

ECONOMIC RESEARCH REPORTS

***PROCEDURAL RATIONALITY AND LEARNING  
IN GAMES: AN EXPERIMENTAL  
STUDY***

BY

**Antonio Merlo  
and  
Andrew Schotter**

RR # 92-33

July, 1992

**C. V. STARR CENTER  
FOR APPLIED ECONOMICS**



NEW YORK UNIVERSITY  
FACULTY OF ARTS AND SCIENCE  
DEPARTMENT OF ECONOMICS  
WASHINGTON SQUARE  
NEW YORK, N.Y. 10003

**PROCEDURAL RATIONALITY AND LEARNING IN GAMES:  
AN EXPERIMENTAL STUDY**

**Antonio Merlo  
University of Minnesota**

**and**

**Andrew Schotter  
New York University**

**July 1992**

The authors would like to thank Jess Benhabib, Yaw Nyarko, Franco Perrachi, Roy Radner, and the participants of the Microeconomics Workshop at New York University for their helpful comments. In addition, the research assistance of Vicky Myroni, Ken Rogoza and Blaine Snyder as well as the financial support of the C.V. Starr Center for Applied Economics are gratefully acknowledged

## Abstract

In this paper we focus on how human experimental subjects in one-person decision problems and two-person games go about learning. It is a paper on procedural rationality. We present evidence that laboratory subjects are somewhat purposeful in their learning, employing simple heuristic learning procedures which change as the environment they are placed in changes. We concentrate on what these heuristics are and how successful they are in guiding our subjects to the full information optimum or equilibrium. As such, our results represent a step on the road toward developing a class of learning models based on observed human behavior. Finally, our results have direct bearing on the methodology of experimental economics.

## Section 1: Introduction

The work of Simon (1976, 1978) and others on human problem solving and procedural rationality sees economic agents as taking an active role in the learning process. For example, it is well known that humans, when faced with a complex problem, either transform that problem into one that is manageable, or break it down in a manageable way so that at least a satisfactory solution is within their cognitive grasp. In this paper we focus on how human experimental subjects in one-person decision problems and two-person games go about learning. It is a paper on procedural rationality. We present evidence that laboratory subjects are somewhat purposeful in their learning, employing simple heuristic learning procedures which change as the payoff environment or institution they are placed in changes.

Unlike most game-theoretical models of learning in which agents are placed in a game of incomplete information, our experiments present laboratory subjects with complete information optimization problems and games which are too complex for them to deductively solve. As a consequence, they must employ some heuristic to aid them in their behavior and we concentrate on what that heuristic is and how successful it is in guiding our subjects to the full information optimum or equilibrium.

What we find interesting is how our subjects decide to structure their learning task when placed in different institutional settings. In our experiments there are two types of decision problems -- individual decision problems and two-person games -- and two payoff conditions which we call learn-while-you-earn and learn-before-you-earn. Hence we specify four different environments. As we will describe more fully later in this section, we find a major difference between how subjects attempt to learn in the learn-while-you-earn and learn-before-you-earn environments on the one hand, and also between the one-person and two-person decision tasks.

Although our interest is primarily in questions of procedural rationality, we do investigate whether the learning procedures created by our subjects lead them to an

optimal action or a Nash equilibrium (questions of substantive rationality), which represents the focus of the recent theoretical work on learning in games (see, e.g., Fudenberg and Kreps (1988), Fudenberg and Levine (1991), Jordan (1991), Kalai and Lehrer (1991), Milgrom and Roberts (1991), and Nyarko (1992)) or learning in individual decision models (see, e.g., Aghion et al. (1991), Easley and Kieffer (1988), and McLennan (1987)). In that literature it is typically assumed that economic agents attempt to maximize their expected earnings by choosing a best response to their recently updated beliefs about the other agents in the game or some unknown parameters in games against nature. The common feature of all of these approaches is that learning is the end result of the adaptive behavior of highly rational agents who have considerable calculating abilities which allow them to compute such things as Bayes–Nash or Bayes–Rational equilibria at each point during the game. In our paper, learning appears to be an act of agents who are boundedly rational, possibly more like the agents studied by Marimon (1990), who base their decisions upon observed average payoffs, or like the agents in the probabilistic learning models of Arthur (1990), Bush and Mosteller (1955), and Estes (1950), who modify their propensities to choose different actions on the basis of past experience of successes and failures.

Previous experimental papers on learning are also relevant. Specifically, Mookerjee and Sopher (1991) investigate how subjects learn in two–person matching pennies games under different information conditions. Like our results, they discover that different environments elicit different types of learning behavior. Furthermore, Merlo and Schotter (1991) investigate how subjects learn through experimentation in individual decision models similar to those studied here.

### **1.1: Environments and Institutions**

We distinguish between two payoff environments which we call learn–while–you–earn and learn–before–you–earn institutions since they span the spectrum of institutions with differing learning costs. A learn–while–you–earn payoff structure is the typical

payoff structure found in laboratory experiments and markets. In this environment time is divided into discrete periods with a known horizon  $T$  and in each period subjects or market participants make decisions. These decisions yield them a payoff at the end of the period, and their final payoff from the experiment or market is the sum of their (possibly discounted) period payoffs. The learn-while-you-earn environment is then one in which payoffs occur each period and cumulate throughout the experiment. The cost to learning is the opportunity cost associated with exploring the environment and a trade-off exists between "exploiting" actions already proven to be satisfactory by using them repeatedly and "exploring" to discover new and possibly better actions.

A learn-before-you-earn environment is a limit case of an institution with low learning costs. It is an artificial environment created by an experimental administrator, although it does have some parallels to real world markets. Here time is again divided into  $T$  discrete periods but no payoffs are awarded during the first  $T-1$  periods. Rather, subjects make decisions and observe what they would have earned if these were played for real. What does count is their period  $T$  decisions and these payoffs are sufficiently large so that their expected payoff from this last round decision is comparable to the expected sum of the payoffs in the learn-while-you-earn environment. In short, in a learn-before-you-earn setting there are  $T-1$  practice rounds and one real and lucrative round so that there is no exploration-exploitation trade-off. In terms of real world markets, we might consider two firms who expected to be infinitely lived and who interact repeatedly in a market in which both have extremely (possibly zero) discount rates. In such a market firms might treat any finite number of periods as free-learning periods since, with zero discount rates, any finite period would have only a negligible influence on their infinite horizon payoff. Under these circumstances, our learn-before-you-earn environment is a reasonable approximation to reality.

## 1.2: Overview of Our Results

Our results indicate that learning is situation and institution specific. For example, in one-person learn-before-you-earn decision problems subjects attempt to learn the structure of the problem they are facing (i.e., attempt to learn the nature of the payoff function facing them) and are successful in that attempt. However, in one-person learn-while-you-earn environments, subjects fail to learn the same things as their colleagues in the learn-before-you-earn experiment. As we will see, it appears that they fail to pay attention to the information they gathered during the experiment and do not use that information when making valuable economic decisions. Rather, subjects behave adaptively with very short memories, using their last period outcomes to guide their current period choices.

In two-person learn-before-you-earn games our results are similar to our one-person results but some modifications have to be made in their presentation. This is true because learning in games, with their strategic uncertainty, is different from learning in one-person environments where there can only be exogenous stochastic uncertainty. More precisely, it may not be practical for subjects to learn the Nash equilibrium of the two-person game they are engaged in since the information they receive each period may be too meager given the noise in the environment. In such low information settings, a live human subject may decide to treat their opponent as a purely random device and may formulate the problem to himself as a one-person decision problem similar to a multi-armed bandit problem. In such a setting, the subject must decide which decision number (or arm) is the "best" without concerning him or herself about the reaction of the other subject to their treatment of the problem. What we find is that subjects in learn-before-you-earn environments reduce the complexity of the problem they face and attempt to find for themselves a profitable mode of behavior or rule of thumb. They do this by behaving like statisticians who test hypotheses as to which decision number is best for them to play, but

these statisticians limit themselves to a small set of decision numbers. At the end of the experiment, our subjects do, in fact, choose that arm which was associated with the highest average payoff for themselves during the first 74 practice rounds. However, these decisions are not Nash equilibrium choices.

In learn-while-you-earn games, while subjects also simplify the decision problem they face by limiting themselves to a small number of decision numbers, they do not seem to choose one of them and certainly not the optimal (profit maximizing) arm at the end of the experiment. What we find is that, like their counterparts in one-person learn-while-you-earn environments, subjects adopt an adaptive behavior with limited memory in which they change their choice depending upon the outcome of the game in the most recent rounds. Furthermore, they do not respond to the decision choices of their opponent.

### **1.3: What Do Our Results Teach Us**

While it is never wise to generalize on the basis of a small number of experimental results, there are a number of lessons that we can learn from these experiments if they hold up to replication elsewhere. To begin, our results add yet another piece of evidence for the growing view that learning is a situation (or institution) specific phenomenon. This view has been recently summarized by Milgrom and Roberts (1991) as follows:

"Taken together, these results [i.e., earlier theoretical results on learning in games, *cfr.*] raise serious doubts about the validity of Nash equilibrium and its refinements as a general model of the likely outcomes of adaptive learning. More fundamentally, they indicate that the 'rationality' of any particular learning algorithm is situation dependent: An algorithm that performs well in some situations may work poorly in others. Apparently, real biological players tailor rules of thumb to their environment and experience: They learn how to learn. Thus, any single, simple specification of a learning algorithm is unlikely to represent well the behavior that actual players would adopt." (p. 84).

Such a view, that learning is situation or institution dependent, creates a problem for economic theory in its quest to present generalizable models of economic behavior. If learning is situation dependent then it raises the possibility that one would have to

construct special learning theories for each and every economic institution — certainly a dismal prospect. This opens the door for experimentalists, however, since if they could classify institutions into equivalence classes across which human learning behavior is similar, then theorists could attempt to characterize these institutions. If successful, a small class of learning theories might be constructed which would explain behavior in a large number of institutions. Our results take a step along this path.

One should not take these remarks as negative or pessimistic since we already consider our results as being supportive of much of the work undertaken by game theorists. For example, in our one-person learn-before-you-earn environments, our subjects' behavior is consistent with the behavior predicted by Aghion et al. (1991) where agents in infinite horizon problems without discounting optimally learn the shape of the payoff function they are facing. Further, in our two-person learn-while-you-earn environment, which most closely mimic the non-cooperative games studied, for example, by Jordan (1991), Kalai and Lehrer (1991), and Nyarko (1992), we find that laboratory subjects adhere to adaptive learning strategies with limited memory. While we see no evidence that our subjects perform the type of calculations called for by the modern game-theoretical literature, they do appear to mimic the adaptive pattern described in Milgrom and Roberts (1991), where adaptive learning is shown to be capable of reaching Nash equilibrium outcomes in certain classes of games. Whether our subjects' behavior would eventually converge to Nash equilibrium or first-best optimal behavior can not be answered by our experiments since those convergence results are asymptotic and our experiments were obviously run with only finite (75 round) time horizons. However, our experiments do provide a clear guide as to the procedural rationality of our subjects in these institutions and the assumption that such subjects act adaptively (and myopically) would seem to be a good guide to theory.

Finally, our results have direct bearing on the methodology of experimental economics. This is true because almost all experiments in economics aim to test static

theories using a repeated game framework. This is typically justified by the claim that doing an experiment once and only once does not allow subjects to learn. Hence, repetition is recommended to foster learning. In most designs, subjects play games repeatedly and earn payoffs each period -- they play in a learn-while-you-earn environment. In performing statistical tests on the data generated by these experiments, experimentalists typically use observations collected at the end of the experiment since those supposedly distill all the information learned during the course of the experiment. Our results, however, indicate that it is exactly in these types of learn-while-you-earn environments that learning is most problematic and in which subjects seem to pay the least amount of attention to the data generated during their multi-period play. Hence, our results seem to lead towards an advocacy of learn-before-your-earn environments in the laboratory in an effort to focus subjects' attention on the data they are generating.

In this paper we will proceed as follows. In Section 2 we describe the game used as the basis for our experiments as well as and our experimental design. In Section 3 we report our results. Finally, in Section 4 we offer some conclusions and speculations.

## **Section 2: Experimental Setting**

All of the experiments performed to investigate learning were of the tournament variety and similar to those of Bull, Schotter and Weigelt (1987) and Schotter and Weigelt (1991). In those experiments, randomly paired subjects must, in each round, each choose a number,  $e$ , between 0 and 100 called their decision number. After this number is chosen a random number is independently generated by each subject from a uniform distribution over the interval  $[-a, +a]$ . These numbers (each player's decision number and random number) are then added together and a "total number" defined for each player. Payoffs are determined by comparing the total numbers of the subjects in each pair and awarding that subject with the largest total number a big payment of  $M$  and that subject with the

smallest total number a small payment of  $m$ ,  $M > m$ . To calculate their final payment a subject would have to subtract the "cost" of their decision number from their payment, where the cost is given by a convex function  $c(e) = k_1 e^2 / k_2$  defined over the interval  $[0,100]$ . Hence in these experiments there is a tradeoff in the choice of decision numbers; higher numbers generate a higher probability of winning the big prize but also define a higher decision cost. If we let  $2a$  be the range of the random variable from which the random numbers are (independently) generated,  $k_1$  and  $k_2$  be two constants in the cost function, and  $M$  and  $m$  be the big and small prizes respectively, then when  $k_1 = 1$ ,  $k_2 = 500$ ,  $a = 40$ ,  $M = 29$ , and  $m = 17.2$ , the two-person tournament defined has a unique symmetric Nash equilibrium at 37. For a description of the theory of tournaments underlying these experiments see e.g. Lazear and Rosen (1981) and Schotter and Weigelt (1992).

We consider these experiments to be good ones for our purposes since they are simple to describe to subjects yet complicated enough so that a deductive solution should be out of the grasp of most experimental subjects.<sup>1</sup> Such complexity would force subjects to learn inductively and it is this process that we are interested in studying. Hence, even though the experiments run are complete information experiments for which optimal actions or Nash equilibria could be calculated a priori, the problem was sufficiently complex so that inductive learning was required. Further, these games involve both strategic uncertainty and exogenous stochastic uncertainty. We find this feature appealing since it might allow subjects to meaningfully treat their opponent as a random device and pursue a multi-armed bandit strategy of the type discussed above.

### 2.1: Experimental Design

We performed a set of 6 different experiments. In Experiments 1a, 1b, and 1b' subjects played the tournament game described above 75 times consecutively against a

---

<sup>1</sup> Such games were also the focus of study by Knez (1992).

computer whose strategy was known to be that of choosing 37 in each period. This experiment then presented our subjects with a one-person decision problem (since the other side of the game was exogenously determined), where the only thing to be learned was the shape of the payoff function they were facing. This experiment was performed under three payoff regimes.

In Experiment 1a subjects did the experiment 74 times without receiving a payoff but were paid for the decisions that they made only in round 75. This was the learn-before-you-earn condition. With the parameters specified above, the (equilibrium) expected amount for this one period choice was \$15.27. (Actually, payoffs were denominated in a fictitious currency called francs and converted into dollars at the rate of \$.75 per franc).

In Experiment 1b subjects performed the same experiment 75 times but received a payoff in each of the 75 rounds. Their final payoff was the sum of their 75 round earnings over the course of the experiment. (Obviously their payoff was scaled down by converting francs into dollars at the rate of \$.01 per franc). This was the learn-while-you-earn condition. It is important to note that in Experiments 1a and 1b the expected earnings of subjects were equal at the unique optimum.

Experiment 1b' was identical to Experiment 1b except for a simple extension. In this experiment, after the 75 rounds were over, subjects were then informed that they would perform the experiment one more time with increased payoffs -- actually, with the same payoffs that were previously used in the learn-before-you-earn environment. (They had not been told about this extra experiment until after they had finished their 75 round experiment). In other words, subjects in Experiment 1b' performed the experiment twice. Once for 75 rounds with small payoffs in each round (the same payoffs as in Experiment 1b), and once for one extra round with a one round payoff equal to the payoff in the last round of the learn-before-you-earn experiment. This extra-round experiment, with

increased payoffs, places subjects in a situation identical to the one faced by learn–before–you–earn subject in round 75 (their only payoff–relevant round). As such, their choice there should serve as a sufficient statistic for all that these subjects have learned during the course of the experiment as does the 75<sup>th</sup> round decision of subjects in the learn–before–you–earn Experiment 1a.

Experiments 2a, 2b, and 2b' increase the learning task of our subjects by having them play against a live human subject for 75 rounds instead of a computer. In each experiment subjects were randomly assigned to an opponent and kept that opponent, whose identity was unknown to them, during the entire course of the experiment. (This fact was common knowledge.) To make the design symmetric we ran Experiment 2a under the learn–before–you–earn condition and Experiment 2b under the learn–while–you–earn condition. Further, we ran Experiment 2b' in which we had the learn–while–you–earn game played 75 times as it was in Experiment 2b and then had an extra round with increased stakes played afterwards just as in Experiment 1b'. Again, this extra round was not announced until the 75 round learn–while–you–earn game was finished. In this experiment subjects played the extra round against that person with whom they were matched during the first 75 rounds (and that information was also common knowledge).

This experimental design is summarized in Table 2.1.

### **Section 3: Results**

We will present our results by first treating the one–person decision experiments and then the two–person game environments. To help us in this task we will formulate our results in the form of facts that we think can be substantiated by the data.

#### **3.1: One–Person Decision Problems**

As was mentioned in the introduction, the learning task our subjects faced in one–person environments can be completely characterized as an attempt to discover the shape of

**Table 2.1: Experimental Design**

<i>Experiment</i>	<i>Opponent</i>	<i>Opponent's Strategy</i>	<i>Payoff Condition</i>	<i>Number of Subjects</i>
1a	Computer	Fixed and Known to be 37	L-B-Y-E	12
1b	Computer	Fixed and known to be 37	L-W-Y-E	12
1b'	Computer	Fixed and known to be 37	L-W-Y-E (+ Extra Round)	12
2a	Human	N.A.	L-B-Y-E	18
2b	Human	N.A.	L-W-Y-E	20
2b'	Human	N.A.	L-W-Y-E (+ Extra Round)	12

---

L-B-Y-E: Learn-Before-You-Earn

L-W-Y-E: Learn-While-You-Earn

---

the payoff function they were facing. Faced with this specific task we investigate exactly how subjects go about learning.

### **3.1.1: Learn–Before–You–Earn Environments**

*Fact 1: In one–person learn–before–you–earn environments, subjects behave as if they were attempting to estimate the payoff function they faced and, in the 75<sup>th</sup> round, choose an action significantly close to the peak of their estimated function. Further, on average the 75<sup>th</sup> round choice of subjects was close to the optimal decision number of 37.*

The claim made by Fact 1 is that subjects acted as if they were econometricians who used the observations generated by their first 74 round actions to estimate the quadratic function that best fit the data observed and then used its shape to guide their behavior in the 75<sup>th</sup> and only payoff–relevant round of the experiment. The support for this claim is presented in Figure 1 and also in Table 3.1.

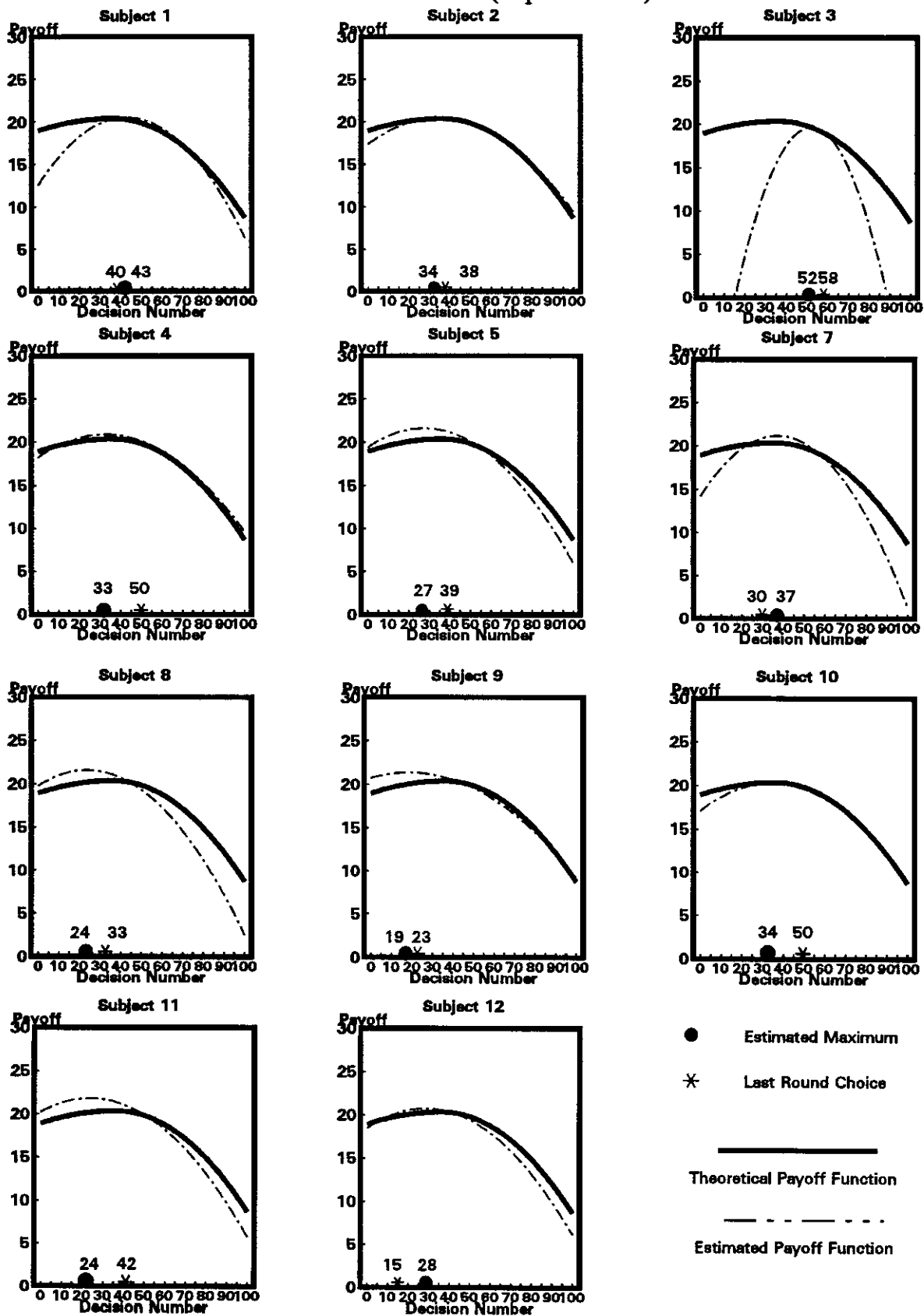
Figure 1 presents the actual payoff function faced by our subjects in their one–person learn–before–you–earn decision problem (Experiment 1a) along with least–squares quadratic approximations to the data they generated during the first 74 rounds of the experiments.<sup>2</sup> Note by inspection how closely each estimated payoff functions are to the theoretical or actual payoff function these subjects faced.<sup>3</sup> This implies that if subjects had paid attention to the data they generated and used that data to estimate a payoff function, choosing the maximum of this estimated payoff function would have provided a good guide to an optimal action in round 75.

---

<sup>2</sup> The estimates are reported in Table A.1 in the Appendix together with the coefficients of the actual (theoretical) payoff function (Table A.0). Note that we dropped Subject 6 who can be classified as an outlier.

<sup>3</sup> This claim is substantiated by the results of F–tests of the hypothesis of equality of the estimated coefficients to the coefficients of the theoretical payoff function (Table A.1 in the Appendix). The null hypothesis of equality can not be rejected at the 5% significance level for 6 out of 12 subjects.

Figure 1: Payoff Functions  
Learn-Before-You-Earn (Experiment 1a)



**Table 3.1: Learning Results - Experiment 1a**

<i>Subject</i>	<i>Estimated Maximum</i>	<i>Standard Error</i>	<i>Last Period Choice</i>	$ (4) - (2) $	$ (4) - 37 $
1	43	16.9	40	3	3
2	34	5.9	38	4	1
3	52	4.5	58	6	21
4 *	33	5.0	50	17	13
5	27	21.3	39	12	2
6 * (a)	-	-	-	-	-
7	37	4.9	30	7	7
8	24	8.1	33	9	4
9	19	11.5	23	4	14
10 *	34	5.4	50	16	13
11	24	10.7	42	18	5
12	28	8.2	15	13	22
<b>Mean</b>	32.3	9.3	38	9.9	9.5

---

(a) Subject 6 is classified as an outlier

---

Table 3.1 provides insights into their behavior by asking whether subjects did, in fact, act as if they chose in period 75 in a manner consistent with their estimated payoff function. In other words, was their round 75 choice significantly close to the maximum of their estimated payoff function (i.e., was it within a 95% confidence interval of the point estimate of the peak) and/or was it close to 37 which was the optimal solution to the problem they faced. In Column 2 of Table 3.1 we see the payoff function maximum that would have been estimated by a subject if he or she used the least-squares quadratic approximation to the data they generated during the first 74 rounds of the experiment depicted in Figure 1. In Column 3 of this table we present the standard error associated with this point estimate (computed, using the Delta Method, from the estimates reported in Table A.1 as illustrated in Table A.3 in the Appendix)<sup>4</sup>, while in Column 4 we present the actual choice made by the subject in round 75. Finally, Column 5 (6) calculates the absolute difference between the last period choice and the estimated maximum (the optimal choice of 37).

Our support for Fact 1 comes from the fact that in Experiment 1a (Table 3.1) 9 out of the 12 subjects acted as if they had estimated the payoff function they faced and in round 75 chose a decision number that was within two standard errors of the point estimate of the maximum of that function. (Subjects who failed this criterion have asterisks placed next to their numbers in Table 3.1). The mean absolute difference between subjects' last period choice and the estimated maximum of their payoff function was 9.9. Further, the mean absolute difference between the last period choice of subjects and the optimal choice of 37 was similarly 9.5. The fact that these means are so similar indicates that subjects estimated their true payoff functions accurately and then chose numbers in period 75, when there was money on the line, that were close to their estimated maxima.

---

<sup>4</sup> For a general description of the Delta Method see, for example, Goldberger (1991).

Finally, note that the mean choice in round 75 of subjects in Experiment 1a was 38 which is remarkably close to the optimum decision number of 37. This can be attributed to the fact that subjects were acting as if they were attempting to estimate the payoff function they faced using an unbiased and consistent estimation procedure. Therefore, even with a finite 75 round history the mean of their estimated maxima should be centered around 37.<sup>5</sup>

### **3.1.2: Learn-While-You-Earn Environments**

***Fact 2:** In one-person learn-while-you-earn environments, subjects do not behave in a manner comparable to learn-before-you-earn subjects. More precisely, at the end of the experiment they do not make choices consistent either with the maxima of their estimable payoff functions or with the optimal choice of 37.*

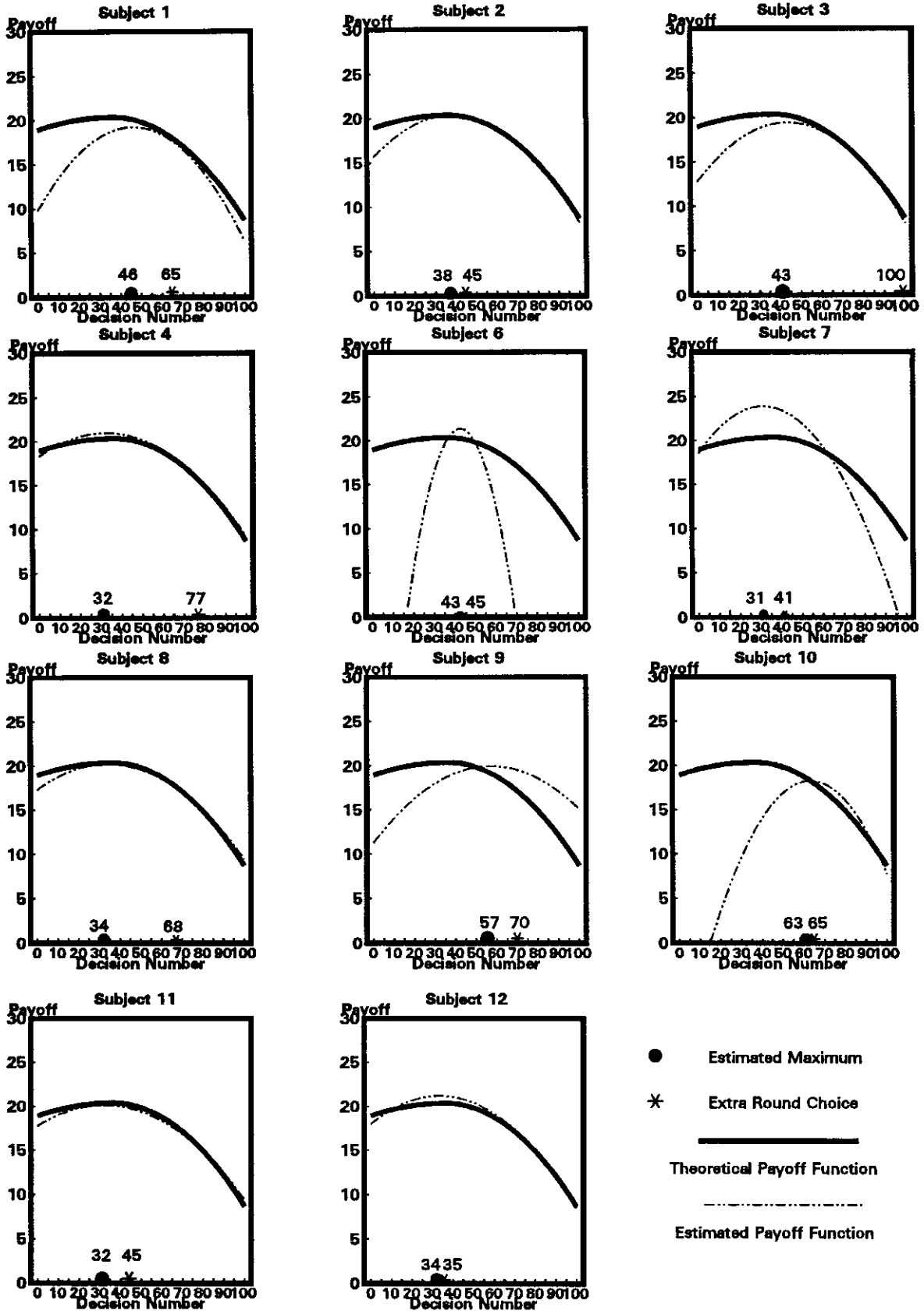
To support this fact we use the data generated by Experiment 1b'. Note that in the extra round of such an experiment subjects have as much information as did the learn-before-you-earn subjects in round 75 of their experiment (Experiment 1a) and the same incentives to use it.

Support for Fact 2 is presented in Figure 2 and in Tables 3.2. Figure 2 presents the actual payoff function faced by our subjects in their one-person learn-while-you-earn decision problem (Experiments 1b') along with least-squares quadratic approximations to the data they generated during the experiment. As we can see in Figure 2, the data generated by subject choices over the first 75 rounds of Experiment 1b' was sufficient to

---

<sup>5</sup> This is not an artifact of the fact that since 37 was announced as the computer's choice, it might act as a focal point. To demonstrate this we ran an experiment which was identical to Experiment 1a except for the fact that subjects were not told the number that the computer was choosing (although they were told that the computer was choosing the same number throughout the length of the experiment). Subjects' behavior here was almost identical to that observed in Experiment 1a. 8 out of 10 subjects in this experiment acted as if they had estimated the payoff function they faced and in round 75 chose a decision number that was within two standard errors of the point estimate of the maximum of that function. Furthermore their mean choice in round 75 was 35.5 which again is remarkably close to the optimal choice of 37. The data relative to this experiment are available from the authors.

Figure 2: Payoff Functions  
Learn-While-You Earn (Experiment 1b')



**Table 3.2: Learning Results - Experiment 1b'**

<i>Subject</i>	<i>Estimated Maximum</i>	<i>Standard Error</i>	<i>Extra Period Choice</i>	$ (4) - (2) $	$ (4) - 37 $
1 *	46	0.6	65	19	28
2	38	20.4	45	7	8
3 *	43	8.6	100	57	63
4 *	32	7.8	77	45	40
5 * (a)	-	-	-	-	-
6	43	2.1	45	2	8
7 *	31	3.8	41	10	4
8	34	36.9	68	34	31
9 *	57	1.6	70	13	33
10	63	8.3	65	2	28
11 *	32	5.1	45	13	8
12	34	7.5	35	1	2
<b>Mean</b>	41.2	9.3	59.6	18.5	23.0

---

(a) Subject 5 is classified as an outlier

---

allow a good quadratic approximation to the true function as was true in Experiment 1a.<sup>6</sup> However, unlike Experiment 1a, in the extra round of Experiment 1b' subjects failed to choose decision numbers significantly close to their estimated maxima. In short, subjects in learn-while-you-earn one-person environments did not act as if they had learned the shape of the payoff function they faced during the experiment and chose decision numbers close to the maximum of these functions even when a considerable amount of money was at stake. This is illustrated in Table 3.2 where we list the estimated maximum for each subject in Experiment 1b' (Column 2) along with the standard error of this estimate (Column 3) and their choice in the extra round (Column 4). Column 5 presents the differences between the extra-round choice and the estimated maximum. Finally, Column 6 computes such differences with respect to the optimal choice of 37.

Note first that only 5 out of the 12 subjects in Experiment 1b' made choices in the extra round which were within two standard errors of their estimable payoff function maxima. (Subjects who failed this criterion have asterisks placed next to their numbers in Table 3.2). This is in contrast to 9 out of 12 subjects in Experiment 1a choosing decision numbers at the end of the experiment within two standard deviations of their estimated maximum. The absolute mean difference between subjects' choices in the extra round of Experiment 1b' and the maxima of the estimable payoff functions was 18.5. In the learn-before-you-earn Experiment 1a this difference was only 9.9. Further, the mean absolute difference between the choice of subjects in the extra round of Experiment 1b' and the optimal choice of 37 was 23 as opposed to 9.5 in Experiment 1a.<sup>7</sup> In addition, the mean

---

<sup>6</sup> These claims are once again substantiated by the results of F-tests of the hypothesis of equality of the estimated coefficients to the coefficients of the theoretical payoff function (Table A.2 in the Appendix). The null hypothesis of equality can not be rejected at the 5% significance level for 7 out of 12 subjects. Note that we dropped Subject 5 who can be classified as an outlier.

<sup>7</sup> A Wilcoxon test on the sample of differences between these two experiments rejects the null hypothesis that they come from the same population at the 5% significance level.

choice of subjects in the extra round of Experiment 1b' was 59.6 which is substantially far from the optimum decision number of 37. (In Experiment 1a it was 38).

Note that in the extra round of Experiment 1b', subjects are able, if they want to, to avail themselves of all of the information generated in the previous 75-round learn-while-you-earn experiment. Since there is a relatively large amount of money on the line, we might think that subjects would take advantage of the information they previously generated. In fact, in this extra round they have as much information as did the learn-before-you-earn subjects in round 75 of their experiment (Experiment 1a) and the same incentives to use it. The interesting thing is that, as we indicated above, they fail to use it efficiently in learn-while-you-earn environments whereas they do not in learn-before-you-earn environments. The explanation we propose for this anomaly is that subjects in learn-before-you-earn environments simply did not pay attention to the data they were generating during the experiment but rather reacted myopically on a round-by-round basis without taking into account the cumulative information they were obtaining. This proposed explanation is described more fully in Fact 3.

*Fact 3: Subjects in learn-while-you-earn one-person environments behave in an adaptive manner. More precisely, they only look one period back to determine their decision in any given period. Furthermore, there appears to be no change in the rule of thumb subjects use to determine their decision over the entire length of the experiment.*

To substantiate this fact we use the data generated by Experiments 1b and 1b'. To characterize the behavior of our subjects in learn-while-you-earn one-person environments we ran two sets of regressions. The first regression ([1]) represents a linear approximation to the rule of thumb used by our subjects to determine their decision number. Here we regressed the decision number chosen by a subject in each round (dec) on his or her lagged decision numbers, lagged payoffs (pay) and lagged dummy variables (win) denoting a win (1) or a loss (0) in the tournament. We used 1, 2, and 3 period lags. The second

regression ([2]) is aimed at characterizing the adjustment rule our subjects used to change their choices from round to round. Here we regressed the change in decision number chosen by subjects in consecutive rounds of the experiment (diff), on lagged values of the same explanatory variables as above with a similar lag structure. We ran such regressions for each individual in Experiments 1b and 1b' and on the pooled sample obtained by combining all individual histories of subjects in these experiments. These panel regressions were run in order to identify behavioral regularities in the population. Also, we included among the regressors round-specific dummy variables to differentiate the early rounds of the experiment (dr25, defined by round  $\leq 25$ ) from the later ones (dr50, defined by round  $> 50$ ). The inclusion of these round dummies was motivated by an attempt to discover whether the decision rule used by subjects changed in the later rounds of the experiment as a result of learning. In the regressions conducted on the pooled sample we also included among the covariates player-specific dummy variables to control for individual heterogeneity.

A truly "myopic" adaptive strategy would be consistent with significant coefficients for the one-period lagged variables but insignificant coefficients for the second and third period lags. The presence of learning would be signalled by significant coefficients for the round dummies. The estimates obtained for the regressions performed using the pooled sample (1728 observations) are the following: (To economize on space we present our results without reporting the estimated coefficients for the individual dummy variables.)

$$\begin{aligned}
 [1] \quad \text{dec}_t^i &= \alpha^i + 0.521 * \text{dec}_{t-1}^i + 0.097 * \text{dec}_{t-2}^i + 0.086 * \text{dec}_{t-3}^i + \\
 &\quad (0.116) \qquad\qquad (0.108) \qquad\qquad (0.105) \\
 &+ 1.485 * \text{pay}_{t-1}^i - 0.027 * \text{pay}_{t-2}^i - 0.390 * \text{pay}_{t-3}^i + \\
 &\quad (0.571) \qquad\qquad (0.524) \qquad\qquad (0.503) \\
 &- 19.384 * \text{win}_{t-1}^i - 0.942 * \text{win}_{t-2}^i + 4.090 * \text{win}_{t-3}^i + \\
 &\quad (6.777) \qquad\qquad (6.241) \qquad\qquad (6.018) \\
 &- 0.704 * \text{dr25} - 0.560 * \text{dr50}, \qquad R^2 = 0.613, \\
 &\quad (1.012) \qquad\qquad (0.841)
 \end{aligned}$$

and

$$\begin{aligned}
 [2] \quad \text{diff}_t^1 = & \alpha^1 - 0.024 * \text{dec}_{t-2}^1 + 0.021 * \text{dec}_{t-3}^1 + \\
 & (0.107) \qquad \qquad \qquad (0.102) \\
 & + 3.619 * \text{pay}_{t-1}^1 - 0.590 * \text{pay}_{t-2}^1 - 0.688 * \text{pay}_{t-3}^1 + \\
 & (0.191) \qquad \qquad \qquad (0.518) \qquad \qquad \qquad (0.493) \\
 & - 45.521 * \text{win}_{t-1}^1 + 5.479 * \text{win}_{t-2}^1 + 7.366 * \text{win}_{t-3}^1 + \\
 & (6.777) \qquad \qquad \qquad (6.241) \qquad \qquad \qquad (6.018) \\
 & - 0.459 * \text{dr25} - 0.551 * \text{dr50}, \qquad \qquad R^2 = 0.388. \\
 & (1.025) \qquad \qquad \qquad (0.853)
 \end{aligned}$$

The numbers in parentheses are robust (heteroskedasticity-consistent) standard errors computed using Huber's formula (Huber (1967)).<sup>8</sup>

These results clearly expose the "adaptive myopia" of our experimental subjects in learn-while-you-earn one-person environments and we present them as substantiation of Fact 3. First, note that in either regression the only coefficients that are significant at the 5% confidence level are the ones associated with one-period lagged variables. None of the coefficients associated with two and three period lagged variables is significant at conventional significance levels. Such findings are consistent with the idea that subjects in these environments only look one period back to determine their decision in any given period and react in a purely adaptive way to its outcome by increasing their choice if either they lose the tournament or if they win with a relatively low decision number -- i.e. they receive a "relatively high" payoff.<sup>9</sup> Furthermore, we do not find any evidence of round heterogeneity since none of the round dummy variables in either regression have

<sup>8</sup> This is also known as White's method (White (1980)).

<sup>9</sup> More precisely, by looking at the coefficients in equation [1] in elasticity form, we see that, on average, our subjects chose their decision numbers in any period by responding to a +1% change in their decision (payoff) in the previous round of the experiment with a +.52% (+.58%) change in their current decision number, and to a win in the tournament with a -.41% change.

significant coefficients attached to them.<sup>10</sup> This suggests that our experimental subjects do not revise their decision rules towards the end of the experiments as a result of learning. Finally, we find significant heterogeneity among players in the intercept terms. Qualitatively similar results both in terms of the sign and the significance level of the estimated coefficients emerge, however, from the regressions conducted at the individual level.<sup>11</sup>

### **3.2: Two–Person Games**

As stated in the introduction, the learning problem faced by subjects in a two–person game experiment is far more complex than in a one–person decision setting for the simultaneous presence of stochastic and strategic uncertainty. Information gathering is a noisier process since the actions of one’s opponent complicate the process. Faced with this complexity we investigate exactly how subjects go about learning and what they decide to learn about.

#### **3.2.1: Learn–Before–You–Earn Environments**

*Fact 4: Subjects in a learn–before–you–earn game environment attempt to find an appropriate mode of behavior for themselves to follow or an appropriate rule of thumb. They do this by treating the two–person decision problem they face as if it were a one–person decision problem similar to a multi–arm bandit problem and limiting their choices during the 74 round trial period to a small set of actions which they play repeatedly in an attempt to gather information. At the end of the experiment, they choose that action, from the set they have limited themselves to, whose mean payoff was highest during these trial runs, i.e. that action which was proven to be the best arm.*

---

<sup>10</sup> Similar regressions run also including round dummies for the slope coefficients confirm this fact. No round dummies enter either regression with coefficients that are significantly different from zero.

<sup>11</sup> These results are available from the authors.

We substantiate Fact 4 by looking at the time series of the choices our subjects made during the 75 rounds of Experiment 2a and their 75<sup>th</sup> round choice. We categorize a choice as a mode if it was chosen at least 5 times during the 74 trial rounds. The claim of Fact 4 is that subjects in this environment will simplify the task facing them by transforming it into a multi-arm bandit problem and then by seriously investigating the payoff properties of a small set of decision numbers (modes or arms). These decision numbers will be used frequently during the experiment and will furnish a set of working hypotheses for the subjects to test during the experiment to see which decision number is best. In the 75<sup>th</sup> round of the experiment, subjects will choose one of the decision numbers they have tested.

Fact 4 is substantiated if subjects choose one of these modal choices in round 75. We will also check to see if their choice is "optimal" or at last optimal for the transformed modal choice problem by calculating their average payoff over the number of times they choose a modal choice and comparing it to their average payoff for choices away from that choice. A decision number in round 75 will be optimal (or boundedly rational with respect to the transformed problem) if its average payoff when used in the experiment was greater than the average payoff when it was not used or when any other mode was used.

Table 3.3 presents each of the 18 subjects in Experiment 2a (the learn-before-you-earn two-person game) along with each decision number that was a mode of behavior for them and the number of times it was used during the first 74 rounds. We also present the mean payoff (and standard deviation) they received when using each of these decision numbers as well as their mean payoff away from it. Finally we list their 75<sup>th</sup> round choice along with the number of times that that decision number was used during the experiment.

Several things are noteworthy. First, for all of the 18 subjects in this experiment (Experiment 2a) we were able to identify at least one mode. (Note that the definition of a

**Table 3.3: Modes of Behavior - Experiment 2a**

<i>Subject</i>	<i>Modal Choices</i>	<i># of Times Chosen</i>	<i>Average Payoff</i>	<i>Standard Deviation</i>	<i>Last Period Choice</i>	<i># of Times Chosen</i>
1	0	9	18.51	3.93	20	8
	20	8	22.30	6.31		
	35	5	21.83	6.46		
	50	7	17.26	6.31		
	80	7	12.83	5.76		
	100	6	9.00	0.00		
	other	33	16.48	6.92		
2	0	62	21.01	5.56	0	62
	other	13	14.95	6.49		
3	15	6	22.65	6.46	30	12
	20	19	25.09	5.34		
	29	6	25.35	4.82		
	30	12	25.23	4.59		
	other	32	22.33	4.84		
4	0	21	21.13	5.70	30	7
	30	7	23.83	5.76		
	40	7	22.43	5.76		
	45	14	21.58	5.53		
	50	17	21.92	4.64		
	other	9	14.12	4.78		
5	65	6	10.72	4.82	66	11
	66	11	14.92	6.16		
	73	6	14.41	6.09		
	other	52	15.11	5.32		
6	88	5	11.15	5.28	100	3
	other	70	13.47	6.28		
7	0	15	19.56	4.89	0	15
	5	40	22.75	5.97		
	25	5	20.67	6.46		
	other	15	18.24	6.65		
8	20	15	20.33	5.76	70	12
	40	7	17.37	5.76		
	50	15	16.13	5.76		
	60	11	14.29	5.95		
	70	12	13.30	6.16		
	other	30	18.74	6.07		
9	0	69	20.28	5.22	15	1
	other	6	19.68	5.04		
10	53	5	16.30	6.46	54	6
	54	6	21.20	4.82		
	56	7	15.98	6.31		
	65	8	14.65	6.31		
	76	5	17.45	0.00		
	87	5	9.14	6.46		
	other	39	14.73	5.03		

(Table 3.3 cont'd)

<i>Subject</i>	<i>Modal Choices</i>	<i># of Times Chosen</i>	<i>Average Payoff</i>	<i>Standard Deviation</i>	<i>Last Period Choice</i>	<i># of Times Chosen</i>
11	43	6	23.34	4.82	54	4
	57	11	11.77	3.56		
	other	58	17.22	6.83		
12	0	20	20.74	5.55	12	9
	10	14	22.06	6.06		
	12	9	26.09	5.20		
	15	10	25.01	5.70		
	22	5	25.67	5.28		
	other	17	20.61	8.51		
13	0	32	20.15	5.19	0	32
	10	6	18.97	4.81		
	40	5	18.72	6.46		
	60	7	21.80	0.00		
	70	6	19.20	0.00		
	other	19	17.63	4.81		
14	25	8	23.32	6.11	11	2
	31	10	25.90	3.73		
	50	5	19.28	6.46		
	other	52	20.49	7.48		
15	0	33	18.27	3.44	23	10
	2	6	17.19	0.00		
	23	10	18.50	4.97		
	100	5	6.64	5.28		
	other	21	13.71	5.07		
16	2	6	17.19	0.00	69	2
	65	5	11.11	5.28		
	67	5	12.94	6.46		
	other	59	14.73	5.26		
17	30	5	27.20	0.00	30	5
	89	5	10.80	5.28		
	other	65	18.93	6.57		
18	69	6	15.54	6.09	72	5
	70	9	15.27	5.90		
	71	5	18.92	0.00		
	72	5	18.63	0.00		
	75	8	14.80	5.46		
	80	7	11.14	6.31		
	other	35	14.59	4.46		

mode here is very strict. If we put a band around each number, the number of choices classified as modal increases dramatically). Second, of these 18 subjects, 13 chose a modal action in the 75<sup>th</sup> round when money was on the line. More strikingly, of these 13 subjects 10 chose a number on round 75 that was associated with the mode of behavior that returned them the highest average payoff during the experiment and 2 chose that mode that was associated with the highest "normalized payoff" (i.e. average payoff divided by standard deviation). Note finally, that in the only payoff relevant round subject choices seem to be far away from the unique pure strategy Nash equilibrium choice of 37. More precisely, their mean absolute deviation from 37 was 25.7.

### **3.2.2: Learn-While-You-Earn Environments**

As was true in our one-person experiments, the results of learn-before-you-earn experiments differ from those of learn-while-you-earn experiments in fundamental ways. These differences are summarized by Fact 5 below.

*Fact 5: In two-person learn-while-you-earn game environments, subjects do not behave in a manner comparable to learn-before-you-earn subjects. More precisely, while they do seem to generate a substantial number of modes of behavior over the length of the experiment, they do not seem to choose one of them (let alone the optimal (profit maximizing) mode) at the end of the experiment.*

To support Fact 5 we look at the data generated by Experiment 2b'. Again, note that in the extra round of such an experiment subjects have as much information as did their counterparts in the 75<sup>th</sup> round of the learn-before-you-earn experiment (Experiment 2a) and the same incentives to use it. To substantiate Fact 5 we repeated the same analysis performed for Fact 4 by identifying modes of behavior from the time series of individual choices in Experiment 2b' and investigating their use in the extra round of this experiment. Table 3.4 presents our results. Note that of the 10 subjects out of 12 who identified modal choices in Experiment 2b', only 3 chose one of them in the extra round of

**Table 3.4: Modes of Behavior - Experiment 2b'**

<i>Subject</i>	<i>Modal Choices</i>	<i># of Times Chosen</i>	<i>Average Payoff</i>	<i>Standard Deviation</i>	<i>Extra Period Choice</i>	<i># of Times Chosen</i>
1	0	29	18.01	3.04	40	12
	22	8	20.66	6.11		
	23	5	18.50	5.28		
	40	12	21.87	5.81		
	50	7	17.26	6.31		
	65	11	17.33	5.51		
	other	3	17.00	6.20		
2	30	9	23.27	5.90	0	0
	39	6	20.06	6.46		
	40	14	19.90	6.12		
	45	11	19.59	6.16		
	50	7	20.63	5.76		
	other	28	22.57	4.50		
3	23	9	21.38	6.22	30	9
	24	8	23.42	6.11		
	25	7	24.38	5.76		
	28	6	25.47	4.82		
	30	9	24.58	5.20		
	35	5	21.83	6.46		
	50	7	22.31	4.46		
	other	24	22.36	6.07		
4	22	5	23.31	6.46	67	5
	45	8	19.05	6.31		
	67	5	12.94	6.46		
	other	57	15.09	6.54		
5	-	-	-	-	63	1
6	0	32	20.52	5.39	10	1
	1	7	23.94	6.31		
	2	8	20.14	5.46		
	5	9	21.08	5.90		
	other	19	18.13	7.02		
7	25	11	20.24	5.95	37	0
	30	20	20.12	5.93		
	31	6	23.14	6.09		
	32	6	24.99	4.82		
	33	8	23.87	5.46		
	35	8	20.65	6.31		
	other	16	18.97	5.34		
8	55	5	22.95	0.00	50	3
	58	5	22.27	0.00		
	65	17	20.55	0.00		
	67	5	20.02	0.00		
	72	5	18.63	0.00		
	other	38	19.92	4.11		
9	-	-	-	-	11	1

(Table 3.4 cont'd)

<i>Subject</i>	<i>Modal Choices</i>	<i># of Times Chosen</i>	<i>Average Payoff</i>	<i>Standard Deviation</i>	<i>Extra Period Choice</i>	<i># of Times Chosen</i>
10	46	6	22.80	4.82	99	0
	65	5	15.83	6.46		
	67	6	16.09	6.09		
	other	58	15.93	6.10		
11	1	6	17.20	0.00	0	4
	3	5	19.54	5.28		
	5	5	17.15	0.00		
	7	5	17.10	0.00		
	12	5	16.91	0.00		
	15	5	16.75	0.00		
	18	5	16.55	0.00		
	other	39	17.09	3.10		
12	85	9	14.55	0.00	41	0
	95	6	5.05	6.46		
	99	17	8.70	2.86		
	other	43	13.20	5.84		

the experiment, and only 1 chose a number in the extra round that was associated with the mode of behavior that returned them the highest average payoff (or highest normalized payoff) during the first 75 round of the experiment. This is in contrast to Experiment 2a where 13 subjects chose modes in the last (payoff relevant) round, and 12 chose that mode which returned them the highest average (or highest normalized average) payoff over the course of the first 74 rounds of the experiment. Note finally, that in the extra round of Experiment 2b' subject choices seem to be far away from the unique pure strategy Nash equilibrium choice of 37 as it was true for the last round choices in Experiment 2a. More precisely, their mean absolute deviation from 37 was 22.7. (It was 25.7 in Experiment 2a).

The explanation we suggest for this behavioral anomaly in two-person learn-while-you-earn environments is similar to that proposed to explain the same phenomenon in one-person environments (see Fact 2). It is that subjects in learn-while-you-earn two-person game environments simply did not pay attention to the data they were generating during the experiment but rather reacted myopically on a round by round basis without taking into account the cumulative information they obtained. This proposed explanation is described more fully in Fact 6.

*Fact 6: Subjects in learn-while-you-earn two-person environments behave in an adaptive manner without responding to the decision choices of their opponent. Their choice in each round is explained exclusively by their previous round choice. However, the round-to-round differences in these choices are explained by the outcomes of the previous two rounds of the experiment. Furthermore, there appears to be no change in the rule of thumb subjects use to determine their decision over the entire length of the experiment.*

To substantiate this fact we use the data generated by Experiments 2b and 2b'. Two similar sets of regressions were run to substantiate this fact as were run in the one-person case; one on the levels of the decision numbers chosen, to explain the decision rule used, and another on the round-to-round differences of these levels, to explain how

subjects alter their previous choices as they receive new information. Here, however, because we were dealing with a two-person game situation, we also included among the covariates the lagged decision numbers of a subject's opponent (opp).

The estimates obtained for the regressions performed using the pooled sample (2304 observations) are the following: (Again, to economize on space we do not report the estimated coefficients for the individual dummy variables.)

$$\begin{aligned}
 [3] \quad \text{dec}_t^i &= \alpha^i + 0.296 * \text{dec}_{t-1}^i - 0.027 * \text{dec}_{t-2}^i + 0.081 * \text{dec}_{t-3}^i + \\
 &\quad (0.094) \qquad (0.090) \qquad (0.083) \\
 &\quad - 0.369 * \text{pay}_{t-1}^i - 0.808 * \text{pay}_{t-2}^i + 0.102 * \text{pay}_{t-3}^i + \\
 &\quad (0.571) \qquad (0.524) \qquad (0.503) \\
 &\quad + 4.152 * \text{win}_{t-1}^i + 7.867 * \text{win}_{t-2}^i - 2.529 * \text{win}_{t-3}^i + \\
 &\quad (5.237) \qquad (5.105) \qquad (4.802) \\
 &\quad + 0.016 * \text{opp}_{t-1}^i - 0.001 * \text{opp}_{t-2}^i + 0.001 * \text{opp}_{t-3}^i + \\
 &\quad (0.022) \qquad (0.022) \qquad (0.020) \\
 &\quad + 1.765 * \text{dr25} + 1.016 * \text{dr50}, \qquad R^2 = 0.605, \\
 &\quad (0.999) \qquad (0.866)
 \end{aligned}$$

and

$$\begin{aligned}
 [4] \quad \text{diff}_t^i &= \alpha^i - 0.248 * \text{dec}_{t-2}^i + 0.034 * \text{dec}_{t-3}^i + \\
 &\quad (0.088) \qquad (0.086) \\
 &\quad + 2.796 * \text{pay}_{t-1}^i - 1.830 * \text{pay}_{t-2}^i - 0.112 * \text{pay}_{t-3}^i + \\
 &\quad (0.156) \qquad (0.425) \qquad (0.406) \\
 &\quad - 34.545 * \text{win}_{t-1}^i + 19.265 * \text{win}_{t-2}^i + 0.245 * \text{win}_{t-3}^i + \\
 &\quad (1.853) \qquad (5.088) \qquad (4.979) \\
 &\quad + 0.000 * \text{opp}_{t-1}^i + 0.005 * \text{opp}_{t-2}^i + 0.000 * \text{opp}_{t-3}^i + \\
 &\quad (0.022) \qquad (0.023) \qquad (0.021) \\
 &\quad + 1.962 * \text{dr25} + 1.317 * \text{dr50}, \qquad R^2 = 0.269. \\
 &\quad (1.024) \qquad (0.891)
 \end{aligned}$$

The numbers in parentheses are robust standard errors computed using Huber's formula.

As in the one-person learn-while-you-earn environment these results again expose an "adaptive myopic" type of behavior. Despite the fact that these subjects are in a game-theoretic situation, they fail to take account of their opponents strategy choice in determining their behavior. None of the coefficients associated with lagged decision choices of the opponents are in fact significant at any significance level in either regression. In determining their decision numbers in each round, subjects tend to rely exclusively on their previous round choice (equation [3]). In determining the changes in these choice levels, however, subjects look one and two periods back and react in an adaptive manner (equation [4]). They do this by increasing their choice if either they lose the tournament or if they win with a relatively low decision number -- i.e. they receive a "relatively high" payoff in the previous round. They also respond to their two-period lagged decision and its outcome in the opposite way. Furthermore, we do not find any evidence of round heterogeneity since none of the round dummy variables have significant coefficients attached to them.<sup>12</sup> This suggests that our experimental subjects do not revise their decision rules towards the end of the experiments as a result of learning. Finally, we find significant heterogeneity among players in the intercept terms. Qualitatively similar results both in terms of the sign and the significance level of the estimated coefficients emerge, however, from the regressions conducted at the individual level.<sup>13</sup>

The results of these regressions then support our assertions in Fact 6. They indicate that, like in one-person learn-while-you-earn environments, subjects appear to adopt myopic adaptive learning procedures which do not get updated as the game they are in proceeds. This behavior is qualitatively different from the type of behavior observed in two-person learn-before-you-earn environments.

---

<sup>12</sup> Again, similar regressions run also including round dummies for the slope coefficients confirm this fact. No round dummies enter either regression with coefficients that are significantly different from zero at conventional significance levels.

<sup>13</sup> These results are available from the authors.

## Section 5: Conclusion

This paper has taken a small step on what is likely to be a long road to discovering how human beings learn when faced with one person decision problems and games. What strikes us as interesting is the variety of ways laboratory subjects structure their learning as the environment they are placed in changes. We have seen that learning in one person decision problems and games are quite different and differ as we change the costs of learning.

If economic theory is going to take account of this rich diversity of behaviors, it is going to have to build theories that are flexible enough to move from environment to environment or else tailor its learning theories to each specific economic institution. We consider this paper a step in guiding theorists toward a more realistic and comprehensive theory of learning -- one which focuses on how subjects learn (procedural rationality) as well as whether that learning leads them to converge to particular outcomes, i.e., Nash equilibria or first best optima (substantive rationality). It is our feeling that a complete descriptive theory of learning and procedural rationality should be constructed first and then the consequences of this learning model investigated for its substantive rationality.

### References

- Aghion, P., P. Bolton, C. Harris, and B. Jullien (1991), "Optimal Learning by Experimentation," Review of Economic Studies, 58, 621-654.
- Arthur, W.B. (1990), "A Learning Algorithm that Mimics Human Learning," Santa Fe Institute, Economic Research Program, Working Paper No. 90-026.
- Bull, C., A. Schotter, and K. Weigelt (1987), "Tournaments and Piece Rates: An Experimental Study", Journal of Political Economy, 95, 1-33.
- Bush, R. and F. Mosteller (1955), Stochastic Models of Learning, New York: John Wiley and Sons.
- Easley, D. and N. Kiefer (1988), "Controlling a Stochastic Process with Unknown Parameters," Econometrica, 56, 1045-1064.
- Estes, W.K. (1950), "Toward a Statistical Theory of Learning", Psychological Review, 57, 94-107.
- Fudenberg, D. and D. Kreps (1988), "A Theory of Learning, Experimentation and Equilibrium in Games," mimeo.
- \_\_\_\_\_, and D.K. Levine (1991), "Steady State Learning and Nash Equilibrium," MIT, Department of Economics, Working Paper No. 594.
- Goldberger, A.S. (1991), A Course in Econometrics, Cambridge: Harvard University Press.
- Huber, P.J. (1967), "The Behavior of Maximum Likelihood estimates under non-standard Conditions," Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, 1, 221-233.
- Jordan, J.S. (1991), "Bayesian Learning in Normal Form Games," Games and Economic Behavior, 3, 60-81.
- Kalai, E. and E. Lehrer (1991), "Rational Learning Leads to Nash Equilibrium," Northwestern University, CMSEMS, Discussion Paper No. 925.
- Knez, M. (1992), "A Model of Sophisticated Learning for a Special Class of Rank Order Games," Ph.D. Dissertation, University of Pennsylvania.
- Lazear, E.P. and S. Rosen (1981), "Rank Order Tournaments as Optimal Labor Contracts," Journal of Political Economy, 89, 841-864.
- Marimon, R. (1990), "Adaptive Learning in Games," mimeo.
- McLennan, A. (1987), "Incomplete Learning in a Repeated Statistical Decision Problem," mimeo.

- Merlo, A. and A. Schotter (1991), "Experimentation and Learning in Laboratory Experiments: Harrison's Criticism Revisited," New York University, C.V. Starr Center, Research Report No. 91-25. Forthcoming, American Economic Review, as "Theory and Misbehavior of First-Price Auctions: Comment."
- Milgrom, P. and J. Roberts (1991), "Adaptive and Sophisticated Learning in Normal Form Games," Games and Economic Behavior, 3, 82-100.
- Mookherjee, D. and B. Sopher (1991), "Learning Behavior in an Experimental Matching Pennies Game," mimeo.
- Nyarko, Y. (1992), "Bayesian Learning without Common Priors and Convergence to Nash Equilibrium," New York University, C.V. Starr Center, Research Report No. 92-25.
- Schotter, A. and K. Weigelt (1992), "Asymmetric Tournaments, Equal Opportunity Laws and Affirmative Action: Some Experimental Results", Quarterly Journal of Economics, 429, 513-539.
- Simon, H.A. (1976), "From Substantive to Procedural Rationality," pp. 129-148 in Method and Appraisal in Economics, S.J. Latsis (ed.), Cambridge: Cambridge University Press.
- \_\_\_\_\_, (1978), "Rationality as Process and as Product of Thought," American Economic Review, 68, 1-16.
- White, H. (1980), "A Heteroskedasticity-Consistent Covariance Matrix Estimator and A Direct Test for Heteroskedasticity," Econometrica, 48, 817-838.

**APPENDIX**

**Table A.0: Coefficients of the Theoretical Payoff Function**

---

payoff function:  $\pi = \alpha + \beta e + \gamma e^2$  , e: effort level

---

$\alpha$	$\beta$	$\gamma$
18.94	0.079	-0.0011

---

**Table A.1: Least-Squares Estimates of the Payoff Function - Experiment 1a**  
(Standard Errors in Parentheses)

<i>Subject</i>	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\gamma}$	<i>F-test*</i>
1	12.56 (14.10)	0.371 (0.509)	-0.0043 (0.0043)	0.78
2	17.37 (1.57)	0.174 (0.728)	-0.0025 (0.0007)	2.06
3	-19.57 (23.97)	1.505 (0.863)	-0.0144 (0.0077)	2.07
4	18.25 (1.20)	0.161 (0.058)	-0.0024 (0.0005)	7.96
5	19.40 (5.67)	0.160 (0.231)	-0.0029 (0.0021)	1.82
6 **	-	-	-	-
7	14.22 (3.41)	0.373 (0.173)	-0.0050 (0.0021)	1.20
8	19.78 (1.99)	0.153 (0.117)	-0.0032 (0.0016)	1.93
9	20.70 (1.08)	0.067 (0.963)	-0.0018 (0.0006)	5.64
10	17.10 (1.63)	0.183 (0.067)	-0.0026 (0.0006)	10.35
11	20.24 (2.23)	0.131 (0.116)	-0.0027 (0.0013)	2.55
12	18.50 (2.19)	0.160 (0.094)	-0.0028 (0.0009)	6.93

\*  $H_0 : \hat{\alpha} = 18.94$  ,  $\hat{\beta} = 0.079$  ,  $\hat{\gamma} = -0.0011$ ;  $F_{.95}(3,72) = 2.15$ .

\*\* Subject 6 could not have estimated a quadratic payoff function.

**Table A.2: Least-Squares Estimates of the Payoff Function - Experiment 1b'**  
(Standard Errors in Parentheses)

<i>Subject</i>	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\gamma}$	<i>F-test*</i>
1	9.79 (10.08)	0.410 (0.452)	-0.0044 (0.0045)	2.29
2	15.67 (8.36)	0.244 (0.427)	-0.0032 (0.0058)	0.08
3	12.79 (5.24)	0.306 (0.159)	-0.0035 (0.0011)	21.74
4	18.29 (2.20)	0.166 (0.084)	-0.0026 (0.0007)	12.61
5 **	-	-	-	-
6	-33.97 (36.93)	2.579 (1.718)	-0.0301 (0.0199)	0.83
7	18.55 (0.78)	0.344 (0.247)	-0.0055 (0.0037)	0.84
8	17.32 (11.86)	0.168 (0.354)	-0.0025 (0.0025)	5.69
9	11.27 (13.80)	0.304 (0.545)	-0.0027 (0.0051)	0.85
10	-12.94 (21.88)	0.987 (0.624)	-0.0078 (0.0044)	3.27
11	17.77 (0.87)	0.147 (0.057)	-0.0023 (0.0007)	3.38
12	17.98 (2.35)	0.192 (0.101)	-0.0029 (0.0009)	8.38

\*  $H_0 : \hat{\alpha} = 18.94$  ,  $\hat{\beta} = 0.079$  ,  $\hat{\gamma} = -0.0011$ ;  $F_{.95}(3,72) = 2.15$ .

\*\* Subject 5 chose the same decision number (0) in all 75 rounds.

**Table A.3: Delta Method**

---

payoff function:  $\pi = \alpha + \beta e + \gamma e^2$  , e: effort level

true maximum:  $e^* = -\beta/2\gamma$

estimated maximum:  $\hat{e}^* = -\hat{\beta}/2\hat{\gamma}$

estimate's standard error:  $\sigma_{\hat{e}^*} = \sqrt{[(1/4\hat{\gamma}^2)\sigma_{\hat{\beta}}^2 - (\hat{\beta}/2\hat{\gamma}^2)\sigma_{\hat{\beta}\hat{\gamma}} + ((\hat{\beta}^2/2\hat{\gamma}^4)\sigma_{\hat{\gamma}}^2]}$

---