# 'Learning' with no feedback in a competitive guessing game

Roberto Weber\*

Social and Decision Sciences

Carnegie Mellon University

Pittsburgh, PA 15213

February 7, 2000

<sup>\*</sup>This research was funded by NSF grant SBR-9631627 to Richard McKelvey and Tom Palfrey. The author thanks Tom Palfrey, Colin Camerer, Linda Babcock, Gary Bolton, Al Roth and participants at the October 1998 Economic Science Association meetings in Tucson and at the 1998 Society for Judgment and Decision Making meetings in Dallas for comments and suggestions.

### **Abstract**

An assumption underlying all of the current models of learning in games is that learning takes place only through repeated observation of and experience with outcomes and payoffs. This paper experimentally tests this assumption using the competitive guessing game first studied by Nagel (1995). The experiments in this paper consist of several periods of repeated play of the game under alternative feedback conditions. The main experimental treatment is a no-feedback condition, in which no information is reported to the players until the end of the experiment. If learning takes place only through observation of outcomes and reinforcement resulting from payoffs, choices in the no-feedback condition should not converge towards the Nash equilibrium through repeated play. While there is less learning under no feedback than when outcomes are revealed, there is convergence towards the equilibrium prediction. Finally, the no-feedback learning phenomenon is explored in two models. In the first model, the learning process is assumed to take place as players observe their own actions, treat these actions as an unbiased estimate of others' actions, and then best respond to this estimate. In the second model, learning is assumed to take place all at once, meaning that the process by which players make their choices changes in only one period.

### 1 Introduction

The study of learning in games has recently received much attention. Since behavior in experimental studies of games changes with experience, the goal of much of this research has been to model this learning process. Several models have been proposed which map feedback on outcomes and payoffs into subsequent choices. Learning in these models takes place only when players receive feedback on prior performance before choosing again. However, much economic activity takes place with delayed or poor feedback concerning performance. This paper explores the possibility that a different form of learning, inconsistent with the models in the literature, may take place with repeated experience. This type of learning occurs

<sup>&</sup>lt;sup>1</sup>Any activity in which a task has to be performed repeatedly prior to obtaining feedback satisfies this description. An example is preparing several proposals (or papers, projects, etc.), one after another, which each take some time for review. Another example is whenever accurate performance feedback can only be obtained from a supervisor's evaluation, which may occur infrequently.

when individuals receive no feedback on prior performance before performing a task again, yet still improve their performance.

There are many different models of learning in games. Examples include: choice reinforcement learning (most recently studied by Erev and Roth, 1998), weighted fictitious play (Cheung and Friedman, 1997), experience-weighted attraction (Camerer and Ho, 1999) and rule learning (Stahl, 1997).<sup>2</sup> In all of these models players' choices are revised only through their observation of previous outcomes. So, for example, in choice reinforcement learning models such as Erev and Roth's a player's propensity to play a particular strategy is updated by the payoff experienced when that strategy is played. In weighted fictitious play and experience-weighted attraction, on the other hand, the propensities are also affected by the foregone payoffs of strategies that were not selected. In Stahl's rule learning model, a rule is a response to observed outcomes and rules are updated using their expected payoff given the actual distribution of other players' choices.<sup>3</sup>

One method for testing the competing models is to vary the information received by subjects concerning outcomes. For instance, choice reinforcement models assume that players only use information about the payoffs to their chosen strategies in making subsequent decisions. Therefore, according to these models, taking away information concerning payoffs to strategies not selected should have no effect on behavior in repeated play. In other models, such as weighted fictitious play and experience-weighted attraction, players use the

<sup>&</sup>lt;sup>2</sup>Experience-weighted attraction includes both choice reinforcement and weighted fictitious play as special cases.

<sup>&</sup>lt;sup>3</sup>For detailed discussion of these models and others, see Camerer (1998, chapter 6).

information on payoffs not received and therefore these models predict that eliminating it will have an effect. This test was performed by Mookherjee and Sopher (1994) and by Van Huyck, Battalio and Rankin (1996) who found that reducing payoff information did have an effect on observed learning paths, which contradicts an important assumption of choice reinforcement learning. A similar study was conducted by Duffy and Feltovich (1998), who found that giving players information on the actions and payoffs of other players as well on as their own outcomes affected repeated play in the ultimatum game. This phenomenon of "observational learning" also provides evidence against simple choice reinforcement models.

While the above models all show that learning can take place when subjects receive feed-back on payoffs and the actions of other players, they ignore the possibility that learning may take place in the absence of this information. Therefore, these models do not account for the possibility that people may learn simply through repeated experience with an environment. Repetition and experience with a set of procedures could lead people to obtain insights concerning how best to perform a task. This paper tests whether or not this type of learning from repetition without feedback can take place in a strategic environment.

The experiments in this paper are similar to previous studies in that they examine an extreme case of payoff information manipulation. This manipulation tests the assumption in all of the above models that learning in games takes place only through reinforcement by the observation of actions and payoffs. Using Nagel's (1995) competitive guessing game (see also, Ho, Camerer, and Weigelt (1998)), experiments are conducted in which players receive no information on the actions of other players or on their own payoff. In these no-feedback

experiments, all of the above models predict that no predictable change in behavior should take place with repeated play. Any convergence towards the Nash equilibrium would indicate that learning does not only take place through experienced payoffs and observed outcomes, but is also affected by experience with an environment and procedures and by repeatedly thinking about a game.

The next two sections discuss the game used in the experiments and the experimental design. Section 4 presents the results and Section 5 develops and estimates two models of no-feedback learning in this particular game. The paper ends with conclusions and a discussion of possibilities for future research.

### 2 The Game

The game used in these experiments is one first studied experimentally by Nagel (1995) to determine the number of steps of iterative reasoning being satisfied by subjects. In the game, each of N subjects simultaneously chooses a number in the interval [0,100]. The average of all the players' choices is then computed and multiplied by a parameter p to determine a target number. The player whose number choice is closest to this target number wins a fixed amount X and all other players receive nothing. If more than one player chooses the number closest to the target number, the amount X is divided equally among the winners. While Nagel conducts experiments using several values of p (including values greater than one), the experiments in this paper all use  $p = \frac{2}{3}$ , which is one of the values studied by Nagel.

It is easy to see that the unique equilibrium in the game where  $p = \frac{2}{3}$  is for everyone to

choose 0. For any average of all players' choices,  $\mu$ , the best response for all players is to choose  $p\mu$ , resulting in a new average, and this process has a unique fixed point at zero.<sup>4</sup>

Nagel conducted four sessions with  $p = \frac{2}{3}$  in each of which the game was repeated four times.<sup>5</sup> At the end of each period, subjects were informed of the choices of all other participants, the average, the target number, and the winning choices. There were 15 to 18 subjects in each experiment. No subjects chose the equilibrium strategy of 0 in the first period. The average in the first play of the game was 36.73 and several subjects chose numbers greater than  $66\frac{2}{3}$ , violating dominance.

With repeated play, the choices converged towards the Nash equilibrium. The average choice was 24.17, 16.14, and 9.52 in periods 2, 3, and 4 respectively. Thus, subjects in Nagel's experiments appeared to learn since their choices reflected a best response to the average in the previous period. The question this study asks is whether similar convergence toward the equilibrium prediction is observed when subjects receive <u>no</u> feedback on payoffs or outcomes. If so, this points to a kind of learning not captured by the current models.

<sup>&</sup>lt;sup>4</sup>Furthermore, a player's choice indicates the number of steps of iterative reasoning that the player is satisfying and thinks others are satisfying. For instance, if a player chooses in the interval  $(66\frac{2}{3}, 100]$ , then this player is violating dominance since the target number can never be in this interval. If a player believes that everyone else is obeying dominance, then she should expect no one else to choose in this interval and should, therefore, not select any number in the interval  $(44\frac{4}{9}, 100]$ . This process can be applied repeatedly to show that the only choice that subjects will make if common knowledge of rationality is satisfied is 0.

<sup>&</sup>lt;sup>5</sup>Nagel's results have been replicated by Ho, et al, using p = 0.7. For a detailed survey of existing research using this game, with varying parameters and payoffs, see Nagel (1998).

# 3 Experimental Design

In this study, Nagel's game with  $p = \frac{2}{3}$  was used in experiments under three information conditions.<sup>6</sup> In each session, the game was repeated ten times with 8 to 10 players.

In the *control* condition (C), the experimenter wrote the average, target number, and participant numbers of the winners on a board at the front of the room at the end of each period. This treatment serves to replicate, with feedback, Nagel's results.

In the no-feedback condition (NF), the game was played ten times, but subjects received no feedback at the end of each period. After the experimenter recorded each subject's choice, he calculated the average and target number and determined who the winner or winners were. The participants were informed that the experimenter had done this (but were not told the results) and were then asked to make a choice for the next period. At the conclusion of the tenth period, subjects were informed of the average, target number, and participant number of the winner or winners for all ten periods.

The final treatment was the same as the NF condition, with one exception. In this meanguess no-feedback condition (MG), subjects were also given no feedback until the end of the
experiment, but after the experimenter calculated the average in each period, participants
were instructed to write down their guess of the value of the average. While this guess did
not provide subjects with any new information and their earnings were not affected by the
accuracy of their guess, it was introduced to aid participants in thinking about the fact that
they wanted to best respond to their expectation of the average.

<sup>&</sup>lt;sup>6</sup>Instructions are available in the Appendix.

The experiments were conducted during September and October 1998 using graduate and undergraduate students at the California Institute of Technology with little or no formal training in game theory. At the end of the experiment, subjects were privately paid their earnings in all ten periods plus a \$7 participation bonus. Three sessions were conducted for each treatment (n = 28 in both MG and NF, and n = 26 in C). Each session lasted 30 to 45 minutes.

Under the null hypothesis that learning does not take place without feedback, there should be no change in subjects' behavior across periods in the NF and MG treatments, but we should see convergence towards the equilibrium prediction in treatment C. On the other hand, if learning does take place by subjects simply gaining experience with the environment and having to think repeatedly about the game, we should see convergence towards 0 in both the NF and MG treatments. Finally, if prompting subjects to think about the value of the mean leads them to perform better the iterative reasoning required for equilibrium behavior to arise, convergence towards the equilibrium should be greater in the MG treatment than in the NF treatment.

### 4 Results

Table 1 presents the mean and median choice by period for each treatment. In period 1, subjects' choices are very similar in both the NF (mean = 31.0) and MG (mean = 31.6) conditions and there is no significant difference in behavior between the two conditions

 $(t_{54} = 0.12)$ . The period 1 choices in the control condition are lower than in the other two treatments (mean = 24.6), but this difference is not significant in a t-test of the means  $(t_{52} = 1.35 \text{ for both C-NF and C-MG comparisons})$ .

Figure 1 displays the cumulative frequency of first period choices in all three conditions. Again, there is no significant difference between the frequencies in all three conditions using a two-tailed Kolmogorov-Smirnov test (C-NF:  $D_{26,28} = 0.28$ ; C-MG:  $D_{26,28} = 0.28$ ; NF-MG:  $D_{28,28} = 0.18$ ). Thus, while initial choices in condition C are slightly lower than in the other two treatments, this difference is not significant.<sup>8</sup>

In all three treatments, the mean and median choices decreased between periods 1 and 10. As expected, the greatest decrease was in treatment C (decrease in mean = 18.1; decrease in median = 20.2). The decrease in the NF treatment (decrease in mean = 8.8; decrease in median = 12.0) was less than the decrease in the MG treatment (decrease in mean = 17.2; decrease in median = 17.9), although this difference is mainly due to the fact that mean and median choices increased slightly in the final periods in the NF treatment, while they continued to decrease in the MG condition.

<sup>&</sup>lt;sup>7</sup>Note that these results are also similar to, but slightly lower than, Nagel's first period results (mean = 36.7).

<sup>&</sup>lt;sup>8</sup>The lower mean in treatment C is due mainly to behavior in one session in which several subjects made low choices in the first period and the resulting mean was 15.2 (the mean initial choice in the other two treatment C sessions was 32 and 29). The robustness of first period behavior in the population is further reflected in the results from an informal class experiment conducted using 17 Caltech undergraduates in which the mean choice was 29.

<sup>&</sup>lt;sup>9</sup>The failure of the mean and median to decrease monotonically in treatment C was due to strategic behavior on the part of at least one subject in each session. Following a rapid initial convergence towards zero, these subjects made high choices (usually 100) in an attempt to cause other subjects to respond with choices that were too high in the following period. That this was the reasoning behind this behavior was determined in informal debriefing at the end of the experiment. However, the success of this strategy is questionable since no subject that chose 100 had the winning choice in the following period.

	C		N	F	MG	
Period	Average	Median	Average	Median	Average	Median
1	24.6	22.9	31.0	30.0	31.6	31.1
2	16.4	10.0	27.2	24.0	28.8	24.4
<b>3</b>	6.7	5.0	24.4	21.0	28.3	24.0
4	6.2	1.2	19.6	20.5	22.4	18.0
, e - g, tea <b>5</b> , , e, e	12.1	2.2	22.5	23.0	22.1	20.5
6	5.4	5.0	18.4	17.5	17.4	16.5
7	9.6	2.0	18.2	17.0	17.1	14.6
8	11.2	2.8	18.0	18.5	17.3	13.0
9	8.4	6.0	19.0	20.0	15.7	12.0
10	6.5	2.7	22.2	18.0	14.4	13.3
diff. = P1 - P10	18.1	20.2	8.8	12.0	17.2	17.9

Table 1. Summary of outcomes by treatment

Figures 2 through 4 present the cumulative frequency of choices in periods 1 and 10 for conditions C, NF, and MG, respectively. In all three treatments, the frequency of lower choices is greater in period 10 than in period 1. This difference is significant for all conditions in a one-tailed Kolmogorov-Smirnov test (C:  $\chi^2(2) = 30.77$ , p < 0.001; NF:  $\chi^2(2) = 7.14$ , p = 0.026; MG:  $\chi^2(2) = 12.07$ , p = 0.002). Thus, subjects are revising their choices towards the equilibrium prediction even when they do not receive any feedback.

As an additional test of whether choices decreased with experience in all three sessions, a regression of choice on period was conducted including both treatment dummy variables and treatment\*period interaction terms. The results are reported in Table 2.<sup>10</sup> Not surprisingly, choice decreases significantly with period in the control treatment.<sup>11</sup> While choices in the

<sup>&</sup>lt;sup>10</sup>The omitted treatment for both the dummy variables and interaction terms is C.

<sup>&</sup>lt;sup>11</sup>If ln(choice) rather than choice is used as the dependent variable the results are similar: the coefficient on period is negative and significant for all three treatments.

two no-feedback treatments (NF and MG) tend to be higher overall, they do not decrease at a significantly lower rate with repetition. In fact, the magnitude of the period coefficient is greater in treatment MG (choice decreases 1.948 per period) than in treatment C (choice decreases 1.166 per period). Though the magnitude of the coefficient for treatment NF is smaller (choice decreases 1.068 per period) than in treatment C, this difference is not significant and the negative relationship between choice and period in treatment NF is significant.

Dependent variable:	Choice	
Period	-1.166	(0.458) **
NF x Period	0.098	(0.658)
MG x Period	-0.782	(0.588)
Treatment NF	10.792	(4.279) **
Treatment MG	15.094	(4.770) ***
Intercept	17.130	(2.956) *** <sup>2</sup>
N	820	
$R^2$	0.128	
Adjusted $\mathbb{R}^2$	0.123	

White-corrected standard errors are in parentheses.

Table 2. Results of choice regression

Additional evidence that subjects' choices are decreasing between periods 1 and 10 can be seen in Table 3. Table 3 presents, for each condition, the number of subjects who changed their choices between periods 1 and 10 in each direction. For instance, between periods 1 and 10, eight subjects in the NF condition increased their choices, 19 decreased their choices,

<sup>\*</sup>p < 0.1; \*\*p < 0.05; \*\*p < 0.01; one-tailed tests, except where noted.

<sup>&</sup>lt;sup>2</sup> Two-tailed test.

and one subject chose the same number. Using a sign test, the null hypothesis that the underlying distribution of choices is unchanged can be rejected (p < 0.026). The changes observed in the MG condition are even more extreme: no subjects increased their choices between periods 1 and 10 while 24 subjects decreased their choices. This is also significant (p < 0.001). As expected, more subjects in treatment C also lowered their choices (22) than increased their choices (3) (p < 0.001).

			C		NF	N	/IG
Choice	e increased	3	11.5%	8	28.6%	0	0.0%
Choice	unchanged	1	3.8%	1	3.6%	4	14.3%
Choice	decreased	22	84.6%	19	67.9%	24	85.7%

Table 3. Direction of changes in subjects' choices between periods 1 and 10

Figure 5 presents the cumulative choice frequencies in period 10 for all three treatments. Note that the frequency of lower choices is higher in treatment C than in the other two treatments. These differences are significant using a one-tailed Kolmogorov-Smirnov test (C-NF:  $X^2(2) = 27.30$ , p < 0.001; C-MG:  $X^2(2) = 21.72$ , p < 0.001).

The above results indicate that convergence towards the equilibrium prediction takes place in all three treatments. This convergence is greater in treatment C than in treatment NF, indicating that while behavior resembling learning does take place without feedback, the process is stronger when outcomes are revealed. The comparison between treatments

<sup>&</sup>lt;sup>12</sup>While it is surprising that any subjects increased their choices in treatment C, two of the three increases were by subjects who initially chose zero. The other increase was by a subject who chose 100 in the final period, possibly out of frustration at not having won in previous periods.

MG and C is less clear and in some tests the decrease in choices between the two conditions does not appear to be significantly different.

The second question these experiments address is whether or not having subjects write down a guess of the value of the mean in each period increases the convergence towards the equilibrium. Tables 1 and 2 provide some evidence in favor of this hypothesis. The changes in the mean and median choices between periods 1 and 10 is greater in the MG treatment than in the NF condition. However, as can be seen in Table 1, this mainly results from the fact that choices increase in the final rounds of the NF condition, rather than from any consistent difference across periods. Additional support can be seen in Table 3, where the number of subjects who lowered their choice is higher in the MG condition (24) than in the NF condition (19). The difference between the NF and MG treatments in the number of subjects who either increased, decreased, or did not change their choices is significant at the p < 0.01 level using a chi-square test.

The graph in Figure 5 shows that the frequency of lower choices in period 1 is higher in the MG condition than in the NF condition, but this difference is not significant using a Kolmogorov-Smirnov test. However, combined with the results from Table 3, there is support for the hypothesis that requiring subjects to write down the mean leads to increased convergence toward the equilibrium. This conclusion finds further support in the next section where two simple models of no-feedback learning in this game are developed and tested.

# 5 Modelling the learning phenomenon

This section presents and tests two simple models of how subjects may be learning in the no-feedback treatments. The models are similar to the adaptive models discussed previously in that subjects are assumed to respond adaptively to beliefs about prior behavior. Therefore, in control treatment C they are assumed to respond to the previous mean choice. In the other treatments, however, they are assumed to respond to an estimate of the mean in the previous period. Given that the only new information subjects had the end of each period in these treatments was their own choice, they are assumed to treat their choice in each period as an unbiased estimate of the group mean and then (partially) respond to this estimate. The difference between the two models lies in whether players are assumed to constantly respond to their previous choice in every period or whether they do so only once.

This type of reasoning on the part of players is consistent with the "false- consensus bias" in social psychology in which individuals' estimates of group characteristics and propensities are biased in the direction of their own characteristics and propensities. Therefore, to the extent that subjects "learn" this way, their learning could be viewed as biased and incorrect. However, as Dawes (1990) has argued, the false-consensus is not a bias if an individual's behavior or characteristics are an unbiased estimate of those of the population and if there is no other relevant information. According to this argument, since a Bayesian observer viewing the behavior of another individual should revise her belief about the likelihood of observing that behavior in the population from which the individual was drawn, the same

<sup>&</sup>lt;sup>13</sup>See, for instance, Ross, Greene and House (1977).

Bayesian observer should similarly use her own behavior as useful information concerning how the population she is drawn from is likely to behave. Therefore, subjects in the two nofeedback conditions in the experiment may be correct in treating their choice in the previous period as an unbiased estimate of the group mean. However, since subjects may be unsure whether their previous choice is such an unbiased estimate, the model allows subjects to only partially best respond to this previous choice.

The first model assumes that players make their choice in every period by such a partial best response to their own previous choice. In the no-feedback case, this "myopic consensus" model assumes that individuals realize at the end of each period – only after choosing – that their choice can be used as an estimate of the population mean. The second model of no-feedback learning similarly assumes that players best respond to their own previous choice, but instead of assuming that they do so in every period, assumes that this learning takes place only at one moment in which subjects become aware of the information contained in their previous choices. <sup>14</sup>

As mentioned above, the first model this section is concerned with is a simple adaptive model in which players partially best respond to what they believe the mean was in the previous period. In the model, the choice of a player  $i \in N = \{1, ..., n\}$  in period  $t \in T = \{1, ..., 10\}$  is represented by  $x_{it}$ . The mean choice in period t is  $y_t = \frac{\sum_{i \in N} x_{it}}{n}$  and player i's estimate of this mean is  $\hat{y}_{it}$ . In treatment C where the mean is announced at the end of each period  $\hat{y}_{it} = y_t$ .

<sup>&</sup>lt;sup>14</sup>This model of no-feedback learning can be thought of as sudden inspiration or serendipity.

In every period after the first, subjects respond to their estimate of the previous period's mean: 15

$$x_{it} = \beta \hat{y}_{it-1}. \tag{1}$$

This model assumes a very simple adjustment rule in which players form a belief concerning the value of  $y_t$  only at the end of period t (in treatment C, this corresponds to when they find out the value of  $y_t$ ) and then partially best respond to this belief in the next period. The response parameter,  $\beta$ , is determined by how players believe the optimal response will change in the next period as a function of the previous period's average choice and their beliefs about how the choices of others will change. Each player knows the value of  $\beta$  but is unsure of whether the other players are using the same value for their adjustment.<sup>16</sup> Players have beliefs over the value of  $\beta$  other players are using, denoted  $\tilde{\beta}$ , and these beliefs are captured by the actual value of  $\beta$ . For instance, if players all believe that other players' behavior will not change in the next period, then they will simply best respond to the choices of others in the previous period and will assume that the only difference between  $y_t$  and  $y_{t-1}$  will be due to the difference between  $x_{it}$  and  $x_{it-1}$ . Therefore, since  $x_{it} = \beta \hat{y}_{it-1}$ , players with

<sup>&</sup>lt;sup>15</sup>Assume that in the first period players best respond to a prior belief of the distribution of choices, k.

<sup>16</sup>If  $\beta$  were commonly known, then the learning process would converge to the equilibrium in one period for any  $\beta$  less than 1.

$$\beta = \left(\frac{p}{n-p}\right) \left(\frac{n\hat{y}_{it-1} - x_{it-1}}{\hat{y}_{it-1}}\right) \tag{2}$$

will be best responding to the beliefs formed at the end of the previous period. Refer to this value of  $\beta$  as  $\beta^{BR}(\hat{y}_{it-1}, x_{it-1})$ .

As mentioned above, one possible explanation of the observed learning phenomenon in the two no-feedback conditions is that in each period players are estimating  $\hat{y}_{it} = x_{it}$ , which is correct as long as their choice is an unbiased estimate of the population mean. Under this myopic consensus model, in which players believe that their choices are representative of the choices of others but fail to take into account that this process implies that others are updating similarly,  $\beta^{BR}(\hat{y}_{it-1}, x_{it-1})$  reduces to  $\frac{pn-p}{n-p}$ . Since p is always  $\frac{2}{3}$  in the experiments,  $\beta^{BR}$  is equal to 0.636 when n=8 and 0.643 when n=10. Therefore, 0.64 provides a reasonable approximation of  $\beta_i^{BR}$  for both group sizes.

If  $\beta$  is equal to 1, then players' beliefs are stationary and players are constantly best responding to the initial belief k, corresponding to a situation where no learning is taking place across periods and players are treating  $x_{it}$  as entirely uninformative. This corresponds to the prediction for the no-feedback treatments of current adaptive learning models. If  $\beta < \beta^{BR}$ , then players believe that other players will also lower their choices (i.e.,  $\sum_{j\neq i} x_{jt-1} > \sum_{j\neq i} x_{jt}$ ). If  $1 > \beta > \beta^{BR}$ , then players are lowering their choices between periods, but are not best responding to their estimates of the previous mean. This sticky best response could result from players being uncertain about whether or not  $\hat{y}_{it} = x_{it}$  is an unbiased estimate

of  $y_t$ .

Assume player i's choice in each period t > 1 is determined by:

$$x_{it} = \beta \hat{y}_{it-1} + \epsilon_{it} \tag{3}$$

with the error term  $\epsilon_{it}$  representing the idiosyncratic component of player i's choice in period t. Further assume that the  $\epsilon_{it}$  are uncorrelated across both i and t. Then the model in Equation 3 can be estimated using least squares.

In treatment C,  $\hat{y}_{it-1}$  is equal to  $y_{it-1}$  for all i since the mean is announced at the end of each period. Therefore, Equation 3 can be rewritten as:

$$x_{it} = \beta y_{it-1} + \epsilon_{it} \tag{4}$$

and estimated using OLS. Under the myopic consensus model,  $\hat{y}_{it-1}$  is equal to  $x_{it-1}$  in treatments NF and MG. For these two treatments, Equation 3 can be rewritten as:

$$x_{it} = \beta x_{it-1} + \epsilon_{it} \tag{5}$$

and similarly estimated. Table 4 reports the results of these regressions. 17

Dependent Variable:					
	C	NF	MG		
$\hat{oldsymbol{eta}}$	0.576 (0.091)	0.859 (0.031)	0.821 (0.024)		
Obs. (#ofsubjects)	234 (26)	252 (28)	252 (28)		
$\sqrt{MSE}$	20.36	12.28	12.05		

White-corrected standard errors are in parentheses.

Table 4. Parameter estimates for myopic consensus model

As Table 4 indicates,  $\hat{\beta}$  is significantly smaller than 1 for all three treatments. The coefficients for treatments MG and NF are not significantly different from each other. Note that  $\hat{\beta}_C < \beta^{BR} = 0.64$ , indicating that in Treatment C players adjusted more quickly than simply best responding to the previous choices of other players. In the other two treatments,  $\hat{\beta}$  is greater than  $\beta^{BR}$  but less than 1 indicating that, according to the model, players are adjusting their choices but not treating their own previous choices entirely as unbiased estimates of the mean.  $^{20}$ 

As mentioned above, a second model of learning in the no-feedback conditions is one in which all the learning takes place at once. This model, referred to as the "one-period" learning model, suggests that players do not revise the process by which they generate their choices for several periods until, in one period  $(t^*)$ , they realize that other players might be

<sup>&</sup>lt;sup>17</sup>Table 4 reports  $\sqrt{MSE}$  instead of R-squared since the omission of a constant term in OLS creates problems with the latter.

<sup>&</sup>lt;sup>18</sup>In a pooled regression, both  $\hat{\beta}_{NF}$  and  $\hat{\beta}_{MG}$  are significantly different from  $\hat{\beta}_{C}$  at the p < 0.05 level.

<sup>&</sup>lt;sup>19</sup>However, this difference is not significant.

The values  $\beta^{BR}$  and 1 lie outside the 95% confidence intervals for both  $\hat{\beta}_{NF}$  and  $\hat{\beta}_{MG}$ .

choosing similarly. In this period players partially best respond to their own previous actions. However, in subsequent periods players' choices remain unchanged, since they now believe they have correctly taken into account what others are doing. This model corresponds to a case of the myopic consensus model in which  $\beta$  is equal to 1 for all periods before  $t^*$  and in which each player realizes only in period  $t^*$  that her previous choices are an unbiased estimate of the population mean. Therefore, the model can be written as:

$$x_{it} = \beta_t k + \epsilon_{it} \tag{6}$$

where  $\beta_t$  is equal to 1 for all  $t < t^*$  and where  $\beta_t = \beta_s < 1$  for all t and s greater than or equal to  $t^*$ . In the above model, k corresponds to the initial choice generated by prior beliefs.

Letting  $\beta_{t^*}$  equal  $\frac{k+\gamma}{k}$ , the model in Equation 6 can be rewritten as:

$$x_{it} = k + \gamma z_t + \epsilon_{it} \tag{7}$$

and estimated using OLS by regressing choice on a constant (k) and an indicator variable  $(z_t)$  equal to zero for all  $t < t^*$  and equal to one otherwise. This was done using fixed effects for individual differences. The value of  $t^*$  was determined by estimating the model for all possible values of  $t^* = 2, 3, ..., 10$  and then using the value of  $t^*$  for each treatment

that provided the best fit.<sup>21</sup> The parameter of interest,  $\beta_{t^*}$ , can then be derived from the results of this regression and is equal to  $\frac{k+\gamma}{k}$ . Therefore,  $\hat{\beta}_{t^*}$  can be obtained from  $\hat{k}$  and  $\hat{\gamma}$ . Table 5 reports the results of these regressions for the two no-feedback treatments and the corresponding estimates of  $\beta_{t^*}$ .

Dependent Variable:	$ x_{it} $					
<i>t</i>	t=4	<b>IF</b>	t = 4			
$\hat{\boldsymbol{k}}$	27.512	(1.256)	29.554	(1.215)		
This had been $\hat{m{y}}$ to the charge $\hat{m{y}}$	-7.805	(1.501)	-11.495	(1.452)		
Obs. (N)	280	(28)	280	(28)		
$R^2$	0.051		0.093			
$\sqrt{MSE}$	15.45		16.56			
$\hat{eta_{t^*}}$	0.716		0.611			

Standard errors are in parentheses. Individual fixed effects included.

Table 5. Parameter estimates for one-period learning model

The first thing to notice in Table 5 is that  $t^*$  is equal to 4 for both treatments, implying that the best fit in the one-period learning model is when the learning is assumed to have occurred after three periods of experience without feedback. The estimates of  $\gamma$  are both significantly less than zero, indicating that, according to the model, learning does take place in the fourth period (i.e.,  $\hat{\beta} < 1$ ). While the values of  $\hat{k}$  are different for the two conditions, this difference is not significant when the two are estimated in a pooled regression. The two values of  $\hat{\gamma}$  are different and this difference is significant at the p < 0.1 level when the

<sup>&</sup>lt;sup>21</sup>Determining the period for each treatment separately allows a test of the hypothesis that learning occurs more quickly in Treatment MG than in Treatment NF, implying that  $t_{NF}^*$  should be greater than  $t_{MG}^*$ .

two treatments are estimated jointly. Specifically,  $\hat{\beta_{t^*}}$  is lower for Treatment MG than for treatment NF indicating that, as hypothesized, learning occurs to a greater extent when players are asked to write down an estimate of the mean.

Comparing the two models (Tables 4 and 5), the myopic consensus model provides a better fit in terms of mean squared error than the one period learning model (Treatment NF: 12.28 vs. 15.45; Treatment MG: 12.05 vs. 16.56). This is in spite of the fact that the latter model has one more parameter.

## 6 Conclusion

The experiments reported in this paper address the question of whether convergence towards equilibrium behavior can occur in repeated play of games without <u>any</u> feedback between periods. The results from both the NF and MG treatments provide strong support for this hypothesis, pointing to a form of learning not captured by the models in the literature. In both conditions, the mean and median choices decreased with repeated play and the number of subjects changing their choices downward (toward the equilibrium) was significantly greater than the number who increased their choices. While the convergence was greater in treatment C, in which subjects received feedback, the fact that it took place at all in the other treatments indicates that learning can take place in the absence of information about outcomes and payoffs.

Parameter estimates for the two models of no-feedback learning also reflect the fact that learning is taking place in both treatment conditions. In both models, the learning parameter

 $(\beta)$  is estimated to be significantly below one. Support for the fact that these models (which are both special cases of the same more general model) capture learning can be found in the fact that the parameter estimate for the myopic consensus model reflects the greatest learning in the control condition.

A second question was whether requiring subjects to write down a guess of the value of the mean would increase convergence towards the equilibrium. The results indicate that this is also true, particularly in that the number of participants whose choices decreased between periods 1 and 10 was greater in the MG condition than in the NF condition. Also, the learning parameter estimates for Treatment MG were lower (implying greater learning) than for Treatment NF in both models and this difference was significant for one of the two models.

Taken together, these results show that something resembling learning can take place when subjects play games repeatedly, even when they receive no feedback on payoffs or the choices of other players. Moreover, there is some evidence that prompting subjects to think more carefully about certain aspects of a game may lead to behavior that is more consistent with equilibrium play.<sup>22</sup> These results are important because they reveal the possibility that the majority of the current learning models are misspecified in that they only take into account learning through adaptation. Since these models ignore the kind of learning that takes place in the absence of feedback, parameter estimates for "feedback learning" are most likely biased.

<sup>&</sup>lt;sup>22</sup>Croson (1999) and Warglien, Devetag and Legrenzi (1998) also report differences in behavior in games when subjects are asked to guess about the choices made by other players.

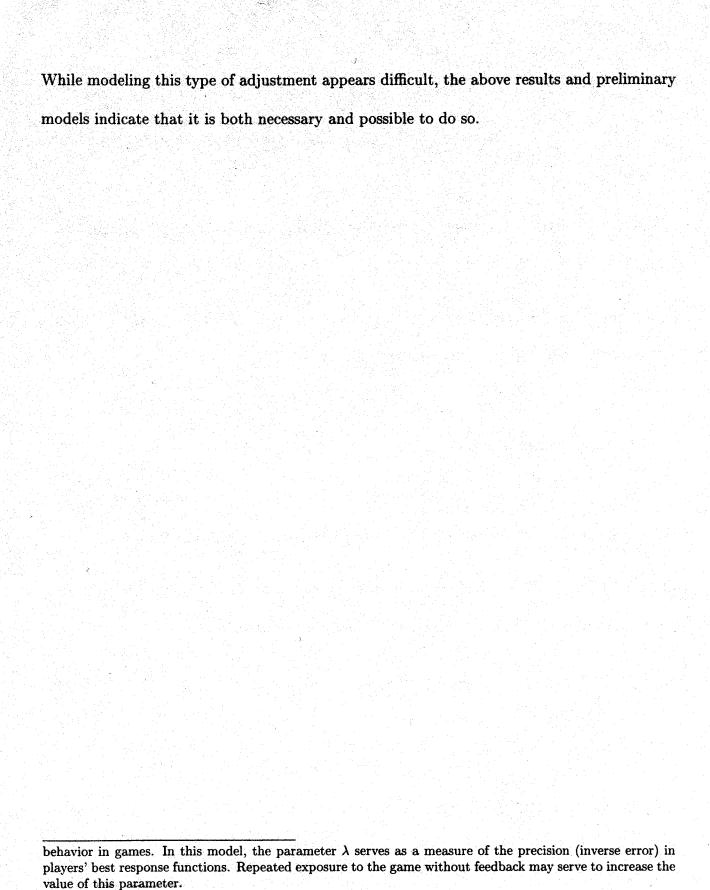
An interesting question meriting further study is to what games this result can be extended. The game used in these experiments has a unique solution on the boundary of the strategy space which is the only strategy to survive iterated deletion of dominated strategies. It is not clear whether similar convergence toward equilibrium behavior would take place in games without this property. In particular, this result might be more difficult to extend to games with multiple equilibria requiring coordinated behavior – such as Battle of the sexes or pure coordination games. On the other hand, subjects may learn to play the equilibrium without feedback in a large number games that require some insight or careful thought to determine the solution, such as the Dirty Faces game (see Weber, 1998) or variants of the Monty Hall problem (see Friedman, 1998).

Another important question has to do with how to further model the "learning" taking place. The two models presented in the paper model the mechanism operating on behavior in these experiments similarly to the standard models of learning through reinforcement of strategies and beliefs, except here the reinforcement comes from a player treating her own action as an estimate of other players' actions. This approach would be proven incorrect if no-feedback learning occurs when no such reinforcement is possible.<sup>23</sup> Instead, it might be the case that subjects are simply finding it easier to solve for equilibrium behavior with repeated experience.<sup>24</sup> For instance, repeated exposure to the game (and the environment in which the game is presented) may lower the cognitive costs of figuring out the equilibrium.<sup>25</sup>

<sup>23</sup>For instance, when players' strategies are not symmetric.

<sup>&</sup>lt;sup>24</sup>Plott (1996) provides an informal theory of subject behavior in experiments and discusses how repeated choices and practice (in addition to feedback) lead subjects to behavior more consistent with the predictions of rational choice.

<sup>&</sup>lt;sup>25</sup>As an example of this, consider McKelvey and Palfrey's (1995) Quantal Response Equilibrium model of



# References

- [1] Camerer, C. 1998. Chapter 6: Learning. In Experiments on Strategic Interaction, in press.
- [2] Camerer, C. and T. Ho. 1999. Experience-weighted attraction learning in normal form games. *Econometrica* 67(4): 827-874.
- [3] Cheung, Y. and D. Friedman. 1997. Individual learning in normal form games: Some laboratory results. *Games and Economic Behavior* 19: 46-76.
- [4] Croson, R. 1999. The disjunction effect and reason-based choice in games. Organizational Behavior & Human Decision Processes 80(2): 118-133.
- [5] Dawes, R. M. 1990. The potential nonfalsity of the false consensus effect. In *Insights in Decision Making*, ed. R. M. Hogarth. Chicago: University of Chicago Press.
- [6] Duffy, J. and N. Feltovich. 1998. Does observation of others affect learning in strategic environments? An experimental study. Unpublished manuscript.
- [7] Erev, I. and A. Roth. Predicting how people play games: Reinforcement learning in experimental games with unique, mixed strategy equilibria. *American Economic Review* 88(4): 848-881.
- [8] Friedman, D. Monty Hall's three doors: Construction and deconstruction of a choice anomaly. *American Economic Review* 88(4): 933-946.

- [9] Ho, T., C. Camerer, and K. Weigelt. 1998. Iterated dominance and iterated best response in experimental p-beauty contests. American Economic Review 88(4): 947-969.
- [10] McKelvey, R. M. and T. R. Palfrey. 1995. Quantal response equilibria for normal form games. Games and Economic Behavior 10: 6-38.
- [11] Mookherjee, D. and B. Sopher. 1994. Learning behavior in an experimental matching pennies game. Games and Economic Behavior 7: 62-91.
- [12] Nagel, R. 1995. Experimental results on interactive competitive guessing. American Economic Review 85: 1313-1326.
- [13] Nagel, R. 1998. A survey on experimental 'beauty contest games': Bounded rationality and learning. Forthcoming in *Games and Human Behavior: Essays in Honor of Amnon Rapoport* eds. D. Budescu, I. Erev, and R. Zwick. Lawrence Erlbaum Associates: New Jersey.
- [14] Plott, C. R. 1996. Rational individual behavior in markets and social choice processes: The discovered preference hypothesis. In *The Rational Foundations of Economic Behavior* eds. K. Arrow, E. Colombatto, M. Perlaman, and C. Schmidt. London: Macmillan.
- [15] Roth, A. and I. Erev. 1995. Learning in extensive form games: Experimental data and simple dynamic models in the intermediate term. Games and Economic Behavior 8: 164-212.

- [16] Ross, L., D. Greene and P. House. 1977. The "false consensus effect": An egocentric bias in social perception and attribution process. *Journal of Experimental Social Psychology* 13: 279-301.
- [17] Stahl, D. O. 1996 Boundedly rational rule learning in a guessing game. Games and Economic Behavior 16: 303-330.
- [18] Van Huyck, J., R. Battalio, and F. W. Rankin. Selection dynamics and adaptive behavior without much information. Unpublished manuscript.
- [19] Warglien, M., M. G. Devetag, and P. Legrenzi. 1998. Mental models and naive play in normal form games. Unpublished manuscript.
- [20] Weber, R. A. 1998. Uncommon knowledge: An experimental test of the 'dirty faces' game. Unpublished manuscript.

# 7 Appendix: Instructions

This is an experiment in decision making. Several research institutions have provided funds for this research. You will be paid for your participation in the experiment. If you follow the instructions, and make good decisions, you may earn an appreciable amount of money. The exact amount you earn will be determined during the experiment and will depend on your decisions and the decisions of others. Your earnings will be paid to you in cash at the conclusion of the experiment. If you have a question during the experiment, raise your hand and an experimenter will assist you. Please do not talk, exclaim, or try to communicate with other participants during the experiment. Participants violating the rules will be asked to leave the experiment and will not be paid.

Please look at the number at the top of this page. This is your participant number for the experiment. This participant number is private and should not be shared with anyone. Your participant number will be the same for the entire experiment.

The experiment will last for 10 rounds. During each round, you and the other participants will be presented with an identical choice problem. There are ten participants in the group, and each of you will simultaneously choose a number from 0 to 100 inclusive. You can choose any number in this range by writing that number in the appropriate place in the table below. This choice is private and will not be revealed to other participants. After each of you has chosen a number, the experimenter will come by to record your choice.

Your payoff will be determined as follows:

First, the average number of all participants' choices (including yours) will be computed. This average will be computed by adding up the numbers chosen by each of the ten participants and then dividing by ten.

A target number will then be determined by multiplying the average of all participants' choices by two-thirds. Thus:

Target number = 2/3 x (Average of everyone's choices)

The person who chose a number closest to this target number will earn \$6. If there are two or more participants whose choices are exactly equally close to the target number, then these participants will equally divide the \$6 prize.

C: At the end of each round, the experimenter will publicly announce the average of the ten choices and the target number and will write these two numbers on the board. The experimenter will also announce the participant numbers of the participants who won all, or part of, the prize.

In each round, please record the number you choose, the average, the target number, and the amount (if any) of the prize that you won in the table below.

NF: At the end of each round, the experimenter will calculate the average of the ten choices and the target number, but these numbers will not be announced.

At the end of the final (10th) round, the ten averages and target numbers will be publicly announced and the experimenter will announce the participant numbers of the participants who won all, or part of, the prize in each round.

At the end of the experiment, please record the average, the target number, and the amount (if any) of the prize that you won in each round in the table below.

MG: At the end of each round, the experimenter will calculate the average of the ten choices and the target number, but these numbers will not be announced. Instead, you will be asked to write down your best guess of the average in the table below. At the end of the final (10th) round, the ten averages and target numbers will be publicly announced and the experimenter will announce the participant numbers of the participants who won all, or part of, the prize in each round.

At the end of the experiment, please record the average, the target number, and the amount (if any) of the prize that you won in each round in the table below.

At the conclusion of the tenth round, your earnings for the experiment will be the sum of your earnings in each of the ten rounds plus the \$7 participation fee.

We are now ready to begin the experiment. There should be no talking from this point on. If you have a question, raise your hand.

