Between-Game Rule Learning in Dissimilar Symmetric Normal-Form Games

by

Ernan Haruvy
University of Texas at Dallas

and

Dale O. Stahl
University of Texas at Austin

July 27, 2010

ABSTRACT. Rule learning posits that decision makers, rather than choosing over actions, choose over behavioral rules with different levels of sophistication. Rules are reinforced over time based on their historically observed payoffs in a given game. Past works on rule learning have shown that when playing a single game over a number of rounds, players can learn to form sophisticated beliefs about others. Here we are interested in learning that occurs between games where the set of actions is not directly comparable from one game to the next. We study a sequence of ten thrice-played dissimilar games. Using experimental data, we find that our Rule Learning model captures the ability of players to learn to reason across games. However, this learning appears different from within-game rule learning as previously documented. The main adjustment in sophistication occurs by switching from non-belief-based strategies to belief-based strategies. The sophistication of the beliefs themselves increases only slightly over time.

Keywords: Learning; Experiments; Transference
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 often on a single repeated game where players receive feedback each period about payoffs and the history of play.\(^1\)

In action-based learning models, players myopically adjust beliefs about other players’ distribution over actions based on the observed past frequency of play. In contrast to action-based learning, Rule Learning (Stahl (1996, 1999, 2000, 2002, 2003) involves player beliefs that allow for others to deviate from past actions. Analogous to iterative elimination of dominated strategies and rationalizability, beliefs have an iterative structure, so a player can learn to switch to higher level rules based on past performance even though the game changes. In past studies, Rule Learning has been shown to improve fit and provide insight in matrix games that were repeated for 15 periods (Stahl (2000), and 12 periods (Stahl, 2003), followed by a second game with a similar horizon. In these studies, players were shown to have evolving beliefs that increase in sophistication over time. A transfer parameter for between-game learning was incorporated and found significant. However, in such designs, transfer of knowledge between games took place only once in the entire experiment.

In contrast to Stahl (2000, 2003), in the present study, transfer between dissimilar games takes place 10 times-- one-third of all the periods to be predicted. This gives us a real empirical test of the ability of alternative learning models to capture learning between games. Moreover, the current design has transfer taking place between 10 games (as opposed to two in Stahl, 2000),

\(^1\)Under 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 (1999) have studied hybrid models that combine reinforcement and belief learning features. See also Anderson, et.al. (2001), Feltovich (2000), Friedman, et.al. (1995), Mookherjee and Sopher (1994, 1997), Nagel (1995), Rapoport,
so the kind of transfer that can take place is much richer. Lastly, the current design has the pairing structure that allows us to corroborate the rule-learning results with pair-wise tests.

By “dissimilar” we mean that there is no re-labeling of the actions that makes the games monotonic transformations of each other (as in Rankin, Van Huyck, and Battalio, 2000). The purpose of this work is to propose approaches to capturing learning in dissimilar games. We pay particular attention to what knowledge is learned from one game to the next and how to best account for such learning. While there have been works that show that humans transfer knowledge from one game to another game (Camerer, Ho and Weigelt, 1998; Stahl, 2000, 2001, 2003; Chong, Camerer and Ho, 2006; Rankin, Van Huyck, and Battalio 2000; Rapoport, Seale and Winter, 2000; Cooper and Kagel, 2003, 2008; Weber, 2003a, 2003b, 2004), the application of the Rule Learning model here allows one to see what is being learned between games as opposed to within games.

We find that the kind of learning that takes place between games, with short horizons for within-game learning, is primarily manifested in an increased propensity to form beliefs rather than in the evolution of beliefs themselves. This is likely due to the short horizon for within-game learning. In previous works on Rule Learning within-games (Stahl, 2000, 2003), within-game learning emerged slowly over about 8 periods in a 15 period run, whereas here, due to the focus on between-game learning, we have only 3 periods to learn within each game.

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 Rule Learning with the experimental data. Section 5 concludes.

2. The Experiment.

We chose a sequence of ten $4 \times 4$ symmetric normal-form games, as displayed in Figure 1. The game payoffs were in tokens, with an exchange rate of $\$1$ per 100 tokens. Six games (3, 4, 5, 8, 9, and 10) are dominance solvable, and all distinguish between the Stahl-Wilson (1995) one-shot Level-1, Level-2 and Nash actions (hereafter SW95 L1*, L2*, and NE actions). The L1* action is the best response to a uniform distribution of actions by other players. The L2* action is the best response to L1*. The NE action is the best response to itself. The asterisk is intended to distinguish these one-shot depths of reasoning from their dynamic rule counterparts (which will be described in Section 3). Such one-shot depths of reasoning have been shown to be empirically supported in the population of human players (SW95).

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. 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. 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 45 second 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 10 seconds to contemplate that information before proceeding to the next period.

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

3. Model

The model we examine is Rule Learning (Stahl, 1996, 1999, 2000). The Rule Learning model is intended to represent how subjects learn to reason about games. This feature cannot be captured in action-based models. Rule Learning was introduced by Stahl (1996) and developed further in Stahl (1999, 2000). Rule Learning entails (1) a specification of behavioral rules, (2) a process for selecting among rules, and (3) a process for updating the likelihood of using these rules.

Behavioral Rules. A behavioral rule maps information about the game and the history of play to a probability distribution over available actions, interpreted as a decision maker’s

2 Since actions do not remain the same from one game to the next, action-based learning cannot account for learning between dissimilar games. However, action-based learning models such as Experience Weighted Attraction (EWA, Camerer and Ho, 1999) can be adapted to learning in a sequence of games by re-labeling the actions as L1*, L2*, NE, and Other. This re-labeling allows for learning predictions about learning over re-labeled actions but still cannot shed light on learning to reason over games, as rule learning does. Since this relabeling approach is unsatisfactory, we do not include it here and refer the interested reader to the earlier working paper version (http://www.utdallas.edu/~eharuvy/papers/LearningTransfer.pdf) of this manuscript for detail and estimates.
probabilistic choice. The basic idea of Rule Learning is that rules which perform well are more likely to be used in the future than worse performers. We denote the set of rules by \( R \). A rule \( \rho \) is an element in the set \( R \).

**Selection Among Rules.** Players select among rules probabilistically. They have propensities to select rules, which are mapped into probabilities, such that higher propensities imply higher probabilities. This is a common feature of most learning models currently in use (e.g., Camerer and Ho, 1999). We let \( \varphi(\rho, t) \) denote the probability of using rule \( \rho \) in period \( t \). Because of the non-negativity restriction on probability measures, it is more convenient to specify the learning dynamics in terms of a transformation of \( \varphi \) that is unrestricted in sign. To this end, we define \( w(\rho, t) \) as the log-propensity to use rule \( \rho \) in period \( t \), such that

\[
\varphi(\rho, t) \equiv \exp\left( w(\rho, t) / \int \exp(w(z, t)dz \right)
\]

(1)

Given a space of behavioral rules \( R \) (described in detail below) and probabilities \( \varphi \), the induced distribution over actions for period \( t \) is

\[
p(t) \equiv \int_R \rho \, d\varphi(\rho, t)
\]

(2)

**The updating process.** The third element of the model is the equation of motion. Rules which perform well are more likely to be used in the future. This is captured by the following dynamic which gives the updated log-propensity after period \( t \):

\[
w(\rho, t^+) = \beta_0 w(\rho, t) + \beta_1 \rho'U'_{\rho}, \text{ for } t \geq 1
\]

(3)

where \( U \) is the payoff matrix, \( \rho' \) is the population distribution of play in period \( t \), \( \beta_0 \) is the inertia parameter, and \( \beta_1 \) scales the expected utility that rule \( \rho \) would have yielded in period \( t \).

The figure below demonstrates how the three aspects of rule-learning come together.
Conceptually, there are infinitely many rules players could use (just about any behavioral model ever published could be thought of as a rule). So the approach here is not to try to specify all possible rules players could use, but rather a limited number of archetypal rules that encompass a larger space of rules via convex combinations of the archetypal rules. Our dynamic archetypal rules are dynamic versions of the one-shot Level-n rules.

We begin with the Level-0 rule which represents a player who has no model of the other players to best-respond to, but instead repeats what he did in the past or imitates what he observes as the population did most recently. Accordingly, the dynamic Level-0 behavior in period \( t \) is classic herd behavior given by

\[
q^t(\theta) \equiv (1-\theta)q^{t-1}(\theta) + \theta p^{t-1}.
\]

Let \( \rho_h \) denote the rule that generates this herd behavior and let \( w_h(t) \) denote the log-propensity to use this herd rule. We assume that \( w_h(t) \) follows the updating process in eq(3). Note that herd behavior is not a belief-based rule. It is the only non-belief-based rule we consider and as such it may be a proxy for other non-belief-based rules.

A Level-1 player believes that all other players are Level-0, and so after period \( t \) the expected utility vector from a Level-1 player’s perspective is
\[ y'_1(\theta) \equiv Uq'(\theta), \]

and the Level-1 rule selects a best-reply, which we denote by \( b(y'_1(\theta)) \).

A Level-2 player believes that all other players are Level-1, and so after period t the expected utility vector from a Level-2 player’s perspective is

\[ y'_2(\theta) \equiv U b(y'_1(\theta)), \]

and the Level-2 rule selects a best-reply, which we denote by \( b(y'_2(\theta)) = b^2(y'_1(\theta)) \).

Note that \( y'_1(\theta) \) and \( y'_2(\theta) \) can be thought of as vectors of evidence for each available action. Similarly, we can define \( y_3(\theta) \equiv U p^{NE} \), where \( p^{NE} \) is the mixed-strategy Nash equilibrium vector, corresponding to a player who believes that all other players are Nash archetypes.\(^3\)

So far we have defined three kinds of evidence based on the Level 1, Level 2 and Nash archetypal rules. Let \( Y(\theta) = \{y_1(\theta), y_2(\theta), y_3(\theta)\} \) denote the matrix of evidence (dimensions are the number of actions × number of archetypes). It is natural to construct other rules by taking a weighted average of this evidence. Let \( \nu_1 \geq 0 \) denote the scalar weight for Level-1 evidence, \( \nu_2 \geq 0 \) for Level-2 evidence, and \( \nu_3 \geq 0 \) for Nash evidence, and let \( \nu \equiv (\nu_1, \nu_2, \nu_3) \). The product \( Y(\theta) \nu \) defines a weighed vector of evidence for a particular rule: namely for a player who believes that a proportion \( \nu_k/(\nu_1+\nu_2+\nu_3) \) of the other players are archetype k given the distributed- lag parameter \( \theta \). Thus, we have a four-dimensional space of evidence-based rules.

---

\(^3\) Since the experiment design entails only games with a unique pure-strategy Nash equilibrium, the issue of multiple Nash equilibria can be ignored. For an approach to equilibrium selection in a similar setting see Haruvy and Stahl (2004).
Each player assesses the weighted evidence with finite precision. For mapping this weighted evidence to actions, we opt for the multinomial logit specification because of its computational advantages when it comes to empirical estimation.\(^4\)

Finally, we allow for uniform trembles by introducing the uniformly random rule. Together with eq(2), the unconditional probability distribution over actions is

\[
\hat{p}(t) \equiv (1 - \varepsilon - \delta_h) \int_\mathbb{R} \rho \, d\varphi(\rho, t) + \delta_h \varphi(\theta) + \varepsilon \bar{p},
\]

(5)

where \(\delta_h = \varphi(\rho_h, t)\). Thus, the base model consists of a four-dimensional space of evidence-based rules, a herd rule, and uniform trembles.

Since this theory is about rules that use game information as input, it can 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,

\[
w(\rho, T+1) = (1 - \tau)w(\rho, 1) + \tau w(\rho, T^+),
\]

(6)

where \(\tau\) is the transference parameter, and \(w(\rho, T^+)\) is the log-propensity that eq(6) would give to rule \(\rho\) if it were to be encountered in period \(T+1\). 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

\(^4\) The sum \((\nu_1 + \nu_2 + \nu_3)\) is equal to the aggregate precision parameter in the logit specification.
\( T+1 \) of the same game. Since \( q^1(\theta) = p^0 \) (the uniform distribution), the herd behavior rule has zero transference, an assumption we will relax later on. The specification of transference extends the model to any number of runs with different games without requiring additional parameters.

The entire Rule Learning framework involves 10 parameters: \( \beta \equiv (\beta_0, \beta_1, \nu, \eta, \nu_1, \nu_2, \nu_3, \Theta, \sigma, \delta_h, \epsilon) \). The first four parameters \( (\nu_1, \nu_2, \nu_3, \Theta) \) define the mean of the participant's initial log-propensity over the evidence-based rules, and \( \sigma \) is the standard deviation of that log-propensity; the next two parameters \( (\delta_h, \epsilon) \) are the initial propensities of the herd and tremble rules respectively; \( \beta_0 \) is an inertia parameter; \( \beta_1 \) is a scaling parameter; and \( \eta \) 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.

Camerer, Ho and Weigelt (1998) suggest that learning transfer (between p-beauty contest games with different p’s) is most prominently manifested in faster speeds of convergence for players with experience with a previous different game\(^5\). We observe a similar pattern here. Figure 2 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 2, 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 the NE action 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 (t=5.84, 24 d.f). The hypothesis that the proportion playing the one-shot L1* action 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 <0.0001 (t=5.39, 24 d.f.).

Figure 3 shows the differences from the first to the second runs in the proportions playing the one-shot L1*, L2*, and NE actions for each pair and each period. The summary of the results in the fourth panel clearly shows the rise of NE play and the decline of L1* play between the two runs. Note that the largest changes in proportions of L1*, L2* and NE occur in period 2, as compared to period 1 and period 3.
The average change in NE play between period 1 and 2 was 0.16 (std. err 0.03) for run 1 and 0.31 (std. error 0.03) for run 2 over 5 games and 5 sessions (N=25 for each of these runs). The average change in NE play between period 2 and 3 was 0.21 (0.03) and 0.14 (0.03) for runs 1 and 2 respectively. A pairwise t-test shows that the rate of change between periods 1 and 2 increased significantly ($t=4.33$, df=24, $p=0.002$) between runs. But the rate of change between periods 2 and 3 increased only marginally ($t=1.97$, df=24, $p=0.061$) between runs. Thus, we conclude that over runs, subjects adjust their choices faster.

We hasten to point out that the manifest increase in NE play does not imply that there is a similar increase in the use of the NE rule or the weight ($\nu_3$) on the NE evidence-vector. In particular, if the herd moves in the right direction, the dynamic Level-1 rule can induce play of the NE action. We will investigate this possibility further after estimating the Rule Learning model.

4.2. Rule Learning Model Estimation

We estimated the 10 parameters of the Rule Learning model using all the data from the five 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. From previous studies (Stahl, 2000, 2001, 2003; Stahl and Haruvy, 2002), we have a strong prior that $\nu_3 = 0$ and $\beta_0 = 1$. Indeed, these parameter values are within the bootstrapped confidence intervals for the 10-parameter model. Imposing our prior decreases the maximized log-likelihood by only 5.96. Since the

---

5Result 4 in Camerer, Ho snd Weigelt (1998) is that experienced players are indistinguishable from inexperienced players in the first round of a new game but converge to equilibrium much faster.
likelihood difference of 5.96 is not significant according to the Bayesian Information Criterion\(^6\), we henceforth focus on the restricted 8-parameter model with \(\nu_3 = 0\) and \(\beta_0 = 1\). The maximum likelihood parameter estimates of the 8-parameter model are given in Table 2. The maximized LL is -3092.12, compared to the entropy\(^7\) of the data of -2768.88 gives a psuedo-R\(^2\) of 0.895. The Pearson Chi-square statistic\(^8\) for the entire data set is 652.43. The Root-Mean Squared Error of the predicted versus actual choice frequencies averaged over all games is 0.0925.

Since the proportion of the Other action is negligible, to compare actual with predictions, we only need to look at the predictions for NE and L1* actions relative to actual choice frequencies. Figure 4 shows the predicted versus actual choice probabilities for each game. The dots represent periods 1-3 in each game and the lines represent the dynamic progression of predicted and actual probabilities of choices of NE and L1* actions in each game. The predictions are t-period-ahead meaning that the prediction for each period is made at the start of the game. The predicted and actual lines are remarkably close suggesting a good model fit.

To gauge the out-of-sample robustness of the estimates, we also estimated the same 8-parameter RL model on games 1-5, and used these parameter estimates to forecast next-period play for games 6-10. The estimated parameters for games 1-5 are very close to the estimated parameters for the 10-game estimation. The LL for the full set of 10 games decreases from the maximized value of -3092.12 to -3117.28, but the RMSE increases only by 0.0033 (from 0.0925

---

\(^6\) The Bayesian Information Criterion score is \(-2 \text{ LL} + (# \text{ of free parameters}) \ln(# \text{ of obs})\). Given any two estimated models, the model with the lower value of BIC is the one to be preferred. With 124 participants and 30 choices each, there are 3720 observations. The BIC scores for the 8 and 10 parameter models are 6250.0 and 6254.5, respectively. Therefore, The BIC favors the 8-parameter model.

\(^7\) This is the likelihood ceiling, computed by plugging in the observed empirical frequencies.

\(^8\) Pearson's chi-square is a test of goodness of fit to establish whether or not an observed frequency distribution differs from a theoretical distribution. The statistic is computed as the sum over periods of the game and possible actions of the squared difference of the observed relative frequency and the predicted relative frequency, divided by the theoretical relative frequency.
to 0.0958). The LL for the last five games decreases from -1367.91 to -1399.93, but the psuedo-$R^2$ decreases only by 0.02 (from 0.878 to 0.858), and the RMSE increases only by 0.0085 (from 0.0959 to 0.1044). Thus, the estimates exhibit good out-of-sample robustness.

Table 1. Parameter Estimates for Rule Learning

<table>
<thead>
<tr>
<th></th>
<th>MLE</th>
<th>2.5%</th>
<th>97.5%</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma$</td>
<td>2.076</td>
<td>1.948</td>
<td>2.230</td>
</tr>
<tr>
<td>$\delta_h$</td>
<td>0.447</td>
<td>0.367</td>
<td>0.524</td>
</tr>
<tr>
<td>$\nu_1$</td>
<td>1.082</td>
<td>1.081</td>
<td>1.082</td>
</tr>
<tr>
<td>$\nu_2$</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>$\theta$</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>0.00656</td>
<td>0.00556</td>
<td>0.01823</td>
</tr>
<tr>
<td>$\tau$</td>
<td>1.000</td>
<td>0.356</td>
<td>1.000</td>
</tr>
<tr>
<td>$\epsilon$</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
</tbody>
</table>

Note that only four of the parameters are determining the fit of the data (the rest are on the boundary): (i) the standard deviation of the initial distribution of propensities ($\sigma$), (ii) the initial probability of the herd rule ($\delta_h$), (iii) the initial precision of the Level-1 rule ($\nu_1$), and (iv) the reinforcement scaling parameter ($\beta_1$). It is also noteworthy that the estimate of the transference parameter ($\tau$) 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 -3151.95. By the likelihood ratio test, twice this difference is distributed Chi-square with 1 degree of freedom, and has a p-value < 10^{-27}. Thus, we can strongly reject the null hypothesis of no transference. Interestingly, without the possibility of transference, rule learning does not help explain the data over simpler models. The behavior in the absence of transference is captured entirely by herding and pure belief learning. This may not be that surprising given the short horizon of the within-game dynamic.

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 -3151.95. Twice this difference is distributed Chi-square with 2 degrees of freedom, and has a p-value < 10^{-26}. 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:

$$q_k(t) \equiv \bar{\nu}_k(t)/[\bar{\nu}_1(t) + \bar{\nu}_2(t) + \bar{\nu}_3(t)]. \quad (10)$$

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, $q(1) = (0.408, 0.110, 0.110)$, and it changes smoothly to $q(30) = (0.392, 0.111, 0.186)$. 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. This modest increase in the weight given to the Nash evidence is not sufficient to explain the major shift in the action choice towards the NE action.

---

9 This is known as the likelihood ratio statistic, which says that twice the likelihood ratio (twice the log-likelihood difference) is distributed chi-square with degrees of freedom equal to the number of restrictions.

10 In the previous test of $\tau = 0$, the maximum likelihood estimate for $\beta_1$ was 0, so the two maximized LL values are the same.
The major dynamic change is a dramatic decrease in the probability of the herd rule – from 0.446 to 0.067, 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 probability to use them increases dramatically. Thus, the major shift towards the NE action is substantially due to the shift away from herd behavior and the resulting cascade induced by the Level-1 rule.

To gauge the importance of the herd rules, we estimated two restricted models. First, we removed the herd rule entirely from the rule space so that performance could never reintroduce herd behavior. Note that the AR1 process described by eq(4) is still part of the Level-1 and Level-2 evidence based rules, but it is not a rule on its own. The LL decreases from -3092.11 to -3131.99, which gives the likelihood ratio statistic (twice the log-likelihood difference) of 79.76. This is above the chi-square critical value (p-value < 0.01). To check if rule-learning is still taking place after removing the herd rule we eliminated rule learning ($\beta_1 = 0$), and re-estimated the model. The LL decreases further from -3131.99 to -3211.52. The likelihood ratio statistic is 159.06 and this is above the chi-square critical value (p-value < 0.01). Hence, rule-learning is taking place with or without the presence of the herd rule.

Herd behavior in the first period of each thrice-played game is specified to be the uniform distribution. It is easy to modify this assumption, by using re-labeling and allowing transference across games analogous to eq(7):

$$q(T+1, \theta) = (1-\lambda)q(1, \theta) + \lambda q(T^*, \theta),$$

(11)

where $\lambda$ is transference parameter for the herd rule. If $\lambda = 0$, there is no transference and re-labeling is superfluous; if $\lambda = 1$, there is 100% transference. Estimating this enhanced Rule Learning model yields a maximized LL of -3085.85, and an estimate of 0.023 for $\lambda$. While the
improvement in LL is statistically significant (p-value = 0.0004) by the likelihood ratio test, the small rate of herd transference indicates that the behavioral effect is almost negligible.

5. Conclusions

In recent years, a growing body of literature (Rankin, Van Huyck, and Battalio 2000\textsuperscript{11}, Rapoport, Seale and Winter, 2000; Cooper and Kagel, 2003, 2008; Weber, 2003a, 2003b, 2004) has introduced settings in which decision makers have experience with a class of non-identical games. When non-identical games are not clear transformations of each other, we call them “dissimilar games.” In the experimental literature, it appears that subjects may learn to reason about dissimilar games and may even increase sophistication over time (e.g., Cooper and Kagel, 2008).\textsuperscript{12} The present framework rigorously demonstrated that subjects indeed learn over dissimilar games and that this learning is substantial.

We investigated how sophistication might change over time and employed rule learning to uncover the learning patterns of between-game learning. We chose the Rule Learning model of Stahl (1996, 1999, 2000) to illustrate this approach. When Rule Learning was used to capture the dynamics, transference of the knowledge learned between games was estimated at 100%. Statistical tests also strongly rejected the null hypothesis of no rule learning.

The model estimation results indicate that the participants in our experiments learned abstract aspects of the games which were transferable to subsequent dissimilar games.

Transference of propensities between games takes the form of a convex combination and reveals

---

\textsuperscript{11} In that paper, games which are monotonic transformations of each other are called “similar games”.

\textsuperscript{12} Increased sophistication with experience has also been shown in learning over similar games. This pattern—first periods seem similar but experienced subjects exhibit faster convergence—has been demonstrated in various settings, including asset markets (Dufwenberg et al. 2005) and prisoner dilemma games (Bereby-Meyer and Roth, 2006).
an increase in “depth of reasoning” as a result of learning. However, the model estimates indicate that while decision makers increase their sophistication over time, they do so largely by employing more belief-based rules in addition to increasing belief sophistication over time.

A more visual demonstration of the same effect comes from pair-wise comparisons of permuted games within the sequence of dissimilar games, demonstrated by Figures 2 and 3. The simple graphical evidence shows that over time subjects reduce L1* actions in favor of NE actions. This is indicative of substantial reduction in herding, accompanied by some increased belief sophistication and transference between games. However, another clear pattern was the steeper adaptive movement between the first period of a game to the second period of the game in the last run of 5 games, relative to the first run of 5 games. Given the reversion to L1* actions in each period 1 combined with an estimated 100% transference, this can only be explained by greater use of belief-based adjustment over time.
References


Figure 1. The Games.\textsuperscript{13}

<table>
<thead>
<tr>
<th>Game 1</th>
<th>Game 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>A 60 15 100 90</td>
<td>A 35 0 15 90 L2*</td>
</tr>
<tr>
<td>B 80 70 80 0</td>
<td>B 5 70 10 65 DOM</td>
</tr>
<tr>
<td>C 90 15 35 0</td>
<td>C 80 0 70 80 NE</td>
</tr>
<tr>
<td>D 65 10 5 70</td>
<td>D 100 90 15 60 L1*</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Game 2</th>
<th>Game 7</th>
</tr>
</thead>
<tbody>
<tr>
<td>A 10 95 20 0</td>
<td>A 45 55 100 90 L1*</td>
</tr>
<tr>
<td>B 90 45 100 55</td>
<td>B 70 95 15 60 NE</td>
</tr>
<tr>
<td>C 95 5 40 15</td>
<td>C 5 15 40 95 L3*</td>
</tr>
<tr>
<td>D 60 15 70 95</td>
<td>D 95 0 20 10 L2*</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Game 3</th>
<th>Game 8</th>
</tr>
</thead>
<tbody>
<tr>
<td>A 25 80 95 15</td>
<td>A 0 15 70 5 DOM</td>
</tr>
<tr>
<td>B 15 80 100 50</td>
<td>B 100 80 80 15 L1*</td>
</tr>
<tr>
<td>C 15 90 70 50</td>
<td>C 50 90 75 15 L2*</td>
</tr>
<tr>
<td>D 5 15 70 0</td>
<td>D 15 80 95 25 NE</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Game 4</th>
<th>Game 9</th>
</tr>
</thead>
<tbody>
<tr>
<td>A 5 55 95 70 L1*</td>
<td>A 80 80 30 15 NE</td>
</tr>
<tr>
<td>B 30 80 10 80 NE</td>
<td>B 10 55 100 10 L2*</td>
</tr>
<tr>
<td>C 0 10 90 50 DOM</td>
<td>C 55 70 5 95 L1*</td>
</tr>
<tr>
<td>D 100 10 15 55 L2*</td>
<td>D 10 50 0 90 DOM</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Game 5</th>
<th>Game 10</th>
</tr>
</thead>
<tbody>
<tr>
<td>A 60 20 0 40 DOM</td>
<td>A 25 30 10 70 L2*</td>
</tr>
<tr>
<td>B 100 65 30 25 L1*</td>
<td>B 70 40 20 50 NE</td>
</tr>
<tr>
<td>C 20 50 40 70 NE</td>
<td>C 40 0 60 20 DOM</td>
</tr>
<tr>
<td>D 10 70 30 25 L2*</td>
<td>D 25 30 100 65 L1*</td>
</tr>
</tbody>
</table>

\textsuperscript{13} “L_n*” denotes the one-shot Level-n action, NE denotes Nash equilibrium action, and DOM denotes a dominated action.
Figure 2. 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.\textsuperscript{14}

\[ y = \frac{\text{Proportion choosing Nash equilibrium in period 2} - \text{Proportion choosing Nash equilibrium in period 1}}{\text{Proportion choosing Nash equilibrium in period 3} - \text{Proportion choosing Nash equilibrium in period 1}}. \]

\textsuperscript{14} y = (Proportion choosing Nash equilibrium in period 2 - Proportion choosing Nash equilibrium in period 1) / (Proportion choosing Nash equilibrium in period 3 - Proportion choosing Nash equilibrium in period 1).
Figure 3. Differences from 1st to 2nd run in proportions playing Level-1(L1), Level-2 (L2), and Nash equilibrium (NE) actions by pair and period.
Figure 4. Predicted versus actual choice probabilities for each game.
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. Each matrix will differ only by the 16 numbers.

How token earnings are computed

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

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>30</td>
<td>100</td>
<td>50</td>
<td>20</td>
</tr>
<tr>
<td>B</td>
<td>40</td>
<td>0</td>
<td>90</td>
<td>40</td>
</tr>
<tr>
<td>C</td>
<td>50</td>
<td>75</td>
<td>20</td>
<td>30</td>
</tr>
<tr>
<td>D</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
</tr>
</tbody>
</table>

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:

<table>
<thead>
<tr>
<th>% of Others’ Choices :</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>Your Row Choice: A</td>
<td>20%</td>
<td>20%</td>
<td>50%</td>
<td>10%</td>
</tr>
<tr>
<td>Product of Each column</td>
<td>6</td>
<td>20</td>
<td>25</td>
<td>2</td>
</tr>
</tbody>
</table>

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:

<table>
<thead>
<tr>
<th>% of Others’ Choices</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>Your Row Choice: B</td>
<td>40</td>
<td>0</td>
<td>90</td>
<td>40</td>
</tr>
<tr>
<td>Product of Each column</td>
<td>8</td>
<td>0</td>
<td>45</td>
<td>4</td>
</tr>
</tbody>
</table>

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 total earnings.

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

Quiz
We are now passing out a Quiz. 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. Raise your hand when you are done, so we can come by and check your answers.

Participant Number ______________
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?

________________________

Practice Session

You will now have a 45 second timed practice session. You will have 45 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 45 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

In the experiment, you will face 10 different matrices, each repeated for 3 periods. In each period you will have 45 seconds to make your choice.

After periods 1 and 2 of each matrix an interim screen will flash by for one second, and then you will see the Choice screen for the next period. In addition to the matrix, the Choice screen will display the results from the previous period. Please look up at the overhead to see an example. Just above the matrix is a line that lists your previous choice, your payoff from the previous period, and the percentages of others choosing A, B, C and D. The payoff numbers displayed will use these actual choices of others from the previous period. Once you enter your own numbers in the hypothesis boxes and click on Calculator, the payoff numbers will then reflect the hypothetical payoffs.

After period 3, you will see a History screen that displays the results from period 3. At the end of 10 seconds, the screen will change to the Choice screen for period 1 of the next matrix. The screens will be displayed automatically until you have made all 30 decisions.