Three Essays on Strategic Behavior

by

James Todd Swarthout

A Dissertation Submitted to the Faculty of the

DEPARTMENT OF ECONOMICS

In Partial Fulfillment of the Requirements
For the Degree of

DOCTOR OF PHILOSOPHY

In the Graduate College

THE UNIVERSITY OF ARIZONA

2002

UMI Number: 3060955

UMI®

ProQuest Information and Learning Company
300 North Zeeb Road
P.O. Box 1346
Ann Arbor, MI 48106-1346

THE UNIVERSITY OF ARIZONA ®
GRADUATE COLLEGE

As members of the Final Examination Committee, we certify that we have read the dissertation prepared by James Todd Swarthout

entitled Three Essays on Strategic Behavior

and recommend that it be accepted as fulfilling the dissertation requirement for the Degree of Doctor of Philosophy

M Walker
Mark Walker — Date July 26, 2002

Steven Gjerstad — Date April 26, 2002

Stanley Reynolds — Date July 29, 2002

Date

Date

Final approval and acceptance of this dissertation is contingent upon the candidate's submission of the final copy of the dissertation to the Graduate College.

I hereby certify that I have read this dissertation prepared under my direction and recommend that it be accepted as fulfilling the dissertation requirement.

M Walker
Dissertation Director Mark Walker — Date July 26, 2002

Statement by Author

This dissertation has been submitted in partial fulfillment of requirements for an advanced degree at The University of Arizona and is deposited in the University Library to be made available to borrowers under rules of the Library.

Brief quotations from this dissertation are allowable without special permission, provided that accurate acknowledgment of source is made. Requests for permission for extended quotation from or reproduction of this manuscript in whole or in part may be granted by the head of the major department or the Dean of the Graduate College when in his or her judgment the proposed use of the material is in the interests of scholarship. In all other instances, however, permission must be obtained from the author.

SIGNED: James T. U. Swarthout

ACKNOWLEDGMENTS

I would like to specially thank to Mark Walker for providing invaluable support to me and my work over the life of this dissertation. Also, I thank the following people for giving me helpful advice and comments: Steve Gjerstad, Stan Reynolds, Jason Shachat, and Ed Zajac. I acknowledge the following organizations for financial support of my research: the Earhart Foundation, the International Foundation for Research in Experimental Economics, and the International Business Machines Corporation. Finally, much of Chapter 1 is joint research with Jason Shachat, and much of Chapter 2 is joint research with Mark Walker.

Dedication

This work is dedicated to my mother, my father, and Shelli. All of you have provided me with much love and support over the years.

Table of Contents

Table of Contents—*Continued*

List of Figures

LIST OF TABLES

ABSTRACT

Each chapter of this dissertation focuses on a different aspect of strategic behavior. The first chapter presents research in which humans play against a computer decision maker that follows either a reinforcement learning algorithm or an Experience Weighted Attraction algorithm. The algorithms are more sensitive than humans to exploitable opponent play. Further, learning algorithms respond to calculated opportunities systematically; however, the magnitudes of these responses are too weak to improve the algorithm's payoffs. Additionally, humans and current models of their behavior differ in that humans do not adjust payoff assessments by smooth transition functions but when humans do detect exploitable play they are more likely to choose the best response to this belief.

The second chapter reports research designed to directly reveal the information used by subjects in a game. Human play is often classified as adhering to reinforcement learning or belief learning. This is typically due to using subjects' observed action choices to estimate the learning models' parameters. We use a different, more direct approach: an experiment in which subjects choose which kind of information they see – either the information required for reinforcement learning, or the information required for belief learning. Results suggest that while neither kind of information is chosen exclusively, subjects most often choose information that is consistent with belief learning and inconsistent with reinforcement learning.

The third chapter discusses the Groves-Ledyard mechanism. In economics we typically rely on continuous analysis, however doing so may not lead to an accurate assessment of a discrete environment. The Groves-Ledyard mechanism is such a case that demonstrates a drastic divergence of results between continuous and discrete analysis. This chapter shows that given quasi-linear preferences, a discrete strategy space will not necessarily yield a single Pareto optimal Nash equilibrium, but typi-

cally many Nash equilibria, not all of which are necessarily Pareto optimal. Further, the value of the mechanism's single free parameter determines the number of Nash equilibria and the proportion of Pareto optimal Nash equilibria.

Chapter 1

LEARNING ABOUT LEARNING IN GAMES THROUGH EXPERIMENTAL CONTROL OF STRATEGIC INTERDEPENDENCE

1.1 Introduction

Identifying how humans respond and adapt their behavior in repeated strategic decision making tasks has emerged as a core, but difficult, question in the social sciences. Most studies that address this question formulate alternative adaptive or "learning" models and estimate model parameters from human experimental data. These estimated models are either evaluated for goodness-of-fit by statistical criteria or they are used to generate simulations which are subsequently compared to human experimental play. Unfortunately, definitive conclusions are difficult to achieve with this approach because current econometric techniques generate exceedingly high rates of Type I and Type II errors when evaluating alternative adaptive models of play (Salmon [29] (2001)).

One source of this difficulty is the nature of the problem. Learning in games is a multivariate stochastic process. One component of this process is a set of latent variables, such as beliefs about opponents or values associated with alternative actions, which each individual uses to select actions and adjusts according to game history. If play doesn't coincide with an equilibrium and players condition actions on the common observed history of play, strong interdependencies are likely to exist among the adjustment rules of the players' latent variables. When a researcher seeks to identify the rules underlying the interdependent latent processes, typically the only observable information is the sequence of realized choices from discrete action sets. Unfortunately, these interdependent and dynamic latent processes, and the resultant sequences of discrete choices are substantial obstacles for current econometric methods.

In this study we adopt a technique that exercises experimental control of strategic interdependence and enables us to gain greater insight into the rules humans use to adjust their play in games. We conduct hybrid experiments in which humans play simple 2×2 games against alternative computer-implemented learning algorithms. Each game has a unique Nash equilibrium in mixed strategies. This technique lets us directly control the nature of the dependence of one of the players. In turn this allows us to more accurately assess human response to opponents adopting particular learning rules. Furthermore, we are able to better evaluate the appropriateness and properties of alternative hypothesized models of human adaptive behavior in games.

Several recent learning models embody two common principles: smooth adjustment rules for values of actions and probabilistic choice. First, for each of the players' actions the models ascribe a latent variable which represents the value of the action. This value is updated after each play of the game according to an adaptive rule and the stage game outcome. Second, in each stage game a player selects his action according to a probabilistic choice rule. This probabilistic choice rule assigns higher probabilities to actions with greater latent values. Typically these models include unobservable parameters whose values are estimated from experimental data. Uniformly across studies, estimated parameter values specify adaptive rules that have significant memory and thus the incremental relative stage game impact on calculated action values is small. In turn, this leads to "smooth" adjustment rules: from period to period, action choices and resultant mixed strategies do not drastically vary.

In this study we evaluate two prominent learning models of this type: Erev and Roth's [8] (1998) reinforcement learning model and Camerer and Ho's [2] (1999) experienced weighted attraction model. There are many other similarly structured models worth studying with our technique, but models in this class tend to generate similar play (Salmon [29]) and consequently most of the potential insights can be gained through evaluation of just one or two models of this type.

We believe our technique reveals properties about both human learning and learning models which cannot be discovered through either pure human experimentation or pure simulation. We present the following summary of our main results. Human play does not significantly vary depending on whether the opponent is a human or a learning algorithm. In contrast, algorithm play markedly differs when playing against a human rather than an identical algorithm in a simulation. When humans' action frequencies deviate from their Nash equilibrium proportions, the algorithms' action choice proportions respond with *systematic adjustments* towards their pure strategy best responses. Adjustments of algorithms in response to their human opponents' play result in a strikingly linear relationship between the learning models' and humans' action frequencies. Moreover, the linear relationship is suggestive of the computer players' best response correspondence. While adjustments by the algorithms are remarkably regular, their linear nature produces quite *muted response magnitudes.* In fact, magnitudes of the adjustments are too small to result in statistically significant gains in their game payoffs.

Our experimental study is just one of several that exploit laboratory control to better measure some of the latent variables underlying human play in games.[1] The composite finding of these studies paints a significantly different picture of human learning in games than the class of models considered by this study. Specifically, experiments with unique mixed strategy Nash equilibrium games have shown that humans' beliefs about opponent play are highly volatile from period to period (Nyarko and Schotter [23](2002)), and correspondingly players' mixed strategies exhibit significant variability with significant amounts of switching between pure and mixed strategy play (Shachat [30](2002)). Furthermore, humans are also quite successful at signif-

[1]For example Camerer, Johnson, Rymon, & Sen [3](1993) and Crawford, Costa-Gomes, & Broseta [7](2000) studied information look-up patterns of subjects. Also, Nyarko & Shotter [23](2002) elicited subjects' beliefs of opponents' future actions.

icantly increasing their payoffs when computerized opponents play either stationary non-equilibrium fixed mixed strategies (Lieberman [17](1961) and Fox [10](1972)) or highly serially correlated action sequences (Messick [19](1967) and Coricelli [6](2001)). In summary, human play is characterized by volatile beliefs, variable mixed strategy choices, and successful exploitation of some strategies. In contrast, the learning models we evaluate generate beliefs that are smooth, make only minor mixed strategy adjustments from period to period, and have an inability to take advantage of calculated payoff-increasing opportunities.

We proceed with a discussion of several past studies incorporating human versus computer game play. Then we present the two learning models adopted in our study. In the fourth section we discuss the games used in our experiments and our experimental procedures. Section 5 covers our experiment results, findings and interpretations. In conclusion, we integrate our results with other experimental results to provide a summary of human play in games and contrast this with current learning models.

1.2 Literature Review

In a number of past studies, researchers have used the technique of humans interacting with computerized decision makers. This technique has been used in various studies to identify social preferences in strategic settings (Houser and Kurzban [15] (2000), and McCabe et al. [18] (2001)), to establish experimental control over player expectations (Roth and Shoumaker [28] (1983) and Winter and Zamir [32] (1997)), and to identify how humans play against particular strategies in games (as in Walker, Smith, and Cox [31] (1987)). In this section, we discuss the last type of study and summarize established results on how humans play against unique minimax solutions, non-optimal stationary mixed strategies, and variants of the fictitious play dynamic (with deterministic choice rules) in the context of repeated constant-sum games with unique minimax solutions in mixed strategies.

All of the studies we discuss incorporated fixed human-computer pairs playing repetitions of one of the zero-sum games presented in Table 1.4.[2] Studies by Lieberman [17] (1961), Messick [19] (1967), and Fox [10] (1972) all contain treatments where humans played against an experimenter-implemented minimax strategy. In these studies, the human participants were not informed of the explicit mixed strategy adopted by their computerized counterparts.[3] All three of these studies reach the same conclusion: human play does not correspond to the minimax prediction, and only in the Fox study does the human play adjust - albeit weakly - towards the minimax prediction. These results are not surprising because when a "computer" adopts its minimax strategy the human's expected payoffs are equal for all of his actions.

However, this indifference is not present when the computer adopts non-minimax mixed strategies. Lieberman [17] and Fox [10] both studied human play against non-optimal stationary mixed strategies and discovered that humans do adjust their play to exploit (though not fully) their opponents. In the relevant Lieberman treatment, subjects played against the experimenter for a total of 200 periods. In the first 100 periods, the experimenter played his minimax strategy of (.25, .75) and then in the final 100 periods the experimenter played a non-minimax strategy of (.5, .5). Humans players were not informed that their opponent had adjusted his strategy. Human play adjusted from best responding approximately 20 percent of the time right after the experimenter began non-minimax play, to best responding approximately 70 percent of the time by the end of the session. However, this experimental design made it difficult to differentiate between the attractiveness of the minimax strategy and the best response since they both lay in the direction of this observed shift.

---

[2]In some of these studies the experimenters implemented stationary mixed strategies by using pre-selected computer generated random sequences in their non-computerized experiments.

[3]When reported, humans were instructed something similar to, "The computer has been programmed to play so as to make as much money as possible. Its goal in the game is to minimize the amount of money you win and to maximize its own winnings." (Messick [19], page 35)

In one of Fox's treatment, humans played 200 periods against a computer which played the non-minimax mixed strategy (.6, .4) for the entire session. This design placed the human's best response, (1, 0), on the opposite side of (.5, .5) from the human's minimax strategy, (.214, .786). Human play started slightly above (.5, .5) and then slowly adjusted towards the pure strategy best response over the course of the experiment. Specifically, humans were best responding approximately 75 percent of the time by the latter stages of the experiment. These experiments established that humans will adjust their behavior to take advantage (but not as much as possible) of exploitable stationary mixed strategies.

Messick [19] and Coricelli [6] (2001) conducted experiments to evaluate how humans respond when playing against variations of fictitious play.[4] These experiments are notable in that the computer's strategy was responsive to the actions selected by its opponent. Messick studied humans matched against two fictitious play algorithms: one with unlimited memory and the other with only a five period memory. Against unlimited memory fictitious play, humans earned substantially more than their minimax payoff level. Humans earned an even greater average payoff against limited memory fictitious play. In the study by Coricelli, there are two treatments (both utilizing the game form introduced by O'Neill [24] (1985)) in which humans play against unlimited memory fictitious play and against the same algorithm that has a bias in the beliefs that subjects tend not to repeat their "P" action. In both treatments humans win significantly more often against the algorithms than they do against humans.[5] Establishing that humans can "outgame" these algorithms is significant, though not surprising. It is well known that in games with unique mixed strategy equilibrium, the fictitious play algorithm can generate strong positively se-

[4]In the original formulations of fictitious play (Brown [1](1951) and Robinson [26](1951)) a player uses the empirical distribution of the entire history of his opponent's action choices as his belief of the opponent's current mixed strategy and then chooses a best response to this belief.

[5]Human versus human data for this conclusion are taken from O'Neill [24] (1985) and Shachat [30](2002).

rially correlated action choices that are easily exploited.[6] It was this speculated vulnerability that partially motivated game theorists to propose and study adaptive learning models which incorporated probabilistic choice as a key component.[7]

To summarize, through the use of experiments pitting humans against algorithms in constant sum games with strictly mixed strategy solutions we have learned (1) that humans do not tend to play their minimax strategy in response to opponents playing their minimax strategy, (2) humans exploit (but not fully) opponents who play mixed strategies significantly different from their minimax strategy, and (3) humans exploit adaptive algorithms which generate highly serially correlated action choices.

1.3 Response Algorithms

A large number of adaptive behavioral models have been recently introduced into the literature on games. Most of these models have similar frameworks with two main components. First, each player retains a latent "score" for each of his available actions, and the score of each action is adjusted after each game iteration based on the outcome. Second, each player chooses an action according to a probability distribution that places higher probability on actions with higher scores. For obvious reasons we limit the number of models we consider. We focus on two of the more popular models in the experimental games literature: Erev and Roth's [8](1998) Reinforcement model and Camerer and Ho's [2](1999) Experience Weighted Attraction model.

1.3.1 Reinforcement Learning

Erev and Roth's model (hereafter ER) is motivated by the reinforcement hypothesis from psychology: an action's score is incremented by a greater amount when it results in a "positive" outcome rather than a "negative" outcome. More formally, let $R_{ij}(t)$

[6] See Jordan [16](1993) and Gjerstad [13](1996).

[7] For example, see cautious fictitious play proposed by Fudenberg and Levine [11] (1995), and the two learning models we utilize in this study.

denote player i's score for his jth action prior to the game at iteration t: let $\sigma_{ij}(t)$ denote the probability that i chooses j at iteration t: and let X_i denote the set of player i's possible stage-game payoffs. The two initial conditions for the dynamical system are (1) that at the initial iteration, each of a player's actions is selected with the same probability of being selected (i.e., in our two games, $\sigma_{ij}(1) = .5$ for each player i and each action j) and (2) that

$$R_{ij}(1) = \sigma_{ij}(1)S(1)\overline{X_i},$$

where $S(1)$ is an unobservable strength parameter, which influences the player's sensitivity to subsequent experience, and $\overline{X_i}$ is the absolute value of player i's payoff averaged across all action profiles.

After each iteration, each action's score is updated as follows

$$R_{ij}(t+1) = (1-\varphi)R_{ij}(t) + \left((1-\varepsilon)I_{(a_i(t)=j)} + \frac{\epsilon}{2}\right)(\pi_i(j, a_{-i}(t)) - \min\{X_i\}),$$

where ϕ is an unobservable parameter that discounts past scores, $I_{(a_i(t)=j)}$ is an indicator function for the event that player i selected action j in period t, ε is an unobservable parameter determining the relative impacts on the scores of the selected versus the unselected action, and $\pi_i(j, a_{-i}(t))$ is i's payoff when he plays action j against the deleted action profile $a_{-i}(t)$. Also player i's minimum possible payoff for any action profile, $\min\{X_i\}$, is subtracted from $\pi_i(j, a_{-i}(t))$ as a normalization to avoid negative scores. The second component of the model, a probabilistic choice rule, is specified as

$$\sigma_{ij}(t) = \frac{R_{ij}(t)}{\sum_k R_{ik}(t)}.$$

For each game we consider, parameters of the model are estimated along the lines suggested by Erev and Roth. We estimate the values of $S(1)$, ϕ, and ε by minimizing the mean square error of the predicted proportions of Left play in 20-period trial blocks for the human versus human treatments. More specifically, for each fixed triple of parameter values from a discrete grid, we proceed as follows: we simulate the play of 500 fixed pairs engaging in 200 iterations. and then we calculate separately the frequency of Left play by the 500 Row players and by the 500 Column players in each 20-period block. These frequencies are the model's predictions for that triple of parameter values. The grid is then searched for the optimal parameters.

1.3.2 Experience-Weighted Attraction

We use the version of EWA developed by Camerer & Ho [2](1999). While the structure of the EWA formulation is similar to ER learning. it adopts a different parametric form of probabilistic choice and it updates actions' scores according to what actions actually earned in past play. and what actions hypothetically would have earned if they had been played.

According to EWA subjects choose stage-game actions probabilistically according to the logistic distribution

$$\sigma_{ij}(t) = \frac{e^{\lambda R_{ij}(t)}}{\sum_k e^{\lambda R_{ik}(t)}},$$

where at stage t player i chooses action j with probability $\sigma_{ij}(t)$. where λ is the inverse precision (variance) parameter, and where $R_{ij}(t)$ is a scoring function. as in the ER model, albeit defined (i.e., updated) differently. The updating of $R_{ij}(t)$ involves a "discounting" factor $N(t)$. which is updated according to $N(t+1) = \rho N(t) + 1$ for $t \geqq 1$. where ρ is an unobservable discount parameter and $N(1)$ is an unobservable parameter. interpreted as the strength of experience prior to the beginning of play. The score $R_{ij}(t)$ is updated as follows:

$$R_{ij}(t+1) = \frac{N(t)\phi R_{ij}(t) + ((1-\varepsilon)I_{(a_i(t)=j)} + \frac{\varepsilon}{2})\pi_i(j, a_{-i}(t))}{N(t+1)}.$$

where $\pi_i(j, a_{-i}(t))$. ϕ. and ε are interpreted as in the Erev and Roth model. Initial scores. $R_{ij}(1)$ for each i and j. are additional unobservable parameters.

Parameters of the EWA model are estimated via maximum likelihood. It is worth noting that EWA is a flexible specification that includes several other models as special cases. For example, a simple reinforcement learning model is generated when $N(1) = 0$, $\varepsilon = 0$, and $\rho = 0$; and probabilistic fictitious play is generated when $\varepsilon = \rho = \phi = 1$.[8]

1.4 Experimental Procedure

There are three basic steps in our experimental methodology. First, we collect baseline data samples consisting of fixed human versus human pairs playing 100 or 200 rounds with one of two 2×2 games. Second, we estimate parameters for the two learning models separately for each of the two games. In the third step, a new sample of humans play one of the two games against an estimated learning algorithm. We proceed by describing the two games we used and then present more details on the outlined steps.

1.4.1 The Two Games

The first game we consider is a zero-sum asymmetric matching penny game called Pursue-Evade. This game was introduced by Rosenthal, Shachat, and Walker [27](2002) (hereafter RSW). The normal form representation of the game is given in Table 1.1. The Nash equilibrium (and minimax solution) of this game is symmetric: each player chooses Left with probability of two-thirds.

There are several reasons why this game is a strong candidate to use in our study. (1) Zero-sum games eliminate social utility concerns often found in experimental studies of games, thereby mitigating some behavioral effects that might arise if a human suspects he is playing against a computer rather than another human. (2) With some standard behavioral assumptions, the repeated game has a unique Nash equilibrium path which calls for repeated play of the stage game Nash equilibrium.

[8] We refer the reader to Camerer and Ho [2](1999) for more discussion of how EWA can emulate various models and for a more complete interpretation of the parameters.

This eliminates potential repeated game effects that the algorithms are not designed to address. (3) Pursue-Evade is a simple game in which the Nash equilibrium predictions differ from equiprobable choice. This generates a powerful test against the alternative hypothesis of equiprobable play.

We selected our second game to pose a more serious challenge to the learning algorithms. We refer to our second game, presented in Table 1.2, as Gamble-Safe. Each player has a Gamble action (Left for each player) from which he receives a payoff of either two or zero and a Safe action (Right for each player) which guarantees a payoff of one. This game has a unique mixed strategy in which each player chooses his Left action with probability one-half, and his expected Nash equilibrium payoff is one. The difference between the Nash equilibrium and the minimax solution makes this game challenging for the learning models. Notice that this game is not constant-sum: therefore the minimax solution need not coincide with the Nash equilibrium. In this game, Right is a pure minimax strategy for both players that guarantees a payoff of one. A game for which minimax and Nash equilibrium solutions differ but generate the same expected payoff is called a non-profitable game.[9] The potential attraction of the minimax strategy can (and does) prove to be difficult for the learning algorithms which, loosely speaking, have best response flavors.

1.4.2 Protocols

Human versus Human Baselines For the human versus human baseline play in the Pursue-Evade game we use the data generated by RSW. In their hand-run experiments, a pair of subjects were seated on the same side of a table with an opaque screen dividing them. The Evader was given an endowment of currency. Each player was given two index cards: one labelled Left and the other labelled Right. At each iteration the players slid the chosen card to the experimenter seated across from them.

[9] Morgan and Sefton [21](2002) present an excellent study of human play in non-profitable games.

Then the experimenter simultaneously turned over the cards, executed the payoffs, and recorded the actions. Twenty pairs of human subjects played this treatment: fourteen for 100 periods and six for 200 periods.

The human versus human baseline experiments for the Gamble-Safe game were executed via computerized interaction. Each subject was seated at a separate computer terminal such that no subject could observe the screen of any other subject. All subjects participated in either 100 or 200 repetitions of the game maintaining a constant role throughout.

The protocol for each period was as follows. At the beginning of each repetition, a subject saw a graphical representation of the game on the screen. Column players had the display of their game transformed so that they appeared to be a Row player. Thus, all subjects selected an action by clicking on a row, and then confirmed their selection. Subjects were free to change row selections before confirmation. Once an action was confirmed, a subject waited until his opponent also confirmed an action. Then a subject saw the outcome highlighted on the game display, as well as a text message stating both actions and the subject's earnings for that repetition. Finally, at all times a history of past play was displayed to the subjects. This history consisted of an ordered list with each row displaying the number of the iteration, the actions selected by both players, and the subject's earnings.

Human versus Algorithm Treatments We conducted our hybrid treatments using both the experimental software and protocol used for the Gamble-Safe game baseline.[10] In these treatments, two human subjects played against each other for the first 23 repetitions of the game. Then, unbeknownst to the human pair, they stopped playing against each other and for the remainder of the experiment they each played against a computer that implemented either the EWA or ER learning algorithm.

We adopted a simple technique to make the "split" seamless from the subjects'

[10]For the Pursue-Evade game, the Evader was given a currency endowment.

perspectives. From period twenty-four on, the two human/computer pairs had no interaction except for the timing of how action choices were revealed. Specifically, although the computers generated their action choices instantly, the computers didn't reveal their choices until both humans had selected their actions. This protocol preserved the natural timing rhythm established by the humans in the first twenty-three stage games.

The non-human opponent treatments began with an initial stage of human versus human play in order to give the algorithms a better chance of successfully "standing in" for the human whose place it takes. Both ER and EWA rely on actions' scores to determine the chosen action in a probabilistic manner. During the first 23 repetitions, we allow these scores to "prime" themselves with the play generated by the subjects. (Although updating of scores is determined by the parameter estimates obtained from the baseline treatments). That is, even though the response algorithms are not selecting actions during the first 23 repetitions, the scores are still being updated according to the specifications of the previous section. For example, consider the 24th repetition of a game. The human Row player is now facing a computer that plays the Column position. Moreover, during the first 23 repetitions, the computer Column player updates the scores associated with Column's actions based on the observed actions of both humans. We conjecture this will de-emphasize the impact of the estimated initial score values of the actions.

In summary, we have two treatment variables, the stage game and the type of opponent. The data samples we have for each treatment cell are given in Table 1.3.[11]

[11] We explain in the next section why we have no observations for the EWA Gamble-Safe treatment.

1.5 Baseline Results, Model Estimation, and Model Simulation

Our experimental baselines are the human versus human play in each of the games we consider. Inspection of the aggregate data reveals that play in the two games departs from the Nash equilibrium and the dynamic features of the data suggest non-stationarity of play. After estimating the unobserved parameters of the learning models, we simulated large numbers of experiments based upon these estimated versions. Simulations reveal that the learning models generate aggregate choice frequencies similar to the experimental data, but only weakly mimic the experimental data time series. Furthermore, the simulations do not reveal striking differences between the two learning models.

We use the data from RSW as the Pursue-Evade game baseline data set. Figure 1.1 shows contingency tables for the data aggregated across subject pairs and stage games. A graph of the time series of the average proportion of Left play for the Row and Column players is shown below each table. Each observation in a series is the average across a twenty period time block. As noted by RSW, the contingency table is distinctly different from the Nash equilibrium predictions (the numbers in parentheses) and Column subjects play Left significantly more often than the Row subjects.[12] In the block average time series, we see that the Column series almost always lies above the Row series and that both series exhibit an increasing trend.

Using this data, RSW estimated the parameters of both the ER and EWA models. As noted by RSW, both models have some success in explaining the deviation. Using the estimated models we simulated 10,000 experiments of twenty pairs playing the Pursue-Evade game for 200 iterations. Averages from the 10,000 simulated experiments were used to construct contingency tables and time series in the same

[12]Moreover, the Column subject plays Left more frequently than his Row counterpart in almost all pairs.

format as those presented for the baseline data. These results are presented alongside the baseline results in Figure 1.1. Unsurprisingly, given the respective objective functions used to select model parameters, casual observation suggests that the EWA model generates an expected contingency very close to the human baseline and the ER model more accurately mimics dynamics in the times series.

We provide a corresponding analysis for the Gamble-Safe game in Figure 1.2. In the contingency table for the baseline data we observe that the Row subjects play Right significantly more than Left, while Column subjects played Left more often. This result partly comes from two pairs in which the Row and Column subjects' action profile sequence eventually converged to the profile (Safe, Gamble). This is evident around the midpoint of the times series for the baseline treatment, where we see the Column and Row subjects' series diverge.

This convergence to minimax play by the Row subjects in these two pairs is problematic for the maximum likelihood estimation used in the EWA model. Specifically, the long strings of Left by Column leads the EWA model to assign a near zero probability to Right (Safe) by Row for any possible parameter values. However, since Row is repeatedly choosing Right in these instances there is a zero likelihood problem in estimating the EWA parameters. Rather than violate the maximum likelihood criterion for parameter selection specified by Camerer and Ho we chose not to conduct a Human versus EWA treatment for this game.

Since the ER model parameter selection does not rely upon maximum likelihood estimation we obtain estimates which generate the best fit for the baseline data. Interestingly we see that the ER contingency table is remarkably similar to the baseline table. However, the predicted ER dynamics are excessively smooth and do not resemble the baseline time series. We believe this failure results from the inability of the model to incorporate the heterogenous behavior the occurs when some players adopt the minimax strategy and other players adopt adaptive strategies.

Comparison of the experimental data to simulations based upon estimated versions of the learning models suggests that the learning models successfully capture some features of the humans' disequilibrium behavior. However, time series views of the simulation data exhibit much smoother and less extreme dynamics than the experiment data, which suggests that learning models are not as responsive as humans and tend to simply "fit" aggregate human choice frequencies.

1.6 Analysis of Human/Algorithm Interaction

In the previous section we used a common technique of comparing experimental data to simulation results to evaluate the appropriateness of alternative learning models. Now we proceed to present analysis of human/algorithm interaction which reveals a significantly different story. Action choice frequencies by the algorithms are more responsive to opponents' play than the humans' action choice frequencies. Moreover, the action frequencies by each algorithm adjust linearly toward its best response to its opponent's non-equilibrium action frequencies. However, the magnitude of these adjustments is too small to generate payoff gains for the learning algorithms. Finally, we see that human play does not vary significantly whether the opponent is another human or a learning algorithm. Examination of the human/human experiments and the model simulations don't reveal these results.

1.6.1 Learning Algorithm Response to Opponents' Play

We now introduce pair-level data to better highlight differences in play across treatments. Inspection of the Row and Column players' proportions of Left play in each pair reveals surprising differences from purely human play and the simulations reported in the prior section. The learning algorithms are quite responsive to human deviations from Nash equilibrium play. Specifically, the algorithms' frequencies of Left play have a striking linear correlation to their human opponents' Left play proportions. Moreover, these linear relationships are consistent with a linear approximation of the algorithms' best response correspondences.

These results are most easily seen in Figures 1.3 - 1.5. Each of these figures is a 2×2 array of scatterplot panels. The rows in the panel array correspond to the decision maker type for the Row player: the top row indicates human decision maker and the bottom row indicates computer decision maker. Similarly the columns of the panel array correspond to the decision maker type for the Column player: the left column for human and the right column for computer. Hence the upper left panel is from the human/human baselines, the lower right panel is from the algorithm/algorithm simulations, and the off diagonal panels are from the human/algorithm and algorithm/human experiments.

The scatterplots show the proportions of Left play by the Row and Column players in each pair after the first 23 iterations. In the simulation panel we only use the data from a single simulated experiment with twenty pairs playing 200 iterations. Also, each of these scatterplots displays a regression line of the Row proportion Left regressed on the Column proportion Left, and a dashed line for the computer's best response correspondence.

Examination of these figures reveals important common results across the two games and learning models. Comparisons between the two main diagonal panels reveal consistent differences and similarities between human/human play and pure simulations of model interaction. Both types of interactions generate uncorrelated "clouds" with the simulations' clouds exhibiting much smaller dispersion.[13] This raises the issue of whether the learning models are quite aggressive in adaptation and quickly converge to an equilibrium or instead the models are quite insensitive to opponents' play and just stubbornly mimic human aggregate frequencies. We can ask a similar question regarding human play. Do the humans' dispersed clouds result from high variance in the humans' propensities to play Left coupled with little response

[13]F-tests reject the significance of the presented regression lines; this gives statistical support for claims of no correlation.

to the opponents' play or is it the result of differential skill in human play in which some humans more successfully exploit other humans' play?

Inspection of the human and learning algorithm interactions answers these questions. In contrast to the model simulations and human/human play, the scatter plots of human and learning algorithm interactions (found in the off-diagonal panels of Figures 1.3 - 1.5) exhibit strongly correlated interactions. This is evident by the tight clustering of the data along the plotted regression lines. Also, in each case the regression line is in the direction of the computer players' best response correspondence (the dashed correspondence given on each scatterplot). In other words, the computer "better" responds instead of best responds. This is best illustrated by an observation in the upper right scatter plot of Figure 1.3. In this scatterplot, Column ER players play against human Row players in the Gamble-Safe game. One of the human players chose his Minimax strategy, Right, exclusively and his computer ER opponent best responded to this only about 70 percent of the time. Hence, we see that (1) the frequency of Left by the learning algorithms move toward (but not all the way to) the best response to their opponents' frequencies, and (2) the magnitude of these responses by the algorithms is described by a surprisingly predictable linear relationship.

Table 1.5 gives some quantitative support for these observations by presenting the OLS results of regressing the learning algorithms' Left frequencies on their human counterparts' Left frequencies.[14] A learning algorithm that is highly sensitive and adjusts systematically to opponents' play should generate regressions that explain a high percentage of the variance of the algorithm's Left frequencies, and the estimated slope coefficient should be consistent with the best response correspondence. These features are found in the Table 1.5 regressions: the slope of each regression has the

[14]Note that the regression lines displayed in the upper-right panel of Figures 1.3 - 1.5 differ from the regression results in Table 1.5. This is because the figures show the plot of Row proportion Left regressed on Column proportion Left, while the table reports Computer proportion Left regressed on Human proportion Left.

correct sign, three of the regressions have exceedingly large adjusted R^2 statistics, and a fourth is still quite large considering the data is cross sectional. These adjusted R^2 results reflect the tight clustering to the fitted regression line observed in the scatterplots and correspondingly the detection and systematic reaction by the learning algorithms to calculated payoff-increasing opportunities. Correspondingly, F-tests for these four regressions do not reject the significance of the regressions at the 5 percent level of significance. Interestingly, the two cases where F-tests reject the regressions are when the EWA and ER algorithms assume the Column role in the Pursue- Evade game. We do not see a reason for the differential performance, but do note that the mean of the computers' data is close to their minimax strategy in this case.

1.6.2 Learning Algorithms' Lack of Effective Exploitation

Previous arguments established that the learning algorithms sensitively detect opponents' exploitable action choice frequencies and then the algorithms respond with a systematic but tempered reaction in the direction of their best response. However, we will now see that these statistically significant responses are too weak in magnitude to generate statistically significant payoff gains. Table 1.6 presents the average stage game winnings for all decision maker types when pitted against a human for each role and game. If the learning algorithms successfully exploit human decision makers we would expect the algorithms in each game and role to have greater winnings than a human when playing against a human in the competing role. The average stage game winnings in Table 1.6 do not exhibit this trait.

The reported average stage game payoff statistics are calculated by first taking the total session payoffs for each decision maker who plays against a human, and dividing by the number of stage games played.[15] Then we partition these decision makers according to the game played, role played, and decision maker type. Finally,

[15] We normalize this way because in the baseline data for Pursue-Evade and Gamble-Safe some pairs played 100 stage games and others 200.

we report the average stage game payoffs across decision makers in each partition. For each game and player role we conduct t-tests with the null hypothesis that on average a non-human decision maker earns the same as a human when the opponent is a human. At a 5 percent level of significance we fail to reject the null hypothesis in four out of the six tests. In the two rejections, the human average exceeds the algorithm average.

Why don't the learning algorithms, which are sensitive and responsive to opponent play, generate higher payoffs than humans? The answer is twofold. First, the two games we consider have fairly flat payoff spaces in the mixed strategy domains presented in Figures 1.3 - 1.5. Thus a pair must be far removed from the Nash equilibrium to generate large payoff deviations from Nash equilibrium payoffs. Second, whenever the algorithm calculates a difference between its two action scores, it adjusts choice probabilities without assessing whether this difference is statistically significant. If this difference is not statistically significant, then there is no adjustment that can generate a real increase in payoff. Alternatively, an adjustment to a statistically significant score difference may also fail to generate a real increase in payoffs. Why? We have already seen that algorithms adjust in statistically significant ways, but these adjustments are relatively small in magnitude. These weak adjustments are the product of probabilistic choice rules, which were adopted to avoid generating transparent serially correlated choice patterns.

1.6.3 Human Play Conditional On Opponent Decision Maker Type

Past studies have demonstrated that humans play differently against Nash equilibrium strategies than they do against other humans. However, we also have presented arguments that play by learning algorithms is more responsive to opponents' decisions than human play is. A natural question to ask is, do humans play differently against learning algorithms than they do against other humans? To answer this question we compare the empirical distributions of the proportions of Left play by humans when facing the different decision-making types as presented in the scatter plots of Figures 1.3 - 1.5. We report a series of Kolmogorov-Smirnov two-sample goodness-of-fit tests (hereafter denoted KS) comparing the distributions of Left play proportions against human opponents to Left play proportions against the alternative algorithms. The main result is that we can't find differences in human play except in the case when the human is the Row player in the Pursue-Evade game.

Figure 1.6 shows the empirical CDFs of proportion of Left play by human Row players as they face human, ER, and EWA Column decision maker types in the Pursue-Evade game. Additionally, the figure reports the results of Kolmogorov-Smirnov tests of whether the Human's distribution of Left play frequencies differs when facing an algorithm opponent as opposed to a human opponent. Previously we have observed that the learning algorithms performed differently in the Column role of the Pursue-Evade game than in any other situation. This trend continues as the proportions of Left by humans in the Row role are significantly different when facing each learning algorithm than when facing another human.

Next we consider the CDFs generated by human Column players when playing against Human, ER, and EWA Row decision maker types in the Pursue-Evade game. We see in Figure 1.7 that play against human opponents is statistically indistinguishable from play against both EWA and ER opponents.

Next, we turn our attention to human play in the Gamble-Safe game. Figure 1.8 shows that human Row players' CDFs of proportion of Left play are not statistically different as they face Human and ER Column decision maker types. Finally, the CDFs and associated KS tests generated by human Column players in the Gamble-Safe game are shown in Figure 1.9. We see that play against human opponents differs from play against ER opponents at the six-percent level of significance.

1.7 Discussion

Through experiments in which humans play games against computer- implemented learning algorithms, we have established that humans do not detect nor exploit the estimated models' non-stationary but rather smooth mixed strategy processes. Furthermore, our experiments provide a unique evaluation of the learning models by establishing that the models are more sensitive than humans in detecting exploitable opponent play. However, the models' corresponding mixed strategy adjustments are systematic but too weak to increase their payoffs.

Recall the common formulation of both the ER and EWA models. We see their adaptive functions generate sequences of action scores which adjust smoothly across periods because stage game outcomes weakly impact action scores. Furthermore, our experiments reveal that the learning algorithms' mixed strategies respond uniformly and linearly to opponents non-equilibrium action choice frequencies. The algorithms' uniform better responses are too weak to generate significant payoff gains.

Our study, in conjunction with other studies, reveals a different depiction of human learning in games. First, through the technique of pitting humans against algorithms we know that humans successfully increase their payoffs (but not as much as possible) against non- optimal but stationary mixed strategy play and against adaptive play that generates highly serially correlated action sequences. On the other hand humans do not exploit the subtle dynamic mixed strategy processes of the learning models examined in this paper.

Some sources of behavioral departure between learning models and humans are identified in experiments that elicit subjects' beliefs (Nyarko and Schotter [23]) or subjects' mixed strategies (Shachat [30]). Elicited beliefs are highly volatile and often times correspond to a belief that one action will be chosen with certainty. Similarly elicited mixed strategies show erratic adjustments and a significant amount of pure strategy play.

This set of stylized facts should set benchmarks which new learning models need to explain. Furthermore, the use of human/algorithm interactions can play an important role in future efforts to identify how humans adapt in strategic environments. First, the technique brings increased power in evaluating proposed models. Second, the adoption of carefully selected algorithms will facilitate further identification of human learning behavior. For example, one could determine the extent of human ability to exploit serially correlated strategies by altering the variance incorporated in the probabilistic choice rule of a cautious fictitious play algorithm. In this instance, the algorithm is not being evaluated: rather it is a carefully chosen stimulus to yield informative measurements of human behavior.

		Column player L	Column player R
Row player	L	1, -1	0, 0
	R	0, 0	2, -2

TABLE 1.1. Pursue-Evade

		Column player L	Column player R
Row player	L	2, 0	0, 1
	R	1, 2	1, 1

TABLE 1.2. Gamble-Safe

	Opponent treatment		
Game treatment	Human	EWA	ER
Pursue-evade	40	30	30
Gamble-safe	34	0	24

TABLE 1.3. Number of subjects that participated in each treatment.

Zero-Sum Games Used In Previous Studies

(Humans are row player. Payoffs are for row player. minimax strategy proportions are next to action names)

Lieberman

	E1 (.25)	E2 (.75)
S1 (.75)	3	-1
S2 (.25)	-9	3

Messick

	A (.556)	B (.244)	C (.2)
a (.400)	0	2	-1
b (.111)	-3	3	5
c (.489)	1	-2	0

Fox

	a1 (.426)	a2 (.574)
b1 (.214)	6	-5
b2 (.786)	-2	1

Coricelli (Introduced by O'Neill)

	G (.2)	R (.2)	B (.2)	P (.4)
G (.2)	-5	5	5	-5
R (.2)	5	-5	5	-5
B (.2)	5	5	-5	-5
P (.4)	-5	-5	-5	5

TABLE 1.4. Zero-sum games used in previous studies.

TABLE 1.5. OLS regression results.

OLS Regression Results

Computer Left Frequency = $\alpha + \beta *$ Human Left Frequency

Game	Algorithm	Algorithm Role	Human Role	α (t-stat)	β (t-stat)	Adjusted R-square	F-Stat	F-Stat P-value
Gamble-Safe	RE	Row	Column	0.07	0.66	0.85	62.40	0.00
				(2.11)	(7.90)			
Gamble-Safe	**RE**	**Column**	**Row**	**0.75**	**-0.69**	**0.96**	**276.54**	**0.00**
				(40.03)	**-(16.63)**			
Persue-Evade	RE	Row	Column	-0.26	1.16	0.85	82.92	0.00
				(-2.89)	(9.11)			
Pursue-Evade	**RE**	**Column**	**Row**	**0.72**	**-0.21**	**0.05**	**1.68**	**0.22**
				(9.40)	**(-1.30)**			
Pursue-Evade	EWA	Row	Column	0.28	0.29	0.29	6.64	0.02
				(3.24)	(2.58)			
Pursue-Evade	**EWA**	**Column**	**Row**	**0.69**	**-0.20**	**0.03**	**1.42**	**0.25**
				(8.85)	**(-1.19)**			

TABLE 1.6. Average decision-maker payoff.

Average Stage Game Payoffs For Decision Makers When Facing A Human Opponent

Game	Human Role	Human's Opponent	Decision Maker Avg. Payoff	T-test Statistic	Approx. d.o.f.	P-value
Gamble-Safe	**Row**	**Human Column**	**1.0776**	***	***	***
Gamble-Safe	**Row**	**RE Column**	**1.0786**	**-0.012**	**23**	**0.990**
Gamble-Safe	Column	Human Row	0.9888	***	***	***
Gamble-Safe	Column	RE Row	0.8983	2.187	25	0.038
Pursue-Evade	**Row**	**Human Column**	**-0.6709**	***	***	***
Pursue-Evade	**Row**	**RE Column**	**-0.6829**	**0.498**	**32**	**0.622**
Pursue-Evade	**Row**	**EWA Column**	**-0.7205**	**2.312**	**33**	**0.027**
Pursue-Evade	Column	Human Row	0.6709	***	***	***
Pursue-Evade	Column	RE Row	0.6395	1.285	31	0.208
Pursue-Evade	Column	EWA Row	0.6395	1.557	32	0.129

All Human Baseline

Row Action	Column Action: Left	Column Action: Right
Left	**.441** (.444)	**.169** (.222)
Right	**.276** (.222)	**.114** (.111)

Estimated RE Model

Row Action	Column Action: Left	Column Action: Right
Left	**.502** (.444)	**.161** (.222)
Right	**.208** (.222)	**.129** (.111)

Estimated EWA Model

Row Action	Column Action: Left	Column Action: Right
Left	**.442** (.444)	**.204** (.222)
Right	**.239** (.222)	**.110** (.111)

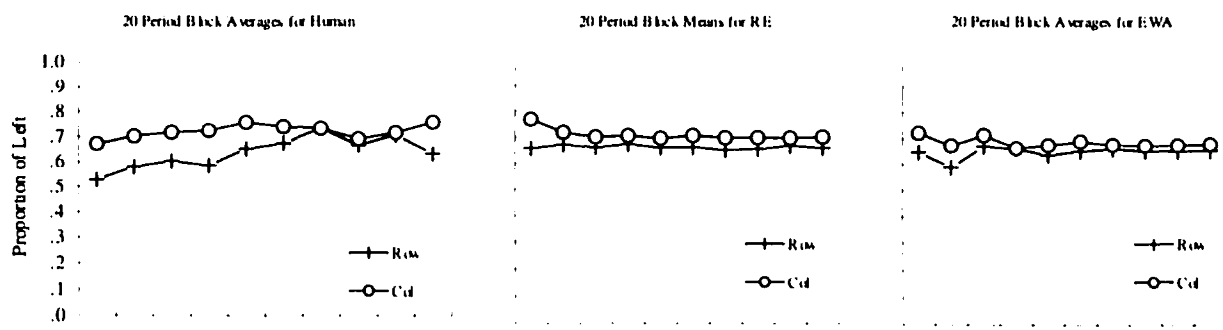


FIGURE 1.1. Baseline Data and Estimated Model Summary for Pursue-Evade Game.

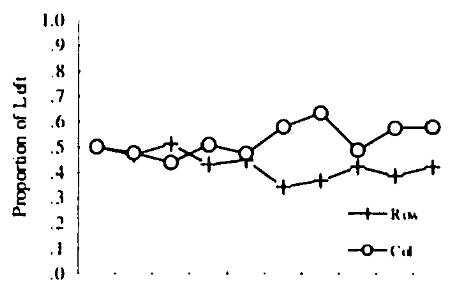
All Human Baseline

Row Action		Column Action: Left	Column Action: Right
	Left	**.212** (.250)	**.220** (.250)
	Right	**.309** (.250)	**.259** (.250)

Estimated RE Model

Row Action		Column Action: Left	Column Action: Right
	Left	**.226** (.250)	**.215** (.250)
	Right	**.297** (.250)	**.262** (.250)

20 Period Block Averages for Human



20 Period Block Means for RE

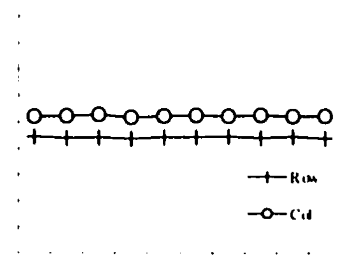


FIGURE 1.2. Baseline Data and Estimated Model Summary for Gamble-Safe Game.

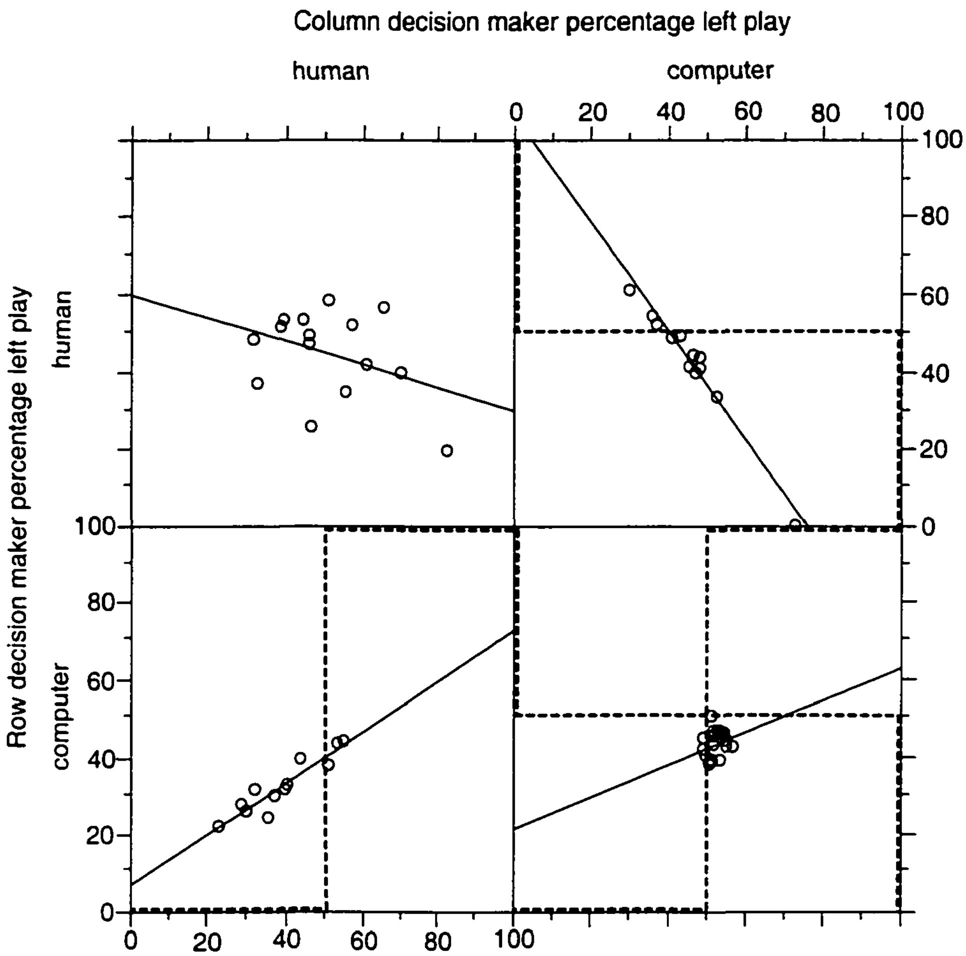


FIGURE 1.3. Gamble-Safe joint densities of proportion Left: ER interactions.

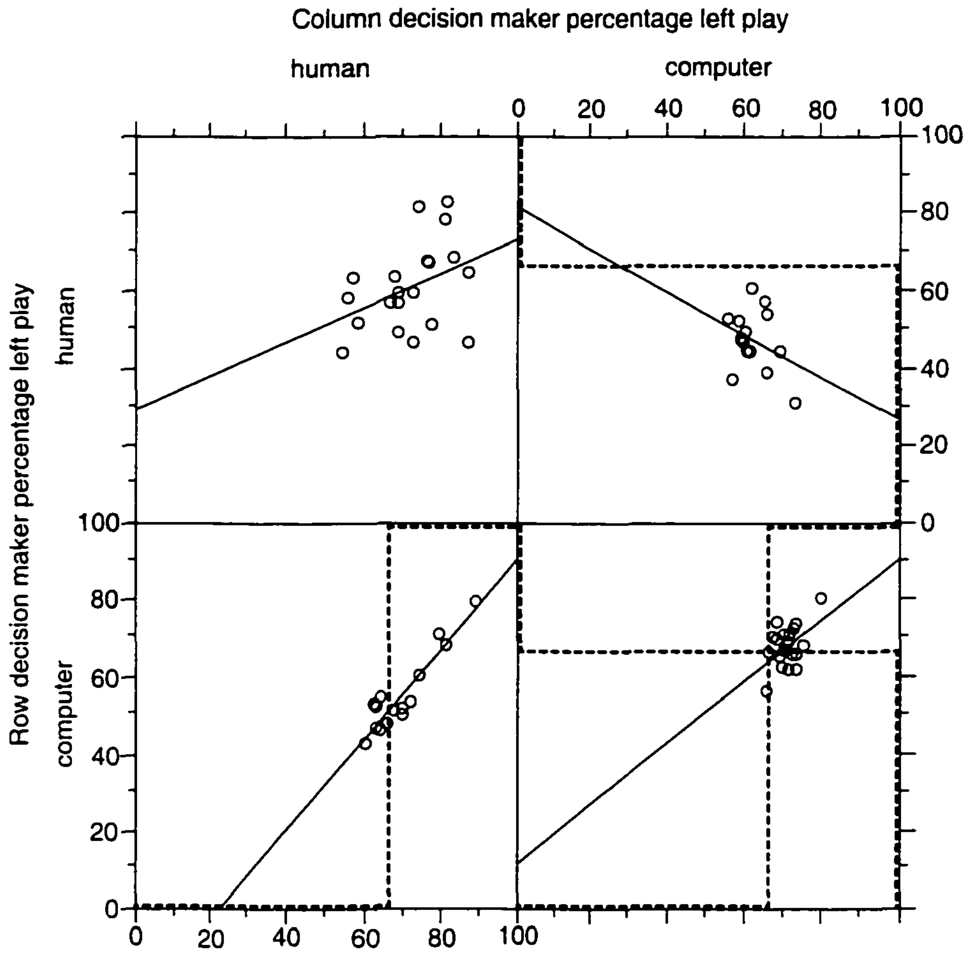


FIGURE 1.4. Pursue-Evade joint densities of proportion Left: ER interactions.

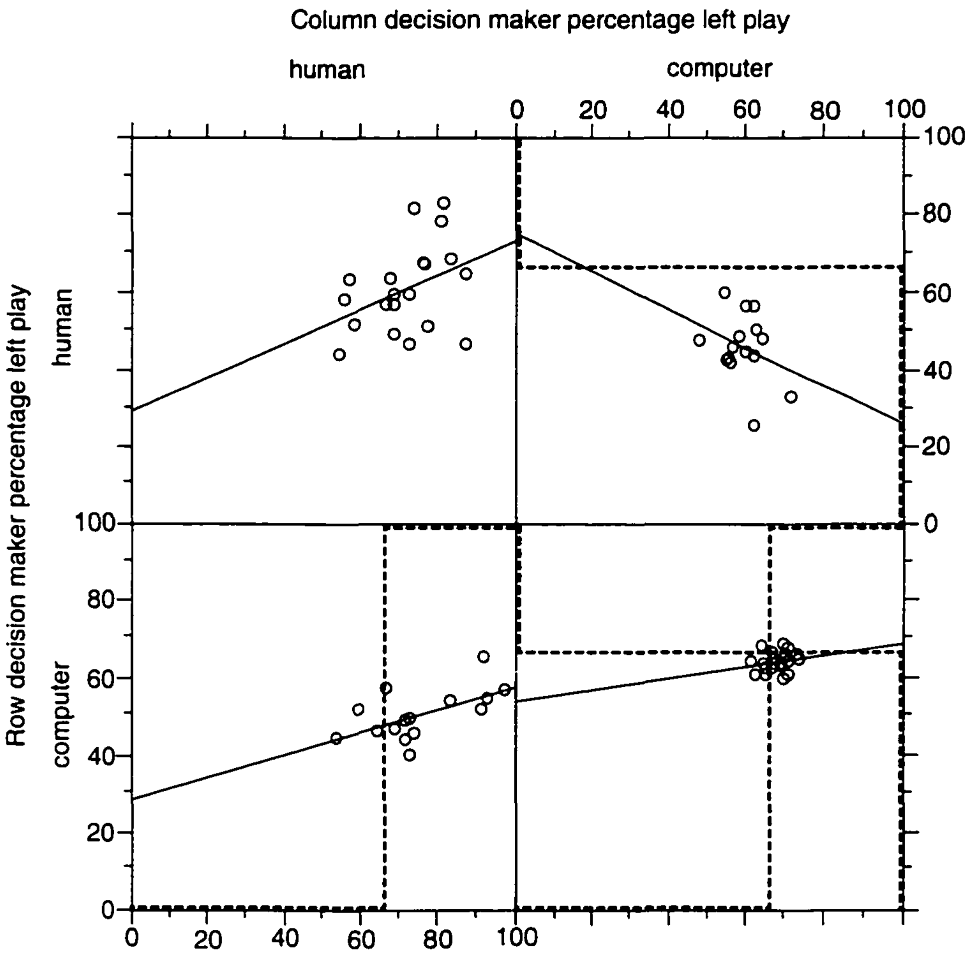
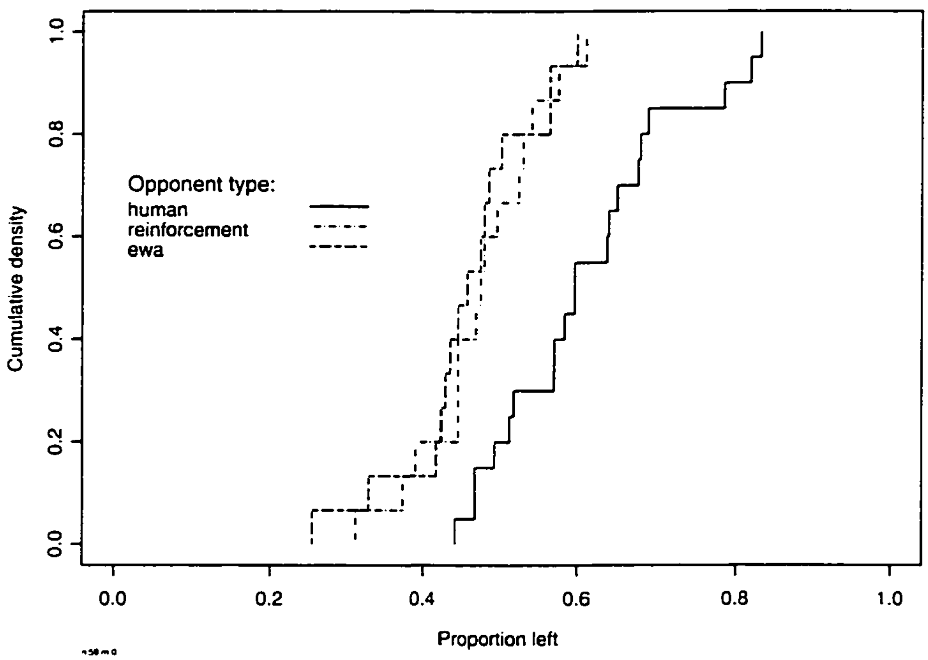
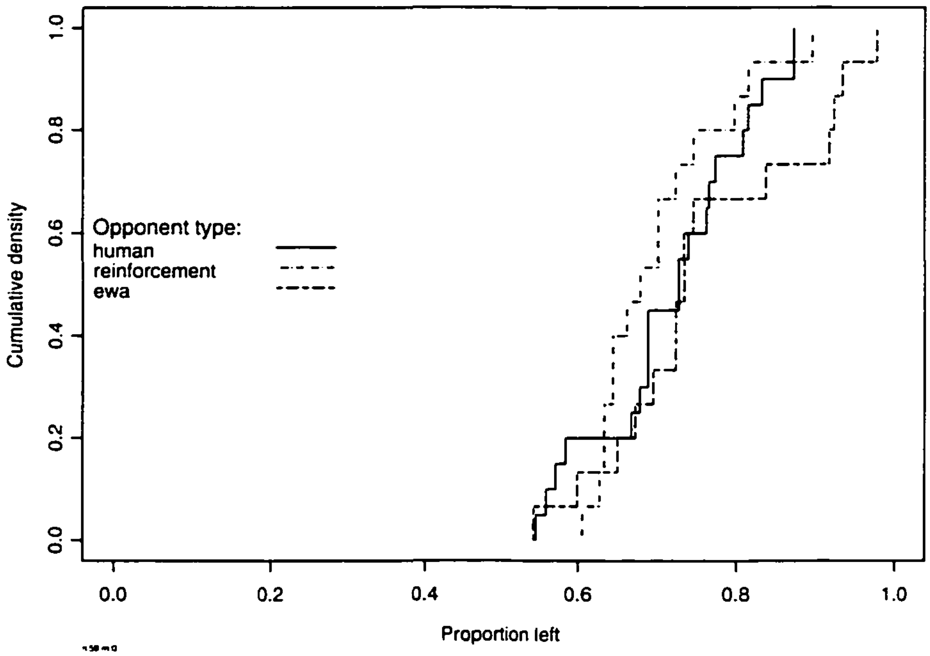


FIGURE 1.5. Pursue-Evade joint densities of proportion Left; EWA interactions.



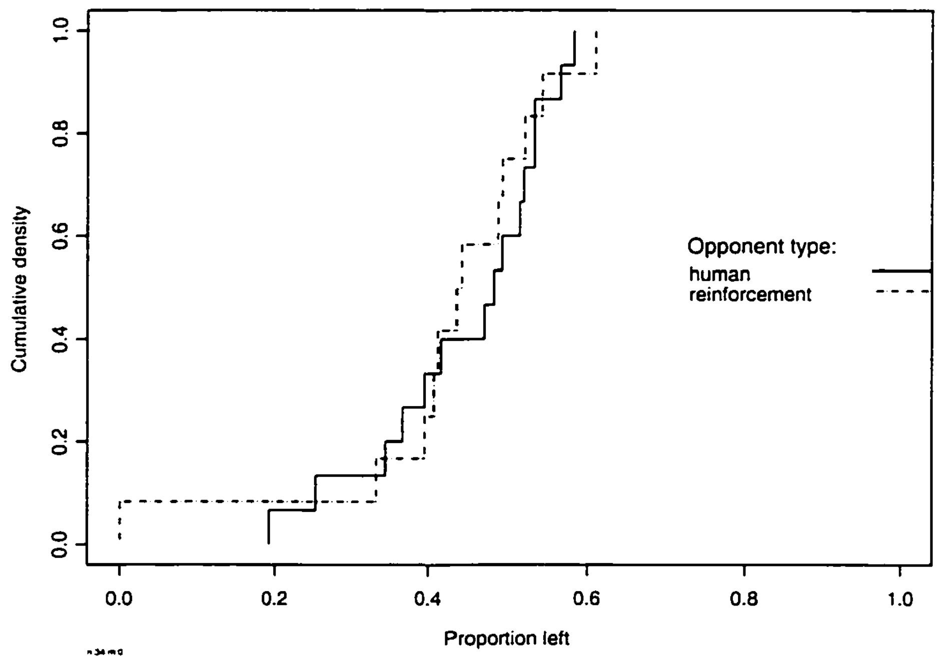
Dist. Left when facing Human tested against dist. Left when facing:	KS statistic	P-value
ER	0.567	0.005
EWA	0.633	0.001

FIGURE 1.6. Distributions of Left by Human Row players in Pursue-Evade.



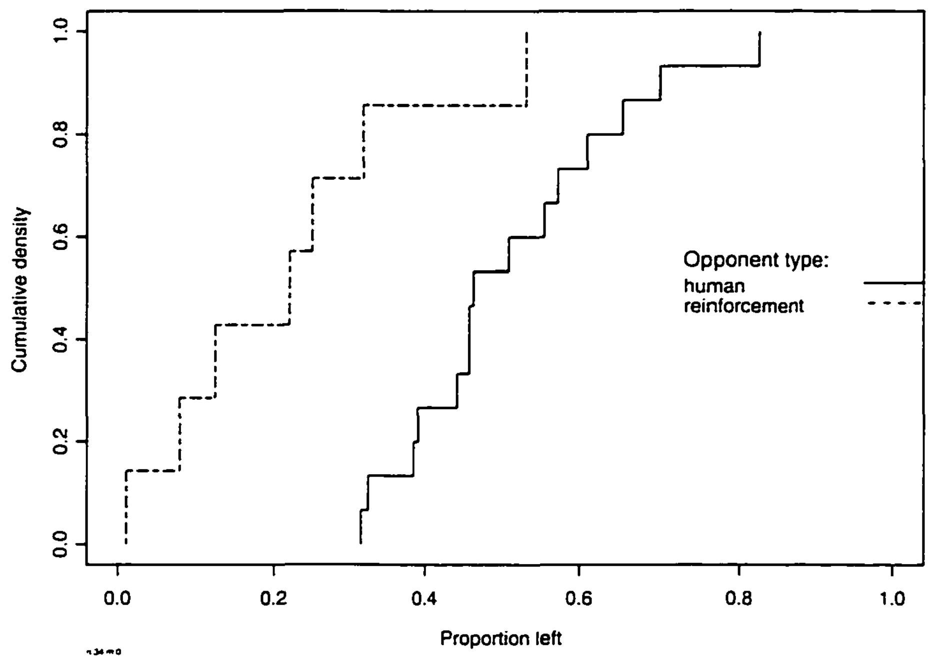
Dist. Left when facing Human tested against dist. Left when facing:	KS statistic	P-value
ER	0.283	0.435
EWA	0.267	0.507

FIGURE 1.7. Distributions of Left by Human Column players in Pursue-Evade.



Dist. Left when facing Human tested against dist. Left when facing:	KS statistic	P-value
ER	0.183	0.952

FIGURE 1.8. Distributions of Left by Human Row players in Gamble-Safe.



Dist. Left when facing Human tested against dist. Left when facing:	KS statistic	P-value
ER	0.483	0.061

FIGURE 1.9. Distributions of Left by Human Column players in Gamble-Safe.

Chapter 2

INFORMATION PROCESSING AND LEARNING MODELS

2.1 Introduction

Theoretical and experimental research on out-of-equilibrium behavior, or "learning," in strategic situations has largely fallen into two classes of models, reinforcement learning models and belief learning models. The two classes of models make quite different assumptions about the information on which participants base their choice of action. Previous experimental research has attempted to discriminate between these two classes of learning models via an indirect, inferential approach that uses subjects' observed action choices to estimate the models' parameters.

We report here on a different, more direct approach: an experiment in which subjects choose which kind of information they wish to see – either the information required for reinforcement learning, or the information required for belief learning. Thus, we can directly observe whether a subject has the information necessary for his actions to be generated by one kind of learning model or the other. Moreover, it is natural to assume that had the subject instead been given both kinds of information, the decision process he would have employed would be more accurately described by the models that are consistent with the information he chose than by the models that are inconsistent with the chosen information.[1] The results of the experiment suggest that while neither kind of information is chosen exclusively, subjects most often choose information that is consistent with belief learning and inconsistent with reinforcement learning. Moreover, the extent to which subjects choose one kind of information or the other is influenced by the amount of "free" information they are provided.

The fundamental character of belief learning models (some examples of which are fictitious play, cautious fictitious play, and Bayesian learning models) is that a player will use his opponent's past actions to form an estimate, or forecast, of the action the opponent will choose at the current play.[2] The player is then assumed to use this forecast, together with his own payoff function, to determine, either deterministically or stochastically, a "good" action for himself. By contrast, reinforcement learning models (a recent example is the model introduced by Erev & Roth [8]), assume that a player chooses his current action by reviewing the payoffs he has obtained from the actions he has taken in the past, and by then choosing with a greater propensity those actions that have produced higher average payoffs.

Notice that the information required in order for a subject to be able to carry out the prescribed decision-making process differs markedly in the two kinds of model. In the belief model the player must know the actions taken by his opponent in the past, and he must know his own payoff function (which includes his own and his

[1]To put it another way, it's as if, when choosing one kind of information in the experiment, a subject is effectively choosing to use the associated learning model.

[2]For simplicity, we couch our description of learning models in the two-person context.

opponent's action sets. as well as his own payoff at each pair of actions). He need not know his own past actions nor the payoffs he has obtained in the past. By contrast. in the reinforcement model the player must know which actions he has taken in the past and the payoffs the actions have produced, but he need not know his own (or his opponent's) payoff function. nor the actions his opponent has taken in the past. Thus. the two kinds of information are mutually exclusive.[3]

The experiment we report uses a simple two-person constant-sum ("strictly competitive") 2×2 game in which there is just one equilibrium. in mixed strategies. The game matrix is depicted in Figure 2.1; the matrix was always available for subjects to view. The game was played repeatedly by fixed pairs of subjects. After every ten plays. each subject was given the opportunity to request information about the history of play. He could request information about the history of his own actions and the payoffs they achieved for him ("Own Inf"). or information about the history of his opponent's actions ("Opp Inf"). Or he could request no information at all.

Note that Own Inf is precisely the information a player needs in order to choose according to reinforcement models. but it is insufficient to enable a player to choose according to belief models. Conversely. Opp Inf is the information (along with his payoff table) a player needs in order to choose according to belief models. and it is inadequate for choosing according to reinforcement models.

Of course. players may be able to remember their own actions without periodically receiving the information again. And if they are also told, at the conclusion of each

[3]This is not to say. of course. that one couldn't use one kind of information to make at least partial inferences about the other kind of information.

play, what action their opponent has taken, then they have (and they may be able to remember) all the information required for either kind of model. Consequently, two alternative variations in treatment were implemented, which limited the amount of "free" information a subject was provided: information that was provided after each play in one treatment (such as the subject's own realized payoff, or his opponent's action) could be obtained in another treatment only as part of the requested information provided at ten-play intervals.

We present the following summary of our main results. Subjects exhibit a strong preference for information required by belief learning, as opposed to that needed by reinforcement learning. While this result is not evident in our full information condition, we find strong support for the finding when subjects are provided with lesser amounts of information. Additionally, we find evidence against Nash equilibrium play. However, we find play to be positively serially correlated (even more when less information is provided to subjects). This finding is at odds with previous work that identified strong negative serial correlation of play.

We describe briefly three of the most salient previous papers that have attempted to discriminate between reinforcement and belief models. All three papers use subjects' observed action choices to estimate parameters of the alternative models. Erev & Roth [8] introduce a multi-parameter model of reinforcement learning and find that the model, when estimated, characterizes observed play better than alternative belief models.[4] Camerer & Ho [2] introduce a multi-parameter model that combines

[4]The following belief models were considered by Erev and Roth: deterministic fictitious play, probabilistic fictitious play, probabilistic fictitious play with an exponential response rule, and simple

elements of both reinforcement and belief learning, and they report that their hybrid model performs better than either of the pure models. Feltovich [9] finds that on some criteria a reinforcement model better reflects observed action choices, and on other criteria a belief model performs better. Clearly, there is no consensus regarding the model type that more accurately explains human behavior.

Using subjects' observed action choices to estimate the parameters of alternative models does not directly reveal the process by which subjects made their decisions, and in particular it does not reveal the information the subjects used. Instead, the decision process and the information it used are inferred. But this inferential approach typically becomes extremely tenuous as soon as there are more than a handful of decision periods, because the models' predictions quickly become much less precise as the number of periods increases.

In a cleverly designed Matching Pennies experiment Mookherjee & Sopher [20] attacked this problem by controlling the information available to subjects in two alternative treatments: in Treatment 1 the Matching Pennies payoff matrix was revealed to the subjects, and in Treatment 2 it was not revealed. Recall that one's own payoff function is required in belief models but not in reinforcement models. In each treatment the subjects were told their own payoffs after each stage of play, but were told nothing else about how play had proceeded. Thus, in Treatment 1 a subject could always infer which action his opponent had just taken, but he had no way of knowing this in Treatment 2. Consequently, each time a subject was required to

best reply.

make a choice in Treatment 1, he had been given, over the course of previous play, all the information he would need in order to choose according to either a belief model or a reinforcement model. In Treatment 2, on the other hand, he had been given all the information required for adhering to a reinforcement model, but none of the information required for belief models.

Mookherjee & Sopher found that in Treatment 2, where information required for belief models was not available, a reinforcement model described observed choices better than belief models. But in Treatment 1, where adequate information for using either model was available, subjects did not play according to reinforcement models *or* belief models, and indeed there was little evidence to reject minimax play – i.i.d. play at the 50-50 equilibrium rate. Play was thus clearly different in the two treatments, and Mookherjee & Sopher suggest that the difference in play "constitutes strong evidence that information regarding the opponent's choices does alter the nature of play." It's not so clear, however, that this is the appropriate interpretation of their results. Perhaps it is knowledge of the *payoff matrix* that influences play, in the following way: in Treatment 1, where subjects know the payoff matrix, perhaps they recognize that they need to play unpredictably, mixing at a roughly 50-50 rate between Heads and Tails. This is the kind of behavior that has been observed in past Matching Pennies experiments, and researchers have often avoided using the pure Matching Pennies game in experiments just because there is evidence that 50-50 unpredictable play is focal.

Our own experiment utilizes a 2×2 game in which the unique equilibrium requires each player to play Left 2/3 of the time and Right 1/3 of the time. Subjects were at all times provided with the complete payoff matrix. Where M&S in one treatment withheld the information required for belief learning (rendering subjects unable to choose according to belief models), and in the other treatment made both kinds of information available (so that it is difficult in this case to determine which model better describes the observed behavior), our experiment instead partitions the two kinds of information exactly and then allows a subject to choose which kind of information he will have. Thus, while we still cannot observe *how* the subjects process their information to reach a decision, we can at least see *what information they choose* and, by extension, which kind of model they *could* be using. Moreover, this approach allows us to identify one model or the other without the need to estimate any parameters.

The remainder of the paper is organized as follows: First, we briefly discuss both the belief-based and reinforcement learning models, and establish the differing informational requirements. Next, we explain our experimental design and describe our information conditions. Finally, we present our experimental results.

2.2 Theory

We now more clearly define both reinforcement and belief learning, and show how the information required by each model is mutually exclusive. To simplify the discussion, we assume two players are repeatedly playing a 2×2 normal form game in which the players' payoff functions are denoted by π_1 and π_2. We index the plays by t, and we denote player i's action at t by a_{it} and his payoff at t by z_{it}: $z_{it} = \pi_t(a_{1t}, a_{2t})$.

According to each theory, a player's action at t is determined by a *mixture* (which is a probability measure) over the player's set of possible actions, and this mixture is written for player i at time t as σ_{it}. Note that this allows for a theory in which players don't actually mix, by specifying that players use only degenerate mixtures. In each theory, a player's mixture at play $t+1$ is somehow determined by the *history of play* over the first t repetitions, $((a_{1\tau}, a_{2\tau}), (z_{1\tau}, z_{2\tau}))_{\tau=1}^{t}$. The difference between reinforcement theories and belief-based theories lies in the way the players' mixtures are determined from histories.

According to belief learning, Player 1 *forecasts* his opponent's next play mixture and formulates his own mixture based on this forecast.[5] He uses his observations of Player 2's past play to generate the forecast, and thus to generate his own mixture in the following manner:

$$\text{Forecast of } \sigma_{2,t+1} : \widehat{\sigma}_{2,t+1} = f_B(a_{2,1}, \ldots, a_{2,t})$$

$$\text{Mixture by Player 1: } \sigma_{1,t+1} = g_B(\widehat{\sigma}_{2,t+1})$$

Note that belief learning implies Player 1's mixture is formed using only $a_{2\tau}$ and not $a_{1\tau}$, $z_{1\tau}$, or $z_{2\tau}$.

In contrast, the idea of reinforcement learning is that actions resulting in higher payoffs tend to be chosen more often. Specifically, each of Player 1's actions has an associated latent *propensity*. These propensities serve as a relative value of each action, and are used to formulate play mixtures. After each play of the game, Player

[5]Without loss of generality, we explain the learning models by using Player 1 as the decision maker.

1's propensities are adjusted according to the relative success of the action. Reinforcement learning is expressed for Player 1 as:

$$\text{Mixture by Player 1:}\sigma_{1,t+1} = h_R(a_{1,1}, \ldots, a_{1,t}; z_{1,1}, \ldots, z_{1,t})$$

We see that reinforcement learning implies Player 1's mixture is formed with $a_{1\tau}$ and $z_{1\tau}$, but not $a_{2\tau}$ or $z_{2\tau}$.

We see that Belief and Reinforcement learning require two distinctly different sets of information to formulate actions. Belief learning requires knowledge of the past actions of one's opponent(s), while reinforcement learning requires the knowledge of one's own past actions and associated payoffs. If either model is in fact an accurate description of how humans actually learn to play strategic games, then we ought to observe systematic preference for the model's required information. The next section explains the design of an experiment whose purpose is to test the existence of these informational preferences.

2.3 Experimental Design

Our goal is to design an experiment that will allow subjects multiple opportunities to select exclusively between information needed for reinforcement learning and information needed for belief-based learning. The opportunity to view either of the two information types will be provided at regular intervals throughout the experiment. The exact nature of the available information will depend on the treatments in use, as described below.

While we are interested in what information is used, we must be aware that perhaps people can "internalize" the needed information from reported stage game results. Because of this possibility, we need to restrict the level of information provided to the point that subjects *must* consult the history reviews in order to obtain needed information. That is, we want to increasingly restrict information to the point that subjects must actively seek it out.

2.3.1 The Game

The game used in this experiment is shown below in Figure 2.1. This game yields a constant sum of 20. Even though the game is not symmetric, it has a unique symmetric mixed-strategy equilibrium of each player selecting 'L' with probability 2/3, and 'R' with probability 1/3. Further, in equilibrium, the expected value of playing the game for each player is 10.

Subjects selected for this experiment were undergraduates at the University of Arizona. Each session involved a subject being randomly and anonymously matched with the same opponent for 200 repetitions of the game. Each experimental session consisted of several pairs of subjects simultaneously playing the game, so that no one would know with whom he was matched.

2.3.2 Review of histories

After every 10 repetitions of the game, each subject was given the opportunity to review history information of either his own actions and associated payoffs, or his opponent's actions. Thus, each subject had the opportunity to review information relevant to belief learning or reinforcement learning. If a subject chose to review Own Inf, then he would see the sequence of his individual actions and associated payoffs, as well as aggregate statistics of his average payoff from choosing both 'Left' and 'Right'. Alternatively, if a subject chose to review Opp Inf, he would see the sequence of his opponent's past moves, as well as the count of his opponent's 'Left' and 'Right' actions. Notice that a subject's own payoff information was only given when he selected Own Inf, as own payoff information is relevant to reinforcement learning but not belief learning.

Each subject had the choice of how many periods back to review. After a subject selected which type of information to review, he was given the option to review the last n periods, where n ranged from zero to the current number of completed periods. The default history length was set at zero periods, as this allowed us to capture subjects who did not actually review either type of information.

While it is possible that selecting to view Own Inf could in fact allow for a subject to deduce his opponent's actions as well, we do not believe this to be likely. According to the two alternative learning models, each model views the information needed by the other as irrelevant. Thus, if a subject was indeed adhering to behavior implied by either reinforcement or belief-based learning and ignoring the information needed by the other model, then it seems likely that the subject would choose to directly view the relevant information, as opposed to viewing irrelevant information simply to deduce needed information.

2.3.3 Stage game information

Because it is possible that subjects can obtain all needed information from reported stage game results and thus will opt not to review either type of information, we employed alternative information environments that would compel subjects to actually seek needed information. Thus, we allowed the reporting of stage-game results, as well as the running balance, to be a treatment. Note that removal of the running balance was necessary, as a subject could easily use it to discern the outcome of each repetition, and infer his opponent's action, by observing the change of his total balance.

By removing the display of stage-game results, the only way a subject could ascertain either his opponent's actions or his own outcomes was to review past history information. While it is true that this procedure in essence forces subjects to select past play information, it is nonetheless the case that if one type of history information is more useful, then we should observe its selection with a higher frequency than the other.

2.3.4 Own payoff information

For the treatment in which stage-game results are not revealed, subjects must review history information in order to obtain play information. Consequently, if a subject opts to review his opponent's history of play, he will be unable to observe his own aggregate performance. While belief-based learning does not formally require any knowledge of one's own payoff information, it is somewhat unusual in game theoretic experiments to not report monetary earnings to a subject after each play.

To address this issue, we introduce the treatment of revealing a subject's own payoff information when viewing Opp Inf. This additional information is extraneous to both belief-based learning as well as reinforcement learning. However, it does allow for a subjects to observe his earnings information while choosing to view opponent actions, and thus perhaps "balances" the two types of history reviews in the eyes of the subjects.

2.3.5 Information Conditions

A 2×2 treatment design was used in this experiment. A total of 164 subjects participated in this study. Figure 2.2 displays how many pairs of subjects participated in each of the four cells of our 2×2 treatment design. Additionally, this table establishes the naming convention used to refer to the different information conditions. Each information condition is referred to as 'Condition n' where n represents the four different cells in the table.

2.4 Experimental Results

The results reported in this section can be summarized as follows: (1) When stage game results are freely revealed, subjects have no clear preference for a specific form of review information. However, when stage game results are not revealed, subjects exhibit a strong preference for reviewing Opp Inf. (2) In general, subjects do not play the Nash equilibrium mixture. (3) Similar to previous studies, we find our subjects exhibiting serially correlated play. However, while previous work has found that subjects tend to generate negatively serially-correlated play, we find our subjects generating positively serially-correlated play.

2.4.1 Information Selection

We will now look at the observed informational preferences of the subjects. Recall that after every tenth repetition each subject had the choice of reviewing either belief learning information or reinforcement learning information. Although subjects were forced to choose only one of these two information types, the subject then had the choice of how many repetitions into the past he wanted to review. The default history review length was zero periods, thus subjects not wishing to review any information could simply exert the least effort and select the default settings in order to not review any information. We categorize subject information selection into three types: Opp Inf, Own Inf, and 'None' which represents selecting a history view containing the default of zero periods.

Figure 2.3 reveals the percentages of information selection by type within each condition. We see that when subjects are shown results after each stage game, there is no definite preference for either information type over the other. Indeed, in both Condition 1 and Condition 2, each of the three review types are selected roughly one-third of the time. A plausible explanation of this is that a significant portion of the subjects are able to sufficiently internalize memory and processing of the results that were presented after each stage game. Figure 2.4 supports this assertion by showing that subjects spend roughly forty-five percent more time viewing history information when stage-game results are not shown, as opposed to when they are shown.

We see that information preferences are dramatically different when subjects are not provided with results after each stage game. The frequency of viewing no information falls from about one-third to near zero. This strongly suggests that subjects were indeed using the data freely provided to them after each stage game. When the data are no longer available, subjects actively seek it out during the review period. Thus, by not revealing the stage-game information, we have induced a state in which the subjects can be expected to rely upon the history review sessions to obtain the information they need for making their strategic decisions.

We will now discuss Condition 3 and Condition 4 in greater detail. We first consider Condition 4, and see that subjects selected Opp Inf 57 percent of the time, while selecting Own Inf only 41 percent of the time. Recall that when a subject selected Opp Inf in this condition, he received no information regarding his own earnings. Thus, we see an overall preference to view information regarding one's opponent, even to the extent that the viewer is not informed of his own performance. One could argue that the subject could remember his own moves, and then by viewing Opp Inf, he could infer his own earnings. However, this information could be obtained *much* more directly by simply choosing to view his own information.

When we allow subjects to see own payoff information while reviewing Opp Inf in Condition 3, we see a drastic change in information preferences. Specifically, the frequency of viewing Opp Inf increases to 74 percent, and the frequency of selecting

Own Inf falls to 22 percent. Perhaps subjects prefer a minimal amount of information regarding their own earnings, even when engaged primarily in belief learning behavior.

While this apparent preference for belief learning information presented in Figure 2.3 is revealing, it does not distinguish whether each subject consistently selected the same type of information, or instead alternated review choices during the nineteen review opportunities. To better address this issue, we determine the number of subjects within each condition who selected primarily one type of information during the course of the experiment. Figure 2.5 reports the number of subjects that selected the same information type at least seventeen of the nineteen times.

It is clear that revealing stage game results discriminates whether subjects prefer primarily one type of review information. Specifically, when stage game results are displayed in Conditions 1 and 2, few subjects relied upon only one review information type across the span of the experiment. Indeed, combining the similar results of both Condition 1 and Condition 2, we see that only fourteen of seventy (20.0%) subjects relied upon primarily one type information. Of the fourteen who did, three selected belief learning information, five selected reinforcement learning information, and six did not use either type of information. Perhaps the requisite information needed for strategic decision making was gathered easily enough from the stage game results.

However, when no stage-game results were displayed, we see an extreme preference for belief learning information. Neither Condition 3 nor Condition 4 has any subjects completely forgoing review information during review opportunities. Further, when looking at Condition 4, we see that fifteen of thirty-eight (40%) subjects relied mainly on belief information, while only six of thirty-eight (16%) relied mainly on reinforcement information. Thus, subjects were two-and-a-half times more likely to review belief learning information over reinforcement information, even when this consistently left them without knowledge of their own earnings during the course of the game.

When subjects are given a means to assess their aggregate earnings while viewing belief information, in Condition 3 we see an extreme increase of the preference for viewing belief information. In this case, thirty-five of fifty-six (63%) subjects primarily viewed belief information, while only four of fifty-six (7%) subjects viewed primarily reinforcement information. Subjects were almost nine times more likely to prefer belief learning information than reinforcement information as a primary form of information to review.

Summarizing, subjects significantly preferred belief information over reinforcement information. However, subjects seem to obtain sufficient strategic information from stage-game feedback to render both belief and reinforcement information review as redundant and somewhat unnecessary. Only when stage-game feedback is withheld do we see an extreme preference for belief information.

2.4.2 Observed Play Mixtures

Tables 2.1 - 2.4 report the pair-level action frequencies from the experiment. Each table corresponds to one of the four treatment conditions, and each row within a table corresponds to a specific pair of subjects. The first column identifies the specific pair.

Columns two through five list the observed outcome frequencies. Columns six and seven give the frequencies with which the Row and Column player, respectively, played Left. Columns eight through ten report the results of statistical tests, to be explained immediately below.

To address the question of whether subjects' play adheres to the predictions of Nash equilibrium theory, we conduct a χ^2 goodness of fit test on the observed outcome frequencies of each game against the expected outcome frequencies according to the equilibrium mixture of each player mixing two-thirds Left and one-third right. Column eight reports the resulting p-value from the χ^2 test on each game session. We see that 66 of the 82 pairs (81%) have their play rejected as being generated by the Nash equilibrium frequencies at the 5% significance level. Further, when we condition the χ^2 test results on whether stage-game results were revealed, we see that 23 of the 35 (66%) pairs that do see stage game results reject the test at the 5% level of significance, while 43 of 47 (92%) pairs that do not see the results reject at the 5% level of significance the null hypothesis that play is generated by the equilibrium frequencies. Clearly, this is support that many subjects are not adhering to the play frequencies predicted by Nash equilibrium theory. Further, it appears that masking stage-game results leads to an even greater number of subjects not playing according to Nash equilibrium frequency predictions.

The aforementioned χ^2 tests jointly consider whether both subjects in a game are playing predicted equilibrium frequencies. To allow for the possibility that one subject may be playing his equilibrium frequency while his opponent is not, we also conduct a binomial test of whether each subject is individually adhering to the equilibrium frequency. Columns nine and ten report the resulting p-values from these binomial tests for the row and column player, respectively. We see that 63 of the 82 row players (77%) have their play rejected at the 5% level of significance as being generated by the row equilibrium frequency of two-thirds Left and one-third Right. Next, we see that 32 of the 82 column players (39%) have their play rejected at the 5% significance level as being generated by the equilibrium frequency of two-thirds Left and one-third Right. Conditioning on whether subjects observe stage-game results, we see that when subjects see stage results, row players have their play rejected 60 percent of the time while column players are rejected 31 percent of the time. Moving to the subjects that do not see stage results, we see that row players have their play rejected 89 percent of the time, while column players have their play rejected 45 percent of the time.

The reported binomial tests indicate we have two factors that influence the subjects' frequency of play. Figure 2.6 summarizes the proportion of equilibrium frequency rejections due to the binomial test, conditioned on seeing stage-game results. First, we observe a role asymmetry, as the frequency with which the row player has his play rejected by the binomial test is approximately double that of the column player. Second, we see that the number of binomial test rejections is increased by about fifty percent when we move from subjects seeing stage-game results to not being able to see stage-game results.

Given that the previous tests establish that many of the subjects are not playing equilibrium frequencies at an individual level. what are the aggregate frequencies we observe? Across all experimental sessions, we see that row players select Left 53 percent of the time, and column players select Left 67 percent of the time. Even though pair level analysis rejected the majority of subjects as not playing equilibrium frequencies, it appears that in aggregate, column players are right on the equilibrium frequency. However, the row players are not; in fact they are much under the equilibrium frequency prediction. Indeed, of the 82 pairs that participated in this study, the column player selected Left at a higher frequency than his row counterpart 76 times, or 93 percent of the time. This result is quite striking, and it is similar to the results reported by Rosenthal, Shachat, and Walker [27].

2.4.3 Runs Analysis

Nash Equilibrium theory holds that in a repeated game, the actions chosen by a player will be serially uncorrelated. One way to test for serially correlated actions is to analyze the *runs* generated by a player.[6] Tables 2.5 - 2.8 report runs tests on each individual. Of the 164 subjects, 89 were observed selecting serially correlated actions according to a two-tailed runs test at the 5 percent level of significance. Of the 89 subjects exhibiting serially-correlated actions, 74 subjects displayed too few runs, and 15 displayed too many runs.

Figure 2.8 displays the empirical distribution of p-values generated by runs tests

[6] "Given an ordered sequence of two or more types of symbols, a run is defined to be a succession of one or more identical symbols which are followed and preceded by a different symbol or no symbol at all." (Gibbons and Chakraborti [12], page 68)

on the actions of each subject. If subjects' actions were indeed uncorrelated. then the distribution plot would be indistinguishable from a 45-degree line. Clearly. this is not the case. In fact. the plot indicates that subjects are generating far too few runs and thus selecting actions in a positively correlated manner.

This evidence of positive serial correlation of actions contrasts with many previous reports of serial correlation: O'Neill [24], Rapoport and Boebel [25]. as well as Mookherjee and Sopher [20] all report *negative* serial correlation of action choices. Only Rosenthal, Shachat, and Walker [27] reports evidence of positive serial correlation of action choices in a two-person normal-form game.

2.5 Discussion

This work does not definitively conclude whether reinforcement or belief learning is the better description of human behavior in strategic games. However, by directly observing subject information selection during the play of a repeated game. we can confidently state that subjects tend to select information that is consistent with belief learning. Although, this tendency exists only when the information is needed: when all information is revealed to subjects. no type of information is preferred. Only when information is withheld from subjects do we see a systematic preference for the information required by belief learning.

		Player 2 L	Player 2 R
Player 1	L	30, -10	-30, 50
	R	-30, 50	90, -70

FIGURE 2.1. Stage game

		Reveal period results: Yes	Reveal period results: No
Reveal own earnings when reviewing Opp Inf	Yes	Condition 2 16 pairs	Condition 3 28 pairs
	No	Condition 1 19 pairs	Condition 4 19 pairs

FIGURE 2.2. Number of subject pairs in each condition.

		Reveal period results: Yes	Reveal period results: No
Reveal own earnings when reviewing Opp Inf	Yes	Condition 2 None 36% Opp Inf 29% Own Inf 36%	Condition 3 None 4% Opp Inf 74% Own Inf 22%
	No	Condition 1 None 27% Opp Inf 36% Own Inf 37%	Condition 4 None 2% Opp Inf 57% Own Inf 41%

FIGURE 2.3. Summary of Review Choices

		Reveal period results	
		Yes	No
Reveal own earnings when reviewing Opp Inf	Yes	Condition 2 Opp: 9.6 Own: 10.3	Condition 3 Opp: 14.7 Own: 13.5
	No	Condition 1 Opp: 9.6 Own: 10.8	Condition 4 Opp: 14.1 Own: 14.7

FIGURE 2.4. Average number of seconds spent reviewing summary information.

Reveal own earnings when reviewing Opp Inf	Reveal period results: Yes	Reveal period results: No
Yes	Condition 2 None 3 Belief 1 Reinf. 2 (32 subjects)	Condition 3 None 0 Belief 35 Reinf. 4 (56 subjects)
No	Condition 1 None 3 Belief 2 Reinf 3 (38 subjects)	Condition 4 None 0 Belief 15 Reinf. 6 (38 subjects)

FIGURE 2.5. Number of subjects who selected the same information type at least 17 of the 19 times.

Reveal period results

Yes	No
Row: 60.0%	Row: 89.4%
Col: 31.4%	Col: 44.7%

FIGURE 2.6. Proportion of subjects whose play is rejected as NE according to the binomial test.

Reveal period results

Yes	No
Row: 56.2%	Row: 50.7%
Col: 66.5%	Col: 67.8%

FIGURE 2.7. Porportion of Left play.

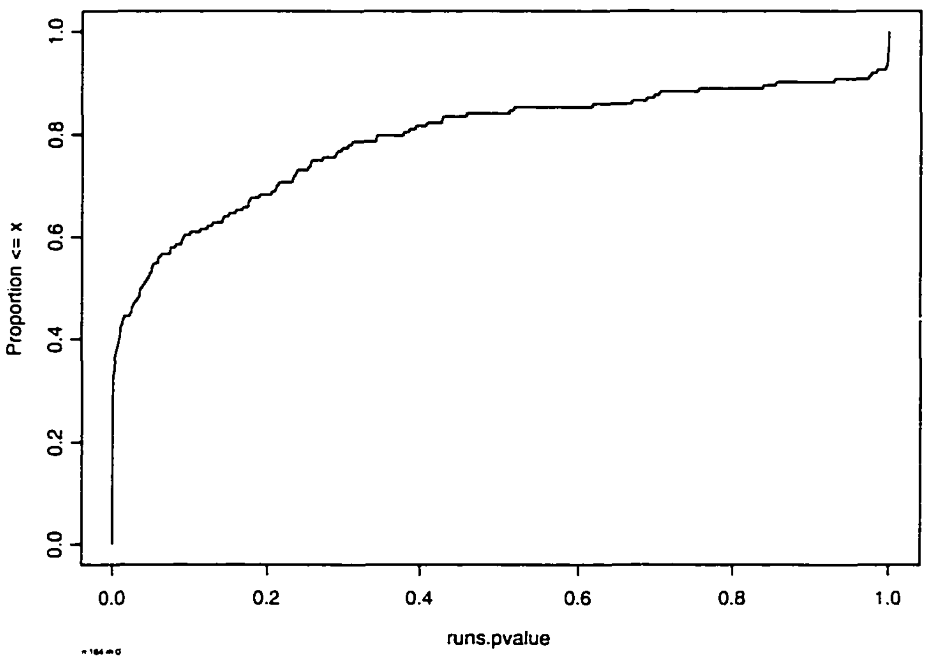


FIGURE 2.8. Empirical distribution of p-values generated by runs tests on each subject.

	Action Profile				Ratio Left		χ^2	Binomial	
Pair	LL	RL	LR	RR	Row	Col	NE test	row	col
101	85	43	48	24	0.665	0.640	0.886	0.507	0.233
102	60	71	32	37	0.460	0.655	0.000	0.000	0.389
103	78	51	55	16	0.665	0.645	0.088	0.507	0.281
104	69	61	36	34	0.525	0.650	0.000	0.000	0.333
105	65	64	39	32	0.520	0.645	0.000	0.000	0.281
106	44	72	35	49	0.395	0.580	0.000	0.000	0.006
107	70	58	43	29	0.565	0.640	0.016	0.002	0.233
108	94	38	38	30	0.660	0.660	0.180	0.447	0.447
109	43	57	42	58	0.425	0.500	0.000	0.000	0.000
110	57	55	41	47	0.490	0.560	0.000	0.000	0.001
111	81	65	31	23	0.560	0.730	0.003	0.001	0.977
112	78	64	29	29	0.535	0.710	0.001	0.000	0.917
113	69	54	39	38	0.540	0.615	0.000	0.000	0.071
114	89	55	45	11	0.670	0.720	0.042	0.566	0.955
115	96	46	36	22	0.660	0.710	0.526	0.447	0.917
116	39	98	20	43	0.295	0.685	0.000	0.000	0.732
117	33	73	37	57	0.350	0.530	0.000	0.000	0.000
118	78	53	43	26	0.605	0.655	0.299	0.039	0.389
119	66	56	40	38	0.530	0.610	0.000	0.000	0.053

TABLE 2.1. Condition 1 pair-level analysis.

	Action Profile				Ratio Left		χ^2	Binomial	
Pair	LL	RL	LR	RR	Row	Col	NE test	row	col
201	72	58	49	21	0.605	0.650	0.049	0.039	0.333
202	92	52	27	29	0.595	0.720	0.016	0.020	0.955
203	86	61	30	23	0.580	0.735	0.012	0.006	0.985
204	71	70	24	35	0.475	0.705	0.000	0.000	0.891
205	84	50	35	31	0.595	0.670	0.092	0.020	0.566
206	88	41	47	24	0.675	0.645	0.904	0.625	0.281
207	54	48	39	59	0.465	0.510	0.000	0.000	0.000
208	87	53	40	20	0.635	0.700	0.502	0.190	0.859
209	94	52	36	18	0.650	0.730	0.263	0.333	0.977
210	124	38	23	15	0.735	0.810	0.000	0.985	1.000
211	86	45	39	30	0.625	0.655	0.322	0.121	0.389
212	91	44	41	24	0.660	0.675	0.927	0.447	0.625
213	88	50	29	33	0.585	0.690	0.010	0.010	0.780
214	92	51	36	21	0.640	0.715	0.432	0.233	0.938
215	99	57	23	21	0.610	0.780	0.002	0.053	1.000
216	54	92	32	22	0.430	0.730	0.000	0.000	0.977

TABLE 2.2. Condition 2 pair-level analysis.

	Action Profile				Ratio Left		χ^2	Binomial	
Pair	LL	RL	LR	RR	Row	Col	NE test	row	col
301	55	94	11	40	0.330	0.745	0.000	0.000	0.993
302	44	74	42	40	0.430	0.590	0.000	0.000	0.014
303	106	53	26	15	0.660	0.795	0.002	0.447	1.000
304	61	66	39	34	0.500	0.635	0.000	0.000	0.190
305	39	50	37	74	0.380	0.445	0.000	0.000	0.000
306	55	84	28	33	0.415	0.695	0.000	0.000	0.822
307	61	62	29	48	0.450	0.615	0.000	0.000	0.071
308	56	77	36	31	0.460	0.665	0.000	0.000	0.507
309	40	119	18	23	0.290	0.795	0.000	0.000	1.000
310	141	38	16	5	0.785	0.895	0.000	1.000	1.000
311	81	58	24	37	0.525	0.695	0.000	0.000	0.822
312	67	80	28	25	0.475	0.735	0.000	0.000	0.985
313	87	55	31	27	0.590	0.710	0.054	0.014	0.917
314	73	79	27	21	0.500	0.760	0.000	0.000	0.998
315	76	52	47	25	0.615	0.640	0.302	0.071	0.233
316	78	67	22	33	0.500	0.725	0.000	0.000	0.968
317	52	78	23	47	0.375	0.650	0.000	0.000	0.333
318	61	76	18	45	0.395	0.685	0.000	0.000	0.732
319	81	55	33	31	0.570	0.680	0.022	0.003	0.680
320	53	71	41	35	0.470	0.620	0.000	0.000	0.094
321	33	50	60	57	0.465	0.415	0.000	0.000	0.000
322	68	76	25	31	0.465	0.720	0.000	0.000	0.955
323	38	78	32	52	0.350	0.580	0.000	0.000	0.006
324	88	71	18	23	0.530	0.795	0.000	0.000	1.000
325	65	55	27	53	0.460	0.600	0.000	0.000	0.028
326	59	60	39	42	0.490	0.595	0.000	0.000	0.020
327	71	65	39	25	0.550	0.680	0.003	0.000	0.680
328	85	67	34	14	0.595	0.760	0.001	0.020	0.998

TABLE 2.3. Condition 3 pair-level analysis.

	Action Profile				Ratio Left		χ^2	Binomial	
Pair	LL	RL	LR	RR	Row	Col	NE test	row	col
401	113	59	24	4	0.685	0.860	0.000	0.732	1.000
402	78	59	32	31	0.550	0.685	0.005	0.000	0.732
403	72	61	46	21	0.590	0.665	0.023	0.014	0.507
404	70	64	30	36	0.500	0.670	0.000	0.000	0.566
405	54	64	48	34	0.510	0.590	0.000	0.000	0.014
406	80	70	38	12	0.590	0.750	0.000	0.014	0.996
407	85	70	24	21	0.545	0.775	0.000	0.000	1.000
408	75	55	29	41	0.520	0.650	0.000	0.000	0.333
409	80	83	21	16	0.505	0.815	0.000	0.000	1.000
410	78	55	28	39	0.530	0.665	0.000	0.000	0.507
411	56	63	35	46	0.455	0.595	0.000	0.000	0.020
412	64	74	21	41	0.425	0.690	0.000	0.000	0.780
413	86	68	23	23	0.545	0.770	0.000	0.000	0.999
414	63	63	46	28	0.545	0.630	0.001	0.000	0.153
415	72	53	45	30	0.585	0.625	0.055	0.010	0.121
416	81	64	43	12	0.620	0.725	0.003	0.094	0.968
417	71	51	50	28	0.605	0.610	0.080	0.039	0.053
418	55	85	24	36	0.395	0.700	0.000	0.000	0.859
419	49	48	56	47	0.525	0.485	0.000	0.000	0.000

TABLE 2.4. Condition 4 pair-level analysis.

	Ratio Left		Runs		Runs Test Pvalue	
Pair	Row	Col	Row	Col	Row	Col
101	0.665	0.640	67	89	0.000	0.288
102	0.460	0.655	94	81	0.202	0.062
103	0.665	0.645	85	87	0.233	0.217
104	0.525	0.650	81	95	0.003	0.708
105	0.520	0.645	120	82	0.997	0.059
106	0.395	0.580	95	92	0.436	0.193
107	0.565	0.640	98	95	0.453	0.641
108	0.660	0.660	73	109	0.003	0.999
109	0.425	0.500	119	89	0.999	0.051
110	0.490	0.560	89	84	0.052	0.015
111	0.560	0.730	73	61	0.000	0.001
112	0.535	0.710	95	79	0.238	0.256
113	0.540	0.615	85	73	0.017	0.000
114	0.670	0.720	87	77	0.379	0.236
115	0.660	0.710	78	89	0.027	0.859
116	0.295	0.685	72	107	0.024	1.000
117	0.350	0.530	98	96	0.843	0.278
118	0.605	0.655	80	92	0.009	0.564
119	0.530	0.610	85	92	0.015	0.291

TABLE 2.5. Condition 1 runs.

	Ratio Left		Runs		Runs Test Pvalue	
Pair	Row	Col	Row	Col	Row	Col
201	0.605	0.650	63	83	0.000	0.094
202	0.595	0.720	62	79	0.000	0.356
203	0.580	0.735	83	73	0.015	0.166
204	0.475	0.705	97	64	0.322	0.000
205	0.595	0.670	88	83	0.095	0.172
206	0.675	0.645	98	88	0.942	0.261
207	0.465	0.510	90	92	0.077	0.115
208	0.635	0.700	83	74	0.060	0.039
209	0.650	0.730	79	73	0.027	0.131
210	0.735	0.810	76	61	0.320	0.409
211	0.625	0.655	95	76	0.546	0.010
212	0.660	0.675	88	87	0.356	0.421
213	0.585	0.690	93	72	0.251	0.010
214	0.640	0.715	90	93	0.338	0.976
215	0.610	0.780	91	61	0.244	0.052
216	0.430	0.730	67	69	0.000	0.034

TABLE 2.6. Condition 2 runs.

	Ratio Left		Runs		Runs Test Pvalue	
Pair	Row	Col	Row	Col	Row	Col
301	0.330	0.745	53	72	0.000	0.195
302	0.430	0.590	43	71	0.000	0.000
303	0.660	0.795	61	35	0.000	0.000
304	0.500	0.635	104	72	0.690	0.001
305	0.380	0.445	83	92	0.039	0.148
306	0.415	0.695	82	77	0.011	0.085
307	0.450	0.615	38	76	0.000	0.002
308	0.460	0.665	68	69	0.000	0.001
309	0.290	0.795	44	33	0.000	0.000
310	0.785	0.895	64	34	0.190	0.060
311	0.525	0.695	29	37	0.000	0.000
312	0.475	0.735	89	47	0.055	0.000
313	0.590	0.710	66	57	0.000	0.000
314	0.500	0.760	99	73	0.416	0.468
315	0.615	0.640	95	71	0.488	0.000
316	0.500	0.725	63	58	0.000	0.000
317	0.375	0.650	67	79	0.000	0.027
318	0.395	0.685	79	78	0.006	0.074
319	0.570	0.680	60	109	0.000	1.000
320	0.470	0.620	83	87	0.007	0.123
321	0.465	0.415	67	81	0.000	0.008
322	0.465	0.720	95	76	0.238	0.179
323	0.350	0.580	23	76	0.000	0.001
324	0.530	0.795	84	79	0.011	1.000
325	0.460	0.600	122	111	0.999	0.984
326	0.490	0.595	104	122	0.692	1.000
327	0.550	0.680	87	77	0.037	0.045
328	0.595	0.760	45	36	0.000	0.000

TABLE 2.7. Condition 3 runs.

	Ratio Left		Runs		Runs Test Pvalue	
Pair	Row	Col	Row	Col	Row	Col
401	0.685	0.860	28	9	0.000	0.000
402	0.550	0.685	104	65	0.740	0.000
403	0.590	0.665	95	70	0.371	0.001
404	0.500	0.670	38	73	0.000	0.006
405	0.510	0.590	72	135	0.000	1.000
406	0.590	0.750	42	35	0.000	0.000
407	0.545	0.775	115	20	0.986	0.000
408	0.520	0.650	77	116	0.000	1.000
409	0.505	0.815	44	54	0.000	0.055
410	0.530	0.665	59	38	0.000	0.000
411	0.455	0.595	93	48	0.170	0.000
412	0.425	0.690	55	69	0.000	0.003
413	0.545	0.770	91	13	0.107	0.000
414	0.545	0.630	90	61	0.083	0.000
415	0.585	0.625	103	86	0.785	0.106
416	0.620	0.725	42	40	0.000	0.000
417	0.605	0.610	49	66	0.000	0.000
418	0.395	0.700	78	76	0.004	0.075
419	0.525	0.485	135	124	1.000	1.000

TABLE 2.8. Condition 4 runs.

Chapter 3

THE GROVES-LEDYARD MECHANISM IN DISCRETE STRATEGIES

3.1 Introduction

When a theoretical model is implemented in the laboratory or the field, careful attention must be paid to any deviation between the implementation and the theory. If any premise of the theory is altered when designing the implementation, then it is possible that a conclusion of the theory will no longer hold. In particular, strategy spaces are continuous in the theoretical specifications of many economic mechanisms. However, when these mechanisms are implemented, strategy spaces are generally discrete. Whether discrete strategy sets at all alter the properties of an implemented mechanism is often not even considered, and the theoretic properties of the mechanism are simply assumed to still hold.

This paper will undertake a detailed examination of the consequences of discretizing the strategy space of the Groves-Ledyard mechanism. Groves and Ledyard [14] has shown that this one-parameter incentive-compatible mechanism allows for financing the production of public goods, such that the Nash equilibria of the mechanism are Pareto optimal. Moreover, given quasi-linear preferences, the mechanism yields a unique Pareto-optimal Nash equilibrium. We find that with discrete strategy spaces, the set of Nash equilibria is often surprisingly different than when the strategies are continuous.

To better understand the impact of allowing agents to have a discrete set of strategies. we take as case studies implementations of the Groves-Ledyard mechanism that have used discrete strategy spaces. More than one thousand computing hours on the University of Arizona supercomputer were used to numerically evaluate whether each discrete strategy profile was a Nash equilibrium. We will see that with discrete strategy spaces and quasi-linear preferences, the mechanism no longer necessarily yields a unique Nash equilibrium. Further, by allowing the mechanism's free parameter to vary, we find that a unique Nash equilibrium is a special case. and not all Nash equilibria are Pareto optimal.

The remainder of this paper is organized as follows. Section two will review the equilibria of the Groves-Ledyard mechanism with a continuous message space, where it is well known that the equilibria are Pareto optimal. Section three will discuss how Nash equilibria are calculated with a discrete message space, and will solve for the Nash equilibria of recent published implementations of the Groves-Ledyard mechanism with discrete strategy spaces. We will see how equilibria are altered once we allow for discrete, rather than continuous, message spaces. In section four, we conclude with a discussion of the implications of these results.

3.2 The Groves-Ledyard Mechanism

In their celebrated 1977 paper, Groves and Ledyard constructed a one-parameter family of incentive-compatible mechanisms for financing the production of public goods, and showed that the Nash equilibria of the mechanisms are Pareto optimal. This section will review the properties of the Groves-Ledyard mechanism and explicitly show the well-known results within the context of constant production costs and quasi-linear preferences. First, the game form will be established. Next, we will determine the Pareto optimal provision level. Further, we will derive the Nash equilibrium, and see that it is both unique and Pareto optimal. Finally, we will characterize the Nash equilibrium as the mechanism's only free parameter is increased.

3.2.1 Game Form

Given the decentralized messages of all participants, the Groves-Ledyard mechanism specifies the level of public good production as well as each participant's share of the public good's production cost. Each individual $i \in \{1, 2, \ldots I\}$ submits a message $x_i \in \mathbb{R}$, the desired increment (or decrement) to the level of public good production: $X = \sum_i x_i$. Each person's preference for the public good is expressed with a quadratic value function. This function tells us how much transferable utility person i will receive if the level of the public good is X and person i incurs none of the production costs. We write person i's value function as

$$V_i(X) = A_i X - B_i X^2 + D_i.$$

The mechanism determines the cost of the public good incurred by each participant as follows. For each participant i we say that x_{-i} is the set of all messages excluding person i's message. Further, we write $\mu_{-i} = \sum_{j \neq i} x_j/(I-1)$ and $\sigma_{-i}^2 = \sum_{j \neq i} (x_j - \mu_{-i})^2/(I-2)$ for the mean and variance, respectively, of all messages, excluding individual i. We let c denote the constant cost of production of the public good. Finally, the mechanism's one free parameter is $\gamma > 0$. We can now write individual i's Groves-Ledyard cost function as

$$C_i(x_i, x_{-i}) = \frac{X}{I}c + \frac{\gamma}{2}\left(\frac{I-1}{I}(x_i - \mu_{-i})^2 - \sigma_{-i}^2\right).$$

Thus, after participating in the mechanism, the payoff to person i is expressed as

$$\pi_i(x_i, x_{-i}) = V_i(X) - C_i(x_i, x_{-i}).$$

3.2.2 Pareto optimal provision level

Given quadratic valuation functions, a Pareto optimal provision level can be characterized by: (1) the marginal social value of a public good being equal to its marginal cost, and (2) no wasted resources. We can express this as

$$\sum_{i \in I} V'(X) = c.$$

Since the valuation functions are quadratic, the Pareto optimal condition can be expressed as

$$\sum_{i \in I}(A_i - 2B_i X) = c.$$

Solving for the provision level in the above condition, we see that the Pareto optimal provision level is

$$X = \frac{\sum_{i \in I} A_i - c}{2\sum_{i \in I} B_i}.$$

3.2.3 Nash equilibrium

To solve for the Nash equilibrium of the mechanism, we first must determine person i's best-response correspondence. Given person i's payoff function π_i, straightforward maximization (given in Appendix A) shows person i's best-response function. $\widehat{x}_i$, as

$$\widehat{x}_i(x_{-i}) = \alpha_i + \beta_i \sum_{j \neq i} x_j.$$

where

$$\alpha_i = \frac{A_i I - c}{2B_i I + \gamma(I-1)}$$

and

$$\beta_i = \frac{\gamma - 2B_i I}{2B_i I + \gamma(I-1)}.$$

To determine the Nash equilibrium, we require all agents to be simultaneously best responding to one another. Thus, we take the I best response functions and construct the following system of equations

$$\begin{bmatrix} 1 & -\beta_1 & \cdots & -\beta_1 \\ -\beta_2 & 1 & \cdots & -\beta_2 \\ \vdots & \vdots & \ddots & \vdots \\ -\beta_I & -\beta_I & \cdots & 1 \end{bmatrix} \begin{bmatrix} \widehat{x}_1 \\ \widehat{x}_2 \\ \vdots \\ \widehat{x}_I \end{bmatrix} = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_I \end{bmatrix}.$$

Assuming the coefficient matrix is non-singular, this system of equations has a unique solution, and thus the mechanism has a unique Nash equilibrium. Appendix C discusses how strictly concave value functions ensure this system of equations has a unique solution.

We next determine the provision level of the unique Nash equilibrium by summing all I best-response functions. Letting $\widehat{X} = \sum_{i \in I} \widehat{x}_i$, Appendix B shows us

$$\widehat{X} = \frac{\sum_{i \in I} A_i - c}{2 \sum_{i \in I} B_i}.$$

It is clearly seen that the provision level of the unique Nash equilibrium is equal to the Pareto optimal provision level. Thus, the unique Nash equilibrium is Pareto optimal.

3.2.4 Nash equilibrium as γ increases

The initial theoretical work on the Groves-Ledyard mechanism was not at all concerned with how different values of γ affect Nash equilibrium. Indeed, so long as $\gamma > 0$, all resulting equilibria are Pareto optimal when message spaces are continuous. In contrast, we will see in the next section that discrete message spaces coupled with large values of γ will result in non-Pareto-optimal equilibria. However, we will first look at the effect of increasing γ when strategy spaces are continuous.

As discussed by Muench and Walker [22], a sufficiently large γ will lead to all mechanism participants submitting essentially identical messages. The Groves-Ledyard cost function is structured such that as γ increases, an individual faces higher tax punishments for deviating from the mean of the others' messages. To be precise, we will analyze what happens to the best-response function as the free parameter γ increases. We write the limit of the best-response function as γ goes to infinity as

$$\lim_{\gamma\to\infty} \widehat{x}_i = \lim_{\gamma\to\infty} \frac{A_i I - c + (2B_i I - \gamma)\sum_{j\neq i} x_j}{2B_i I + \gamma(I-1)}.$$

By applying L'Hospital's rule, we see that

$$\lim_{\gamma\to\infty} \widehat{x}_i = \frac{\sum_{j\neq i} x_j}{(I-1)} = \mu_{-i}.$$

Thus, as γ increases, each person's best response approaches the mean of everyone else's messages. Thus, sufficiently large values of γ will lead to all participants disregarding their own preferences and simply striving to match one another's messages, in order to avoid the large taxation due to deviation. Indeed, large enough values of γ highlight the consequence of the mechanism not being individually rational: the slightest 'wrong' message by an agent could result in a nearly infinite loss of the agent's welfare.

3.3 GL Equilibria with Discrete Strategies

The previous section shows the Nash equilibrium of the Groves-Ledyard mechanism is unique and Pareto optimal, given quasi-linear preferences. Recall that one of the premises was each agent's strategy set was the set of all real numbers. To better understand the consequences of discretizing the mechanism's strategy space, we will now calculate the Nash equilibria of several Groves-Ledyard implementations that have utilized discrete strategy spaces and have been previously reported in the literature.

Since the implementations under consideration did not have continuous strategy spaces, we can not use the calculus reviewed in the previous section to determine Nash equilibria. Instead, more than one thousand computing hours were utilized on the University of Arizona's supercomputer to numerically solve the results reported in this section. Software was developed to consider every possible discrete strategy profile and determine whether the profile was indeed a Nash equilibrium. Specifically, given a set of agents, each agent's quasi-linear preferences, the considered discrete set of messages, the constant cost of production, and the value of γ, we determine each agent's payoff for each message profile. If no agent's payoff can be increased by changing its own message, then the message profile is a Nash equilibrium.

Two recent publications that report the implementation of the Groves-Ledyard mechanism with discrete message spaces are Chen and Plott [4] (hereafter CP) as well as Chen and Tang [5] (hereafter CT). Both CP and CT report experiments that were designed, in part, to study the consequences of using different values of γ with the Groves-Ledyard mechanism. Both used five agents ($I = 5$). The constant cost of production in CP was 5, while in CT it was 100. Also, each subject had a different quadratic value function for the public good, as reported in Table 3.1.

The strategy spaces used by each of these implementations were as follows. Each agent in CP each had a strategy set consisting of 9 different messages, namely $\{-2, -1, \ldots, 6\}$. Thus, CP had $9^5 = 59,049$ distinct strategy profiles. Each agent in CT had a strategy set consisting of 51 different messages, specifically $\{-4, -3.8, \ldots, 6\}$.[1] This yields $51^5 = 345,025,251$ different strategy profiles in the CT implementations.

Both CP and CT used the value of γ as a treatment variable, and allowed γ to be either 1 or 100. In contrast, we will allow γ to assume an entire range of values while holding all other parameters constant, so that we can better observe the properties of the Nash equilibria as the value of γ is changed. Therefore, within the CP framework we will use $\gamma \in \{0.01, 0.02, \ldots, 250.00\}$, while within the CT framework we will use $\gamma \in \{0.01, 0.02, \ldots, 50.00\}$ and $\gamma \in \{51, 52, \ldots, 5500\}$.[2] Thus, given the quasi-linear preferences, discrete strategy sets, and constant production costs used by both CP and CT, we will vary γ and calculate every Nash equilibrium associated with each stated

[1] In the CT implementations, agents actually selected from a message set where each message was multiplied by 5, namely $\{-20, -19, \ldots, 30\}$, and all formulas were appropriately scaled. This was done to avoid the use of fractional messages. We will ignore this transformation and work only with the message set that corresponds to the published formulas and parameters by CT.

[2] A larger γ interval was used for larger values of γ within the CT implementations because of two reasons: (1) the Nash equilibria began to adhere to a regular pattern as γ increased, and (2) calculation of the Nash equilibria requires about six minutes for each value of γ, and thus to continue using an interval of 0.01 would have required an inordinate amount of time.

value of the parameter, as well as the surplus generated by each Nash equilibrium.

In general, the values of γ under consideration do not result in a specification of the Groves-Ledyard mechanism with a unique Nash equilibrium. Figures 3.1 and 3.2 graph the number of Nash equilibria as the CP and CT implementations[3] assume all values of γ under consideration, respectively. Logarithmic axes allow us to better distinguish the number of equilibria as gamma changes value. Both figures exhibit the same pattern of the number of Nash equilibria as γ is increased from zero. We see that the number of Nash equilibria spike upward when γ is initially increased from zero. Then, as γ continues to be increased, the number of Nash equilibria decrease to zero. Note that the two apparent discontinuities in Figure 3.1 are due to the inability to graph zero Nash equilibria on a logarithmic-scaled axis. Finally, as γ is increased even more, the number of Nash equilibria begin to steadily increase, although not surpassing the number of equilibria associated with the initial spike.

To more clearly understand this pattern, we break it into two phases. The two phases are partitioned by γ^*, where γ^* is defined as the lowest value of γ such that all greater values of γ result in at least one Nash equilibrium. Phase one corresponds to all γ less than γ^*, and phase two corresponds to all γ greater than or equal to γ^*. The value of γ^* associated with the CP implementations is 5.00, while the γ^* associated with the CT implementations is 46.67.

Phase one is characterized by both the Pareto optimality of all Nash equilibria and the sheer number of these equilibria. As γ is increased from zero, the number of

[3]For the remainder of this section, we will use the term 'CP implementations' to refer not only to the two experimental implementations used by Chen and Plott where $\gamma = 1$ and $\gamma = 100$, but also to all implementations using the previously mentioned values of γ for which we solve the Nash equilibria. The same also holds for the term 'CT implementations'.

Nash equilibria rapidly increases. Indeed, this initial increase in the number of Nash equilibria results in a maximum of 29 equilibria in the CP implementation (at $\gamma = 0.59$ and $\gamma = 0.60$), and 1955 equilibria in the CT implementation (at $\gamma = 0.85$). As γ continues to be increased, the number of Nash equilibria decreases to zero. When $\gamma \in \{3.34, 3.35, \ldots, 4.99\}$ for the CP implementations and $\gamma \in \{35.01, 35.02, \ldots, 46.66\}$ for the CT implementations, no Nash equilibria exist. Thus, the number of Nash equilibria in phase one varies from 0 to 29 in the CP implementation, and more dramatically, from 0 to 1955 in the CT implementations.

In Phase two, all Nash equilibrium profiles consist of all agents submitting the same message. When γ is sufficiently large and messages are discrete, an agent's only utility-maximizing message is the mean of the other agents' messages, and the only equilibrium profiles are therefore the ones in which every agent chooses the same message. Further, since all phase two Nash equilibria are profiles containing identical messages from all agents, the Nash equilibria are no longer generally Pareto optimal. In fact, every phase two implementation has exactly one Pareto-optimal Nash equilibrium. Namely, the profile where all five agents are submitting a message equal to one-fifth of the Pareto-optimal provision level.

The phase two Nash equilibria follow a very regular pattern, in contrast to the seemingly erratic behavior observed in phase one. Initially (i.e., for small values of γ), phase two has only one Nash equilibrium, and it is Pareto optimal. As γ is increased, additional equilibria appear: each additional Nash equilibrium provides a non-increasing level of surplus. Further, once a specific value of γ results in a specific Nash equilibrium, all larger values of γ will also result in the same Nash equilibrium.

Tables 3.2 and 3.3 display the specific Nash equilibria observed in phase two. We see that the common message associated each non-Pareto-optimal Nash equilibrium progresses steadily away from the common message associated with the single Pareto optimal Nash equilibrium as γ is increased. Likewise, the surplus generated by the additional Nash equilibria is steadily decreased as γ is increased. Indeed, large enough values of γ result in negative surplus levels. Eventually, a level of γ is reached such that the set of Nash equilibria consists of every strategy profile in which all agents are submitting an identical message, with the number of Nash equilibria determined by the number of discrete messages contained in the message set.

3.4 Discussion

Given these seeming general results, care must be taken when selecting the Groves-Ledyard parameter value to be used with discrete strategies. Indeed, an arbitrary parameter value could yield a multitude of Nash equilibria (including non-pareto-optimal ones), or even no Nash equilibrium.

Problems that may arise when one does not account for the effects of discrete strategy spaces can be seen by the results of past Groves-Ledyard experiments. Both CP and CT report that the Groves-Ledyard mechanism performs better with a γ of 1 as opposed to a γ of 100.[4] CT proposes that this better performance is due to supermodularity: when γ is 100 in the CT setting the game is supermodular, while the game is not supermodular when γ is 1. While this may be true, it is also true

[4]Better performance was measured in various ways, such as realized efficiency and frequency of obtaining "the" Nash equilibrium.

that in the CT sessions when γ is 1 there are 1445 Nash equilibria, and when γ is 100 there is only 1 Nash equilibrium. Further, in the CP sessions a γ of 1 results in 9 Nash equilibria, while a γ of 100 results in 5 Nash equilibria. In the context of both the CP and CT experiments, the higher value of γ is associated with fewer Nash equilibria. The better performance associated with γ equal to 100 may not at all be due to supermodularity, but instead may well be due to the difference in the number of Nash equilibria.

Clearly, it is imperative that we understand the implications of deviating from theoretical premises when implementing a model that has previously existed only in theory. In the case of the Groves-Ledyard mechanism, we see that discrete strategies can result in both a multitude of Pareto-optimal Nash equilibria, as well as the presence of non-Pareto-optimal Nash equilibria. Neither of these results are to be expected given casual acceptance of the theoretical results, yet nonetheless the problems exist.

$$V_i(X) = A_iX - B_iX^2 + D_i$$

	CP			CT		
Subject i	A_i	B_i	D_i	A_i	B_i	D_i
1	-1	0.0	55	26	1	200
2	5	0.5	35	104	8	10
3	10	0.9	20	38	2	160
4	20	1.8	0	82	6	40
5	15	1.2	5	60	4	100

TABLE 3.1. Value functions used in the CP and CT implementations.

Minimum integer value of γ needed for profile to be a NE	New profile (common message)	Resulting surplus
5	1	225
32	2	115
44	0	115
83	3	-215
88	-1	-215
128	4	-765
133	-2	-765
173	5	-1535
218	6	-2525

TABLE 3.2. CP phase two Nash equilibria.

Minimum integer value of γ needed for profile to be a NE	New profile (common message)	Resulting surplus
47	1	1035
130	1.2	1014
230	0.8	1014
330	1.4	951
430	0.6	951
530	1.6	846
630	0.4	846
731	1.8	699
830	0.2	699
930	2	510
1030	0	510
1130	2.2	279
1230	-0.2	279
1331	2.4	6
1430	-0.4	6
1530	2.6	-309
1631	-0.6	-309
1730	2.8	-666
1831	-0.8	-666
1930	3	-1065
2031	-1	-1065
2130	3.2	-1506
2231	-1.2	-1506
2331	3.4	-1989
2431	-1.4	-1989
2531	3.6	-2514
2631	-1.6	-2514
2730	3.8	-3081
2830	-1.8	-3081
2931	4	-3690
3030	-2	-3690
3131	4.2	-4341
3231	-2.2	-4341
3330	4.4	-5034
3431	-2.4	-5034
3531	4.6	-5769
3630	-2.6	-5769
3730	4.8	-6546
3830	-2.8	-6546
3931	5	-7365
4030	-3	-7365
4130	5.2	-8226
4231	-3.2	-8226
4330	5.4	-9129
4431	-3.4	-9129
4531	5.6	-10074
4630	-3.6	-10074
4730	5.8	-11061
4831	-3.8	-11061
4930	6	-12090
5031	-4	-12090

TABLE 3.3. CT phase two Nash equilibria.

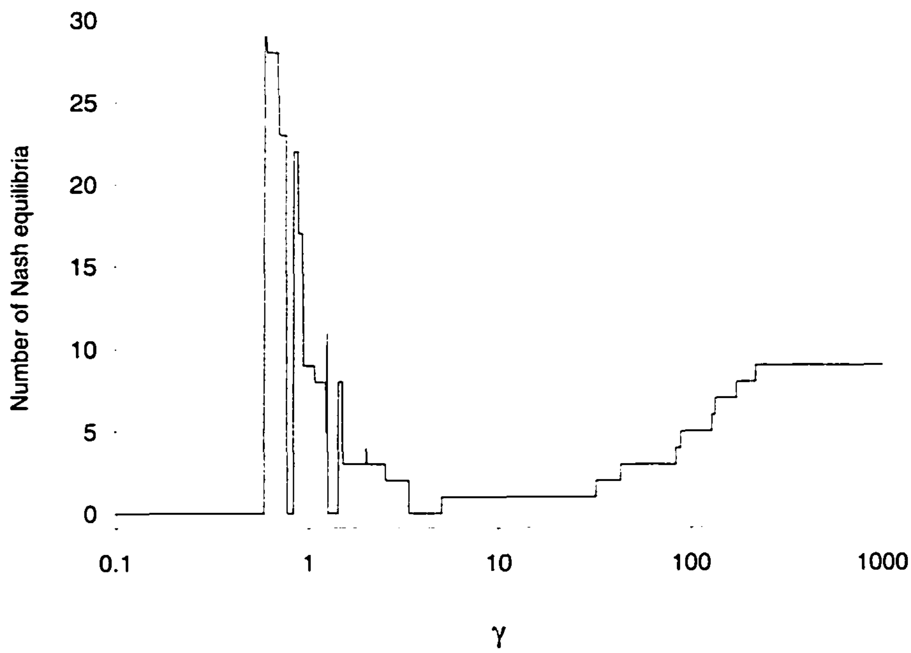


FIGURE 3.1. Number of Nash equilibria in the CP implementations.

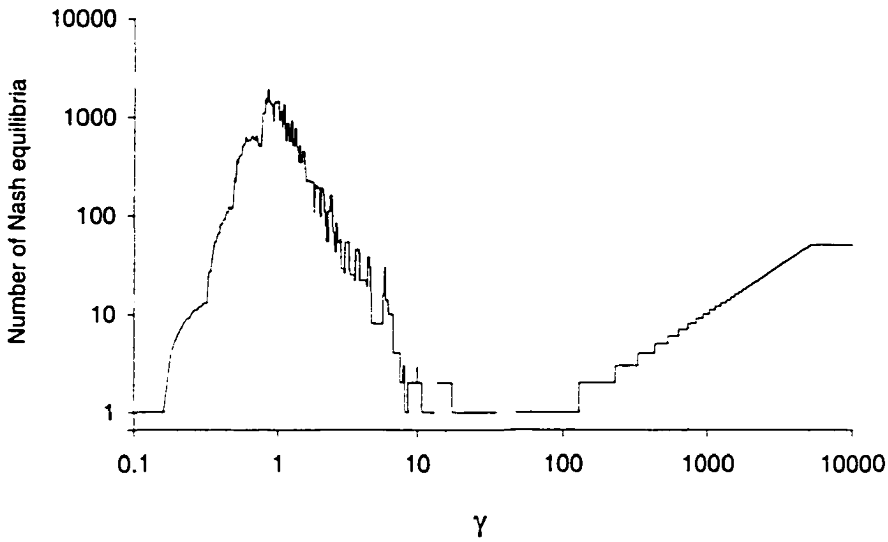


FIGURE 3.2. Number of Nash equilibria in the CT implementations.

Appendix A

BEST-RESPONSE CORRESPONDENCE

Given person i's quadratic value function, we can express his Groves-Ledyard payoff maximization problem as

$$\max_{x_i} \left(A_i X - B_i X^2 + \alpha_i - \frac{X}{I}c - \frac{\gamma}{2}\left(\frac{I-1}{I}(x_i - \mu_{-i})^2 - \sigma_{-i}^2 \right) \right).$$

Taking the derivative of person i's payoff function and setting it equal to zero, we see that the maximization condition is given by the linear equation

$$A_i - 2B_i X - \frac{c}{I} - \gamma \frac{I-1}{I}(x_i - \mu_{-i}) = 0.$$

Denote the best response for person i as $\widehat{x}_i$. Thus, to determine i's best response, we will change x_i in the above equation to $\widehat{x}_i$ and solve for this term. First we will isolate all terms containing $\widehat{x}_i$, noting that X also contains $\widehat{x}_i$:

$$2B_i X + \gamma \frac{I-1}{I}\widehat{x}_i = A_i + \gamma \frac{I-1}{I}\mu_{-i} - \frac{c}{I}.$$

Next, separate $\widehat{x}_i$ from X to get

$$2B_i \widehat{x}_i + 2B_i \sum_{j \neq i} x_j + \gamma \frac{I-1}{I}\widehat{x}_i = A_i + \gamma \frac{I-1}{I}\mu_{-i} - \frac{c}{I}.$$

Now, isolate $\widehat{x}_i$ and gather its terms:

$$\left(2B_i + \gamma \frac{I-1}{I} \right) \widehat{x}_i = A_i + \gamma \frac{I-1}{I}\mu_{-i} - \frac{c}{I} - 2B_i \sum_{j \neq i} x_j.$$

Note that $\mu_{-i} = \sum_{j \neq i} x_j / (I-1)$, so that we can write

$$\left(2B_i + \gamma \frac{I-1}{I} \right) \widehat{x}_i = A_i + \frac{\gamma}{I} \sum_{j \neq i} x_j - \frac{c}{I} - 2B_i \sum_{j \neq i} x_j.$$

Collect the terms of $\sum_{j \neq i} x_j$ to get

$$\left(2B_i + \gamma \frac{I-1}{I}\right) \widehat{x}_i = A_i + \left(\frac{\gamma}{I} - 2B_i\right) \sum_{j \neq i} x_j - \frac{c}{I}.$$

By dividing both sides by $2B_i + \gamma \frac{I-1}{I}$, we see that i's best response is

$$\widehat{x}_i = \frac{A_i + \left(\frac{\gamma}{I} - 2B_i\right) \sum_{j \neq i} x_j - \frac{c}{I}}{2B_i + \gamma \frac{I-1}{I}}.$$

Multiplying through by I/I gives a slightly reduced form of

$$\widehat{x}_i = \frac{A_i I + (\gamma - 2B_i I) \sum_{j \neq i} x_j - c}{2B_i I + \gamma (I-1)}.$$

Appendix B

Nash equilibrium provision level

To determine the provision level of the unique Nash equilibrium, we will simply sum the best response functions as defined in Appendix A. In order to sum all I best responses, we first rearrange the form of the best response function. Recall that person i's best response function is:

$$\widehat{x}_i = \frac{A_i + \left(\frac{\gamma}{I} - 2B_i\right)\sum_{j\neq i} x_j - \frac{c}{I}}{2B_i + \gamma\frac{I-1}{I}}.$$

Multiplying both sides of Equation B.1 by $2B_i + \gamma\frac{I-1}{I}$ we now have

$$\left(2B_i + \gamma\frac{I-1}{I}\right)\widehat{x}_i = A_i + \left(\frac{\gamma}{I} - 2B_i\right)\sum_{j\neq i} x_j - \frac{c}{I}$$

Moving all terms with a message to the left side we see

$$\left(2B_i + \gamma\frac{I-1}{I}\right)\widehat{x}_i + \left(2B_i - \frac{\gamma}{I}\right)\sum_{j\neq i} x_j = A_i - \frac{c}{I}.$$

By collecting the $2B_i$ terms on the left side we get

$$2B_i\left(\widehat{x}_i + \sum_{j\neq i} x_j\right) + \gamma\frac{I-1}{I}\widehat{x}_i - \frac{\gamma}{I}\sum_{j\neq i} x_j = A_i - \frac{c}{I}.$$

Note that $\mu_{-i} = \sum_{j\neq i}\frac{x_j}{I-1}$ and gather the $\gamma\frac{I-1}{I}$ terms so that we can rewrite the expression as

$$2B_i\left(\widehat{x}_i + \sum_{j\neq i} x_j\right) + \gamma\frac{I-1}{I}(\widehat{x}_i - \mu_{-i}) = A_i - \frac{c}{I}.$$

Assume that we are in Nash equilibrium and all persons are simultaneously best responding. Further, let $\widehat{\mu}_{-i} = \sum_{j \neq i} \widehat{x}_j / (I-1)$. The expression can now be rewritten as

$$2B_i\left(\widehat{x}_i + \textstyle\sum_{j\neq i}\widehat{x}_j\right) + \gamma\frac{I-1}{I}(\widehat{x}_i - \widehat{\mu}_{-i}) = A_i - \frac{c}{I}.$$

Recall that $\widehat{X} = \sum_{i\in I}\widehat{x}_i$. We now write

$$2B_i\widehat{X} + \gamma\frac{I-1}{I}(\widehat{x}_i - \widehat{\mu}_{-i}) = A_i - \frac{c}{I}.$$

We now aggregate over all participants. Taking the above expression for person i, and summing over all I people, we get

$$\sum_{i\in I}\left(2B_i\widehat{X} + \gamma\frac{I-1}{I}(\widehat{x}_i - \widehat{\mu}_{-i})\right) = \sum_{i\in I}\left(A_i - \frac{c}{I}\right).$$

Rearranging this expression, we see

$$2\widehat{X}\sum_{i\in I}B_i + \gamma\frac{I-1}{I}\left(\sum_{i\in I}\widehat{x}_i - \sum_{i\in I}\widehat{\mu}_{-i}\right) = \sum_{i\in I}A_i - MC_X.$$

Since $\sum_{i\in I}\widehat{\mu}_{-i} = \sum_{i\in I}\widehat{x}_i$, the term with this expression drops out, leaving us with

$$2\widehat{X}\sum_{i\in I}B_i = \sum_{i\in I}A_i - c.$$

Thus, we see that the unique Nash equilibrium provision level, $\widehat{X}$, can be expressed as

$$\widehat{X} = \frac{\sum_{i\in I}A_i - c}{2\sum_{i\in I}B_i}.$$

Appendix C

Existence and Uniqueness of Nash Equilibrium

This appendix will show that if $\beta_i = \beta_j$ for all i and j, then a unique Nash equilibrium exists. To prove this, we first show in a lemma that β_i has absolute bounds. These bounds are based on: (1) the definition of the Groves-Ledyard mechanism which specifies that γ is positive and I is an integer greater than 2, and (2) the assumption of strictly concave quadratic value functions which implies the quadratic term, B_i, is always positive. Then, using the lemma, we will see that a unique Nash equilibrium always exists.

Lemma 1. *If $\gamma > 0$, $I \in \{3, 4, \ldots\}$, and $B_i > 0$ for all $i \in \{1, 2, \ldots, I\}$, then* $-1 < \beta_i < \dfrac{1}{I-1}$.

Proof. By substituting for β_i, we can write the expression as

$$-1 < \frac{\gamma - 2B_iI}{2B_iI + \gamma(I-1)} < \frac{1}{I-1}.$$

Since the restrictions on γ, B_i, and I imply $2B_i + \gamma(I-1)$ is greater than 0 for all γ, B_i, and I, we can multiply the expression by $2B_i + \gamma(I-1)$ and see that

$$-2B_i - \gamma(I-1) < \gamma - 2B_iI < \frac{2B_iI + \gamma(I-1)}{I-1}.$$

We now express this compound inequality as two separate inequalities

$$-2B_i - \gamma(I-1) < \gamma - 2B_iI \quad \textit{and} \quad \gamma - 2B_iI < \frac{2B_iI + \gamma(I-1)}{I-1}.$$

Considering the first inequality, we see that adding $2B_i - \gamma$ to both sides results in $-\gamma I < 0$. It is obvious that this is always true, as γ and I are both always positive. Moving to consider the second inequality, we see that multiplying both sides by $I - 1$ yields

$$\gamma(I-1) - 2B_i I(I-1) < 2B_i I + \gamma(I-1).$$

Subtracting $\gamma(I-1)$ from both sides simplifies the expression to

$$-2B_i I(I-1) < 2B_i I.$$

Since both B_i and I are always positive, the above expression is always true. ∎

Assuming $\beta_i = \beta$ for all i, The determinant of the $I \times I$ Nash equilibrium coefficient matrix is expressed as

$$(1 - (I-1)\beta)(1+\beta)^{I-1}.$$

The Nash equilibrium system of equations has a unique solution if and only if the determinant of the coefficient matrix is nonzero. It is clear from the expression above that the determinant is zero only if $\beta = -1$ or $\beta = \dfrac{1}{I-1}$. However, Lemma 1 proved that β is never equal to either of these values. Thus, strictly concave value functions result in the existence of a unique Nash equilibrium.

References

[1] G. W. Brown, *Iterative solutions of games by fictitious play*, Activity Analysis of Production and Allocation (T. C. Koopmans, ed.), John Wiley, 1951.

[2] Colin F. Camerer and Teck-Hau Ho, *Experience-weighted attraction in games*, Econometrica **67** (1999), 827–874.

[3] Colin F. Camerer, Eric J. Johnson, Talia Rymon, and Sankar Sen, *Cognition and framing in sequential bargaining for gains and losses*, Frontiers of Game Theory (Ken Binmore, Alan Kirman, and Piero Tani, eds.), MIT Press, 1993.

[4] Yan Chen and Charles Plott, *The groves-ledyard mechansim: An experimental study of institutional design*, Journal of Public Economics **59** (1996), 335–364.

[5] Yan Chen and FangFang Tang, *Learning and incentive compatible mechanisms for public good provision: An experimental study*, Journal of Political Economy **106** (1998), 633–662.

[6] Giorgio Coricelli, *Strategic interaction in iterated zero-sum games*, Tech. Report 01–07, University of Arizona, 2001.

[7] Miguel Costa-Gomes, Vincent Crawford, and Bruno Broseta, *Cognition and behavior in normal-form games: An experimental study*, Econometrica **69** (2001), 1193–1235.

[8] Ido Erev and Alvin E. Roth, *Predicting how people play games: Reinforcement learning in experimental games with unique, mixed strategy equilibria*, American Economic Review **88** (1998), 848–881.

[9] Nick Feltovich, *Reinforcement-based vs. belief-based learning models in experimental asymmetric-information games*, Econometrica **68** (2000), 605–641.

[10] John Fox, *The learning of strategies in a simple, two-person zero-sum game without saddlepoint*, Behavioral Science **17** (1972), 300–308.

[11] Drew Fudenberg and David Levine, *Consistency and cautious fictitious play*, The Journal of Economic Dynamics and Control **19** (1995), 1065–1089.

[12] Jean Dickinson Gibbons and Subhabrata Chakraborti, *Nonparametric statistical inference*, Marcel Dekker, Inc., 270 Madison Avenue, New York, New York 10016, 1992.

[13] Steven Gjerstad, *The rate of convergence of continuous play*, Economic Theory **7** (1996), 161–178.

[14] Theodore Groves and John Ledyard, *Optimal allocation of public goods: A solution to the 'free rider' problem*, Econometrica **45** (1977), 783–809.

[15] Daniel Houser and Robert Kurzban, *Revisiting confusion in public good experiments*, American Economic Review (2002), Forthcoming.

[16] James S. Jordan, *Three problems in learning mixed-strategy nash equilibria*, Games and Economic Behavior **5** (1993), 368–386.

[17] Bernhardt Lieberman, *Experimental studies of conflict in some two-person and three-person games*, Mathematical Methods in Small Group Processes (Joan H. Criswell, Herbert Solomon, and Patrick Suppes, eds.), Stanford University Press, 1962, pp. 203–220.

[18] Kevin McCabe, Daniel Houser, Lee Ryan, Vernon Smith, and Theodore Trouard, *A functional imaging study of cooperation in two-person reciprocal exchange*, Proceedings of the National Academy of Sciences, U.S.A. **98** (2001), no. 20, 11832–11835.

[19] David M. Messick, *Interdependent decision strategies in zero-sum games: A computer-controlled study*, Behavioral Science **12** (1967), 33–48.

[20] Dilip Mookherjee and Barry Sopher, *Learning behavior in an experimental matching pennies game*, Games and Economic Behavior **7** (1994), 62–91.

[21] John Morgan and Martin Sefton, *An experimental investigation of unprofitable games*, Games and Economic Behavior (2002), Forthcoming.

[22] Thomas Muench and Mark Walker, *Are groves-ledyard equilibria attainable?*, Review of Economic Studies **50** (1983), 393–396.

[23] Yaw Nyarko and Andrew Schotter, *An experimental study of belief learning using real beliefs*, Econometrica (2002), Forthcoming.

[24] Barry O'Neill, *Nonmetric test of the minimax theory of two-person zerosum games*, Proceedings of the National Academy of Sciences, U.S.A. **84** (1987), 2106–2109.

[25] A. Rapoport and R. Boebel, *Mixed strategies in strictly competitive games: A further test of the minimax hypothesis*, Games and Economic Behavior **4** (1992), 261–283.

[26] Julia Robinson, *An iterative method of solving a game*, Annals of Mathematics **54** (1951), 296–301.

[27] Robert W. Rosenthal, Jason Shachat, and Mark Walker, *Hide and seek in arizona*, Tech. report, IBM TJ Watson Research Laboratory, 2002.

[28] Alvin E. Roth and Francoise Schoumaker, *Expectations and reputations in bargaining: An experimental study*, American Economic Review **73** (1983), 362–372.

[29] Timothy C. Salmon, *An evaluation of econometric models of adaptive learning*, Econometrica **69** (2001), 1597–1628.

[30] Jason Shachat, *Mixed strategy play and the minimax hypothesis*, Journal of Economic Theory (2002), Forthcoming.

[31] James W. Walker, Vernon L. Smith, and James C. Cox, *Bidding behavior in first price sealed bid auctions: Use of computerized nash competitors*, Economic Letters **23** (1987), 239–244.

[32] Eyal Winter and Shmuel Zamir, *An experiment with ultimatum bargaining in a changing environment*, Tech. Report 159, The Hebrew University of Jerusalem, December 1997, Center for Rationality and Interactive Decision Theory.