract and bundle will be used interchangeably for the rest of this chapter.

Assumptions

1. For all $\theta \in \Theta$ the demand price function $p(q,\theta)$ is decreasing in q for $q \leq q^{fb}(\theta)$ $\left(\frac{dp(q,\theta)}{dq} \leq 0 \right)$ and $p(q,\theta) \geq c$ iff $q \leq q^{fb}(\theta)$.

2. $p(q,\theta)$ is strictly increasing in θ $\left(\frac{dp(q,\theta)}{d\theta} > 0, \text{ or equivalently } \frac{dU((q,T),\theta)}{d\theta} > 0 \right)$ for all $\theta \in \Theta$.

The first assumption says that for all agents, the marginal utility of an additional good is non-increasing in the quantity of goods already purchased. Also it claims that there are always possible gains from trade to be had and that there exist a unique first-best quantity to be traded with each agent. The second assumption says that agents of higher types are associated with higher demand. These conditions and the way utility function has been set up imply that the slope of an agent's indifference curve is

$$-\frac{\partial U}{\partial q} \Big/ \frac{\partial U}{\partial T} = p(q,\theta). \tag{2.2}$$

Thus, for any bundle (q,T), the indifference curve for a buyer of a higher type has a higher slope. This is the Spence-Mirrlees condition or the Single Crossing Condition. Also note that $U((0,0),\theta) = 0$ which imply that staying out is normalized to the $(0,0)$ bundle.

2. The Principal

The principal's objective is to maximize the profits made by selling the non-linearly priced bundles to the agents.

i.e.

$$\max_{(q_i, T_i), i \in I} \Pi = \sum_{i=1}^{n} T_i - c \sum_{i=1}^{n} q_i \tag{2.3}$$

a. The Incomplete Information Setting

The principal is not aware of the possible characteristics of an individual agent but would have information on the realizable distribution $F(\theta)$ of characteristics in the population. Let $F(\theta)$, also denoted by $(\mu(\theta_1), ..., \mu(\theta_r))$, represent a possible realizable distribution of the agent characteristics, where $\mu(\theta_s)$ is the fraction of agents whose characteristic is θ_s in the c.d.f. $F(\theta)$. Let the principal's knowledge of the actual distribution (or the sample distribution) of characteristics of the population $F(\theta)$ be, that it is going to be one from the set of possible distributions $\{F_1(\theta), ..., F_k(\theta)\}$. This will be common knowledge[6]. Thus, if the realized type distribution of the population is $F_a(\theta) = (\mu_a(\theta_1), ..., \mu_a(\theta_r))$ then $\mu_a(\theta_1)$ fraction of individuals of the population are of type θ_1, $\mu_a(\theta_2)$ fraction of type θ_2 and so on.

---

[6]Thus, the agents are aware of the principal's information. For the settings considered here, results would hold even if the agents were fully aware of each agent's type. In such a case, the only requirement is that the principal have the realized distribution.

In Piketty (1993), this set of possible type distributions would be a singleton. Also, if the agents were drawn randomly from some initial distribution, $G(\theta)$, the set of possible distributions, $\{F_1(\theta), ..., F_k(\theta)\}$, would be all possible type distributions that could be realized from drawing agents from the initial distribution $G(\theta)$.

b. The New Incomplete Information Setting (NIIS)

Let the New Incomplete Information Setting (NIIS) be such that in the set of realizable type distributions $\{F_1(\theta), ..., F_k(\theta)\}$, no two distributions first order stochastically dominate the other.

   i.e.

$\nexists$ two distinct distributions $F_a(\theta), F_b(\theta) \in \{F_1(\theta), ..., F_k(\theta)\}$

$$\text{where } F_a(\theta_s) \geq F_b(\theta_s) \text{ for all } \theta_s \in \Theta.$$

This chapter will design a mechanism (a game) where first-best allocations will result in the above setting while the principal extracts all the possible surplus. In the NIIS, the principal has less information than if he had known the realized sample distribution with certainty. However, in the NIIS, the principal has more information than if the agents were randomly drawn from some distribution. The NIIS can be viewed as a case where the agents are drawn from some distribution where certain realizable type profiles have been ruled out until the above criterion is met.

3. The Mechanism and Implementation

The timing of the sequence of events is shown in Figure 1 is similar to that of the standard principal-agent model.
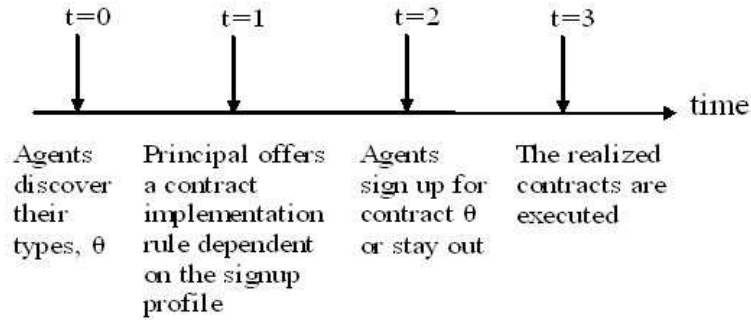
Fig. 1. Timing of Events

The principal moves first. The principal lists r $(= |\Theta|)$ possible contracts, conveniently indexed $\theta_1, ..., \theta_r$, for the agents to sign up for. The principal also announces the contract implementation rule that is dependent on the final sign up distribution of the population. The principal would intends contract θ_s for agents of type θ_s.

Next is the agent's move. Each agent signs up for a contract. This act can be viewed as a message sent or as a type revelation by the agents. Let m_i represent the contract signed up for (or type revealed/the message sent) by agent i. Therefore, $m \in (m_1, ..., m_n) \in M \equiv [\Theta \cup \{0\}]^n$ is the sign up profile of the agents. This profile will thus determine the resulting contracts according to the announced implementation

rule. Thus, the contract implementation rule announced by the principal would be some function $h : M \to Z$ where Z is a set of profiles of bundles (q, T) for all agents. Let $m_i = 0$ correspond to opting to stay out. If the agent opts to stay out, the principal can no longer influence him. This is the outside option of the agent. Also, let m_i^{tr} denote truthful revelation by agent i (i.e. $m_i^{tr} = \theta(i)$).

Let $h_i (m_i, m_{-i})$ denote the bundle obtained by agent i when his revelation is m_i and the revelations of the others is the vector m_{-i}. Once the agent makes a revelation $m_i \neq 0$, having opted not to stay out, he must go through with the contract/bundle, $h_i (m_i, m_{-i})$, designated by the mechanism.

Also, let the contract $\left(q^{fb}(\theta_s), T^{fb}(\theta_s)\right)$ denote the first-best bundle for an agent with the characteristic θ_s where $q^{fb}(\theta_s)$ is the first-best quantity for the agent (i.e. $p\left(q^{fb}(\theta_s), \theta_s\right) = c$) and the payment, $T^{fb}(\theta_s)$ is such that

$$U \left(\left(q^{fb}(\theta_s), T^{fb}(\theta_s)\right), \theta_s\right) = U \left((0, 0), \theta_s\right) = 0. \qquad (2.4)$$

Thus, the bundle $\left(q^{fb}(\theta_s), T^{fb}(\theta_s)\right)$ is the perfect-price discriminating bundle a monopoly with complete information would choose for an agent with the characteristics θ_s.

a. The Sign-Up Game

Let $m \in M$ be the revealed type profile or the sign up profile of the agents. Let this

sign up profile (distribution) of the population be represented by

$P = (P(\theta_1), ..., P(\theta_r))$ where

$$P(\theta_j) = \frac{\# \{i \in I, \text{ s.t. } m_i = \theta_j\}}{n} \tag{2.5}$$

That is, $P(\theta_s)$ is the fraction of the population signing up for (revealing them-

selves as) θ_s. Let the principal allocate bundle $\left(q^{fb}(\theta_s), T^{fb}(\theta_s)\right)$ to all agents revealing

(signing up) themselves to be of type θ_s, if $P(\theta_s) = \mu_j(\theta_s)$ for all $\theta_s \in \Theta$ for some j

s.t. $F_j(\theta) \in \{F_1(\theta), ..., F_k(\theta)\}$. Otherwise, the principal offers no bundles to all agents

(or in other words offers the bundle (0,0)).

Thus, bundles will only be implemented if the sign up (revelation) profile matches

with one of the possible characteristics profiles. And if bundles are ever implemented,

they will be implemented so that each agent revealing themselves to be type θ_i, will

get the corresponding first-best bundle for agent of type θ_i .

Remark 1:   In the Sign-Up Game, by signing up as a type θ_s, an agent might either

end up getting the bundle $\left(q^{fb}(\theta_s), T^{fb}(\theta_s)\right)$ or $(0,0)$.

Thus, the name Sign-Up Game is derived from the fact that agents can be viewed

as signing up for the first best bundle θ_s when they reveal themselves as such and

since the game's outcome uses this final sign up (revelation) profile of the agents to determine the final outcome.

As mentioned earlier, the set of possible realizable type distributions being a singleton implies that the principal is aware of the actual realized type distribution of the agents as with Piketty (1993). In such a setting the Sign-Up Game is similar to the mechanism in Piketty (1993).

b. The ϵ-Sign-Up Game

The ϵ-Sign-Up Game is similar to the Sign-Up Game in all aspects except for all $\epsilon > 0$ the principal allocates bundles $\left(q_i^{fb}, T_i^{fb} - \epsilon'\right)$ where he was allocating $\left(q_i^{fb}, T_i^{fb}\right)$ in the Sign-Up Game with

$$\epsilon' = \min\left[\epsilon, \min_{\theta_s \in \Theta - \{\theta_r\}} U\left(\left(q^{fb}(\theta_{s+1}), T^{fb}(\theta_{s+1})\right), \theta_s\right)\right] \qquad (2.6)$$

Thus, in this setting, agents have an incentive ($\epsilon' > 0$) to obtain their corresponding bundle than to staying out. The restriction of ϵ' (eqn: 2.6) is to guarantee that the incentive is not too big so as to make an agent prefer the bundle of a higher type to staying out. As $\epsilon \to 0$ this game becomes the Sign-Up Game. As always, the

agents will still prefer the bundle of a lower type to his own.

C.   Results

Proposition 1:   If the principal implements a Sign-Up Game in the NIIS, truth revealing is a Nash Equilibrium.

Proof   If all the agents from a realized distribution $F_a(\theta)$ are truth revealing, the fractions signed up for each bundle, $P(\theta)$, will match the fractions of the population of $F_a(\theta)$ and thus the bundles they have signed up for will be implemented in their first-best state.

i.e.

If $m_i = m_i^{tr}$ for all i

$h_i\left(m_i^{tr}, m_{-i}^{tr}\right) = \left(q^{fb}(\theta(i)), T^{fb}(\theta(i))\right)$ and thus

$$U\left(\left(q^{fb}(\theta(i)), T^{fb}(\theta(i))\right), \theta(i)\right) = 0 \text{ for all } i.$$

For this not to be a Nash Equilibrium, an agent $i \in I$ would find it beneficial to, sign up for some other bundle θ_j.

i.e. If this is not a Nash Equilibrium,

$$\exists\, \theta_j \text{ s.t. } h_i\left(\theta_j, m_{-i}^{tr}\right) \succ_i h_i\left(m_i^{tr}, m_{-i}^{tr}\right) \text{ for some } i \in I.$$

We know that for any agent $i \in I$, signing up for a bundle θ_j such that $\theta_j > \theta(i)$, would always result in the agent getting a non-positive payoff (zero if the bundle is not implemented and a negative payoff if the bundle is implemented[7]).

i.e. From Remark 1

For $\theta_j > \theta(i)$

$$U\left(h_i\left(\theta_j, m_{-i}^{tr}\right), \theta(i)\right) \leq 0$$

$$\text{since} \begin{cases} U\left(\left(q^{fb}(\theta_j), T^{fb}(\theta_j)\right), \theta(i)\right) < 0 & \text{if the bundle is implemented} \\ \\ U\left((0,0), \theta(i)\right) = 0 & \text{if not} \end{cases}$$

Thus, it is never beneficial for an agent to reveal himself as a higher type.

Now, we only have to show that for all agents $i \in I$, it is not better to reveal themselves as $\theta_j < \theta(i)$, if everyone else is truthfully revealing themselves.

If an agent $i \in I$ is better off signing up for a bundle θ_j, $(\theta_j < \theta(i))$, then it must be the case that this bundle is implemented (since not being implemented would yield the same payoff of 0).

i.e.

If $h_i\left(\theta_j, m_{-i}^{tr}\right) \succ_i h_i\left(m_i^{tr}, m_{-i}^{tr}\right)$ for some agent $i \in I$ where $\theta_j < \theta(i)$

$$\Rightarrow h_i\left(\theta_j, m_{-i}^{tr}\right) = \left(q^{fb}(\theta_j), T^{fb}(\theta_j)\right)$$

$$\text{since } U\left(\left(q^{fb}(\theta_j), T^{fb}(\theta_j)\right), \theta(i)\right) > 0 \text{ for } \theta_j < \theta(i).$$

---

[7]Due to Assumption (ii) of U(.).

Let this new sign up distribution[8] be $F_b(\theta)$. If the bundles are being implemented in the Sign-Up Game, this implies that the new sign up (revelation) distribution, $F_b(\theta)$, must be represented by one of the other possible distributions $(F_b(\theta) \in \{F_1(\theta), ..., F_k(\theta)\})$.

Now let us look at distributions that correspond to $F_a(\theta)$ and $F_b(\theta)$.

The only difference is that one agent signing for bundle $\theta(i)$ in distribution $F_a(\theta)$ has signed up for bundle θ_j to yield distribution $F_b(\theta)$ where $\theta_j < \theta(i)$.

Therefore,

$F_a(x) = F_b(x)$ for $x < \theta_j$, distributions are identical for options $x, x < \theta_j$.

$F_a(x) < F_b(x)$ for $\theta_j \le x < \theta(i)$, an extra agent has signed up for option θ_j.

$F_a(x) = F_b(x)$ for $\theta(i) \le x$, since all options $k, k > \theta(i)$ have the same number of agents signed up

This implies that $F_b(\theta)$ stochastically dominate $F_a(\theta)$. A contradiction. Therefore, there is no way an agent can be better off by not truthful revealing himself if all the other agents are. Thus, truth revealing is a Nash Equilibrium. ∎

In this case, since the principal is implementing first-best bundles and agents are truth revealing, the first-best outcome is obtained and the principal is extracting the entire surplus.

---

[8]When agent i is revealing himself as type θ_j while the rest are truthfully revealing.

Proposition 2: If the principal implements a Sign-Up Game in the NIIS, it is not a Nash Equilibrium for the population from the realized distribution $F_a(\theta)$ to sign up according to one of the other possible distribution $F_b(\theta) \in \{F_1(\theta), ..., F_k(\theta)\}$.

Proof   If the population reveals itself to be of a distribution of $F_b(\theta)$, all bundles will be in their implemented state. We know then that, in this case, for this to be an equilibrium, all agents of type θ_s must have to signed up for bundles θ_j such that $\theta_j \leq \theta_s$,

$$\text{since } U\left(\left(q^{fb}(\theta_j), T^{fb}(\theta_j)\right), \theta_s\right) > 0 \text{ for } \theta_j < \theta_s$$

$$\text{and } U\left(\left(q^{fb}(\theta_j), T^{fb}(\theta_j)\right), \theta_s\right) < 0 \text{ for } \theta_j > \theta_s.$$

Therefore, the fraction of bundles θ_j, s.t. $\theta_j \leq \theta_s$ available to be signed up in $F_b(\theta)$, must at least accommodate agents of type θ_j, s.t. $\theta_j \leq \theta_s$ in $F_a(\theta)$, for all $\theta_s \in \Theta$.

i.e. $F_b(\theta_s) \geq F_a(\theta_s)$ for all $\theta_s \in \Theta$, with the relation being strict for some θ_s since $F_b(\theta)$ and $F_a(\theta)$ are different distributions.

This implies that $F_b(\theta)$ (first order) stochastically dominate $F_a(\theta)$. A contradiction.

Therefore, it can never be the case that a population from a realized distribution $F_a(\theta)$, would be in equilibrium imitating to be of a distribution $F_b(\theta) \in \{F_1(\theta), ..., F_k(\theta)\}$. ∎

This implies that in the NIIS, there can be no equilibrium in which agents can do better than by truth revealing. Once the appropriate Sign-Up Game is announced, unless there is a redistribution of wealth/payoff among the agents, the agents can not do better even by colluding among a coalition of them. Thus, the truth revealing equilibrium is Pareto superior to any other equilibrium.

Remark 2:   This means that in the NIIS, if a population is in equilibrium with the bundles in their implemented state, then it must be the case that the agents have revealed their types truthfully. This is because bundles will only be implemented if it corresponds to one of the possible distributions and the above proposition showed that it is not an equilibrium for any population to have had revealed itself to be another distribution of the possible realizable distributions $\{F_1(\theta), ..., F_k(\theta)\}$. So if they are in equilibrium with the bundles implemented, it must be the case that they have revealed their types truthfully.

Remark 3:   It follows from Remark 2 that in the NIIS, for the Sign-Up Game, in all possible Nash Equilibria, the agents' payoff will be the same as the outside option. This is simply because in the equilibrium when the bundles are implemented, the principal extracts all the rent from each agent and in equilibria where bundles are not implemented the agent's payoff is the same as opting to stay out.

Thus, if the principal decides to offer $\epsilon > 0$ for each agent if the bundles are implemented, then the truth revealing equilibrium Pareto dominates (Pareto superior

to) all other possible equilibria, since in all other possible equilibria, bundles being not implemented, agents and the principal get a payoff of 0.

Proposition 3: If the principal implements an ϵ-Sign-Up Game, in the NIIS, (two rounds of) iterative elimination of weakly dominated strategies yield the equilibrium in which all agents truthfully reveal themselves. Thus, first-best outcomes result and as $\epsilon \to 0$ the principal extracts all the possible surplus.

Proof   For all agents $i \in I$, the outside option (staying out) weakly dominates obtaining a first-best bundle designed for a higher type θ_j, s.t. $\theta_j > \theta(i)$. Therefore, all strategies $m_i = \theta_j > \theta(i)$ can be eliminated.

In the ϵ-Sign-Up Game if all agents are truth revealing, all agents get an $\epsilon' > 0$. From Proposition 2 we know that if agents are not truth revealing, for bundles to be implemented, at least one agent $i \in I$ must be revealing (signing up) himself to be of a higher type ($m_i = \theta_j > \theta(i)$) whose possibility we just eliminated.

Thus, the only bundle implementable strategy profile left is the truth revealing one where agents get non-zero payoff ($\epsilon > 0$), while all other strategies left ($m_i = \theta_j < \theta(i)$) always yield a zero payoff. Therefore, all strategies, $m_i = \theta_j < \theta(i)$, can be eliminated in favor of truth revealing through weak dominance. Thus, we arrive at truth revealing through iterative elimination of weakly dominated strategies. ∎

Thus, not only is the truth revealing equilibrium Pareto-superior it is also attainable by two rounds of iterative elimination of weakly dominated strategies.

Example 1:

Consumers utility function $= q(\theta - q) - T$

Monopolist's marginal cost $= 0$

Table I. The First-Best Bundles for Agents of Different Types (θ), $U((q, T), \theta(i)) = q(\theta(i) - q) - T$

θ	First best q^{fb}	T^{fb}
1 (low)	$\frac{1}{2}$	$\frac{1}{4}$
2 (mid)	1	1
3 (high)	$1\frac{1}{2}$	$2\frac{1}{4}$

Table II. Payoff for Different Agents for the Various First-Best Bundles

Bundle (q, T)	$U((q,T), \theta_i = 1)$	$U((q,T), \theta_i = 2)$	$U((q,T), \theta_i = 3)$
($\frac{1}{2}, \frac{1}{4}$)	0	$\frac{1}{2}$	1
(1, 1)	-1	0	1
($1\frac{1}{2}, 2\frac{1}{4}$)	-3	$-1\frac{1}{2}$	0

Possible Distributions

Let us assume that there are two agents and they will either both be of type $\theta = 2$ or one each from $\theta = 1$ and $\theta = 3$. We can denote this by two distributions, D1 and D2 as

D1 = $(\mu_1(\theta = 1), \mu_2(\theta = 2), \mu_3(\theta = 3)) = (0, 1, 0)$ and D2 = $\left(\frac{1}{2}, 0, \frac{1}{2}\right)$

Each type of agent's first-best bundles are presented in Table I while their payoffs from obtaining each of these bundles are presented in Table II. Note that the

characteristics profile of the agents in the two scenarios do not first order stochastically dominates the other. The principal only has this information and will not know whether the realized state is D1 or D2. Let us assume the principal implements a $\epsilon-$Sign-Up Game here.

Let us look at the ϵ-Sign-Up Game that the agents will be playing. That is, the principal announces that the bundles will be implemented (in their first-best state) only if the fractions signed up for all options correspond to either D1 or D2.

Table III. Case 1: The Sign-Up Game when the Realized State Is D1

Player 2 (mid type) $\theta = 2$

		Stay Out	$m_2 = 1$	$m_2 = 2$	$m_2 = 3$
Player 1 (mid type) $\theta = 2$	Stay Out	0, 0	0, 0	0, 0	0, 0
	$m_1 = 1$	0, 0	0, 0	0, 0	$\frac{1}{2} + \epsilon', -1\frac{1}{2} + \epsilon'$
	$m_1 = 2$	0, 0	0, 0	ϵ', ϵ'	0, 0
	$m_1 = 3$	0, 0	$-1\frac{1}{2} + \epsilon', \frac{1}{2} + \epsilon'$	0, 0	0, 0

Table IV. Case 2: The Sign-Up Game when the Realized State Is D2

Player 2 (high type) $\theta = 3$

		Stay Out	$m_2 = 1$	$m_2 = 2$	$m_2 = 3$
Player 1 (low type) $\theta = 1$	Stay Out	0, 0	0, 0	0, 0	0, 0
	$m_1 = 1$	0, 0	0, 0	0, 0	ϵ', ϵ'
	$m_1 = 2$	0, 0	0, 0	$-1 + \epsilon', 1 + \epsilon'$	0, 0
	$m_1 = 3$	0, 0	$-3 + \epsilon', 1 + \epsilon'$	0, 0	0, 0

In the example above, if $\epsilon' < 1$, one can deduce how the agents could arrive at the truth revealing equilibrium through iterative elimination of weakly dominated strategies. However, in this example, when an agent realizes his type, he could figure out the other agent's type and thus the realized agent distribution. Therefore, this is indeed a very special example, but the results hold for any case as long as the realizable type distributions do not first order stochastically dominate each other.

Notice that in the above design, the probability of realizing the state D1 (Table III) or the state D2 (Table IV) did not matter. This would not be the case with traditional mechanism design. Traditionally, the principal would need to know the underlying probability of realizing each of the possible types to go about creating the

second-best outcomes. However, as long as the realizable type distributions do not first order stochastically dominate another, a Sign-Up Game can be designed. In this Sign-Up Game, if the agents choose strategies that survive iterative elimination of weak dominance, first-best outcomes result and the principal obtains first-best rent.

D.   Discussion

The Sign-Up Game offers an avenue to overcome loss of efficiency in adverse selection problems in certain settings. This chapter shows that for the ϵ-Sign-Up Game, if the principal's information is a setting of the NIIS, rational agents choosing strategies that survive iterative elimination of weakly dominated strategies result in a unique Nash Equilibrium with first-best outcomes while the principal extracts all the surplus.

In traditional solutions of adverse selection problems a dominant strategy is available to the agents. Also, given the contract schedule, the outcome does not depend on the actions of the other agents in such solutions. However, an agent's final payoff in the Sign-Up Game depends on the actions of other agents. For first-best outcomes to result in the ϵ-Sign-Up Game, agents are required to choose strategies that survive iterative elimination of weak dominance. Therefore, agent's rationality in not choosing weakly dominated strategies need to be common knowledge. There has been many discussions in the implementation theory literature on the use of different solution concepts: dominant strategy equilibrium (Gibbard (1973), Satterthwaite

(1975)), Nash equilibrium (Maskin (1999)), sophisticated equilibrium (Farquharson (1969)), undominated Nash equilibrium (Palfrey and Srivastava (1991)) etc. Having a dominant strategy is ideal since in that case the agent's assumption of other agents' rationality is not necessary. However, such a condition restricts implementation. The use of iterative elimination of strictly dominated strategies to obtain a unique equilibrium is also suitable since it need only assume that agents do not choose strictly dominated strategies and assume that other agents also act likewise. This chapter uses iterative elimination of weakly dominated strategies to arrive at the desired outcome. Thus, agents here need to choose weakly dominant strategies and assume others follow suit. The use of the solution concept of iterative elimination of weakly dominated strategies is questionable if it could make agents choose a Pareto-inferior equilibrium. Looking at the Sign-Up Game, one can deduce that the equilibrium thus obtained is Pareto dominant to any other achievable Nash Equilibrium. Thus, in these Sign-Up Games, agents face no dilemma in choosing equilibria through strategies that survive iterative elimination of weakly dominance. Thus, the assumption on agents' behavior used is not unnecessarily restrictive.

A crucial feature for the success of the Sign-Up Game is the information setting it is used in. In the appropriate settings, if an agent truthfully reveal his type, the principal is able to better infer the types of the remaining agents. Thus, each agent's revelation gives the principal not only information about that particular agent but also information on other agents. The principal thus uses this fact in designing

the appropriate Sign-Up Game to manipulate the self-interested, non-cooperative behavior of the agents.

In this chapter the Sign-Up Game in the NIIS setting does not implement any contracts when the principal realizes that a single agent has deviated from truthful revelation. Even in Piketty (1993), the principal could end up extracting less than second-best rent even when a single agent does not choose his dominant strategy. Hamilton and Slutsky (2004, 2007) propose a mechanism, in Piketty's setting, where the principal extracts rent even in off-equilibrium outcomes, while still extracting first-best rent in the truth-revealing equilibrium. They have observed that the off-equilibrium outcomes that would result due to a single 'noise' player are very harsh both to the principal and the other agents. They try to account for this 'noise' players' irrational behavior, but in a theoretical setting, assuming rational behavior, such behavior should be of little concern. Using the directives of Hamilton and Slutsky, it could be possible to ensure that the principal does extract rent even in off-equilibrium outcomes in the NIIS. However, the exercise of this chapter is in showing settings in which full rent can be extracted in the presence of rational agents.

In Cremer and McLean (1985, 1988), they derive conditions on the incomplete information structure in a finite population in which full extraction of surplus is possible. They assume risk neutral agents and the result hinges on the principal's ability to construct lotteries which the agents have to accept when participating. Due do the presence of the lottery, agents maximize their expected payoff and it is

quite possible for some of them to end up with a lower payoff than if they had not dealt with the principal (status quo) at all. Also, they require that the agents have the same information about the agents' types as the principal[9]. The results of the current setting is quite different. Risk attitudes of agents are irrelevant and agents in the mechanisms proposed are guaranteed to, ex post, prefer the truthful revealing outcome to the status quo. Additionally, in the mechanism proposed, agents will prefer the truth revealing outcome to the status quo regardless of the actions of the other agents. Therefore, if we assume agents have limited liability, the results of the proposed mechanism would hold. Also, the results of Cremer and McLean and other literature dealing with incomplete information depend on the ability of the principal to attach probabilities to possible outcomes. However, to use the Sign-Up Game in the NIIS, the principal needs only to know the possible realizable distributions and does not need to be aware of any probability associated with any of them being realized. This approach of analyzing incomplete information via possible realizable states without taking their associated probability into account is indeed a novel one.

What does this say about a population of agents randomly drawn from a distribution? Take a simple example of a population of n agents being randomly drawn from a distribution, $G = (\Pr(\theta = 1), \Pr(\theta = 2)) = (p, 1-p)$. One can easily see that the possible realizable distributions first order stochastically dominate each other.

---

[9]In Cremer and McLean, if the principal's information on the agents' types is $F(\Theta)$, then agent i after realizing his type, θ_i, would have the information $F(\Theta|\theta(i) = \theta_i)$, on the other agents.

Therefore, the Sign-Up Game with the above implementation rule would not be of use here. In such settings, the question of whether the principal can better his rent obtained by second-best allocations, through an implementation method in the vein of the Sign-Up Game is ambiguous.

CHAPTER III

SOPHISTICATED LEARNING IN REPEATED COORDINATION GAMES

A.   Introduction

Life is full of repeated interactions with the same people. Our behavior in such situations is influenced by factors such as precedents, conventions, norms and an anticipation of future interactions. To understand the dynamics of repeated interactions, laboratory experiments on an effective way. For example, coordination games have multiple Nash equilibria. Coordination failure results when subjects do not implement a payoff-dominant equilibrium in the game. Studies of behavior in coordination games have enabled us to understand coordination failure.

This chapter investigates sophisticated learning in coordination games. Coordination games describe many social interactions from behavior on the road to behavior at home, school or work. It could model productivity in an assembly line to trade negotiations among national economies. Coordination failure and the resulting efficiency lost is the root of underperformance in many situations. Thus, understanding how people behave in such situations or how people choose or converge to an equilibrium is critical for the success and survival of organizations and to better design mechanisms, institutions and public policy so as to overcome and avoid coordination

failure.

If the same coordination game is played repeatedly, adaptive learning models predict that players end up playing an equilibrium strategy. In this chapter, a model of sophisticated learning is proposed and estimated on data from repeated coordination game experiments from Van Huyck, Cook and Battalio (1997). The estimations find that a fraction of the population can be classified as sophisticated. Furthermore, the model of sophisticated learning has the ability to explain other behaviors observed in repeated games.

Reinforcement learning and belief learning are the two common forms of learning studied. In reinforcement learning, (Erev & Roth (1998), Borgers & Sarin (2000)), the player learns according to the history of his payoffs for each of his actions. In belief learning, as in fictitious play (Brown 1951), players try to learn the behavior of the opponent in order to best respond to it. Thus, players in these learning models try to maximize the payoff in the next period according to a learning rule exclusively based on the history of play.

A drawback of such models is that players are assumed to be myopic, that is, to not anticipate future interaction. They imply that players are unaware of the influence of their current actions on future payoffs. Another drawback of these models is the mutual inconsistencies in players' beliefs about their opponent. For example, if two players are playing a game repeatedly, and they are both fictitious play learners, their assessment that the opponent is playing a fixed mixed strategy is incorrect (in most

cases) since each player's strategy will be a series of best responses and not a fixed mixed strategy.

The learning behavior analyzed in this chapter allows a fraction of the population to be sophisticated. Sophisticated learning has been studied before[1] in models of "level-k" learning (Stahl 2000) and sophisticated EWA learning (Chong, Camerer and Ho 2006). Thus, the term sophistication has generally been used in instances where the agent has complex reasoning ability in regard to anticipation, recursive thinking, better forecasting methods, etc.

Chong, Camerer and Ho (2006) consider a model of learning where a fraction of the population learns according to self-tuning EWA, and the rest play sophisticatedly. The sophisticated players assume that a certain fraction of the population are self-tuning EWA learners, and the rest are sophisticated as themselves and best respond to this belief taking into consideration the future effects of current play for the rest of the game.

The model used in this chapter is similar to that of Chong, Camerer and Ho (2006). As in Chong et al. (2006), a fraction of the population is assumed to follow the self-tuning EWA learning rule. The self-tuning EWA learning rule is similar to the EWA learning rule but has one free parameter. The rest of the parameters are self-tuned according to the dynamics of play. Ho et. al. (2007) show that the self-

---

[1]see Selten (1991) and Milgrom and Roberts (1991) . For a review of sophisticated learning see the chapter on Sophisticated Learning in 'The Theory of Learning in Games' by Fudenberg and Levine (1998).

tuning EWA model does as well as the EWA model in predicting behavior in new games[2]. In Chong et. al., the sophisticated population assumes the same fraction of sophisticated players throughout the game. However, if players' behavior is different depending on their type, then players may be able to deduce, with repeated play, the type of their opponent from the actions they have taken thus far. Thus, a player could have a more accurate belief about his opponent's type later in the game. Since a player's chosen action is dependent on the beliefs about his opponent, it is important that the model capture what the player learns about his opponent's type. In the model below, the sophisticated population updates their belief about the fraction of the population who are sophisticated by observing the history of play in a Bayesian manner.

In this chapter a model of sophisticated learning is fit to data from the continental divide game (Van Huyck, Cook & Battalio 1997) to test for sophisticated behavior. It finds that a fraction of the population can be classified as sophisticated.

―――――――

[2]They also show that self-tuning adds the most economic value, that is, subjects would have earned more in an experimental session if they had followed the recommendations of its theory.

1.   Teaching in Coordination Games

Teaching is when a player influences his opponent to take actions that are beneficial to either of the players or to the group as a whole. Coordination games are ideal for testing teaching, because the best response in a coordination game is dependent on the action of the opponent and if someone is playing with an adaptive learner, he could, with enough play, influence his opponent's most attractive action.

There is evidence of such behavior. A fraction of the population trying to drag the population out of an inferior equilibrium is observed in several experiments on coordination games, which has been called as teaching or leading behavior. For example, Brandts and Cooper (2006) observe this in their experiment on minimum-effort games,

> we often observe that a subset of the employees act as leaders, raising their
> effort levels following a bonus rate hike and guiding the other employees
> to higher effort levels.

This is evidence of teaching. The leaders here correspond to sophisticated players while the laggards may be adaptive learners. Brandts and Cooper go on further to say that

> The success of this leadership by example depends both on the persistence
> of leaders and on whether laggards, employees who do not initially increase
> their effort, eventually respond by raising their efforts.

This is consistent with the idea that the rest of the population are adaptive learners.

2. Speed of Convergence to Equilibria of Different Efficiency

Van Huyck, Cook and Battalio (1997) notes that

> the resistance to dynamics is most pronounced in the low sessions. Naturally, subjects in a low session are more likely to resist the logic of the myopic best response and fictitious play dynamics than the subjects in a high session, since the low sessions are converging to less efficient outcomes.

This fact suggests that players are acting sophisticatedly, that is, sophisticated players realize that the faster they converge to the inferior equilibrium the more periods they will earn the low payoffs of the inferior equilibrium. Therefore, attempting to resist the inevitable convergence might increase the overall payoff of the repeated game since the payoffs obtained while converging to the inferior equilibrium could be higher than that of the payoff of the inferior equilibrium.

3. Other Insights of Sophistication

Sophisticated players update their beliefs about the likelihood of their opponents being sophisticated through the observed actions of their opponents. Thus, the ability to observe everyone's actions would influence the observed behavior in a repeated setting. The ability to observe the groups' actions not only facilitates belief formation but also helps a sophisticated player to effectively signal his presence to fellow sophisticated players. In many coordination games experiments, a player's payoff is dependent on an order statistic of his group (minimum effort games, median games, etc.). Presumably, even though it is only the order statistic that is required to determine one's payoff, observing the action profile of all of the opponents would help in overcoming coordination failure in repeated settings because it helps in belief formation and as a way to signal sophistication.

Brandts and Cooper (2006) investigate weak-link (minimum effort) games where they test full-feedback (players observing all the actions of opponents) versus limited feedback (players observing only the minimum effort of the game). They find that when starting from coordination failure, the use of full feedback improves subjects' ability to overcome coordination failure. Reinforcement learning certainly can not explain such a phenomenon. However, this can be easily explained by the fact that full feedback gives sophisticated players the ability to signal one's intent and presence while also assisting in teaching the adaptive learners.

Sophisticated players are aware of their ability to influence the future dynamics of the game through their current actions. Increasing the number of players in the population would weaken such ability and could hinder overcoming coordination failure. This prediction is validated through the results of Van Huyck, Battalio and Rankin (2007). The design of Van Huyck, Battalio and Rankin (2007) varies the order statistic used in the coordination game's payoff matrix, to either 2 or 4 (second lowest and fourth lowest), and the group size to either 5 or 7. Lower sized groups, with the same order statistic, are observed to be more likely to overcome coordination failure. This would be in accordance with the predictions of a sophisticated model since sophisticated players in a smaller group would have comparatively more influence than one from a larger group.

If the intention of a sophisticated player is to decrease the effort level it could be easily accomplished in a minimum effort game. However, in the same game, it would be harder to increase the level of effort in the group. Thus, if higher effort levels corresponded to higher payoff equilibria, overcoming coordination failure would be tougher in games with lower order statistics. Van Huyck, Battalio and Rankin (2007) find that groups playing games with high order statistics have a higher chance of converging to more efficient equilibria.

B. The Model

We start with some standard notation. There are n players indexed by i ($i \in \{1, ..., n\}$) playing the game repeatedly for T periods indexed by t ($t \in \{1, ..., T\}$). The strategy profile of player i, $S_i = \{s_i^1, ..., s_i^{m_i}\}$, consists of m_i discrete choices. Thus, $S = S_1 \times \cdots \times S_n$ is the strategy space of the game and $s = \{s_1, ..., s_n\} \in S$ is a strategy combination which consists of n strategies. A strategy combination for all players except i is represented by $s_{-i} = \{s_1, ..., s_{i-1}, s_{i+1}, ..., s_n\} \in S_{-i}$, which thus has a cardinality of $m_{-i} = \prod_{j=1, j \neq i}^{n} m_j$. Let s_{-i}^k denote the kth vector in S_{-i} and $\left(s_{-i}^k\right)_j$ denote the strategy of player j in the strategy vector s_{-i}^k. Let $s_i(t)$ denote the strategy chosen by player i in period t while $s_{-i}(t)$ the strategy vector chosen by the rest of the players. $\pi_i(s_i, s_{-i})$ is the scalar valued payoff function for player i and thus player i's payoff in period t would be $\pi_i(s_i(t), s_{-i}(t))$.

The model assumes two types of players playing the game. A fraction α of the population are sophisticated players while the rest are self-tuning EWA learners. Self-tuning EWA is chosen as the default since it is a hybrid of reinforcement and belief learning. In self-tuning EWA, most of the parameters are deterministically tuned by the experience of the players. Unlike EWA with its fixed parameters, self-tuning EWA has the ability to change with the dynamics of repeated play. Thus, if indeed players' learning behavior were different in the latter stages from the outset of the game, self-tuning EWA would tune itself to this difference.

1. Self-tuning EWA Learners

The self-tuning EWA model is derived from the Experience-Weighted Attraction (EWA) model of Camerer and Ho (1999). Each player i has a numerical attraction, $A_i^j(t)$, for strategy j after updating experience of period t. A logistic stochastic response function, $P_i^j(t+1) = \frac{e^{\lambda A_i^j(t)}}{\sum_{k=1}^{m_i} e^{\lambda A_i^k(t)}}$, determines the choice probabilities in period $t+1$. In the parametric EWA model, the attractions are updated as

$$A_i^j(t) = \frac{N(t-1) \cdot \phi \cdot A_i^j(t-1) + \left[\delta + (1-\delta) \cdot I\left(s_i^j, s_{-i}^j(t)\right)\right] \cdot \pi_i\left(s_i^j, s_{-i}^j(t)\right)}{N(t-1) \cdot \phi \cdot (1-\kappa) + 1}$$

$$(3.1)$$

The free parameters of the EWA model that are to be estimated are δ, λ, κ and ϕ. However, in the self-tuning EWA model, κ is set to zero and specific functions are defined for δ and ϕ. The parameter ϕ is replaced by $\phi(t)$, the change-detector function which is defined by

$$\phi(t) = 1 - \frac{1}{2}S_i(t) \qquad (3.2)$$

Here, $S_i(t)$ is the surprise index which captures the degree of change of the most recent observation from the historical average. Thus, its defined by

$$S_i(t) = \sum_{k=1}^{m_{-i}} \left(h_i^k(t) - I(s_{-i}^k, s_{-i}(t)) \right)^2 \tag{3.3}$$

Here, $h_i^k(t)$ $(= \frac{\sum_{\tau=1}^{t} I(s_{-i}^k, s_{-i}(t))}{t})$ is the historical frequencies of choices by other players. In the self-tuning EWA[3], the parameter δ of the EWA model is replaced by the attention function, $\delta_{ij}(t)$, defined by

$$\delta_{ij}(t) = \begin{cases} 1 & if \quad \pi(s_i^j, s_{-i}(t)) \geq \pi_i(t) \\ 0 & otherwise \end{cases} \tag{3.4}$$

As the choice probabilities could be found through the attractions calculated this way, the likelihood of observing a particular action profile by a self-tuning EWA learner could thus be calculated.

2. Sophisticated Learners

A sophisticated learner chooses the action that maximizes his payoff in the repeated game. With probability α a person in the population is sophisticated. Under rational expectations, a fraction α of the population is expected to be sophisticated. In a pool of players, if $\alpha = 0$, then all players are adaptive learners. If $\alpha = 1$, and the

---

[3]Refer to Ho, Camerer & Chong (2007) for a detailed account of the self-tuning EWA model.

sophisticated players are aware of this fact, it will result in a variant of an agent quantal response equilibrium (AQRE). The sophisticated players are aware that a fraction of the population is sophisticated as they are. The fact that sophisticated players take the action in the current period that maximizes the payoff of the entire repeated game implies that they take into account the effect of their present action on the future behavior of adaptive learners. Thus, they could become teachers to the adaptive learners.

Adaptive learners choose the action that maximizes the payoff in the next period. Since sophisticated players are maximizing the payoff of the entire length of the repeated game, they need not take the action that maximizes the payoff in the next period. Sophisticated players take an action so as to influence the adaptive learners in the expectation of getting higher payoffs in the future. This type of behavior, where sophisticated players take an action which does not give them the highest payoff in the next period, but do so to influence adaptive learners, will be referred to as teaching behavior. Since teaching does not yield the higher payoff at present, teaching is costly in the short run. So we will only observe teaching if the sophisticated players think it is worthwhile in the long run.

If a cohort tends to drift towards a Pareto inferior equilibrium, there might be a required critical fraction, α_{cr}, which needs to be sophisticated in order to effectively teach the rest of the population to prevent converging to that inferior equilibrium. This is so because a small fraction of teachers might not be enough to influence the

large adaptive learning population to turn the current trend. Thus, the belief that a particular sophisticated player has about the fraction of the sophisticated players in the population plays a big role in his behavior. If sophisticated player i's belief about the fraction of sophisticated players, α_i, is less than this critical fraction, α_{cr} ($\alpha_i < \alpha_{cr}$), then sophisticated player i might not use a teaching strategy since he would not believe that there are enough sophisticated players to have a significant influence on the adaptive learners' behavior. However, if player i believes there is a higher fraction of sophisticated players ($\alpha_i > \alpha_{cr}$), he might use a teaching strategy since he would believe that there are enough teachers to overcome coordination failure.

The sophisticated players are not aware of the identity of the other sophisticated players. At the beginning of the game, they are aware that in general a fraction α of the population is sophisticated. This is the probability the sophisticated players associate with each player being sophisticated. The sophisticated players update their belief of the fraction of sophisticated players in the population depending on the outcomes observed in a Bayesian manner.

Unfortunately, as teaching is costly in the short-run, if a sophisticated player believed that there were more than enough players to teach adaptive learners, he might be tempted not to teach and try to free-ride and enjoy the benefits of others teaching the adaptive learners. If we assume that players start the game with the same priors, they would be starting with the same beliefs. Therefore, by symmetry, they should have the same inclination for each action at the outset, and so it cannot

be the case that a certain sophisticated player realizes that he could free-ride while the others do not. The likely outcome is that the sophisticated players would choose an action path that teaches the adaptive learners while making free-riding not worthwhile for other sophisticated players. It would be a certain equilibrium in sophisticated players' actions. That is, given the identical action paths of the sophisticated players, a sophisticated player's best response would be to choose that action path itself.

Let $\alpha_{ij}(t)$ be the probability with which sophisticated player i anticipates player j to be sophisticated at time period t by observing the history of play. Since the model assumes only two types of players, $(1 - \alpha_{ij}(t))$ would be player i's probability that player j is a self-tuning EWA learner. Let H_t be the history of the game up to time period t. Let $L^j_{Soph}(H_t)$ be the probability (likelihood) of observing the history of play of player j in history H_t if he was sophisticated. Let $L^j_{EWA}(H_t)$ be the probability (likelihood) of observing the sequence of play by player j in history H_t if he was a self-tuning EWA learner. Since, by assumption, player j has to be either a sophisticated or a self-tuning EWA learner, player i's belief of player j being sophisticated at time t, $\alpha_{ij}(t)$, will be, by Bayes' rule,

$$\alpha_{ij}(t) = \frac{\alpha L^j_{Soph}(H_t)}{\alpha L^j_{Soph}(H_t) + (1 - \alpha) L^j_{EWA}(H_t)} \tag{3.5}$$

Sophisticated players have an attraction to each action depending on their belief about its expected payoff. Letting $A^j_i(t)$ denote the attraction player i has for strategy

j in period t, it is defined by,

$$A_i^j(t) = \sum_{k=1}^{m_{-i}} \tilde{P}_{-i}\left(s_{-i}^k, t+1\right) \left[\pi\left(s_i^j, s_{-i}^k\right) + V_i(t+1)\right].\qquad (3.6)$$

$\tilde{P}_{-i}\left(s_{-i}^k, t+1\right)$ is the probability player i associates with observing s_{-i}^k in period $t+1$, while $V_i(t+1)$ is the ex ante value of future payoffs the sophisticated players believes to be attainable at time t given his current strategy choice. Letting $\tilde{P}_j(s_j, t+1)$ be the subjective probability of observing s_j by player j in period $t+1$, we get the relation:

$$\tilde{P}_{-i}\left(s_{-i}^k, t+1\right) = \prod_{j=1, j\neq i}^{n} \tilde{P}_j\left(\left(s_{-i}^k\right)_j, t+1\right).\qquad (3.7)$$

Given that $\alpha_{ij}(t)$ is the probability player i associates player with j to be sophisticated, let $P_{Soph}(s_j, t+1)$ and $P_{EWA}(s_j, t+1)$ be the probability of observing s_j in period $t+1$ by a sophisticated player and an adaptive learner respectively. Then $\tilde{P}_j(s_j, t+1)$ can be specified as

$$\tilde{P}_j(s_j, t+1) = \alpha_{ij}(t)\, P_{Soph}(s_j, t+1) + (1 - \alpha_{ij}(t))\, P_{EWA}(s_j, t+1)\qquad (3.8)$$

Thus, a player can deduce the probability of observing a strategy profile in the popu-

lation by tracking the probability with which each player is a sophisticated learner or
an adaptive learner. Going back to eqn. 3.6, the term $V_i(t+1)$, which represents the
present value of future payoffs from the current action, can be specified recursively as

$$V_i(t) = \max_{J_t = \left\{ s_i^{j_t}, \ldots, s_i^{j_T} \right\}} \sum_{k=1}^{m_{-i}} \left(\tilde{P}_{-i}\left(s_{-i}^k, t\right) \left[\pi\left(s_i^{j_t}, s_{-i}^k\right) + V_i(t+1 \mid S(t)) \right] \right). \qquad (3.9)$$

$J_t = \left\{ s_i^{j_t}, s_i^{j_{t+1}}, \ldots, s_i^{j_T} \right\}$ is an action sequence from the current period to the end
of the game that is available to the sophisticated player. This is the sequence of actions
the sophisticated player believes would result in him obtaining the highest payoff for
the game. It is the presence of this value function that makes a player sophisticated.
The function enables a player to evaluate the possible benefits of teaching, free-riding,
myopically best responding and so on, given his current beliefs about his opponents.

The attraction to each action computed through the use of the recursive value
function determines sophisticated player i's choice probability according to the logit
rule,

$$P_i\left(s_i^j, t+1\right) = \frac{e^{\lambda_s A_i^j(t)}}{\sum_{k=1}^{m_i} e^{\lambda_s A_i^k(t)}} \qquad (3.10)$$

Given the rules of the sophisticated player's actions the model can now be esti-

mated through a maximum likelihood procedure. The likelihood L, of observing the actions of players in the repeated game thus would be

$$L = \prod_{i=1}^{n} \left[\alpha \prod_{t=1}^{T} P_{Soph}\left(s_i\left(t\right)\right) + \left(1 - \alpha\right) \prod_{t=1}^{T} P_{EWA}\left(s_i\left(t\right)\right) \right] \qquad (3.11)$$

Estimations of the MLE were done using GAUSS with the help of the MAXLIK procedure. As explained, a sophisticated player contemplates all possible strategy combinations in future periods however unlikely they maybe. Using combinatorics one could see that this could mean looking into millions of possibilities each period for each player depending on the number of periods left and the number of actions available to the players. To resolve this problem, several assumptions were made so that a sophisticated player would narrow the possibilities to those that were more likely to occur without hindering the estimation process[4].

C. Data

The data used to find evidence of sophisticated learners is from the continental divide experiments by Van Huyck, Cook & Battalio (1997). Undergraduate economics students at Texas A&M University played game G in Table V repeated for 15 periods.

---

[4]Refer to the Appendix A for an overview and discussion of these assumptions.

Each cohort consisted of 7 subjects and there is data available for 8 cohorts.

The stage game G has many interesting features. It has two symmetric equilibria, one in which all players are choosing action three while the other is when all the players are choosing action twelve. The equilibrium in which all players are choosing action 3 yields each player 60 cents each period, while the other equilibrium yields 112 cents each period. Thus, the high action equilibrium Pareto dominates the low action equilibrium. If players could choose the equilibrium (instead of their action), every player would choose the high action equilibrium. However, due to risk considerations, precedents set by initial play, and so on players may be unable to coordinate on the efficient equilibrium. An interesting feature of game G is that most learning rules have two basins of attractions. Medians of {1, 2, 3, 4, 5, 6, 7} are attracted to the inefficient equilibrium while medians of {8, 9, 10, 11, 12, 13, 14} are attracted to the efficient equilibrium.

Table V. Payoff Table of Game G

Median choice	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	45	49	52	55	56	55	46	−59	−88	−105	−117	−127	−135	−142
2	48	53	58	62	65	66	61	−27	−52	−67	−77	−86	−92	−98
3	48	54	60	66	70	74	72	1	−20	−32	−41	−48	−53	−58
4	43	51	58	65	71	77	80	26	8	-2	-9	−14	−19	−22
5	35	44	52	60	69	77	83	46	32	25	19	15	12	10
6	23	33	42	52	62	72	82	62	53	48	43	41	39	38
Your 7	7	18	28	40	51	64	78	75	69	66	64	63	62	62
Pick 8	−13	-1	11	24	37	51	69	83	81	80	80	80	81	82
9	−37	−24	−11	3	18	35	57	88	89	91	92	94	96	98
10	−65	−51	−37	−21	-4	15	40	89	94	98	101	104	107	110
11	−97	−82	−66	−49	−31	-9	20	85	94	100	105	110	114	119
12	−133	−117	−100	−82	−61	−37	−5	78	91	99	106	112	118	123
13	−173	−156	−137	−118	−96	−69	−33	67	83	94	103	110	117	123
14	−217	−198	−179	−158	−134	−105	−65	52	72	85	95	104	112	120

D.   Results

Results are reported in Table VI. The column 'Sophisticated Learning' reports the parameter estimates of the sophisticated model, while the 'Self-tuning EWA' column reports the results when the data is fit assuming that all the players are adaptive learners. Results indicate evidence in favor of the sophisticated model[5]. Sixteen

---

[5]$\chi^2(2) = 90$, p-value = 0.00.

percent of the population are characterized as playing sophisticatedly. The λ of the groups of players indicate the ability of the players to best respond given their attractions to each action. The λ of the sophisticated players are significantly higher than that of the adaptive learners indicating that the sophisticated players are better at choosing their desired action than the adaptive learners.

Table VI. Parameter Estimates Using Van Huyck, Cook & Battalio (1997)

	Self-tuning EWA	Sophisticated Learning	Soph. Learning w. discounting
λ_{EWA}	5.37*** (0.0011)	6.00*** (0.229)	6.24*** (0.000)
α		0.160* (0.111)	0.341*** (0.0889)
λ_{Soph}		15.0*** (2.59)	11.1*** (1.54)
δ			0.720*** (0.0959)
Log Likelihood	-1357	-1312	-1302
No. of observations	840	840	840

* 10% significance, ** 5% significance, ***1% significance

Discounting future payoffs is a common practice in economics. The model used could also be modified such that in eqn. 3.6, the value function is discounted by $0 \leq \delta \leq 1$.

$$A_i^j(t) = \sum_{k=1}^{m_{-i}} \tilde{P}_{-i}\left(s_{-i}^k, t+1\right)\left[\pi\left(s_i^j, s_{-i}^k\right) + \delta V_i(t+1)\right]. \qquad (3.12)$$

If $\delta = 0$, it would correspond to a fictitious player while $\delta = 1$ would represent the original model. Any other value of $\delta, 0 < \delta < 1$, would represent a sophisticated player discounting his predicted future payoffs in calculating his attractions to each action. Such a model was estimated and its parameter estimates are reported in the last column of Table VI. The estimates indicate that the presence of the discounting factor fits the data better[6]. δ could indeed be capturing the fact that sophisticated players discount their predicted future payoffs in calculating their attractions.

The λs of the modified model are consistent with the original sophisticated model. Another observation is that the λs are consistently higher than in the basic adaptive learning model. This suggests that there are two types of players evaluating their attractions to actions differently[7].

---

[6] $\chi^2(1) = 20$, p-value $= 0.00$.

[7] If players were (close to) best-responding using different rules of finding attractions to each action, a correctly specified model should be able to deduce the fractions of players using each rule. Also, the λ estimates of such a model would be expected to be high. If indeed players were using different rules, estimating the model using a representative model would result in low values of λ, since there would be too much 'noise' in the data which would correspond to $\lambda = 0$ players.

E.   Discussion

The results provide evidence of sophisticated players. Adding sophistication into the model increases the number of free parameters to be estimated by two, one being the fraction of sophisticated players in the population and the other free parameter is the sophisticated player's precision parameter.

In the game investigated, players are modeled to strategize and teach using the actions of the stage game. However, depending on the game being played, one can think of many kinds of complicated, sophisticated strategy combinations or rules players could learn to play. A nice example of such rule learning is learning to play tit-for-tat in a repeated prisoner's dilemma game. Such behavior will require sophistication and the degree of sophistication in such a scenario could be more complicated than in the model used here.

The fraction of sophisticated players estimated in the population is lower than the estimates obtained in Chong et. al. (2006). This could be due to several different features of the game used here. The continental divide game was chosen since it had many actions available to the players, which meant that it is easier to distinguish sophisticated players from the rest. However, the continental divide game uses the median as the statistic to determine payoffs. It is difficult to influence the median of seven choices[8]. Sophisticated players distinguish themselves when they can influence

----

[8]For example, it is possible that any given player might not be able to make any influence at all, as when the choices are { 1, 2, 3, 3, 3, 4, 5}.

other players' actions by influencing other players' payoffs. If a sophisticated player cannot influence the median (and/or believes so) his behavior could resemble that of an adaptive learner. Having somewhat lost that ability it is quite possible that the behavior of many 'would be sophisticated' players, believing they could not influence the dynamics in the current setting, would resemble that of adaptive learners. Therefore, this estimate is an indication of the fraction of the population who still believed they could effect the trend of the game through their actions even in the median action game.

Decreasing the number of players in the group increases the influence a particular member has on the order statistic of the group, and thus would increase the influence of a sophisticated player. It would be interesting to see if it is more likely for players to reveal their sophistication when the influence they have is greater.

In the model used, the same degree of sophistication is assumed for all the sophisticated players. Indeed the level of sophistication might vary among each individual due to factors such as cognitive capacity, experience and a model could be specified where sophisticated players differ in their level of sophistication similar to that of (Stahl (2000), Stahl & Wilson (1995)). However, given that the primary objective was to find evidence of sophisticated behavior and given the low estimates of the fraction of sophisticated players in the population such a differentiation was not warranted.

Since the model can be extended or refined in many ways, it opens the door for many further investigations. Possible investigations include deciphering sophisticated

players beliefs about their fellow opponents beliefs and sophisticated players belief about their influence on others behavior in the game. It would also be interesting to investigate whether sophistication increases with experience and whether certain experiences increase one's propensity to be sophisticated.

In future studies it would also be important to design experiments in which the model predicts a marked difference between sophisticated behavior and adaptive behavior. This would make studying sophisticated learning much easier.

CHAPTER IV

LEARNING ABOUT VARIABLE DEMAND IN A DUOPOLY

A. Introduction

This chapter investigates the dynamics involved in a setting where firms of an oligopoly have to learn their demand environment while taking into account the competitive nature of the firms. It extends the work of Rustichini and Wolinsky (1995), where they deal with a monopoly that is uncertain about the demand it faces and learns about the fluctuating demand over time through its pricing experience. The related literature in monopoly demand learning include the work of Aghion, Bolton, Harris and Jullien (1991), McLennan (1984), Easley and Kiefer (1988), Balvers and Cosimano (1990, 1993), and Nyarko and Olson (1991). This literature mostly analyze learning a fixed demand curve and thus the learning process boils down to the initial uncertainty about it making them similar to the bandit problems studied by Berry and Fristedt (1985) and Rothschild (1974).

In the model discussed in this chapter, demand is not a constant and it changes over time in a Markovian fashion as in Rustichini and Wolinsky (1995). The state of demand changes with time and so the firms have to devise a strategy to check (learn) their demand environment and thus the process of learning demand never ceases.

In the case of a duopoly considered here, the firms also have to take the strategic implications of the setting into consideration.

The optimal pricing strategy for oligopolies has been studied extensively. The impact of varying demand (business cycles) on collusive behavior has been of interest in industrial organization and macroeconomics. First, largely based on pre-World War II case studies, it was believed that collusion was more difficult during economic downturns, a position weakly supported by Suslow (1988) through data from many case studies. Formal theoretical studies done by Rotemberg and Saloner (1986) under a setting where demand is subject to (observable) i.i.d. shocks imply that it is more difficult to collude during booms (when the level of demand is high). This is due to the simple reason that there are more economic profits to be earned from deviating during a boom than when demand is low, since, in the model, the expected future profits from punishment (from the grim trigger strategy) is the same regardless of the level of demand in the period of defection.

Haltiwanger and Harrington (1991) extend the fluctuations in demand so that demand follows a predetermined cycle. From their simulations, they conclude that the oligopolies have the highest tendency to deviate from collusion when demand is falling and a lower incentive to deviate when demand is rising. Thus, in the equilibrium pricing strategy under collusion, given the same level of demand, price would be higher when demand is rising than when demand is falling. The reason for this conclusion can be attributed to the fact that there are more economic profits to be

earned by colluding when demand is rising which increases the incentive to collude and possible prices that could be sustained. On the other hand, incentive to collude decreases when demand falls since expected future profits from collusion are now relatively bleak which imply that prices might have to be shaded to be sustained. One feature of their demand function is that it is fully deterministic. That is, though demand is cyclic, firms are fully aware of the future demand at any given time with certainty.

This chapter extends both the literature on monopolistic learning by studying the impact of introducing an additional firm, and analyzes dynamic models of oligopolies by introducing an environment where firms learn the fluctuating demand through their pricing policy.

B.   The Model

The model is an extension of Rustichini and Wolinsky (1995), who considered learning by a monopoly. The model proposed has two competing firms in the market: a duopoly market. By adding another firm to the setup, firms have to consider the pricing policy for learning consumer demand characteristics against the pricing policy under competition with the other firm. In the setup, demand can only be learned through increasing prices. Thus, the desire to learn would drive prices up while

considerations of possible free-riding and competition would drive prices down. Since the two factors affect pricing in opposing directions, it is interesting to see what the equilibrium strategy would be.

The process is an infinitely repeated game where time is discrete, labeled $t = 1, 2, ...$. The demand the duopoly faces varies over time in a Markovian fashion. In each period the duopoly faces a demand of two-units (normalized to two for convenience) where the buyer's reservation price is d_t. At the beginning of the period, the two firms, firm 1 and firm 2, simultaneously set their prices p_{1t} and p_{2t}. The consumer will buy his two-units from the firm quoting the lowest price if it is less than or equal to his reservation price. If the prices quoted by the two firms are equal and less than or equal to the reservation price, the buyer will buy one-unit from each of the two firms. Thus, only if $min\,(p_{1t}, p_{2t}) \leq d_t$ will the two units be sold. If we let I_{it} be the number of units sold by firm i at time t, then the revenue for firm i at time t will be $I_{it}p_{it}$. The firm's marginal cost of producing the good is normalized to zero.

The reservation price, d_t, is assumed to take two values 1 and $D > 1$. It follows a Markov process with transition probabilities[1]

$$Prob[d_{t+1} = 1|d_t = D] = Prob[d_{t+1} = D|d_t = 1] = \alpha.$$

---

[1]Also represented in Figure 2.
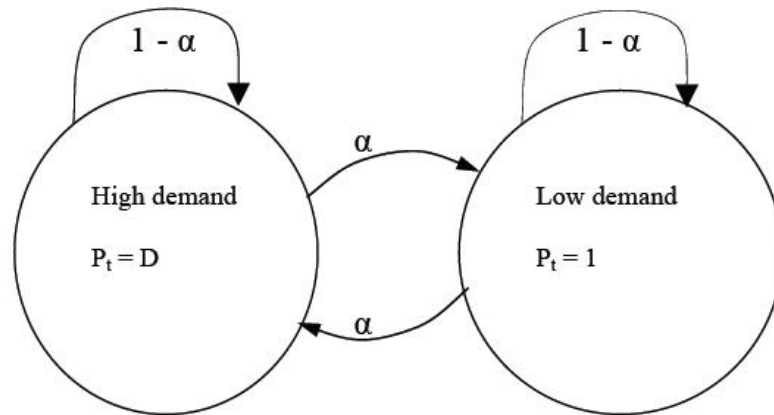
Fig. 2. Transition of Demand States

Here α is assumed to be less than $\frac{1}{2}$ which makes it more likely for the next period's demand to be the same as the current's. Thus, this increases the value of getting to know the current demand. The firms have perfect monitoring regarding pricing and sales of the other firm, so their beliefs on demand will be the same. If we let w_t denote the probability with which the firms believe that $d_t = 1$ at time t, it will evolve as:

$$
w_{t+1} = \begin{cases} \alpha & if \quad min\,(p_{1t}, p_{2t}) \in (1, D] \quad \& \quad \sum_i I_{it} = 2 \\[2em] 1 - \alpha & if \quad min\,(p_{1t}, p_{2t}) \in (1, D] \quad \& \quad \sum_i I_{it} = 0 \\[2em] (1 - \alpha)\,w_t + \alpha\,(1 - w_t) & if \qquad min\,(p_{1t}, p_{2t}) \leq 1 \end{cases}
$$

Properties of the Markov process

1. The stationary probability of the above Markov process is $(\frac{1}{2}, \frac{1}{2})$. Therefore, if the process's state is not observed for N consecutive periods, $w_{t+N} \to \frac{1}{2}$.

2. Also, since $\alpha < \frac{1}{2}$, if $w_t > (<)\frac{1}{2}$ and the process's state is not observed for N consecutive periods, $w_{t+N} > (<)\frac{1}{2}$.

Firm i's discounted profit is $\sum \delta^t p_{it} I_{it}$, where $\delta < 1$ is the discount rate, for a sequence of prices p_{it} and sales realizations I_{it}. At the beginning of period t the firm knows the history $h_t = [(p_1, I_1), ..., (p_{t-1}, I_{t-1})]$, where $p_t = (p_{1t}, p_{2t})$ and $I_t = (I_{1t}, I_{2t})$. Firm i chooses a pricing policy $p_{it}(h_t)$ so as to maximize $E[\sum \delta^t p_{it} I_{it}]$.

At any time t, the firms engage in a price competition game. The only equilibrium of this stage game is for firms to quote a price of zero (the marginal cost), and this will result in them obtaining zero profits. However, in repeated play, many equilibria are attainable. In particular, a grim trigger strategy profile with positive pricing

(and profits) through tacit collusion is sustainable when deviating results in minmax payoffs of zero (the equilibrium of the stage game).

Claim 1. In repeated play, for $\delta \geq \delta^{crit}$, there exists an equilibrium pricing policy p^* such that $D \geq p_{it}^*(h_t) = p^*(h_t) \geq 1$ for all i (therefore $I_{it} = I_t$).

Proof.

$2p^*(h_t)$ is the maximum possible profit from deviating at time t. To sustain the equilibrium, future discounted profits from collusion must be higher than that of the profits obtained from deviation. So,

$$2p_{it}^* \leq \sum_{T=t}^{\infty} \delta^{T-t} p_{iT}^* I_{iT} \Rightarrow p_{it}^* \leq \sum_{T=t+1}^{\infty} \delta^{T-t} p_{iT}^* I_{iT}$$

which is satisfied for all $\delta \geq \delta^{crit}$ since $p_t^* \leq D$ and the right hand side is an infinite sum with $p_t^* \geq 1$ for all t.

In the equilibrium pricing policy suggested above, the symmetric pricing imply that the firms will sell no units or a maximum of one unit each at any time.

Claim 2. The optimal policy is characterized by a cutoff belief W.

If $w_t \leq W \Rightarrow D \geq p_t > 1$. If $w_t > W \Rightarrow p_t = 1$.

This claim follows from Claim 1 of Rustichini and Wolinsky (1995). The result implies that if it is optimal to experiment with prices at the current belief of high demand, then it should also be optimal to price experiment for any belief which has a higher belief of high demand. Also, if it is not optimal to price experiment at the current belief of high demand, then it should also be not optimal to price experiment for any lower belief of high demand.

Therefore, three possible pricing schemes or policies result from W.

$$W \leq \frac{1}{2} \Rightarrow p_t = \begin{cases} 1 & \delta \geq \frac{1}{2} \\ 0 & otherwise \end{cases} \quad N = \infty$$

$$W \geq (1-\alpha) \Rightarrow p_t = \begin{cases} p_H & w_t = 1-\alpha \\ p_L & w_t = \alpha \end{cases} \quad N = 0$$

$$\frac{1}{2} < W < (1-\alpha) \Rightarrow p_t = \begin{cases} 1 & W < w_t \\ p_H & W \geq w_t > \alpha \\ p_L & w_t = \alpha \end{cases} \quad 0 < N < \infty$$

In the first scheme, for the given D, α and δ, it is never optimal to experiment, which may be due to two reasons

1. Even for a monopoly, for the given D, α and δ, whenever $w_t > \frac{1}{2}$, the future expected profits from experimentation is less than the future discounted profits from pricing 1 forever. (Note: $w_t = 1 - \alpha > \frac{1}{2}$ when demand is observed to be low in the previous period. By the second property of the Markov process, if no experimentation was done for N consecutive periods after $w_t = 1 - \alpha$, it will still be the case that $w_{t+N} > \frac{1}{2}$).

2. Even if demand is found to be high through experimentation, the firms cannot sustain a sufficiently high price to make experimentation worthwhile.

In the second scheme, D, α and δ are such that it is optimal to experiment forever. In the third scheme can be easily characterized by N, where

$$N = \ \arg_N \max \psi_N \quad s.t. \quad \psi_N = \Pr\left[d_{t+N+1} = 1 | d_t = 1\right] \leq W$$

Standard calculations yield

$$\psi_N = \Pr\left[d_{t+N+1} = 1 | d_t = 1\right] = \left[1 + (1 - 2\alpha)^{N+1}\right]/2 \ \leq W \text{ (See Feller, 1968)}.$$

Thus, N is the minimum number of periods pricing at 1 (no experimentation) is to be done after low demand is observed, at which time $w_t \leq W$, the cut-off for experimentation will be achieved. Therefore, after N periods of pricing at 1, the firms will charge p_H and with that, if demand is found to be high, p_L will be charged

until low demand is observed. After low demand is observed (both at pricing p_H or p_L), firms will start pricing at 1 for N periods and this process will continue. Also, note that the pricing scheme can be characterized by N, where the first two schemes correspond to the special cases of $N = 0$ and $N = \infty$ respectively.

1. Solving for p_H and p_L

For the monopoly case (Rustichini and Wolinsky 1995), $p_H = p_L = D$, but in the case of the duopoly, considerations of the possibility of deviation could lower the optimal collusive price in the given environment as in Haltiwanger and Harrington (1991). p_H and p_L are such that it is not optimal for the firms to deviate at any time.

By Claim 2, since the optimal collusive pricing policy at a given is time is determined by w_t, let $P(w_t)$ be the price charged by the firms for belief w_t. Let Y be the expected future discounted profits (at the end of a period) after a price greater than one was accepted and let Z be the future discounted profits (at the end of a period) after a price great than one was rejected. Therefore, Z is the expected profits generated starting with N consecutive periods of quoting the price to be 1, while Y is the expected profits generated from quoting the price p_L in the next period.

$$Y = \delta \left[(1 - \alpha) \left[P(\alpha) + Y \right] + \alpha Z \right]$$

$$Z = \frac{\delta - \delta^{N+1}}{1 - \delta} + \delta^{N+1} \psi_N Z + \delta^{N+1} (1 - \psi_N) (P(\psi_N) + Y)$$

Claim 3. Given $p_t \geq 1$ for all t in the optimal pricing policy, the future expected profits must be greater than or equal to that of an infinite sequence of one-unit sales priced at 1 for all periods.

Proof. (By contradiction)

If it is not the case, then the firms would be better off in the equilibrium where they price at 1 and get a guaranteed sale and obtain the expected profits of an infinite sequence of unit sales at price 1, which contradicts that the given was an optimal policy.

It follows that the firms would only experiment if in the long run they get at least as much profits as from pricing at 1 forever.

It follows from Claim 3 that $Z \geq \frac{\delta}{1-\delta}$ and $(1 - \alpha)P(\alpha) \geq 1$. Therefore it can be seen that $Y \geq Z$.

Claim 4. $1 < p_L = P(\alpha) = \min\left[D, Y + \frac{\alpha}{1-\alpha}Z\right], p_H = P(\psi_N) = \min\left[D, Y + \frac{\psi_N}{1-\psi_N}Z\right]$

At any point, the firms must have higher expected profits through collusion than

through deviating for the pricing profile to be sustainable.

Case 1. Condition for firms not to deviate in periods when $p_t = P(\alpha) = p_L$ (that is, after $p_{t-1} > 1$ was accepted)

$$2\left(1-\alpha\right)P\left(\alpha\right) \leq \left(1-\alpha\right)\left(P\left(\alpha\right)+Y\right)+\alpha Z$$

$$\Rightarrow \left(1-\alpha\right)P\left(\alpha\right) \leq \left(1-\alpha\right)Y+\alpha Z$$

$$\Rightarrow P\left(\alpha\right) \leq Y + \frac{\alpha}{1-\alpha}Z$$

Case 2. Condition for firms not to deviate in periods when $p_t = P(\psi_N) = p_H$ (that is, the period of experimentation after N periods of pricing at 1)

$$2\left(1-\psi_N\right)P\left(\psi_N\right) \leq \left(1-\psi_N\right)\left(P\left(\psi_N\right)+Y\right)+\psi_N Z$$

$$\Rightarrow \left(1-\psi_N\right)P\left(\psi_N\right) \leq \left(1-\psi_N\right)Y+\psi_N Z$$

$$\Rightarrow P\left(\psi_N\right) \leq Y + \frac{\psi_N}{1-\psi_N}Z$$

Firms want to maximize expected profits given the above conditions for p_H and p_L which is to quote the maximum possible price. Since, for experimentation, it is required that p_H and $p_L \leq D$ at all times, it follows that

$$1 < p_L = P(\alpha) = \min \left[D, Y + \frac{\alpha}{1 - \alpha} Z \right]$$

$$p_H = P(\psi_N) = \min \left[D, Y + \frac{\psi_N}{1 - \psi_N} Z \right]$$

$$\Rightarrow 1 < p_L \leq p_H \leq D$$

Therefore, the possible optimal pricing schemes will be the following.

The first scheme is as follows:

$$W \leq \frac{1}{2} \Rightarrow p_t = \begin{cases} 1 & \delta \geq \frac{1}{2} \\ 0 & otherwise \end{cases} \quad N = \infty$$

In this scheme, for the given D, α and δ, W is such that experimentation is never possible since firms will not be able to collude on a higher price.

The second scheme is as follows:

$$W \geq (1-\alpha) \Rightarrow p_t = \begin{cases} \min\left[D, Y + \frac{\psi_N}{1-\psi_N}Z\right] & W \geq w_t = 1 - \alpha \\ \\ \min\left[D, Y + \frac{\alpha}{1-\alpha}Z\right] & w_t = \alpha \end{cases} \quad N = 0$$

In this scheme, for the given D, α and δ, W is such that experimentation is always worthwhile.

The third scheme is as follows:

$$\frac{1}{2} < W < (1-\alpha) \Rightarrow p_t = \begin{cases} 1 & W < w_t \\ \\ \min\left[D, Y + \frac{\psi_N}{1-\psi_N}Z\right] & W \geq w_t > \alpha \\ \\ \min\left[D, Y + \frac{\alpha}{1-\alpha}Z\right] & w_t = \alpha \end{cases} \quad 0 < N < \infty$$

This is the scheme where experimentation ceases for a finite number of periods following the observation of a low demand period. Figure 3 contrasts the pricing policy of such a case with that of a monopoly with the same cut-off belief W, $\frac{1}{2} < W < (1-\alpha)$. If low demand was observed in the last period, then $w_t = 1 - \alpha$ and price would be set at one. Since nothing would have been learnt about the state of demand by pricing at one, w_t would tend towards the stationary probability of $\frac{1}{2}$. Thus, the firms would continue pricing at one until w_t falls below the threshold W. In the presented figure, firms are to price at D (p_H) after crossing the threshold. If

demand was found to be high, the firms would set a price of p_L and would continue
to price at p_L as long as high demand is observed. At any point, if low demand is
observed, then in the next period, w_t will equal $1 - \alpha$ and then firms will switch to
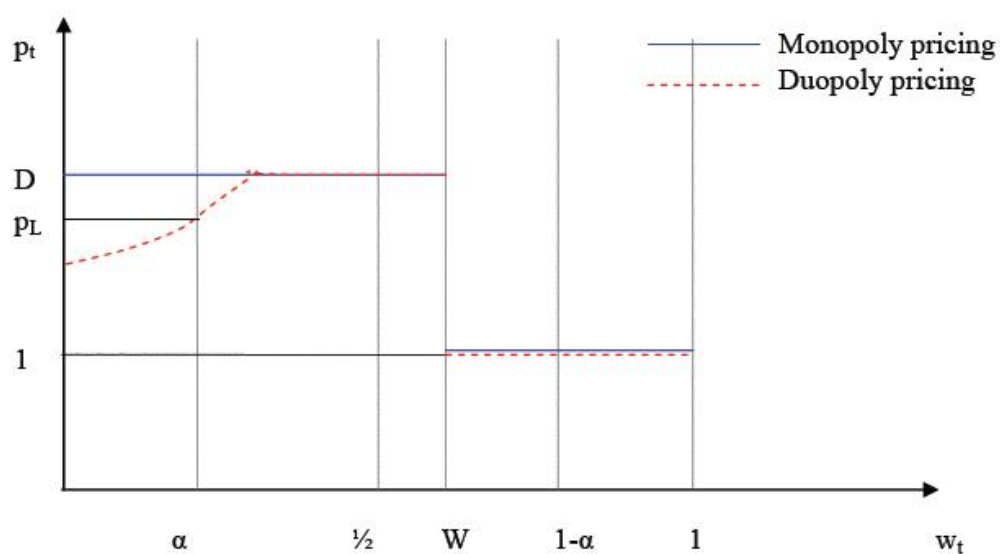pricing at one.

Fig. 3. A Possible Price Profile for a Duopoly and a Monopoly with the Same Cutoff
W, $\frac{1}{2} < W < (1 - \alpha)$

C. Discussion

Observations:

1. When the belief of low demand is high the duopoly is able to collude perfectly

on pricing low (at 1). The firms set a low price (at 1) for N consecutive periods after observing a low demand period just as in the monopoly case. The belief of low demand being high corresponds to a recession in a business cycle. Therefore, this suggests that firms hold off experimentation for some length of time after observing a recession.

2. After pricing low (at 1) for a certain number of periods, the belief of low demand crosses a certain threshold, W. At this point the firms start to experiment. There is a sudden jump in the prices charged at this point (after N periods of pricing at 1), and this price that the duopoly charges might be higher than that would charge after they find demand to be high. That is, in periods with price greater than one, the possible colluding price weakly increases with the belief in low demand. This observation is quite startling at first. It is due to fears of cheating that the firms would not be able to collude on higher prices when the belief of high demand is high. However, there are many interesting implications of pricing behavior that would come out of this observation.

This implies that there is a threshold belief in future demand at which firms drastically change their pricing profile and experimentation in anticipating a boom in the business cycle. On realizing a boom however, the firms vigorously compete.

Therefore, firms are able to collude perfectly at times when optimal price policy is to charge p = 1. This is because these are the periods with the lowest profits from deviation. When charging a price greater than one, the optimal price weakly

increases with the belief in low demand. This is because these are the periods when the conditions for deviating from the pricing policy would be binding. Therefore, when the probability of higher demand is high, the price has to be shaded to decrease the possible profits from deviation.

These predictions are quite different from predictions from the oligopoly pricing literature. In Rotemberg and Saloner (1986) the results came about because high demand and low demand was equally likely which made future demand always unrelated to the past and current market conditions. Haltiwanger and Harrington (1991) extend the fluctuations in demand so that demand follows a predetermined cycle. However, using a predetermined cycle also has drawbacks. Though business cycle seem to consistently go through booms and busts, their occurrences are not predetermined. Haltiwanger and Harrington (1991) derive that the oligopolies have the highest tendency to deviate from collusion when demand is falling and a lower incentive to deviate when demand is rising. This is quite simply due to the predetermined cycle they are in.

This chapter extends these models by making demand change in a Markovian fashion and avoids the drawbacks of the business cycle model in previous work on oligopoly pricing. Therefore, the observations and conclusions drawn should be more realistic.

CHAPTER V

SUMMARY

A.   Implementing First-Best Allocations in the Principal-Agents Model

This research extends the work done by Piketty (1993). Piketty shows that when the number of agents is finite and the characteristics profile of the population is known, the principal can design a game whose unique Bayesian Nash Equilibrium (through iterative elimination of strictly dominated strategies) yields First-Best allocations. There is no loss in efficiency and the principal is able to extract all the rent in this case. This research proposes the Sign-Up Game which the principal can use in certain other incomplete information settings (NIIS) to implement first-best outcomes while extracting all the rent.

The Sign-Up Game offers an avenue to overcome loss of efficiency in adverse selection problems in certain settings. This research shows that for the ϵ-Sign-Up Game, if the principal's information is a setting of the NIIS, rational agents choosing strategies that survive iterative elimination of weakly dominated strategies result in a unique Nash Equilibrium. There are many properties of the Sign-Up Game which make the result very unique. The unique equilibrium is self-selecting, ex-post preferred to status quo, independent of risk attitudes, truth-revealing and Pareto

efficient. The Sign-Up Game and its approach to the principal-agent problem offers new insight into mechanism design. It opens up the possibility of improving upon second-best solutions in many finite principal-agent settings.

B. Sophisticated Learning and Learning Demand

A sophisticated learning model was developed which assumed that a fraction of the population was adaptive learners. The rest of the population (the sophisticated learners) was aware of this fact and took these adaptive learners' behavior into account. Thus, when playing a repeated game the sophisticated learners would in each period choose a strategy that would maximize their payoff for the entire length of the game.

The sophisticated learning model was tested using the data from the continental divide game experiments by Van Huyck, Cook & Battalio (1997). When tested against pure adaptive learning (self-tuning EWA), the results are in favor of sophisticated learning and 34% of the population was found to be sophisticated learners.

The final research investigates the dynamics involved in a setting where firms of an oligopoly have to learn their demand environment while taking into account the competitive nature of the firms. It extends both the literature on monopolistic learning by studying the impact of introducing an additional firm, and analyzes dynamic models of oligopolies by introducing an environment where firms learn the fluctuating

demand through their pricing policy.

The predictions made are quite different from predictions from the oligopoly pricing literature (Rotemberg and Saloner (1986), Haltiwanger and Harrington (1991)). In Rotemberg and Saloner (1986), the results came about because high demand and low demand was equally likely which made future demand always unrelated to the past and current market conditions. Haltiwanger and Harrington (1991) model demand so that it follows a predetermined cycle. However, using a predetermined cycle also has its own drawbacks. This research extends these models by making demand change in a Markovian fashion and avoids the drawbacks of the business cycle models in previous work on oligopoly pricing. Therefore, the observations and conclusions drawn are more realistic.

REFERENCES

[1] P. Aghion, P. Bolton, C. Harris, and B. Jullien, Optimal learning by experimentation, *Review of Economic Studies* **58** (1991), 621-654.

[2] R. Balvers and T. Cosimano, Actively learning about demand and the dynamics of price adjustment, *Economic Journal* **100** (1990), 882-898.

[3] R. Balvers and T. Cosimano, Periodic learning about a hidden state variable, *Journal of Economic Dynamics and Control* **17** (1993), 805-827.

[4] D. A. Berry and B. Fristedt, "Bandit Problems: Sequential Allocation of Experiments," Chapman and Hall, London, 1985.

[5] T. Borgers and R. Sarin, Naive reinforcement learning with endogenous aspirations, *International Economic Review* **41** (2000), 921- 950.

[6] J. Brandts and D. J. Cooper, Observability and overcoming coordination failure in organizations: An experimental study, *Experimental Economics* **9** (2006), 407-423.

[7] G. W. Brown, Iterative solution of games by fictitious play, *in* "Activity Analysis of Production and Allocation" (T. C. Koopmans, Ed.), 374-376, Wiley, New York, 1951.

[8] C. F. Camerer, T.-H. Ho, Experience-weighted attraction learning in normal-form games, *Econometrica* **67** (1999), 827-874.

[9] C. F. Camerer, T.-H. Ho and J. K. Chong, Sophisticated EWA learning and strategic teaching in repeated games, *Journal of Economic Theory* **104** (2002), 137-188.

[10] J. K. Chong, C. F. Camerer and T.-H. Ho, A learning-based model of repeated games with incomplete information, *Games and Economics Behavior* **55** (2006), 340-371.

[11] J. Cremer and R. P. McLean, Optimal selling strategies under uncertainty for a discriminating monopolist when demands are interdependent, *Econometrica* **53** (1985), 345-361.

[12] J. Cremer and R. P. McLean, Full extraction of surplus in bayesian and dominant strategy auctions, *Econometrica* **56** (1988), 1247-1257.

[13] C. Derman, "Finite State Markovian Decision Processes," Academic Press, New York, NY, 1970.

[14] E. Dierker and H. Haller, Tax systems and direct mechanisms in large finite economies, *Journal of Economics* **52** (1990), 99-116.

[15] D. Easley and N. Kiefer, Controlling a stochastic process with unknown parameters, *Econometrica* **56** (1988), 1045-1064.

[16] I. Erev and A. Roth, Predicting how people play games: Reinforcement learning in experimental games with unique mixed strategy equilibria, *American Economic Review* **88** (1998), 848-881.

[17] W. Feller, "An Introduction to Probability and its Application," (3rd ed.), Wiley, New York, NY, 1968.

[18] D. Fudenberg and D. K. Levine, "The Theory of Learning in Games," MIT Press, Cambridge, MA, 1998.

[19] M. Goldman, H. Leland, and D. Sibley, Optimal nonuniform prices, *Review of Economic Studies* **51** (1984), 305-319.

[20] R. Guesnerie, "On Taxation and Incentives: Further Reflections on the Limits to Redistribution," Discussion Paper #89, Department of Economics, University of Bonn, 1981.

[21] J. Haltiwanger and J. E. Harrington, The impact of cyclical demand movements on collusive behavior, *Rand Journal of Economics* **22** (1991), 89-106.

[22] P. Hammond, Straightforward incentive compatibility in large economies, *Review of Economic Studies* **46** (1979), 263-282.

[23] O. Hart, Optimal labour contracts under asymmetric information: An introduction, *Review of Economic Studies* **50** (1983), 3-36.

[24] T.-H. Ho, C. F. Camerer, and J. K. Chong, Self-tuning experience weighted attraction in games, *Journal of Economic Theory* **133** (2007), 177-198.

[25] G. Keller and S. Rady, Optimal experimentation in a changing environment, *Review of Economic Studies* **66** (1999), 475-507.

[26] M. Knez and D. Simester, Firm-wide incentives and mutual monitoring at Continental Airlines. *Journal of Labor Economics* **19** (2002), 743-772.

[27] E. Maskin, Nash equilibrium and welfare optimality, *Review of Economic Studies* **66** (1999), 23-38.

[28] E. Maskin and J. Riley, Monopoly with incomplete information, Rand *Journal of Economics* **15** (1984), 171-196.

[29] A. McLennan, Price dispersion and incomplete learning in the long run, *Journal of Economic Dynamics and Control* **7** (1984), 331-347.

[30] P. Milgrom and J. Roberts, Adaptive learning and sophisticated learning in repeated normal form games, *Games and Economic Behavior* **7** (1991), 62-91.

[31] J. Mirrlees, An exploration in the theory of optimal income taxation, *Review of Economic Studies* **38** (1971), 175-208.

[32] M. Mussa and S. Rosen, Monopoly and product quality, *Journal of Economic Theory* **18** (1978), 301-317.

[33] Y. Nyarko and L. Olson, Optimal growth with unobservable resources and learning, *Journal of Economic Behavior and Organization* **29** (1996), 465-491.

[34] T. Palfrey and S. Srivastava, Nash implementation using undominated strategies, *Econometrica* **59** (1991), 479-501.

[35] T. Piketty, Implementation of first-best allocations via generalized tax schedules, *Journal of Economic Theory* **61** (1993), 23-41.

[36] M. H. Riordan and D. E. M. Sappington, Optimal contracts with public ex post information, *Journal of Economic Theory* **45** (1988), 189-199.

[37] K. Roberts, Welfare implications of nonlinear prices, *Economics Journal* **89** (1979), 66-83.

[38] J. J. Rotemberg and G. Saloner, A supergame-theoretic model of price wars during booms, *American Economic Review* **76** (1986), 390-407.

[39] M. Rothschild, Two-armed bandit theory of market pricing, *Journal of Economic Theory* **9** (1974), 185-202.

[40] A. Rusthichi and A. Wolinsky, Learning about variable demand in the long run, *Journal of Economic Dynamics and Control* **19** (1995), 1283-1292.

[41] R. Selten, Anticipatory learning in 2-person games, *in* "Game Equilibrium Models I" (R. Selten, Ed.), pp. 98-154, Springer-Verlag, Berlin, 1991.

[42] A. Spence, Nonlinear prices and welfare, *Journal of Public Economics* **8** (1977), 1-18.

[43] D. O. Stahl and P. Wilson, On players models of other players: Theory and experimental evidence, *Games and Economic Behavior* **10** (1995), 213-254.

[44] D. O. Stahl, Rule learning in symmetric normal-form games: Theory and evidence, *Games and Economic Behavior* **32** (2000), 105-138.

[45] J. Stiglitz, Monopoly non-linear pricing and imperfect information: Insurance market, *Review of Economic Studies* **44** (1997), 407-430.

[46] V. Y. Suslow, "Stability in International Cartels: An Empirical Survey, Domestic Studies Program," Hoover Institution, Stanford University, Working Papers in Economics E887, 1988.

[47] J. Van Huyck, J. Cook, and R. Battalio, Adaptive behavior and coordination failure, *Journal of Economic Behavior & Organization* **32** (1997), 483-503.

[48] J. Van Huyck, R. Battalio and F. Rankin, Selection dynamics and adaptive behavior without much information, *Economic Theory* **33** (2007), 53-65.

[49] J. Van Huyck, J. Wildenthal and R. Battalio, Tacit cooperation, strategic uncertainty, and coordination failure: Evidence from repeated dominance solvable games, *Games and Economic Behavior* **38** (2002), 156-175.

APPENDIX A

ON MODELING THE SOPHISTICATED PLAYER IN THE CONTINENTAL

DIVIDE GAME

In the continental divide game in Table V, each player has 14 actions. Since there are 7 players, there are $14^7 = 105,413,504$ possible outcomes each period. If a sophisticated player were to evaluate all possible scenarios at the outset of the 15 periods, there would be $(105,413,504)^{15}$ possible outcomes. After each of the 7 players go through those calculations they would have to re-evaluate $(105,413,504)^{14}$ possible outcome in the next period. To overcome this problem several assumptions are made that significantly reduce the number of possibilities needed to be analyzed. In making these reductions the attractions to particular actions are used so that only the very likely outcomes are analyzed.

In the continental divide game, players are only aware of the cohort median. The sophisticated player needs to be aware of the probabilities of future outcomes. It is assumed that the sophisticated player is aware of the most likely action(s) of each player in each period if the player was an adaptive learner[1].

Sophisticated players form beliefs about the other players being sophisticated. The sophisticated players are assumed to be aware of the number of players who are

---

[1]It is as if the sophisticated player is tracking the attractions of an adaptive learner through the medians that have materialized thus far in the game and that he is assuming the adaptive learner chooses the action with the highest attraction ($\lambda = \infty$).

more likely to be sophisticated (That is, $\#\{j|\alpha_j > 0.5\}$), and sophisticated players would believe that these many players would be acting as sophisticated players. Each sophisticated player would assume that their beliefs are the same as the other sophisticated players. This makes the attraction calculation of sophisticated players in eqn. 3.9 much easier since sophisticated players would know that the payoff maximizing action sequence, $J_t = \left\{s_i^{j_t}, s_i^{j_{t+1}}, ..., s_i^{j_T}\right\}$ would be the same for all sophisticated players.

The likelihood was first estimated by making the sophisticated player look one period into the future, and after its successful estimation, two periods were done and so on. This was done to ensure that the estimation converged and did so in reasonable time. The player extrapolates his future earnings based on the payoff he got from this last period he iterated into. It was observed that the likelihood and parameter estimates did not change significantly beyond three periods of iteration into the future. This could be due to the fact that sophisticated players only looked so far into the future due to their cognitive capabilities or due to some kind of future payoff discounting that made payoffs more than three periods into the future not as attractive. Another explanation is that the extrapolating technique itself gave a good estimate of his future earning that would have been obtained in the full iterative procedure. An additional period of reasoning into the future period means that each sophisticated player, in each period of the repeated game, has to consider all possible action sequences of additional length while also taking into account all possible action se-

quences of his opponents for an additional period. Thus, the estimation time increases exponentially (with a large base) with the increase in the number of periods iterated into. Since the likelihood and parameter estimates were not observed to change after iterating for more than three periods into the future, the iteration for three periods was used, and is reported here. Thus, in the reported estimates sophisticated players maximize their payoff by calculating their payoff through an action sequence for the next three periods.

APPENDIX B

MAXIMUM LIKELIHOOD ESTIMATION FOR SELF-TUNING EWA

Following is the code used to find the maximum likelihood of observing the data assuming that all the players are self-tuning EWA learners.

```
/*  Author : Megha W. Watugala

Version : 4

Maximum likelihood estimation for self-tuning EWA

using the continental divide game */

new;

library maxlik;

SFile = "P:/gauss/datalongd.txt";

load data[840,6]=^SFile;

num_sessions = 8;

// periods in each session /

periods_session = {15, 15, 15, 15, 15, 15, 15, 15};

// subjects in each session /

subjects_session = {7, 7, 7, 7, 7, 7, 7, 7};


// number of actions available to the subjects

// assume actions are indexed 1 .. actions
```

```
actions = 14;

game_matrix =

{45    49   52   55   56 55 46 -59 -88 -105 -117 -127 -135 -142,

 48    53   58   62   65   66   61 -27 -52 -67 -77   -86 -92  -98,

 48    54   60   66   70   74   72   1 -20 -30 -41   -48 -53 -58,

 43    51   58   65   71   77   80   26   8   -2   -9   -14 -19 -22,

 35    44   52   60   69   77   86   46   32   25   19   15   12   10,

 23    33   42   52   62   72   82   62   53   47   43   41   39   38,

  7    18   28   40   51   64   78   75   69   66   64   63   62   62,

-13    -1   11   23   37   51   69   83   81   80   80   80   81   82,

-37    -24  -11   3   18   35   57   88   89   91   92   94   96   98,

-65    -51  -37 -21   -4   15   40   89   94   98 101 104 107 110,

-97    -82  -66 -49 -31   -9   20 85   94 100 105 110 114 119,

-133 -117 -100  -82 -61 -37   -5 78   91   99 106 112 118 123,

-173 -156 -137 -118 -96  -69 -33 67 83   94 103 110 117 123,

-217 -198 -179 -158 -134 -105 -65 52 72 85 95 104 112 120};


game_matrix = game_matrix/100;


session_index=1;

period_index=2;

otheract_index=3;

choice_index = 5;


total_periods=0;
```

```
total_subject_periods=0;


k=1;

do while k <= num_sessions;

    total_periods = total_periods + periods_session[k];

    total_subject_periods = total_subject_periods +

periods_session[k]*subjects_session[k];

    k=k+1;

endo;

p = {1.3};


/*maxclr; __row=0; _max_CovPar = 0; _max_algorithm = 2; */


{x1,f,grad,cov,ret}= maxlik(data,0,&selftuning,p);


print "  results  ";

print " lambda   " x1;

print "  avg likelihood  "  f;

print "  cov matrix "  cov;


proc selftuning(p,data);

/* variables used in likelihood estimation */

local ylog,i,j,k,m,attractions,Probase,ave_act,change_act;

local delta,shi,current_session,current_period;

local actions_done, temp_actions;
```

```
local
predicted_median,temp_prob,likely_action,best_response,periods_a
ction,temp_payoff;


local x;

//y = p[2];

x= exp(p);      // x is lambda

// attrations at each session,period,subject, action

attractions = zeros(total_subject_periods,actions);

// used to track the denominator of the probability calculation

Probase = zeros(total_subject_periods,1);

// likelihood of each action taken


ylog = zeros(total_subject_periods,1);

// average of times a particular action is taken

ave_act = zeros(total_subject_periods,actions);

// the change in behavior stored here. ave_act used to calculate
this

change_act = zeros(total_subject_periods,1);

// attention function, whether a particular action needs to be

reinforced by the payoff it would have got or not


delta = zeros(total_subject_periods,actions);

// the change-detector function

shi = zeros(total_subject_periods,1);
```

```
predicted_median = zeros(total_subject_periods,1);

likely_action = zeros(total_subject_periods,1);

best_response = zeros(total_subject_periods,1);

temp_payoff = zeros(actions,1);



current_period = 1;

current_session = 1;

// initial attractions counter

local f;

// long data

f={0.3,0.3,0.3,1,2,15,16,7,5,8,5,5,2,4};



// main loop

k=1;



do while k < total_subject_periods;     // for all subject
periods
// in one loop a whole periods data is taken care of
// the while j loop goes through all the subjects in that period
of that session


    if data[k,period_index] == 1;  // new session
        // initialize attractions, etc.
        current_session = data[k,session_index];
            current_period = 1;
```

```
        actions_done =

zeros(subjects_session[current_session],1);

        temp_actions =

zeros(subjects_session[current_session],1);

        periods_action =

zeros(subjects_session[current_session],1);

        j = 1;


        do while j <= subjects_session[current_session]; //

setting initial attractions

            i =1;

                Probase[k+j-1] = 0;    // the base of the prob of

an action just making sure it starts at 0


            do while i <= actions;

                    attractions[k+j-1,i]= (1/x)*ln(f[i]./f[5]) +

20;    // attraction of 5 normalized to 20

                Probase[k+j-1]=Probase[k+j-

1]+exp(x*attractions[k+j-1,i]);

                    i=i+1;

            endo;


                i=1;

                temp_prob =0;

                do while i <= actions;
```

```
                temp_prob = temp_prob +

exp(x*attractions[k+j-1,i])/Probase[k+j-1];

                    if temp_prob > 0.5;

                        predicted_median[k+j-1]=i;

                        i = actions;

                    endif;

                i=i+1;

            endo;


            ylog[k+j-1]= x*attractions[k+j-1,data[k+j-

1,choice_index]] - ln(Probase[k+j-1]);


            likely_action[k+j-1] = maxindc(attractions[k+j-

1,.]');

            periods_action[j]=maxindc(attractions[k+j-1,.]');

                j = j+1;

        endo;


        j=1;

        do while j <= subjects_session[current_session];

            i=1;


                do while i <= actions;

                temp_actions = periods_action;

                temp_actions[j]=i;
```

```
temp_payoff[i]=game_matrix[i,median(temp_actions)];

                i=i+1;

        endo;


        best_response[k+j-1]=maxindc(temp_payoff);


        j=j+1;
    endo;
  endif;   // first period taken care of


    j=1;
    do while j <= subjects_session[current_session];
        actions_done[j]=data[k+j-1,choice_index]; // action
taken by each subject this period
        if current_period != 1;   // setting up ave_act
                i =1;
            do while i <= actions;
                ave_act[k+j-1,i] = ((current_period-
1)*ave_act[k+j-subjects_session[current_session]-
1,i])./current_period;
                i = i+1;
            endo;
        endif;
```

```
        ave_act[k+j-1,data[k+j-1,choice_index]] =

ave_act[k+j-1,data[k+j-1,choice_index]]+ 1/current_period;

        j = j+1;

    endo;



    j=1;



    do while j <= subjects_session[current_session];  // for

all subjects in this period deducing their logl

            m=1;



        do while m <= subjects_session[current_session];  //

to calculate change in action

            if m != j; // of opponents

                i=1;



                do while i <= actions;

                    if i == actions_done[m];

                        change_act[k+j-1] = change_act[k+j-1]

+ ((1-ave_act[k+m-1,i])^2)./(subjects_session[current_session]-

1);

                    else;

                        change_act[k+j-1] = change_act[k+j-1]

+ (ave_act[k+m-1,i]^2)./(subjects_session[current_session]-1);

                    endif;

                    i = i+1;
```

```
                    endo;

                  endif;

              m = m+1;

          endo;



          i=1;

          do while i <= actions;   // calculate attention
function


              temp_actions = actions_done;

              temp_actions[j]=i;

                  if game_matrix[i,median(temp_actions)] >=
game_matrix[data[k+j-1,choice_index],data[k+j-
1,otheract_index]];

                      delta[k+j-1,i] = 1;

                  else;

                      delta[k+j-1,i] = 0;

                  endif;


            i=i+1;

          endo;


              shi[k+j-1] = 1-0.5*change_act[k+j-1];    //
change detector function
```

```
        shi[k+j-1] = (0.5 + (current_period-1)*shi[k+j-

1])./current_period; // adjustment to better update initially



        if current_period !=

periods_session[current_session];  // unless it is the last

period

            Probase[k+j+subjects_session[current_session]-1]

=0;



            i=1;

        do while i <= actions;

            // set the attractions for next period

            temp_actions = actions_done;  // actions taken

this period

            temp_actions[j]=i;



attractions[k+j+subjects_session[current_session]-

1,i]=(current_period*shi[k+j-1]*attractions[k+j-1,i]+delta[k+j-

1,i]*(game_matrix[i,median(temp_actions)]))./(current_period*shi

[k+j-1]+1);



Probase[k+j+subjects_session[current_session]-

1]=Probase[k+j+subjects_session[current_session]-
```

```
1]+exp(x*attractions[k+j+subjects_session[current_session]-

1,i]);

                        i=i+1;

                endo;



                    i=1;

                    temp_prob =0;



                    do while i <= actions;



                        temp_prob = temp_prob +

exp(x*attractions[k+j+subjects_session[current_session]-

1,i])/Probase[k+j+subjects_session[current_session]-1];

                            if temp_prob > 0.5;



    predicted_median[k+j+subjects_session[current_session]-

1]=i;

                                i = actions;

                            endif;

                            i=i+1;

                    endo;



                ylog[k+j+subjects_session[current_session]-1]=

x*attractions[k+j+subjects_session[current_session]-

1,data[k+j+subjects_session[current_session]-1,choice_index]] -

ln(Probase[k+j+subjects_session[current_session]-1]);
```

```
likely_action[k+j+subjects_session[current_session]-1] =

maxindc(attractions[k+j+subjects_session[current_session]-

1,.]');


periods_action[j]=maxindc(attractions[k+j+subjects_session[curre

nt_session]-1,.]');
                endif;
            j = j+1;
        endo;
        if current_period != periods_session[current_session];
// unless it is the last period


            j=1;
            do while j <= subjects_session[current_session];
                    i=1;
                    do while i <= actions;
                    temp_actions = periods_action;
                    temp_actions[j]=i;


temp_payoff[i]=game_matrix[i,median(temp_actions)];
                        i=i+1;
                    endo;
```

```
best_response[k+j+subjects_session[current_session]-

1]=maxindc(temp_payoff);

                j=j+1;

         endo;

      endif;


  current_period = current_period +1;

  k = k + subjects_session[current_session];


endo;

retp(ylog);

endp;


proc max(x,y); // procedure to return maximum

    if x>y;

         retp(x);

    else;

         retp(y);

    endif;

endp;
```

APPENDIX C

MAXIMUM LIKELIHOOD ESTIMATION FOR SOPHISTICATED LEARNING

Following is the code used to find the maximum likelihood of observing the data assuming that a fraction of the players are sophisticated learners while the rest are self-tuning EWA learners.

```
/*  Author : Megha W. Watugala
Maximum likelihood estimation for Sophisticated Learning
using the continental divide game */
new;
library maxlik;
//  long datad set header
SFile = "P:/gauss/datalongd.txt";
load data[840,6]=^SFile;
num_sessions = 8;
// periods in each session /
periods_session = {15, 15, 15, 15, 15, 15, 15, 15};
// subjects in each session /
subjects_session = {7, 7, 7, 7, 7, 7, 7, 7};
// 1 if converging to low action and 2 if converging to high
action
session_end = {1, 1, 2, 2, 2, 1, 1, 2};
```

```
session_index=1;

period_index=2;

otheract_index=3;

choice_index = 5;

// number of actions available to the subjects

// assume actions are indexed 1 .. actions

actions = 14;

//payoff table

game_matrix =

{45    49   52   55   56 55 46 -59 -88 -105 -117 -127 -135 -142,

 48    53   58   62   65   66   61 -27 -52 -67 -77  -86 -92  -98,

 48    54   60   66   70   74   72   1 -20 -30 -41  -48 -53 -58,

 43    51   58   65   71   77   80   26   8  -2  -9  -14 -19 -22,

 35    44   52   60   69   77   86   46   32   25   19   15   12   10,

 23    33   42   52   62   72   82   62   53   47   43   41   39   38,

  7    18   28   40   51   64   78   75   69   66   64   63   62   62,

-13    -1   11   23   37   51   69   83   81   80   80   80   81   82,

-37    -24  -11   3   18   35   57   88   89   91   92   94   96   98,

-65    -51  -37 -21  -4   15   40   89   94   98 101 104 107 110,

-97    -82  -66 -49 -31  -9    20 85   94 100 105 110 114 119,

-133 -117 -100  -82 -61 -37  -5 78   91   99 106 112 118 123,

-173 -156 -137 -118 -96  -69 -33 67 83   94 103 110 117 123,

-217 -198 -179 -158 -134 -105 -65 52 72 85 95 104 112 120};

// convert to dollar amounts
```

```
game_matrix = game_matrix/100;


total_periods=0;

total_subject_periods=0;

k=1;

do while k <= num_sessions;

    total_periods = total_periods + periods_session[k];

    total_subject_periods = total_subject_periods +

periods_session[k]*subjects_session[k];

    k=k+1;

endo;

p = {1,0,0,1};

/* p is the passed in values of the variable

x1 the values of the variable at maximum likelihood */

/*maxclr; __row=0; _max_CovPar = 0; _max_algorithm = 2; */

 _max_MaxIters =15;

{x1,f,grad,cov,ret}= maxlik(data,0,&selftuning,p);

print "  results  ";

print " lambda   " x1;

print "  avg likelihood  "  f;

print "  cov matrix "  cov;

end;

proc selftuning(p,data);

/* variables used in likelihood estimation */
```

```
local
ylog,yylog,yy,i,j,k,m,attractions,Probase,ave_act,change_act;
local delta,shi,current_session,current_period;
local actions_done, temp_actions;
local
predicted_median,temp_prob,likely_action,best_response,periods_a
ction,temp_payoff;
local r, Vmax, Vtemp, alphabar, attractions_block,
ave_act_block;
local x,y,alpha,randnum,selflog,sophlog,ifsoph;


x= exp(p[1]);      // x is lambda for selftuning EWA
y=exp(p[4]);     // y is lambda for sophisticated players
alphabar = exp(p[2])/(1+exp(p[2]));
alpha = exp(p[3])/(1+exp(p[3]));


local numsoph;
local x1,fm,grad,cov,ret,alphatable,logtable,templogtable;


logtable =
zeros(maxc(periods_session),2*maxc(subjects_session));


alphatable = zeros(maxc(subjects_session),1);


// attrations at each session,period,subject, action
```

```
attractions = zeros(total_subject_periods,actions);

// used to track the denominator of the probability calculation

Probase = zeros(total_subject_periods,1);

// likelihood of each action taken

ylog = zeros(total_subject_periods,1);

// average of times a particular action is taken

ave_act = zeros(total_subject_periods,actions);

// the change in behavior stored here. ave_act used to calculate

this

change_act = zeros(total_subject_periods,1);

// attention function, whether a particular action needs to be

reinforced by the payoff it would have got or not

delta = zeros(total_subject_periods,actions);

// the change-detector function

shi = zeros(total_subject_periods,1);

predicted_median = zeros(total_subject_periods,1);

likely_action = zeros(total_subject_periods,1);

best_response = zeros(total_subject_periods,1);

temp_payoff = zeros(actions,1);


current_period = 1;

current_session = 1;


// initial attractions counter

local f;
```

```
// long data

//f={0.1,0.1,0.1,1,4,16,18,7,6,9,6,6,2,5};

f={0.3,0.3,0.3,1,2,15,16,7,5,8,5,5,2,4};

// f is set according to players behavior in the first period


// main loop

k=1;

do while k < total_subject_periods;      // for all subject
periods

// in one loop a whole periods data is taken care of

// the while j loop goes through all the subjects in that period
of that session

    if data[k,period_index] == 1;  // new session


      // initialize attractions, etc.

      current_session = data[k,session_index];

          current_period = 1;

      actions_done = zeros(subjects_session[current_session],1);

      temp_actions = zeros(subjects_session[current_session],1);

      periods_action =
zeros(subjects_session[current_session],1);

      selflog = zeros(subjects_session[current_session],1);

      sophlog = zeros(subjects_session[current_session],1);

      ifsoph = zeros(subjects_session[current_session],1);
```

```
      attractions_block =

zeros(subjects_session[current_session],actions);

      ave_act_block =

zeros(subjects_session[current_session],actions);


      logtable =

zeros(maxc(periods_session),2*maxc(subjects_session));

      templogtable = zeros(maxc(periods_session),2);


      alphatable = zeros(maxc(subjects_session),1);


      j = 1;

      do while j <= subjects_session[current_session]; //

setting initial attractions

         i =1;

               Probase[k+j-1] = 0;    // the base of the prob of

an action just making sure it starts at 0

         do while i <= actions;

           //attractions_block[j,i] =

(1/x[session_end[current_session]])*ln(f[i]./f[5]);

                  attractions[k+j-1,i]= (1/x)*ln(f[i]./f[5]);


           Probase[k+j-1]=Probase[k+j-1]+exp(x*attractions[k+j-

1,i]);

               i=i+1;
```

```
        endo;


                ylog[k+j-1]= x*attractions[k+j-1,data[k+j-
1,choice_index]] - ln(Probase[k+j-1]);

          selflog[j] = x*attractions[k+j-
1,choice_index]] - ln(Probase[k+j-1]);

          sophlog[j] = x*attractions[k+j-
1,choice_index]] - ln(Probase[k+j-1]);

          logtable[current_period,2*j-1] = x*attractions[k+j-
1,data[k+j-1,choice_index]] - ln(Probase[k+j-1]);

          logtable[current_period,2*j] = x*attractions[k+j-
1,data[k+j-1,choice_index]] - ln(Probase[k+j-1]);

          alphatable[j] = alpha;

          periods_action[j]=maxindc(attractions[k+j-1,.]');


                j = j+1;

      endo;


      j=1;

      do while j <= subjects_session[current_session];     //
creating the best_response vector
          i=1;

              do while i <= actions;
        temp_actions = periods_action;

        temp_actions[j]=i;
```

```
                temp_payoff[i]=game_matrix[i,median(temp_actions)];

                    i=i+1;

            endo;

            best_response[k+j-1]=maxindc(temp_payoff);

            j=j+1;

        endo;


    endif;    // first period taken care of


    j=1;

    do while j <= subjects_session[current_session];

        actions_done[j]=data[k+j-1,choice_index]; // action
taken by each subject this period


        if current_period != 1;    // setting up ave_act

                    i =1;

            do while i <= actions;

                ave_act[k+j-1,i] = ((current_period-
1)*ave_act[k+j-subjects_session[current_session]-
1,i])./current_period;

                ave_act_block[j,i] = ave_act[k+j-1,i];

                i = i+1;

            endo;

        endif;
```

```
        ave_act[k+j-1,data[k+j-1,choice_index]] = ave_act[k+j-

1,data[k+j-1,choice_index]]+ 1/current_period;

        ave_act_block[j,data[k+j-1,choice_index]] =

ave_act[k+j-1,data[k+j-1,choice_index]];

        j = j+1;

    endo;



    j=1;



    do while j <= subjects_session[current_session];  // for

all subjects in this period deducing their logl

            m=1;



        do while m <= subjects_session[current_session];  //

to calculate change in action

            if m != j; // of opponents

            i=1;

            do while i <= actions;

              if i == actions_done[m];

                    change_act[k+j-1] = change_act[k+j-1]

+ ((1-ave_act[k+m-1,i])^2)./(subjects_session[current_session]-

1);



                else;
```

```
                          change_act[k+j-1] = change_act[k+j-1]

+ (ave_act[k+m-1,i]^2)./(subjects_session[current_session]-1);

                endif;

                i = i+1;

            endo;

          endif;

      m = m+1;

    endo;



    i=1;

    do while i <= actions;  // calculate attention
function

        temp_actions = actions_done;

        temp_actions[j]=i;

            if game_matrix[i,median(temp_actions)] >=
game_matrix[data[k+j-1,choice_index],data[k+j-
1,otheract_index]];

                    delta[k+j-1,i] = 1;

                else;

                    delta[k+j-1,i] = 0;

                endif;

        i=i+1;

        endo;


            shi[k+j-1] = 1-0.5*change_act[k+j-1];
```

```
                                    // change detector function


        shi[k+j-1] = (0.5 + (current_period-1)*shi[k+j-
1])./current_period; // adjustment to better update initially


        if current_period != periods_session[current_session];
// unless it is the last period
            Probase[k+j+subjects_session[current_session]-1] =0;
                    i=1;
            do while i <= actions; // set attractions
                // set the attractions for next period
                temp_actions = actions_done;
                            // actions taken this period
                temp_actions[j]=i;


attractions[k+j+subjects_session[current_session]-
1,i]=(current_period*shi[k+j-1]*attractions[k+j-1,i]+delta[k+j-
1,i]*(game_matrix[i,median(temp_actions)]))./(current_period*shi
[k+j-1]+1);


            attractions_block[j,i] =
attractions[k+j+subjects_session[current_session]-1,i];
            Probase[k+j+subjects_session[current_session]-
1]=Probase[k+j+subjects_session[current_session]-
```

```
1]+exp(x*attractions[k+j+subjects_session[current_session]-
1,i]);

                i=i+1;

        endo;


        ylog[k+j+subjects_session[current_session]-1]=
x*attractions[k+j+subjects_session[current_session]-
1,data[k+j+subjects_session[current_session]-1,choice_index]] -
ln(Probase[k+j+subjects_session[current_session]-1]);



periods_action[j]=maxindc(attractions[k+j+subjects_session[curre
nt_session]-1,.]');
             endif;
        j = j+1;
     endo;


    if current_period != periods_session[current_session];  //
unless it is the last period

        j=1;


        do while j <= subjects_session[current_session];
                 i=1;
          numsoph = 0;
          Vmax = zeros(actions,1);
```

```
      do while i <= actions;

    temp_actions = periods_action;


    temp_actions[j]=i;


  m=1;

  do while m<=subjects_session[current_session];

    if m !=j;


        if  ifsoph[m] >= 0.5;

          temp_actions[m]=i;

          numsoph = numsoph +1;

        endif;


      endif;

      m=m+1;

    endo;




temp_payoff[i]=game_matrix[i,median(temp_actions)];


      Vmax[i] = temp_payoff[i] +

alphabar*FVmax(i,temp_actions,alphabar,current_period,min(2,peri
```

```
ods_session[current_session]-current_period-

1),attractions_block,j,subjects_session[current_session],actions

,ave_act_block,p,periods_session[current_session]-

current_period-1,ifsoph);


            i=i+1;

        endo;


        selflog[j] = selflog[j] +

ylog[k+j+subjects_session[current_session]-1];


        sophlog[j] = sophlog[j] +

ln(exp(y*Vmax[data[k+j+subjects_session[current_session]-

1,choice_index]])./sumc(exp(y*Vmax)));


        logtable[current_period+1,2*j-1] =

ylog[k+j+subjects_session[current_session]-1];


        logtable[current_period+1,2*j] =

ln(exp(y*Vmax[data[k+j+subjects_session[current_session]-

1,choice_index]])./sumc(exp(y*Vmax)));


        ylog[k+j+subjects_session[current_session]-1]=

ln((1-ifsoph[j])*exp(ylog[k+j+subjects_session[current_session]-

1]) +
```

```
((ifsoph[j]*exp(y*Vmax[data[k+j+subjects_session[current_session

]-1,choice_index]])).sumc(exp(y*Vmax)))));

          j=j+1;

       endo;

       j=1;

       do while j <= subjects_session[current_session];

         ifsoph[j] =

exp(sophlog[j])*alpha/(exp(sophlog[j])*alpha+exp(selflog[j])*(1-

alpha));


          j=j+1;

       endo;

    endif;


   current_period = current_period +1;

   k = k + subjects_session[current_session];

endo;

retp(ylog);

endp;


proc max(x,y); // procedure to return maximum

    if x>y;

         retp(x);

    else;

         retp(y);
```

```
        endif;

endp;


proc min(x,y); // procedure to return maximum

      if x>y;

            retp(y);

      else;

            retp(x);

      endif;

endp;


// procedure used by the sophisticated players to iterate into

the future periods

proc

FVmax(Jt,periods_act,alphabar,T,Tleft,attracts,player,subjects,a

cts,ave_acts,p,Tleftact,ifsoph);

/* variables used in FVmax */

local r,Vmax,Vtemp,i,j,m,fattractions,fProbase,fchange_act;

local fdelta,fshi;

local acts_done, temp_acts;

local

fbest_response,periods_acts,ftemp_payoff,randnum,yy,Vmaxyy;


local x,y;

x= exp(p[1]);      // x is lambda
```

```
// attrations for each subject for each action

fattractions = zeros(subjects,acts);

// used to track the denominator of the probability calculation

fProbase = zeros(subjects,1);

// the change in behavior stored here. ave_act used to calculate

this

fchange_act = zeros(subjects,1);

// attention function, whether a particular action needs to be

reinforced by the payoff it would have got or not

fdelta = zeros(subjects,acts);

// the change-detector function

fshi = zeros(subjects,1);


fbest_response = zeros(subjects,1);

//randnum = zeros(subjects,1)

ftemp_payoff = zeros(acts,1);

acts_done = zeros(subjects,1);


    if Tleft <= 0;

      Vmax = 0;

      retp(Vmax);

    else;


      T = T + 1;

      j=1;
```

```
        do while j <= subjects;

            acts_done[j]=periods_act[j]; // action taken by each
subject this period

            i =1;

              do while i <= acts;

                  ave_acts[j,i] = ((T-1)*ave_acts[j,i])./T;

                  i = i+1;

              endo;

            if acts_done[j] >acts;

              print acts_done;

            else;

              ave_acts[j,acts_done[j]] = ave_acts[j,acts_done[j]]+
1/T;

            endif;

            j = j+1;

        endo;

        j=1;

        do while j <= subjects;  // for all subjects in this
period deducing their logl

                m=1;

            do while m <= subjects;  // to calculate change in
action

                if m != j; // of opponents

                i=1;
```

```
            do while i <= acts;

                if i == acts_done[m];

                        fchange_act[j] = fchange_act[j] +

((1-ave_acts[m,i])^2)./(subjects-1);

                    else;

                        fchange_act[j] = fchange_act[j] +

(ave_acts[m,i]^2)./(subjects-1);

                    endif;

                    //change_act_block[j] = change_act[k+j-1];

                    i = i+1;

                 endo;

                endif;

          m = m+1;

        endo;

        i=1;

        do while i <= acts;  // calculate attention function

           temp_acts = acts_done;

           temp_acts[j]=i;

               if game_matrix[i,median(temp_acts)] >=

game_matrix[acts_done[j],median(acts_done)];

                        fdelta[j,i] = 1;

                    else;

                        fdelta[j,i] = 0;

                    endif;

            i=i+1;
```

```
     endo;

          fshi[j] = 1-0.5*fchange_act[j];    // change
detector function


     fshi[j] = (0.5 + (T-1)*fshi[j])./T; // adjustment to
better update initially

     //if Tleft != 1;   // unless it is the last period

       fProbase[j] =0;

             i=1;

       do while i <= acts;

          // set the attractions for next period

          temp_acts = acts_done;  // actions taken this
period

          temp_acts[j]=i;


fattractions[j,i]=(T*fshi[j]*attracts[j,i]+fdelta[j,i]*(game_mat
rix[i,median(temp_acts)]))./(T*fshi[j]+1);

          fProbase[j]=fProbase[j]+exp(x*fattractions[j,i]);

          i=i+1;

        endo;

                periods_act[j]=maxindc(fattractions[j,.]');

     j = j+1;

   endo;

       j = player;

          i=1;
```

```
                do while i <= acts;

            //temp_acts = round(periods_act);

            temp_acts = periods_act;

            temp_acts[j]=i;



ftemp_payoff[i]=game_matrix[i,median(temp_acts)];

                 i=i+1;

              endo;

            fbest_response[j]=maxindc(ftemp_payoff);

            r = max(1,fbest_response[j]-1);



          Vmax = 0;

           if Tleft == 1;



             //Vmax = (Tleftact-

Tleft+2)*maxc(ftemp_payoff)./2;

               Vmax = (1-

alphabar^Tleftact)*maxc(ftemp_payoff)./(1-alphabar);

               retp(Vmax);

             else;



             do while r <= min(acts, fbest_response[j]+1);

               //temp_acts = round(periods_act);

               temp_acts = periods_act;

               temp_acts[j]=r;
```

```
                    m=1;

                    do while m<=subjects;

                         if m !=j;

                             if ifsoph[m] ==1;

                                 temp_acts[m]=r;

                             endif;

                         endif;

                         m=m+1;

                    endo;

                    //temp_acts = round((1-
alphabar)*temp_acts+alphabar*r);

                    Vtemp = maxc(ftemp_payoff) +
alphabar*FVmax(r,temp_acts,alphabar,T,Tleft-
1,fattractions,player,subjects,acts,ave_acts,p,Tleftact-
1,ifsoph);


                     if Vtemp > Vmax;

                         Vmax = Vtemp;

                     endif;

                     r = r+1;

                  endo;

               endif;

      retp(Vmax);

    endif;

    //retp(0);
```

```
endp;


proc findalpha(alp,tlogtable);

// first column of log table is selflog the second is sophlog

local ylog, i, size, alpha;

size = rows(tlogtable);

ylog = zeros(size,1);

alpha = exp(alp)/(1+exp(alp));


i=1;

do while i <= size;

    ylog[i] = ln(alpha*exp(tlogtable[i,1])+(1-

alpha)*exp(tlogtable[i,2]));


    i = i+1;

endo;

retp(ylog);

endp;
```

VITA

Name: Megha Weerakoon Watugala

Address:

1, Wijerama Road,

Gangodawila,

Nugegoda,

Sri Lanka

Education:

Ph.D., Texas A&M University, 2008

B.S. (Honors), California Institute of Technology, 2003