

Learning Transference Between Dissimilar Symmetric Normal-Form Games

by

Ernan Haruvy
University of Texas at Dallas

and

Dale O. Stahl
University of Texas at Austin

March, 2008

ABSTRACT. Action-based learning models are not directly applicable to learning that occurs between dissimilar games. One way to account for the transference of learning between games is to re-label actions based on some common properties across games. In this work, we examine the level- n framework for such purpose, and combine it with two learning dynamics—Experience Weighted Attraction and Rule Learning—to arrive at predictions for a sequence of ten thrice-played dissimilar games. Using experimental data, we find that with simple action re-labeling, both learning models perform well in capturing the between-game transference as it affects initial play in each new game. However, only Rule-Learning captures the ability of players to learn to reason across games.

1. Introduction.

The field of learning in games has evolved considerably, with many models able to make robust predictions in a variety of interesting games. The focus is typically on a single repeated game where players receive feedback each period about payoffs and the history of play.¹ Different models, although different in details, derive their dynamic predictive power from the fact that the same game is played over and over again. In contrast, if the players encounter different games, it is unlikely that the performance of action labeled “A” in one game has any relevance whatsoever to the performance of the action labeled “A” in a different game. Indeed, if the players encounter a sequence of “sufficiently dissimilar” games, so there is no obvious linkage between the actions in subsequent games, then the above-mentioned learning models would predict no learning. On the other hand, it is not unreasonable to expect human subjects to learn something from their experience in a sequence of dissimilar games. For example, they might learn to identify and eliminate dominated strategies, and perhaps even to iterate such eliminations².

It might be possible to bridge between different games by re-labeling their actions according to their properties. For example, one could capture learning to avoid dominated actions by labeling dominated actions as “D” in all games and looking at reinforcement over “D” in the sequence of games. While such re-labeling, as we show in this work, can improve the fit of action-based learning models, it cannot account for players increasing their sophistication over time in playing dissimilar games. That is, after playing several different games, players may learn to reason about the games. Such increased sophistication cannot be captured by action-based learning models. In contrast to action-based learning, the Rule Learning framework of Stahl (1996, 1999, 2000, 2003; henceforth “Rule Learning”) and Stahl and Haruvy (2002), by explicitly allowing for *transference* across games, would predict learning in a sequence of

¹Under action reinforcement learning (Roth and Erev, 1995; Erev and Roth, 1998), each action is reinforced according to the payoff received relative to a dynamic aspiration level. Under belief learning (Fudenberg and Levine, 1998; Cheung and Friedman, 1997, 1998), each player updates his/her belief about the action the other player will choose, and then chooses a (possibly noisy) best-reply to that updated belief. Camerer and Ho (1997, 1998, 1999) have studied hybrid models that combine reinforcement and belief learning features. See also Anderson, et.al. (2001), Camerer, et.al. (2000), Feltovich (2000), Friedman, et.al. (1995), Mookherjee and Sopher (1994, 1997), Nagel (1995), Rapoport, et.al. (1997), Sarin and Vahid (1999), Selten (1990, 1991), Stahl (2003), Tang (1998), and Van Huyck, et.al. (1991, 1994, 1996).

²Weber (2004) shows that such “reflective learning” can occur even without feedback. That learning has to do with the ability to reason about a game and to learn strategies from reflection over time. In this paper, we focus on cognitive learning with feedback.

dissimilar games. Analogous to iterative elimination of dominated strategies and rationalizability, the level-n rules have an iterative structure, so a player can learn to switch to higher level rules based on past performance even though the game changes.

The purpose of this paper is twofold. First, we present an experiment design to gather data on how human players behave when faced with a sequence of dissimilar games. By “dissimilar” we mean that there is no obvious re-labeling of the actions that makes the games monotonic transformations of each other (as in Van Huyck *et al*, 2001, Cooper and Kagel, 2004). We include a permutation of each payoff matrix with 12 periods of dissimilar games (4 games with 3 periods each) between occurrences of a permuted game. The use of permuted games allows us to provide easily identifiable evidence of behavioral differences which can be attributed to learning.

Section 2 describes the experimental design and the data. Section 3 presents the Rule Learning model adapted to the environment of dissimilar games. Section 4 confronts experienced weighted attraction and the enhanced Rule Learning model with the experimental data. Section 5 concludes.

2. The Experiment.

We chose a sequence of ten 4×4 symmetric normal-form games, as displayed in Figure 1. The game payoffs were in tokens, with an exchange rate of 100 tokens for \$1. Four games (3,5,8,10) are dominance solvable, and all distinguish between Level-1, Level-2 and Nash actions. Note that game $i \in \{1,2,3,4,5\}$ is the same as game $i + 5$, but with the rows (and columns) permuted so the identity is not obvious. Each game was played for three periods before proceeding to the next game. After each period (play of the game), the participants were shown the results of the most recent period. We will call the first 15 periods the “first run”, and the second 15 periods with the permuted games the “second run”.

A “mean-matching” protocol was used. In each period, a participant’s token payoff was determined by her choice and the percentage distribution of the choices of all *other* participants, as follows: the row of the payoff matrix corresponding to the participant’s choice was multiplied by the choice distribution of the *other* participants. The lotteries that determined final monetary payoffs were conducted following the completion of all thirty periods. Payment was made in cash immediately following the session.

Participants were seated at private computer terminals separated so that no participant could observe the choices of other participants. The relevant game, or decision matrix, was presented on the computer screen. Each participant could make a choice by clicking the mouse button on any row of the matrix, which then became highlighted. In addition, each participant could make hypotheses about the choices of the other players. An on-screen calculator would then calculate and display the hypothetical payoffs to each available action given each hypothesis. Participants were allowed to make as many hypothetical calculations and choice revisions as time permitted. Following each one-minute period, each participant was shown the payoff matrix, her choice, the percentage distribution of the choices of all other participants, and her payoff, and was given 30 seconds to contemplate that information before proceeding to the next period.

The experiment consisted of three sessions of 24, 25, and 25 participants playing this sequence of thirty games. The subjects were upper division undergraduate students and non-economics graduate students from the University of Texas.

3. Models

In many works on learning, initial play has been all but ignored. In some works (e.g., Camerer and Ho, 1999), initial propensities are estimated as free parameters; in others (Erev and Roth, 1998), initial play is fixed. Either approach is reasonable when the focus is on the dynamics from the initial starting point. However, in predicting play in a sequence of games, dismissing initial play is not an option as this would result in eliminating much of the history of play (one third in our case). Hence, a model of initial play in each game is now central to the modeling of learning.

Hence, a complete learning model has to meet two goals: (1) predicting initial play in each new game, and (2) predicting dynamics in each new game. One advantage of studying a learning model as opposed to simple hypothesis testing is that we can look at how the entire past history—not just the most recent period—affects initial game play in a new game. The second advantage is that we can look at how the past affects within game learning and reasoning. As our data show, learning between periods 1 and 2 is more substantial in the second game in a pair. We conjecture that this faster game dynamic is due to increased sophistication on the part of the subjects and we try to capture this pattern with a learning model.

The first learning model we examine is Experience Weighted Attraction (hereafter EWA, Camerer and Ho, 1999). This model is widely used in the literature and claims to encompass other common representations. We adapt the model to learning in a sequence of games by re-labeling the actions according to the taxonomy employed by Stahl and Wilson (1995, hereafter SW95). Specifically, we relabel the actions so the first action is the Nash equilibrium strategy, the second is the Level-1 strategy of SW95, the third is the Level-2 strategy of SW95, and the fourth is the remaining strategy.

The second model we examine is Rule Learning (Stahl, 1996, 1999, 2000). Rule Learning model is intended to represent how subjects learn to reason about games. This feature cannot be captured in a re-labeled strategy EWA, but will be the main focus of a rule-learning model.

3.1. Experience Weighted Attraction

Camerer and Ho (1996) put forth an innovative propensity-based reinforcement model that combines several features of the Erev-Roth reinforcement learning and belief learning models. The idea behind the EWA model is that decision makers evaluate the performance of each possible action in the last period and update their propensities to use each action accordingly. However, the action actually chosen by each decision maker receives greater attention in the evaluation process. Hence, actions are reinforced according to past performance, but actions actually selected receive some additional reinforcement.

EWA is an individual learning model, so we aggregate to derive a population version. We also restrict attention to the logit form. The underlying state variable, $A_j(t)$, which Camerer and Ho call “attractions,” are updated according to the following dynamic:

$$A_j = \{ \theta N(t-1) A_j(t-1) + g_j(t-1) \} / N(t),$$

$$\text{where } g_j(t) \equiv \alpha e_j' \lambda U_p(t) + (1-\alpha) p_j(t) \lambda [e_j' U_p(t) n / (n-1) - U_{jj} / (n-1)],$$

$$(1) \quad N(t) = \gamma N(t-1) + 1, \text{ and } N(0) \equiv N_0.$$

The parameter α is the probability that an individual evaluates the past performance of all actions. Camerer and Ho call this the “imagination parameter.”

The $N(t)$ term is called the “experience weight.” If $N_0 < (1-\gamma)$ and $\gamma \in (0,1)$, then $N(t)/N(t+1)$ declines, putting less weight on new evidence as time passes (the power law of practice), and vice versa. On the other hand, if $\gamma = 0$ or $\gamma = (N_0-1)/N_0$, then $N(t)$ is constant for all t , and since N_0 and λ are then not separately identifiable in the population model, we can eliminate $N(t)$ entirely.

For initial conditions, Camerer and Ho let $\{A_j(1)\}$ be free parameters. Rather than increase the number of free parameters, we take the approach of insufficient reason, specifying $A_j(0) = 0$ for all j and $p(0) = p^0$.

To adopt EWA to dissimilar games, we re-label the actions so the first action is the Nash equilibrium strategy, the second is the precise Level-1 strategy, the third is the precise best-reply to the Level-1 strategy, and the fourth is the remaining strategy. To allow for transference, for games $g = 2, \dots, 10$, we specify the initial attractions as

$$A_{jg}(1) = \tau A_{j,g-1}(4) + (1-\tau)A_{j,g-1}(1). \quad (2)$$

Note that if $\tau=1$, EWA learning proceeds across games without interruption, while if $\tau=0$, no transference at all occurs across games.

3. 2. Rule Learning

We provide a brief overview of rule learning here. A detailed description is provided in appendix B and in Stahl (1996, 1999, 2000, 2003) and Stahl and Haruvy (2002). A *behavioral rule* maps from information about the game and the history of play to a strategy. By the law of effect, behavioral rules which perform well are more likely to be used in the future. Since we

specify a model of population averages rather than individual responses, the reinforcement function represents the cumulated effects of experience throughout the population. Our approach to specifying the space of rules is to define a finite number of empirically relevant discrete rules that can be combined to span a much larger space of rules. In S96, S99 and S00, the family of "evidence-based" rules was introduced as an extension of the Stahl-Wilson (1995) [hereafter SW95] level-n rules. Evidence-based rules are derived from the notion that a player considers evidence for and against the available actions and tends to choose the action which has the most net favorable evidence based on the available information.

The first kind of evidence comes from a "null" model of the other players. In the first period, the null belief, p^0 , is that all actions are equally likely. In later periods ($t > 1$), the null model uses simple distributed-lag forecasting:

$$q^t(\theta) \equiv (1-\theta)q^{t-1}(\theta) + \theta p^{t-1}. \quad (3)$$

where p^0 is the uniform distribution and p^1 is the empirical distribution of play in period 1. The "level-1" evidence in favor of action j is the expected utility payoff to action j given belief $q^t(\theta)$. The second kind of evidence is based on the Stahl & Wilson (1995) "level-2" player who believes all other players are level-1 players. This is a second order theory of mind in which other minds are believed to be level 1. The "level-2" evidence in favor of action j is the expected utility payoff given this belief..

The third mode of behavior is Nash equilibrium. Letting p^{NE} denote a Nash equilibrium of G , $y_3 \equiv Up^{NE}$ provides a third kind of evidence on the available actions.

So far we have defined three kinds of evidence: $Y \equiv \{y_1, y_2, y_3\}$. Each type of evidence is weighted to arrive at the overall evidence. Each player assesses the weighted evidence with some error, and chooses the action that from his/her perspective has the greatest net favorable

evidence. For the mapping between evidence and action, we opt for the multinomial logit specification because of its computational advantages when it comes to empirical estimation.

Next, we represent behavior that is random in the first period and "follows the herd" in subsequent periods. Following the herd does not mean exactly replicating the most recent past, but rather following the past with inertia as represented by $q^t(\theta)$. Hence, Eq.(3) represents herd behavior as well as the beliefs of level-1 types.

Finally, we allow for uniform trembles by introducing the uniformly random rule. Thus, the base model consists of a four-dimensional space of evidence-based rules (v, θ) , where $v_k \geq 0$ denotes a scalar weight associated with each evidence y_k , a herd rule characterized by θ , and uniform trembles.

Since this theory is about rules that use game information as input, we should be able to predict behavior in a temporal sequence that involves a variety of games. For instance, suppose an experiment consists of one run with one game for T periods, followed by a second run with another game for T periods. How is learning about the rules during the first run transferred to the second run with the new game? A natural assumption would be that the log-propensities at the end of the first game are simply carried forward to the new game. Another extreme assumption would be that the new game is perceived as a totally different situation so the log-propensities revert to their initial state. We opt for a convex combination, with $(1-\tau)$ weight on the initial state and τ weight on the end of the previous game. If $\tau = 0$, there is no transference, so period $T+1$ has the same initial log-propensity as period 1; and if $\tau = 1$, there is complete transference, so the first period of the second run has the log-propensity that would prevail if it were period $T+1$ of the first run (with no change of game). This specification extends the model to any number of runs with different games without requiring additional parameters.

The entire model involves 10 parameters: $\beta \equiv (\bar{v}_1, \bar{v}_2, \bar{v}_3, \bar{\theta}, \sigma, \delta_h, \varepsilon, \beta_0, \beta_1, \tau)$. The first four parameters $(\bar{v}_1, \bar{v}_2, \bar{v}_3, \bar{\theta})$ represent the mean of the participant's initial propensity over the evidence-based rules, and σ is the standard deviation of that propensity; the next two parameters (δ_h, ε) are the initial propensities of the herd and tremble rules respectively; β_0 is an inertia parameter; β_1 is a scaling parameter; and τ is the transference parameter for the initial propensity of the subsequent runs.

4. Results.

4.1. Pairwise Comparisons of Games

Recall that the experiment design repeated each of the first five games in permuted form, with four games (12 periods) in-between. This allows us to compare behavior for a game in the first run with behavior in the (permuted) game in the second run. By "pair i " we mean games i and $i+5$. We first look, in Figure 2, at the first period in each pair. With respect to Level-1 and NE, a noticeable difference in pairs 2, 3, and 5 is the rise of NE and the decline of Level-1 play. Note that this observation does not hold for every pair in every session.

The aggregate choices expressed as choices of level-1, level-2 and Nash choices are displayed in Figure 3. Figure 4 shows the change in Nash equilibrium play between periods 1 and 2 of each game as a percentage of the change between periods 1 and 3. From figure 4, we see that in each pair, relatively more change takes place from period 1 to period 2 in the second run. This suggests faster learning and possibly increased sophistication. The hypothesis that the proportion playing NE strategy in the last period of the second run of a pair is greater than the last period of the first run has a p-value of 0.0001 (14 df). The hypothesis that the proportion playing L1 strategy in the last period of the second run of a pair is smaller than the last period of the first run has a p-value of 0.0014 (14 df).

Figure 5 shows the differences from the first to the second runs in the proportions playing L1, L2, and NE for each pair and each period. The summary of the results in the fourth panel clearly shows that the largest changes in proportions of L1, L2 and NE occur in period 2, as

compared to period 1 and period 3. These differences are not significant between period 2 and period 3, but are statistically significant at the 6% and 5% level (2-sided t-test) between period 1 and period 2 for L1 and NE, respectively (but not L2). The difference between period 1 and combined period 2 and 3 is significant at under 4% for both L1 and NE.

4.2. Experience Weighted Attraction

Our maximum likelihood estimation of the population EWA model yielded an estimate for γ of exactly 0, which renders N_0 and β not separately identifiable. Therefore, we dropped γ and N_0 from the population EWA model. To give population EWA its best chance, we introduced a tremble parameter ε in that model which is the probability of a uniform error each period. The following table gives the ML estimates of the individual and population EWA models.

Table 1 Population EWA model with re-labeling and transference

	Pop EWA w/ re-labeling – without transference ($\tau=0$)	Pop EWA w/ re-labeling and no re- initialization ($\tau=1$)	Pop EWA w/ re-labeling and transference
θ	0.100	0.829	0.720
α	0.746	0.479	0.900
ε	0.040	0.007	0.041
λ	0.092	0.044	0.065
τ	0	1	0.096
LL	-1985.18	-2098.59	-1947.61

To test the hypothesis of no-transference, we set $\tau=0$ and re-estimate the model. The maximized LL decreases to -1985.18. Twice this difference is distributed Chi-square with 1 degree of freedom, and has a p-value $< 10^{-17}$; thus, we reject the null hypothesis of no transference in the EWA framework. For completeness we also tested the hypothesis of 100% transference by setting $\tau=1$ and re-estimating the model. Not surprisingly, the maximized LL dramatically decreases to -2098.59, thereby strongly rejecting the hypothesis of 100% transference in the EWA framework.

4.3. Rule Learning

We estimated the 10 parameters of the basic Rule Learning model using all the data from the 3 sessions. We pool over games in order to try to identify regular features of learning dynamics that are general and not game-specific, so we can be more confident that these features will be important in predicting out-of-sample behavior. The ML parameter estimates are given in Table 2.

Table 2. Parameter Estimates for RLRN

σ	1.915
δ_h	0.484
\bar{v}_1	1.082
\bar{v}_2	0.000
\bar{v}_3	0.000
θ	1.000
β_0	1.000
β_1	0.00648
τ	1.000
ε	0.000

The maximized LL is -1855.12 compared to an entropy likelihood of -1651.02. The Pearson Chi-square statistic for the entire data set is 428.17. The Root-Mean Squared Error of the predicted versus actual choice frequencies averaged over all games is 0.095.

Note that six of these 10 ML estimates are on the boundary of their respective theoretical

domain. Thus, *only four* of the parameters are determining the fit of the data: (i) the standard deviation of the initial distribution of propensities (σ), (ii) the initial precision of the Level-1 rule (\bar{v}_1), (iii) the initial probability of the herd rule (δ_h), and (iv) the reinforcement scaling parameter (β_1). It is also noteworthy that the ML estimate of the transference parameter (τ) is exactly 1. In other words, there is 100% transference of rule propensities across the dissimilar games.

The central hypothesis to consider is whether there is any transference and/or learning of rules whatsoever. To test the hypothesis of no-transference, we set $\tau=0$ and re-estimate the model. The maximized LL decreases to -1889.83. Twice this difference is distributed Chi-square with 1 degree of freedom, and has a p-value $< 10^{-16}$; thus, we can strongly reject the null hypothesis of no transference. To test the hypothesis of no-rule-learning, we set $\beta_0 = 1$ and $\beta_1 = 0$, and re-estimate the model. The maximized LL decreases to -1889.83.³ Twice this difference is distributed Chi-square with 2 degrees of freedom, and has a p-value $< 10^{-15}$; thus, we can strongly reject the null hypothesis of no rule learning.

One way to assess what is learned is to compute the implied beliefs over types:

$$q_k(t) \equiv \bar{v}_k(t) / [\bar{v}_1(t) + \bar{v}_2(t) + \bar{v}_3(t)]. \quad (4)$$

For $k \in \{1, 2, 3\}$, $q_k(t)$ can be interpreted as a representative participant's belief (probability) that the other participants use the Level-0, Level-1, and Nash rules respectively. At the beginning of period 1, $q(1) = (0.442, 0.101, 0.101)$, and it changes smoothly to $q(30) = (0.421, 0.104, 0.175)$. There is a slight decline in the belief that others are Level-0 types, and a corresponding increase in the belief that others are Nash types.

The major dynamic change is a dramatic decrease in the propensity of the herd rule – from 48.4% to 7.3%, and the corresponding increase in the evidence-based rules (Level-1, Level-2 and Nash). Thus, while the distribution over these evidence-based rules does not change substantially, the aggregate propensity to use them increases dramatically.

³ In the previous test of $\tau = 0$, the ML estimate for β_1 was 0, so the two maximized LL values are the same. Curiously, without the possibility of transference, rule learning does not help explain the data: all the behavior change appears to be due to herding and belief learning.

5. Conclusions

We selected a sequence of 10 normal-form games. By design, action-reinforcement learning models would predict no learning between games, but in contrast Rule Learning predicts learning. We examined a sample consisting of 74 participants. Statistical tests also strongly rejected the null hypothesis of no rule learning. Therefore, the participants appear to learn abstract aspects of the games which are transferable to subsequent dissimilar games. The implied change in the distribution over rules reveals a substantial increase in the “depth of reasoning”.

An interesting insight comes from pairwise comparisons of permuted games within the sequence of dissimilar games. A majority of subjects exhibit level-1 behavior in the initial period of a game, but not in later periods. While there is some learning towards NE in both runs of a game, the bulk of the increase in NE behavior occurs earlier in the second period.

This pattern—first periods seem similar but experienced subjects exhibit faster convergence-- is not uncommon in the experimental literature, with instances in various settings, including asset markets (Dufwenberg et al. 2005) and prisoner dilemma games (Bereby-Meyer and Roth, 2006). This is the first attempt at providing a model of adaptive dynamics that can capture such patterns and as such it provides important economic insights.

References

- Anderson, S., J. Goeree, and C. Holt (2001), "Minimum-Effort Coordination Games: Stochastic Potential and Logit Equilibrium," *Games and Economic Behavior* **34**, 177-199.
- Barron, G., and I. Erev (2000), "Toward a General Descriptive Model of One Shot and Repeated Decision Making Under Risk and Uncertainty," Technion Working Paper.
- Bereby-Meyer, Yoella, and Alvin E. Roth (2006) "Learning in Noisy Games: Partial Reinforcement and the Sustainability of Cooperation." *American Economic Review* 96(4), 1029-1042.
- Cheung, Y. and D. Friedman (1997), "Individual Learning in Normal Form Games: Some Laboratory Results," *Games and Economic Behavior*, **19**, 46-76.
- Cheung, Y-W, and D. Friedman (1998), "Comparison of Learning and Replicator Dynamics Using Experimental Data," *Journal of Economic Behavior and Organization*, **35**, 263-280.
- Camerer, C. and T. Ho (1997), "EWA Learning in Games: Preliminary Estimates from Weak-Link Games," in D. Budescu, I. Erev, and R. Zwick (eds), *Games and Human Behavior: Essays in Honor of Amnon Rapoport*.
- Camerer, C. and T. Ho (1998), "EWA Learning in Coordination Games: Probability Rules, Heterogeneity, and Time-variation," *Journal of Mathematical Psychology*, **42**:2, 305-326
- Camerer, C. and T. Ho (1999), "Experience-Weighted Attraction Learning in Normal Form Games," *Econometrica*, **67**, 827-874.
- Camerer, C., T-H. Ho, and J-K. Chong (2000), "Sophisticated EWA Learning and Strategic Teaching in Repeated Games," Working Paper #00-005, Wharton, Univ. of Penn.
- Cooper, David and John Van Huyck (2004), Transfer in Random Order Statistic and p-Beauty Contest Games, working Paper Case Western Reserve University.
- Cooper, David and John Kagel (2004), Learning and Transfer in Signaling Games, working paper.
- Dufwenberg, Martin, Lindqvist & Evan Moore (2005). Bubbles & Experience: An Experiment, *American Economic Review* 95 (2005), 1731-37.
- Erev, I. and A. Roth (1998), "Predicting How People Play Games: Reinforcement Learning in Experimental Games with Unique Mixed Strategy Equilibria," *American Economic Review* **88**, 848-881.
- Feltovich, N. (2000), "Reinforcement-Based vs. Beliefs-Based Learning in Experimental Asymmetric-Information Games," *Econometrica*, 60, 605-641.
- Friedman, D., D. Massaro and M. Cohen (1995), "A Comparison of Learning Models," *J. of Mathematical Psychology*, 39, 164-178.
- Fudenberg, D. and D. K. Levine (1998), *The Theory of Learning in Games*, MIT Press.
- Haruvy, E., and D. O. Stahl (1999), "Empirical Tests of Equilibrium Selection based on Player Heterogeneity."

- Haruvy, E., D. Stahl and P. Wilson, (1999), "Evidence for Optimistic and Pessimistic Behavior in Normal-Form Games," *Economics Letters*, **63**, 1999, 255-260.
- Mookherjee, D. and B. Sopher (1994), "Learning Behavior in an Experimental Matching Pennies Game," *Games and Economic Behavior*, **7**, 62-91.
- _____ (1997), "Learning and Decision Costs in Experimental Constant Sum Games," *Games and Economic Behavior*, **19**, 97-132.
- Nagel, R. (1995). "Unraveling in Guessing Games: An experimental study," *American Economic Review* **85**, 1313-1326.
- Rapoport, A., I. Eren, E. Abraham, and D. Olson (1997), "Randomization and Adaptive Learning in a Simplified Poker Game," *Organizational Behavior and Human Decision Processes*, **69**, 31-49.
- 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.
- Sarin, R. and F. Vahid (1999) "Payoff Assessments without Probabilities: A Simple Dynamic Model of Choice," *Games and Economic Behavior*, **28**, 294-309.
- Selten, R. (1990), "Anticipatory Learning in Games," in *Game Equilibrium Models. Vol. I: Evolution and Game Dynamics* (R. Selten, ed), Springer-Verlag.
- Selten, R. (1991), "Evolution, Learning, and Economic Behavior," *Games and Economic Behavior*, **3**, 3-24.
- Stahl, D. (1996), "Boundedly Rational Rule Learning in a Guessing Game," *Games and Economic Behavior*, **16**, 303-330.
- Stahl, D. (1999), "Evidence Based Rules and Learning in Symmetric Normal Form Games," *International Journal of Game Theory* **28**, 111-130.
- Stahl, D. (2000), "Rule Learning in Symmetric Normal-Form Games: Theory and Evidence," *Games and Economic Behavior*, **32**, 105-138.
- Stahl, D. (2003), "Action Reinforcement Learning versus Rule Learning," *Greek Economic Review*, **22**.
- Stahl, D. and E. Haruvy (2002), "Aspiration-based and Reciprocity-based Rules in Learning Dynamics for Symmetric Normal-Form Games," *J. of Math. Psych*, **46**, 531-553.
- Stahl, D. and P. Wilson (1994), "Experimental Evidence of Players' Models of Other Players," *J. of Econ. Behavior and Org.*, **25**, 309-327.
- Stahl, D. and P. Wilson (1995), "On Players Models of Other Players: Theory and Experimental Evidence," *Games and Economic Behavior*, **10**, 218-254.
- Tang, F-F. (1998), *Anticipatory Learning in Two-Person Games: An Experimental Study*, Lecture Series in Mathematical and Economic Systems, Springer-Verlag (in press).
- Van Huyck, J., R. Battalio and R. Beil (1991), "Strategic Uncertainty, Equilibrium Selection Principles, and Coordination Failures in Average Opinion Games," *Quart. J. of Economics*, **106**, 885-911.

Van Huyck, J., J. Cook, and R. Battalio (1994), "Selection Dynamics, Asymptotic Stability, and Adaptive Behavior," *J. of Political Economy*, 102, 975-1005.

Van Huyck, J., R. Battalio and F. Rankin (1996), "Evidence on Learning in Coordination Games," mimeo.

Van Huyck, J., R. Battalio and L. Samuelson, (2001), "Optimization Incentives and Coordination Failure in Laboratory Stag Hunt Games," *Econometrica*, 69, 749-764.

Weber, Roberto (2004), Reflective Learning and Transfer of Learning in Games Played Repeatedly without Feedback, working paper

Figure 1. The Games

Game 1						Game 6					
	A	B	C	D			A	B	C	D	
A	60	15	100	90	L1	A	35	0	15	90	L2
B	80	70	80	0	NE	B	5	70	10	65	DOM
C	90	15	35	0	L2	C	80	0	70	80	NE
D	65	10	5	70	DOM	D	100	90	15	60	L1
Game 2						Game 7					
	A	B	C	D			A	B	C	D	
A	10	95	20	0	L2	A	45	55	100	90	L1
B	90	45	100	55	L1	B	70	95	15	60	NE
C	95	5	40	15	L3	C	5	15	40	95	L3
D	60	70	15	95	NE	D	95	0	20	10	L2
Game 3						Game 8					
	A	B	C	D			A	B	C	D	
A	25	80	95	15	NE	A	0	15	70	5	DOM
B	15	80	80	100	L1	B	100	80	80	15	L1
C	15	90	75	50	L2	C	50	90	75	15	L2
D	5	15	70	0	DOM	D	15	80	95	25	NE
Game 4						Game 9					
	A	B	C	D			A	B	C	D	
A	5	55	95	70	L1	A	80	80	30	15	NE
B	30	80	15	80	NE	B	10	55	100	10	L2
C	0	10	90	50	DOM	C	55	70	5	95	L1
D	100	10	15	55	L2	D	10	50	0	90	DOM
Game 5						Game 10					
	A	B	C	D			A	B	C	D	
A	60	20	0	40	DOM	A	25	30	10	70	L2
B	100	65	30	25	L1	B	70	40	20	50	NE
C	20	50	40	70	NE	C	40	0	60	20	DOM
D	10	70	30	25	L2	D	25	30	100	65	L1

Figure 2. First period differences of population choices for each pair (proportion in the second run minus proportion in the first run).

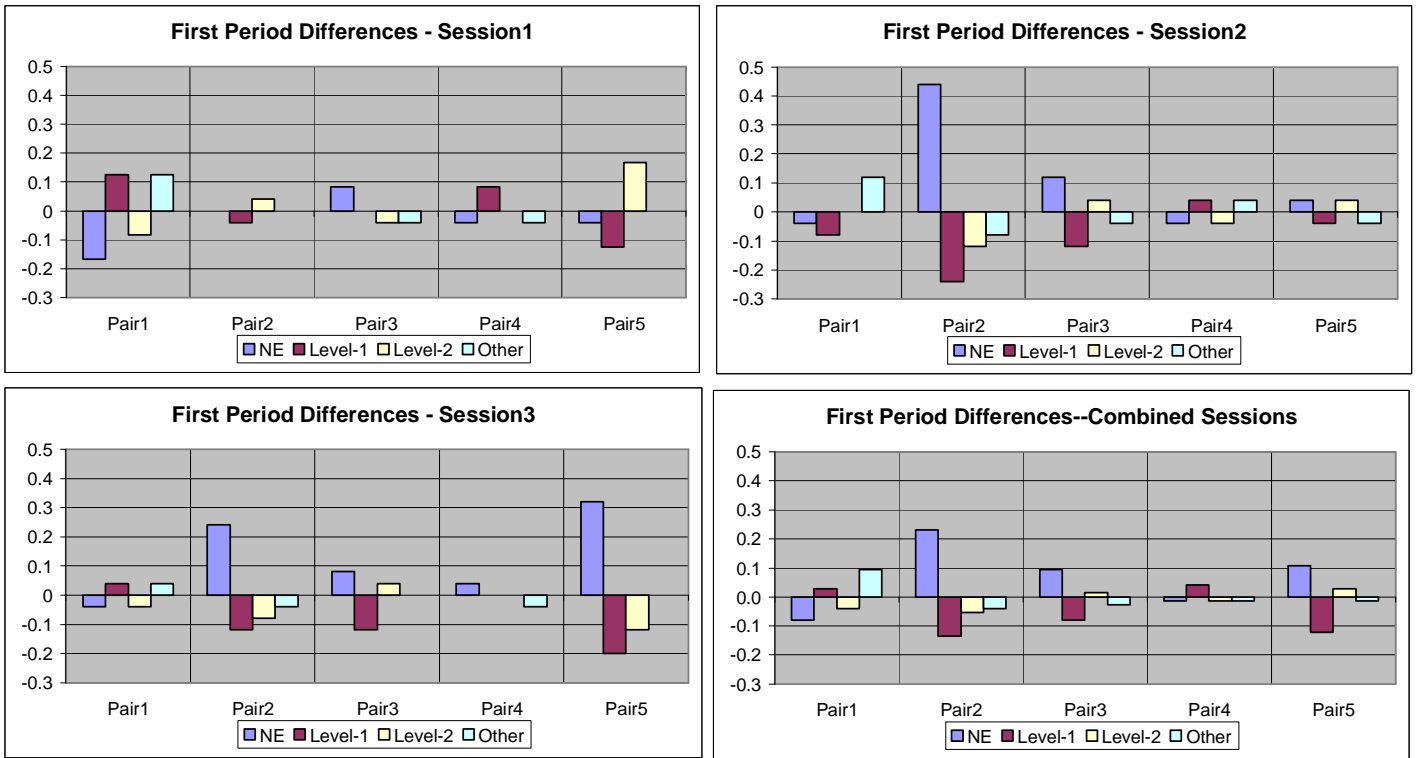


Figure 3. Population choice percentages of Level-1, Level-2 and NE play for all three periods and each run.

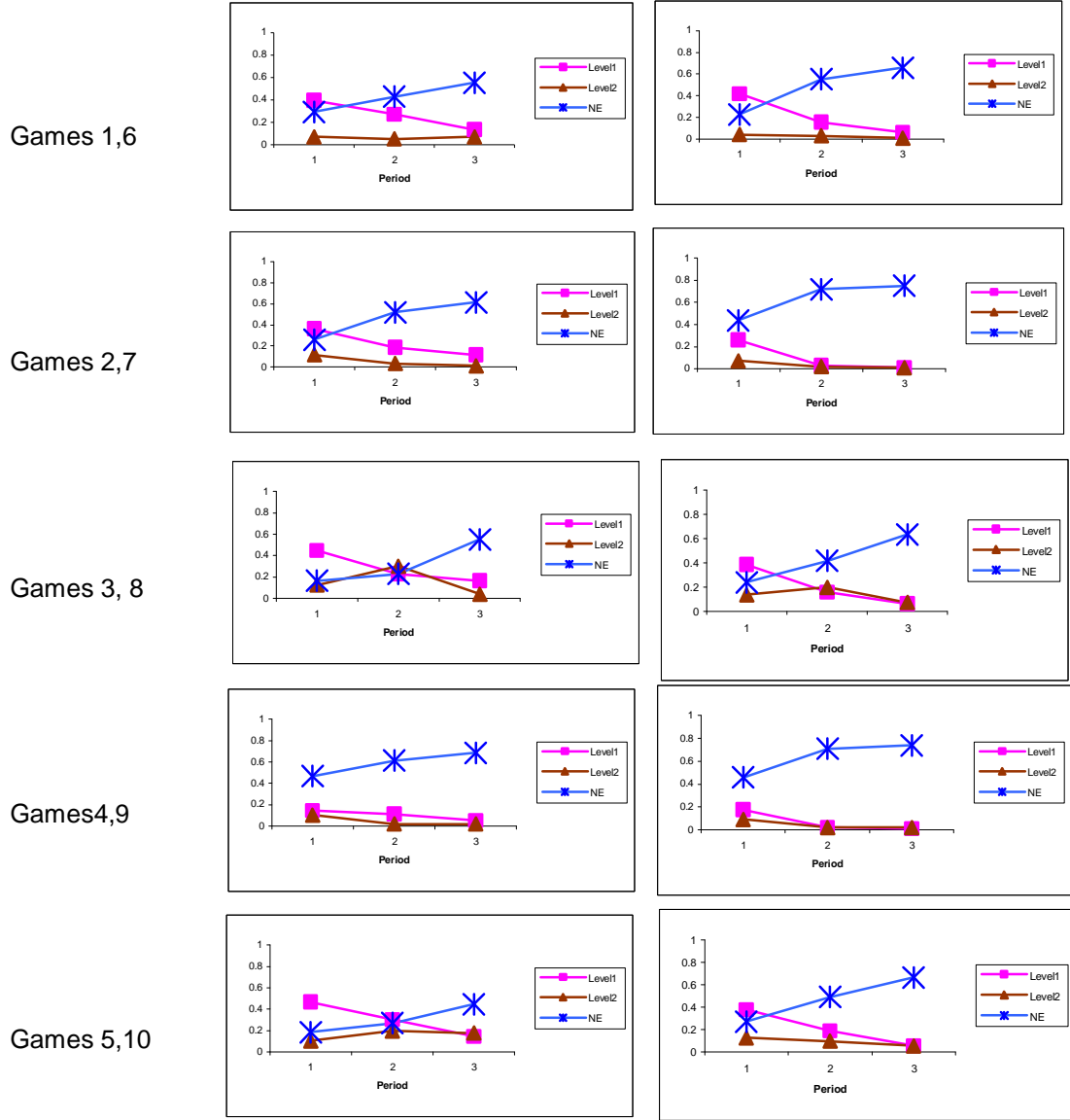


Figure 4. Change in proportion of population choosing Nash equilibrium play between periods 1 and 2 as a proportion of the overall change from period 1 to 3. $y = (\text{Proportion choosing NE in period 2} - \text{Proportion choosing NE in period 1}) / (\text{Proportion choosing NE in period 3} - \text{Proportion choosing NE in period 1})$. This shows that by the second time a game is played, most of the change to NE happens by period 2.

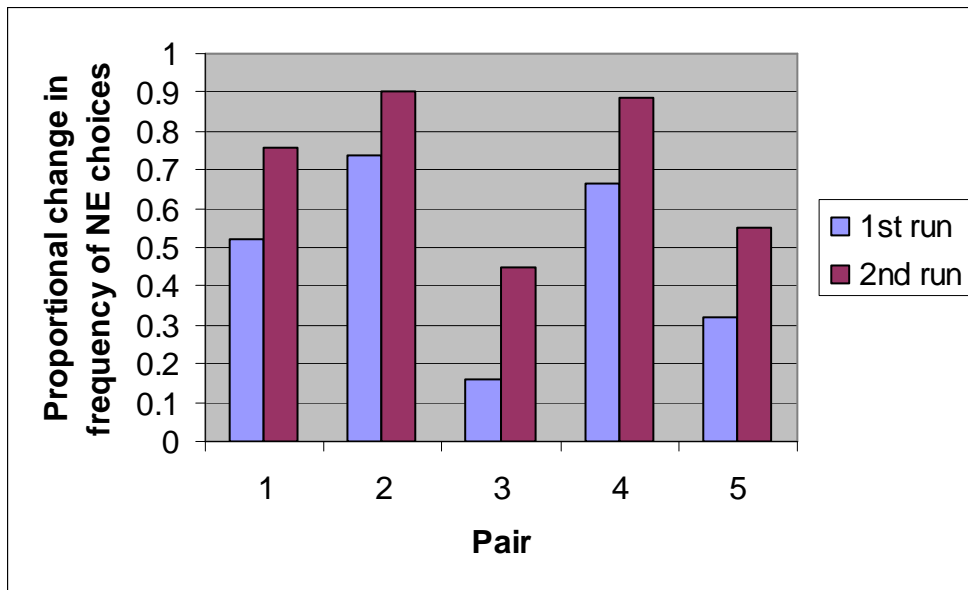
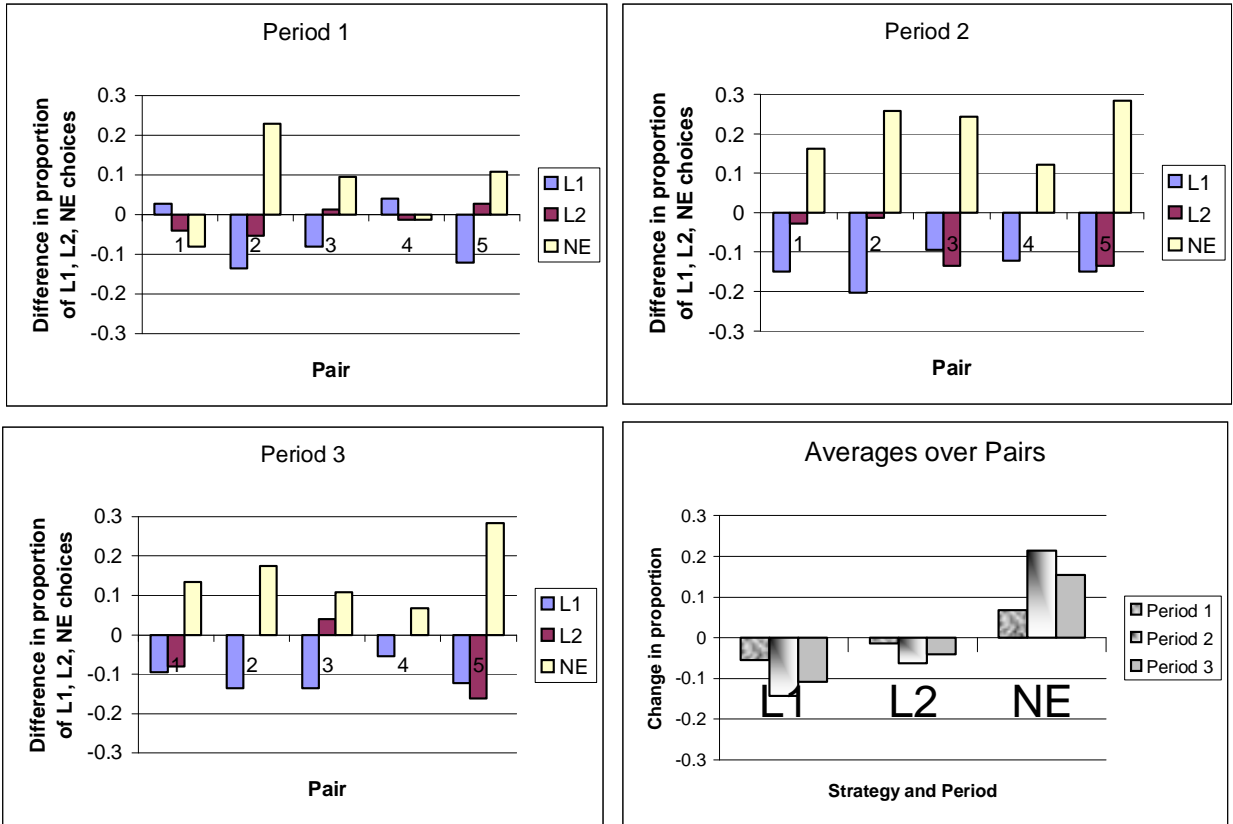


Figure 5. Differences from 1st to 2nd run in proportions playing L1, L2, and NE by pair and period. The summary of the results in the fourth panel clearly shows that the largest changes in proportions of L1, L2 and NE occur in period 2, as compared to period 1 and period 3.



Appendix A Instructions

Welcome. This is an experiment about economic decision making. If you follow the instructions carefully you might earn a considerable amount of money which will be paid at the end of the experiment in private and in cash. It is important that during the experiment you remain silent. If you have questions or need assistance, raise your hand but do not speak. During the experiment, you and all other participants will make 30 decisions, each worth up to \$1.00 (exchange rate is 100 tokens = \$1). Hence, it is possible to earn up to \$30 in this experiment. Each decision you face will be described by a MATRIX, consisting of 16 numbers arranged in 4 rows and 4 columns.

The rows indicate your possible choices; the columns indicate the possible choices of all other participants in this room. The numbers in the MATRIX, along with your choices and the choices of all OTHER participants in this session, determine your TOKEN earnings for each decision. Each participant will face exactly the same matrices and will have the same information. Your decision and those of all the other participants will determine your TOKEN earnings.

How token earnings are computed

Press QUIT and click on the DEMO button. Suppose the matrix you are facing is the following:

	A	B	C	D
A	30	100	50	20
B	40	0	90	40
C	50	75	20	30
D	10	10	10	10

Suppose 20% chose A, 20% chose B, 50% chose C, and 10% chose D. You chose Row A. We first write down the choice labels A, B, C, and D. Underneath them we write down the percentage choices of others. And underneath the percentage choices we write down the numbers of row A (your choice). We then multiply each column. Finally, we add up the results:

	A	B	C	D
% of Others' Choices :	20%	20%	50%	10%
Your Row Choice: A	30	100	50	20
Product of Each column	6	20	25	2

Sum of the bottom row = $6 + 20 + 25 + 2 = 53$. Hence, your payoff for choosing A given the other participants' percentages would be 53 tokens.

The payoffs you would have earned for the other row choices are calculated the same way. We will quickly work out your payoff had you chosen row B. We first write down the choice labels A, B, C, and D. Underneath them we write down the percentage choices of others (same as before). And underneath the percentage choices we write down the numbers of row B (your choice). We then multiply each column. Finally, we add up the results:

	A	B	C	D
% of Others' Choices :	20%	20%	50%	10%
Your Row Choice: B	40	0	90	40
Product of Each column	8	0	45	4

Sum of the bottom row = $8 + 0 + 45 + 4 = 57$. Hence, your payoff for choosing B given the other participants' percentages would be 57 tokens.

Entering Hypotheses

During the experiment you will have a computer interface to calculate hypothetical payoffs.

To demonstrate this, enter 20, 20, 50 and 10 in the four white boxes in the bottom labeled A, B, C, D and click on CALCULATOR.

Four numbers now appear to the right of the matrix under the word PAYOFF, corresponding to your payoffs for each row exactly as we computed.

Suppose the actual choice percentages were as in this example but suppose your hypothesis was that all other participants would choose B. Hence, you entered (0, 100, 0, 0) as your hypothesis. Do so and click on CALCULATOR.

Notice that the computed payoffs indicate that choice A would give you the largest payoff (i.e. 100). In reality, entering this hypothesis cannot change anyone else's ACTUAL choices.

Therefore, given the actual choices of everyone else your payoff from choosing A would be ONLY 53, not 100.

*The point is that **the more your hypothesis differs from the actual percentage of other participants, the more the computed hypothetical payoffs will differ from the actual token earnings, row by row.***

What are the hypothesis boxes good for?

1. *By entering different hypotheses and calculating hypothetical payoffs to these hypotheses, you can explore how the actual choices (including your own) will affect your token earnings. In other words, you can answer "what if" questions.*
2. *You can enter your best guess about the percentage of others choosing each row and use the computed token earnings to guide your choice.*
3. *Between periods you can enter the actual choice frequencies of others, which you will be given, and use the calculator to verify your token earnings.*

Making a Choice

We will now demonstrate **how you make a choice**. Move the mouse cursor to the row you wish to choose in the yellow matrix and click the left mouse button. The row you clicked on will change color to an orange/pink color indicating your choice.

Make a choice now by clicking on ANY row of the yellow matrix.

Change your choice now by clicking on ANY OTHER row of the matrix.

Notice that it is not necessary for you to do any hypothetical calculations before making a choice.

Summary

To Enter or Change a Hypothesis: click inside a white box under the Matrix. Use the keyboard to enter a number. All hypotheses are in terms of percentages and hence must sum to 100. Caution: The hypothetical payoffs will NOT match your hypothesis UNLESS you click on CALCULATOR.

To Calculate Hypothetical Payoffs: Once the white boxes contain your hypothesis and total 100, click on CALCULATOR.

To Make a Choice: Click on the desired row of the matrix and that row will turn pink indicating your choice. To change your choice, simply click on a different row.

To Review Instructions: click on INSTRUCTIONS. To return to the main screen, click on "QUIT INSTRUCTIONS" and move the mouse a bit.

Warning: If you fail to make a choice for any period, you will earn \$0 in that period and be penalized \$5 in Stage II.

At any time during the experiment, you can display this summary page by clicking the INSTRUCTIONS button, and then QUIT INSTRUCTIONS to return.

Practice Session

*You will now have a 40 second timed practice session. You will have 40 seconds to make choices and practice making hypothetical calculations on the Demo matrix. The clock at the bottom right of your screen will count down from 40 seconds to 0 seconds. A **15-Second** warning will appear when only 15 seconds remains for you to make your decision. Otherwise the screen will look exactly as during the Demo session. You should practice making hypothetical calculations, making choices, and revising choices.*

QUIT the Instructions now. Click on the password box on top of your screen, and enter "555". If you do not have 555 entered yet, please raise your hand. The clock starts counting down immediately after you click the DEMO button. Click on the DEMO button now please.

History Screen and the remainder of the experiment

*After each period you see a History Screen, like the one displayed below. Your choice will still be highlighted, and the percentage choices of all OTHER participants will be listed above the Matrix. To the right of the Matrix is the computed actual payoff for each row using the **actual** choices of the OTHER participants. Note that your payoff as displayed above the Matrix corresponds to the computed payoff to the right of the highlighted row.*

After each period, you will have about 30 seconds during which you can ponder the results and do more hypothetical calculations on this Matrix before proceeding.

In the experiment, you will face 6 different matrices, each repeated for 5 periods. Both the matrix and the period numbers will be displayed on your screen.

HISTORY SCREEN
PERIOD 1

Your choice	Payoff	% Others choosing A	% Others choosing B	% Others choosing C	% Others choosing D	Payoff
A	33.0	25.0	25.0	25.0	25.0	33.0
Your Choice B	80	0	70	12		40.5
C	40	100	65	12		54.25
D	35	25	60	12		33.0

A: B: C: C: Total is: CALCULATOR CLOCK: 0:00

Quiz

Participant Number _____

Make sure you put your participant number on the quiz. Your participant number is located on the very top of your screen. Please read the questions carefully, follow the directions exactly. You must use the Demo screen to answer some of these questions. You have 4 minutes in which to complete this quiz.

1. Suppose 50% of the other people in the room chose B and 50% chose D, what would be your payoff if you chose A?

2. With the same percentages above, which choice would give you the highest possible payoff?

3. With the same percentages above, which choice would give you the lowest possible payoff?

4. If the actual percentages of people's choices are as above but you change your hypothesis, can you earn more money?
