STUDYING LEARNING IN GAMES USING EYE-TRACKING

Daniel T. Knoepfle
Stanford University

Joseph Tao-yi Wang
National Taiwan University

Colin F. Camerer
California Institute of Technology

Abstract

We report results from an exploratory study using eye-tracking recording of information acquisition by players in a game theoretic learning paradigm. Eye-tracking is used to observe what information subjects look at in 4x4 normal-form games; the eye-tracking results favor sophisticated learning over adaptive learning and lend support to anticipatory or sophisticated models of learning in which subjects look at payoffs of other players to anticipate what those players might do. The decision data, however, are poorly fit by the simple anticipatory models we examine. We discuss how eye-tracking studies of information acquisition can fit into research agenda seeking to understand complex strategic behavior and consider methodological issues that must be addressed in order to maximize their potential.

10/13/08. This paper was prepared for an invited session at the European Economics Association, 2008. We thank Andrew Caplin, Vince Crawford, Rosemarie Nagel, and an anonymous referee for helpful comments. This research was supported by NSF-HSD, HFSP, National Science Council of Taiwan (NSC 96-2415-H-002-034-MY2), and the Gordon and Betty Moore Foundation (grants to CFC).

1. Introduction

Many theories have been proposed about how players learn in games, leading to a large experimental literature comparing theories econometrically (see Camerer, 2003, chapter 6). Here, we explore using eye-tracking recording of information acquisition to inform this empirical debate. The maintained hypothesis underlying this approach is that theories can be taken as algorithms that use specific information about historical actions and payoffs in a precise way to guide choices. Furthermore, the extent to which people look at the relevant information is likely to correlate with how that information influences choices. Eye-tracking data can be useful because it is difficult to identify learning rules precisely from decision data alone, especially once we consider heterogeneity across subjects (or even across time, within subjects, as in "rule learning" models). In an observational study of information acquisition, the goal is to gain insights about behavior that would be difficult to establish using choices alone (and which is inefficient to test only by exogenously varying what information is available).

Eye-tracking and mouse-tracking (MouseLab) methods for measuring information acquisition have been used previously to study various topics such as backward and forward induction (Johnson et al., 2002) and to estimate decision rules used in normal-form games (Costa-Gomes et al, 2001) and two person p-beauty contest games (Costa-Gomes and Crawford, 2006).

The core of our analysis consists of (a) measures of how well models predict the actual choices of players and (b) measures of how often players look at information relevant to executing a particular learning rule.

An ideal outcome would be that one theory is supported by both decision and information lookup choices. Unfortunately, that is not the case. Instead, we found that in predicting choices, adaptive models in which people learn by generalized reinforcement (EWA and self-tuning EWA) predict more accurately than simple "sophisticated" (or "anticipatory learning") models in which agents anticipate that others are learning. However, the lookup data suggest that players often look at payoffs as required by sophisticated models, about equally as often as they look at adaptive-model payoffs. The fact that behavioral fits favor adaptive models but lookups favor sophisticated ones suggest there are some models of sophisticated learning (in which lookups overlap with those in the models that do not fit well here) which could fit both lookup and choice data well. We leave development of such models to future research.

Details and further analyses are reported in our online working paper, Knoepfle, Wang, and Camerer (2008), denoted KWC.

2. Design

2.1 Experiment Structure

Subjects play four asymmetric non-zero-sum two-player 4x4 normal form games (see KWC A1 for the game matrices and A2 for design details). Each of the four games was played 10 times in a random-matching protocol with feedback. The order of games was fixed. Experiments were conducted in groups of six; in each group, two subjects were eye-tracked. Periods ended when all subjects had entered their strategy choices.

2.2 Models

We consider five models of learning: reinforcement (Re), experience-weighted attraction learning (EWA), self-tuning EWA (stEWA), and anticipatory response level one (C1) and two (C2).

EWA is a general adaptive model that nests types of reinforcement and belief-based learning as special cases. In EWA, strategies have numerical attractions which begin as subjective initial attractions $A_i^j(0)$, with initial weight N(0), and are updated each period after receiving choice feedback, according to

$$(1) \qquad A_i^j(t+1)= \varphi N(t)A_i^j(t)/N(t+1) + [\delta+(1-\delta)I(s_i^j,s_i(t))]\pi_i(s_i^j,s_{-i}(t))]/N(t+1)$$

with $N(t+1)=\varphi(1-\kappa)N(t)+1$. Attractions are then mapped into strategy choice probabilities using a logit form $P_i^j(t+1)=\exp(\lambda A_i^j(t))/(\sum_k \exp(\lambda A_i^k(t)))$. The EWA parameters δ, φ, and κ correspond, respectively, to the weight placed on the payoffs one would have received had they chosen a different strategy, the weight of old attractions, and the degree to which attractions cumulate rather than average. Of these parameters, only δ has implications for information search.

For the adaptive models, we assume that subjects must look up the information required to update their attractions each period but have perfect recall of their attractions in the previous period. If we constrain $\delta=0$, we get a reinforcement-type model (Re) in which the only relevant information is the payoff received; reinforcement models have the own payoff matrix cell $\pi_i(s_i(t-1),s_{-i}(t-1))$ as their lookup area. If $\delta>0$, as in unconstrained EWA, the payoffs that would have been received are also relevant information; thus, EWA has as its lookup area the own payoff matrix column $\pi_i(\cdot,s_{-i}(t-1))$ corresponding to $s_{-i}(t-1)$, the opponent's strategy choice in the previous period.

We also consider a "self-tuning" form of EWA (stEWA) proposed by Ho, Camerer, and Chong (2007), in which two psychologically uninteresting parameters are removed from EWA and the remaining free parameters are replaced by functions of experience. The reinforcement weight δ on un-chosen strategies is replaced by 1 for strategies which are weakly better responses than the chosen strategy (i.e., $\pi_i(s_i^j,s_{-i}(t-1)) \geq \pi_i(s_i(t-1),s_{-i}(t-1))$) and 0 for worse responses. In addition, the weight on old experience φ is specified as a function of observed choices. The lookup area for stEWA is the set of cells in the own payoff matrix column corresponding to $s_{-i}(t)$ with positive weight.

The two "sophisticated" models C1 and C2 are myopic anticipatory models that predict opponents' play. Let $BR_i(s)$ denote subject i's best-response to his opponent's strategy s.[1] Consider the familiar Cournot rule in which subjects simply best-respond to their opponent's previous choice; that is, they choose $BR_i(s_{-i}(t-1))$. In C1, a player assumes his opponent is using a Cournot rule and best-responds accordingly, choosing $BR_i(BR_{-i}(s_i(t-1)))$. The relevant information for determining this choice is $\pi_{-i}(s_i(t-1),\cdot)$ and $\pi_i(\cdot,BR_{-i}(s_i(t-1)))$.

The C2 model is one level above the C1 model: in C2, a player assumes his opponent is using the C1 rule and best-responds accordingly. A C2 player chooses $BR_i(BR_{-i}(BR_i(s_{-i}(t-1))))$. The relevant information for determining this choice is $\pi_i(\cdot,s_{-i}(t-1))$, $\pi_{-i}(BR_i(s_{-i}(t-1)),\cdot)$, and $\pi_i(\cdot,BR_{-i}(BR_i(s_{-i}(t-1))))$.

2.3. Screen Display

[FIGURE 1 ABOUT HERE]

A mock-up of an experimental task screen is shown in Figure 1 (specifically, for a row player in period 5 of game 3). Some key design features can be seen. First, following Costa-Gomes et al. (2001), we separate the subject's payoffs and his opponent's payoffs into two matrices: the left matrix contains the subject's own payoffs and the right matrix contains his opponent's payoffs. Rows in the matrices correspond to the subject's own strategies, and columns correspond to his opponent's strategies. Second, we display the history of the subject's choices, his opponents' choices, and his own received payoffs at the bottom of the screen. Another screen (not shown here) appears at the conclusion of each period, displaying the subject's choice, his opponent's choice, and his received payoff for the period; we discuss a possible bias related to this screen in KWC A6.2.2.

2.3. Model-Relevant Information and Lookup Areas

At this point, it is worthwhile to discuss and give examples of the model lookup areas. The model lookup areas are areas of the payoff matrices corresponding to the information relevant to determining the model choice, defined with respect to the choices of a subject and his opponent in the previous period. Consider the game and its on-screen display in Figure 1, and suppose that in the previous period the player chose strategy 1 and his opponent chose strategy 1. The reinforcement cell (the player's experienced payoff) is row 1, column 1 in his own payoff matrix; this is Re's lookup area. His corresponding payoffs for un-chosen strategies are in the other cells of column 1 in his own payoff matrix; EWA's lookup area is this entire column. His strategies 1 and 2

[1] For simplicity, assume best responses are always unique.

are weakly better responses to his opponent's previous choice than his own previous choice; thus, stEWA's lookup area are the cells in row 1 and 2 of column 1 in his own payoff matrix. Observe that Re and stEWA's lookup areas are nested within EWA's; this is always the case.

Recall that C1 plays $BR_i(BR_{-i}(s_i(t-1)))$. Thus, C1 must first find his opponent's best response to his own previous strategy choice. This entails finding the maximum payoff in the row of his opponent's payoff matrix corresponding to his own previous strategy choice, strategy 1. Thus, C1's lookup area includes row 1 of his opponent's payoff matrix. Here, his opponent's best response is strategy 4. A C1 player must then find his best response to his opponent's strategy 4; to do so, he looks at the corresponding column of his own payoff matrix (column 4) and finds the maximum payoff; the strategy yielding this maximum is his strategy choice. Thus, C1's lookup area also includes column 4 of his own payoff matrix.

The C2 lookup area comes from a similar process. Recall that C2 plays $BR_i(BR_{-i}(BR_i(s_{-i}(t-1))))$; thus, a C2 player must first find his best response to his opponent's previous strategy choice by looking for the maximum payoff in his corresponding own payoff matrix column, column 1. Therefore, C2's lookup area includes column 1 of his own payoff matrix. The maximum payoff in this column corresponds to strategy 2; that is, $BR_i(s_{-i}(t-1)) = BR_i(1) = 2$. A C2 player must then find his opponent's best response to this strategy, $BR_{-i}(BR_i(s_{-i}(t-1))) = BR_{-i}(2)$; this is accomplished by looking for his opponent's maximum payoff in row 2 of his opponent's payoff matrix. Thus, C2's lookup area also includes row 2 of his opponent's payoff matrix. Here, the maximum payoff corresponds to his opponent's strategy 4; that is, $BR_{-i}(2) = 4$. Finally, a C2 player finds his best response to this strategy; that is, $BR_i(BR_{-i}(BR_i(s_{-i}(t-1)))) = BR_i(4)$. He does so by looking for his maximum payoff in the corresponding column of his own payoff matrix, column 4; this maximum, strategy 1, is his choice. Thus, C2's lookup area includes column 4 of his own payoff matrix.

3. Results

Due to practical and technical reasons, some subjects were excluded prior to analysis; we use decision data from 44 subjects and lookup data from 12 subjects. In some cases, players choose the same strategy across all ten periods of a game. As learning is the topic of interest, all of the results we report herein exclude these games and use only the "learning games" in which players do not make the same choice every time (see KWC A5.3 and A6.3 for results computed using all games).

3.1 Behavior

[FIGURE 2 ABOUT HERE]

Choices in the games appear to converge to Nash equilibrium relatively rapidly; the frequency of Nash play tends to increase over the ten periods of each game and averages 79.7% in the final period (see KWC 5.1).

We fit the learning models to observed behavior using maximum likelihood. As a non-learning benchmark, we also estimate a model in which players are predicted to choose their Nash strategy. The adaptive models have likelihoods specified by the logit-form probability map; for the Ck and Nash-type models, we estimate the likelihood maximizing p such that players choose the predicted strategy with probability p and randomize uniformly over the other strategies. To control for different numbers of parameters, model parameters are fit on periods 1-7 in each game and geometric mean likelihoods of observed choices in the hold-out periods 8-10 are reported. Parameters are estimated at the subject level (see KWC A5.2 for details on the behavioral estimation).

Out of sample goodness-of-fit measures are reported in Figure 2 separately for eye-tracked subjects and for all subjects. The best models put predicted probability of about 0.5 on the actual choices, well above a random guessing benchmark of 0.25. The results are similar across the four games (see KWC A5.3). Eye-tracked subject fits are worse for some models but not significantly so (and we have no sensible explanation for the differences). Pooling tracked and untracked subjects, within subject signed rank tests between models find that EWA and stEWA significantly outperform C1, C2 and Nash at $p<0.01$ (all p-values two-tailed, uncorrected for multiple comparisons). Re significantly outperforms C1 ($p=0.024$) but is not significantly better than C2 or the Nash benchmark ($p=0.067$ and $p=0.196$, respectively). EWA and stEWA are better than Re at $p=0.055$ and $p=0.065$, respectively. The C1 and C2 models are clearly the worst-fitting overall, and are strongly so within-subjects.

3.2 Eye-tracking

Eye-tracking produces a huge amount of data. With this increase in data volume comes an increase in noise and uncertainty in interpretation. We emphasize that our results are conditional on a set of assumptions and that the numbers we report here derive from data which has been preprocessed, filtered, and transformed in a number of ways. We find that our qualitative results are generally robust to different preprocessing and analysis choices; discussion of these preprocessing methods and analysis choices can be found in the online appendix (KWC A6).

The eye-trackers recorded gaze location and pupil dilation samples at 250 Hz. These gaze location samples are processed into fixations and saccades (which are rapid movements between fixations). Fixation locations are compared with screen areas corresponding to cells in the payoff matrices and other on-screen items; we consider a fixation in a payoff matrix cell's area to be a "lookup" of that cell's information. Since we have limited prior knowledge regarding the relationship between gaze and information acquisition, we take all fixations within the payoff matrices' cells as representing information lookups. A notable analysis choice is our use of fixation counts rather than durations as the atoms of analysis; see KWC A6.4.5 for discussion of this issue. In addition, as our on-screen display did not visually separate different history components, data on history lookups cannot help identify learning models without undesirably strong assumptions. Furthermore, history lookups represent a small fraction of the observed lookups. We consider only payoff matrix lookups here, and discuss observed history lookup behavior in KWC A6.5.

The median (mean) number of payoff matrix lookups (henceforth, "lookups") per trial was 32 (42.24). There is a sharp decline in the average number of lookups from the first period to the second and a more gradual decline across the following nine periods of each game (see KWC A6.3).

Our first result is very simple but important: the mean share of payoff matrix lookups corresponding to the opponent's payoff matrix is 46.1%; while there is subject-wise heterogeneity, eight of the twelve subjects have shares above 40% and all have shares above 25%. This observation is in stark contrast to the class of adaptive learning models, in which information about the opponent's payoffs is *never* relevant. Unless we assume an unrealistic level of noise, we must conclude that many subjects are deliberately acquiring information about their opponents' strategic situation. This result demonstrates the power of observing information acquisition: with minimal assumptions, we establish subjects' interest in opponents' payoff structures, suggesting that adaptive learning models ignore important determinants of choice.

To move from statements about classes of models to statements about specific models, we make additional assumptions about memory and information relevance. We measure the degree to which lookups conform to theories by examining lookup rates in screen areas corresponding to the information which is relevant in *updating* underlying attractions (or beliefs) in each particular period. That is, we are implicitly assuming subjects recall previous-period attractions and only look up the information needed to update these attractions. (Note that we did display choice and own-payoff history at the bottom of the screen, but subjects did not look at it frequently.)

Under this memory assumption, the theories predict that subjects will look at a screen area containing certain target cells. We use a linear measure to score how well lookups correspond to model predictions. Let x equal the "hit rate", the proportion of lookups in a period that fall in the target cells, and let a equal the proportional area of the target cells. The linear measure (LM) equals x-a, the proportional hit rate minus the proportional area. Under uniform random looking, the LM has expectation zero. The LM has a number of desirable properties (cf. Selten, 1991).[2] Areas with positive LM scores are looked at more than average. Thus, the skeptical can take our LM scores for the various models as mere indications of what information subjects look at more than would be predicted by uniform looking.

[FIGURE 3 ABOUT HERE]

Figure 3 shows the mean linear measure scores for the set of learning models. The highest score is for the sophisticated learning rule C2, followed by its sibling C1, then by EWA and stEWA, and finally by Re. All of the scores are well above a uniform-looking benchmark (zero LM score).

Signed rank tests using subject-wise mean linear measure scores (learning games only) find that EWA and stEWA have higher scores than Re at p=0.09 and p=0.001, respectively; there is no significant difference between stEWA and EWA (p=0.91). C1 is insignificantly superior to Re, EWA, and stEWA, with p=0.07, p=0.13, and p=0.13, respectively. In contrast, C2 is significantly superior to these adaptive models, with p=0.001, p=0.016, and p=0.002, respectively. Furthermore, C2 is superior to C1 at the marginally significant level p=0.052.

Taking the learning models as cognitive algorithms, the adaptive models do not predict any particular order in which their relevant payoffs are looked up. In contrast, the Ck models do suggest ordered stages of lookups. In the Ck models, the player must find certain best responses in order to know what other information to look at. For example, in C1, the player must first find $BR_{-i}(s_i(t-1))$ in "stage 1" in order to determine $BR_i(BR_{-i}(s_i(t-1)))$ in "stage 2". We impose a simple order restriction requiring at least one lookup in a stage's lookup area before lookups in the next stage's area count as hits, and adjust the model areas accordingly to maintain zero expected LM score under uniform looking (see KWC 6.2.1 for additional details). Figure 3 shows the resulting LM scores for the restricted Ck models (labeled C1r and C2r). Adding the order restriction, the lookup scores drop sharply for C1r and slightly for C2r, so that C1r now scores much worse on lookups but C2r still scores better than the adaptive models. Signed rank tests find that applying this order restriction significantly decreases the corresponding scores for C1 (p=0.007) and C2 (p=0.012). Restricted C2 remains superior to the adaptive models Re, EWA, and stEWA with p=0.003, p=0.077, and p=0.034, respectively; restricted C1 is

[2] An oft-suggested alternative measure is the ratio measure x/a. This ratio averages around 2 for Re and stEWA and 1.4 for all other theories. Selten (1991) notes that the ratio measure favors point theories (given a set of equal-sized areas, the ratio measure is maximized by a model including *only* the area(s) with the highest hit-rate). We consider it a poor summary of lookups (see KWC A6.2 for discussion).

insignificantly superior to these models with p=0.110, p=0.266, and p=0.301. If C1 and C2 are accurate models of subjects' cognitive processes, the simple lookup order restriction discussed above should not decrease their scores. That the scores significantly decrease suggests that C1 and C2 include relevant information lookups but are inaccurate descriptions of true cognitive processes.

4. Discussion

We think our qualitative results imply that learning has a large component of sophistication (as the lookups suggest) but the behavioral fits make it clear we have not yet found the appropriate model. Our fundamental conclusion is that researchers should work towards specifying sophisticated learning rules that can fit both choices and lookups; a model that does so is likely to be a good approximation of true cognitive processes.

It is important to note that different learning processes can have identical lookup areas. Given that C1 and C2 perform worse than the adaptive models in predicting behavior, one wonders why they have superior lookup scores. The true learning process may have a similar or identical set of relevant information but map this information into choices differently. We explore this possibility in KWC A5, varying the behavioral specifications corresponding to the Ck models.

Future studies could improve on our design in myriad ways. Convergence to equilibrium is rapid here, so games with slower convergence might provide richer data. Introducing variation into the location and orientation of on-screen elements could help us gain an understanding of biases in looking patterns and permit more sophisticated analyses of eye-tracking data.

Fundamental changes in the observational paradigm could help us to differentiate genuine information lookups from noise fixations with confidence. Eye-tracking is very noisy. While mouse-tracking designs have much less noise, they introduce exogenous costs for information acquisition that may bias subjects' lookups and behavior. As a complement to eye-tracking designs, however, they offer a range of alternative balances between noise and bias. Future studies could use a hybrid design combining mouse-tracking features with eye-tracking recording. For instance, the payoff matrices' structure could be freely visible but subjects might be required to press and hold a key in order to view the numerical payoffs. If this were combined with a strategically irrelevant cost rate, we could infer that subjects' fixations while the key is depressed are genuine information lookups. With our current design, we get fixations, which we take to be lookups, so long as subjects have their eyes open. This does not invalidate our results, but does add noise and constrain what we might hope to accomplish with the data. In eye-tracking studies, we accept an increase in noise and uncertainty in interpretation in our pursuit of the ideal of naturalistic observation; hybrid designs could lessen these tradeoffs.

References

Camerer, Colin (2003). Behavioral game theory: Experiments on strategic interaction. Princeton, Princeton University Press.

Camerer, Colin and Teck Ho (1999). Experience-weighted attraction learning in normal-form games. *Econometrica*, 67, 827-74.

Costa-Gomes, Miguel A. and Vincent P. Crawford (2006). Cognition and Behavior in Two-Person Guessing Games: An Experimental Study. *American Economic Review*, 96, 1737-1768.

Costa-Gomes, Miguel A., Vincent P. Crawford and Bruno Broseta (2001). Cognition and Behavior in Normal-Form Games: An Experimental Study. *Econometrica*, 2001, 69, 1193-1235.

Ho, Teck, Colin Camerer and Juin-Kuan Chong (2007). Self-tuning experience weighted attraction learning in games. *Journal of Economic Theory,* 133, 177-198.

Johnson, Eric J., Colin Camerer, Sankar Sen, and Talia Rymon (2002). Detecting Failures of Backward Induction: Monitoring Information Search in Sequential Bargaining. *Journal of Economic Theory*, 104, 16-47.

Knoepfle, Daniel, Joseph Tao-yi Wang, Colin Camerer (in preparation). Using eyetracking to study learning in games. SSRN working paper.

Selten, Reinhard (1991), Properties of a measure of predictive success, *Mathematical Social Sciences,* 21, 153-167.

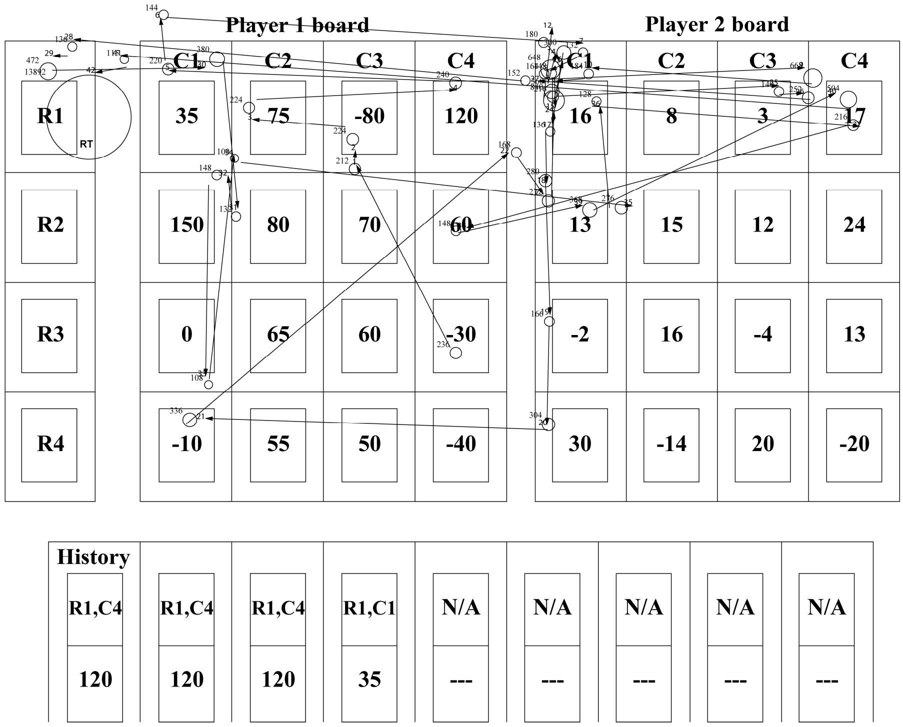

Figure 1. An example of the experimental task display.

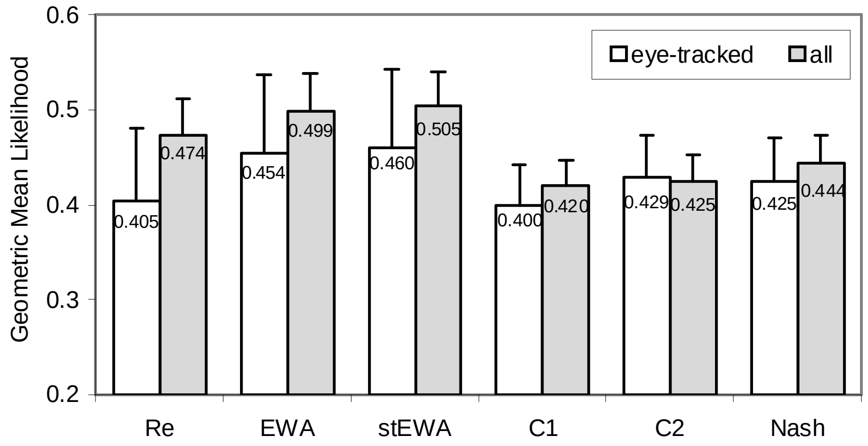

Figure 2. Geometric mean predicted probabilities of actual choices for various learning models. Models fit on trials 1-7; results shown are out-of-sample cross-validation predictions. Standard errors computed across subjects.

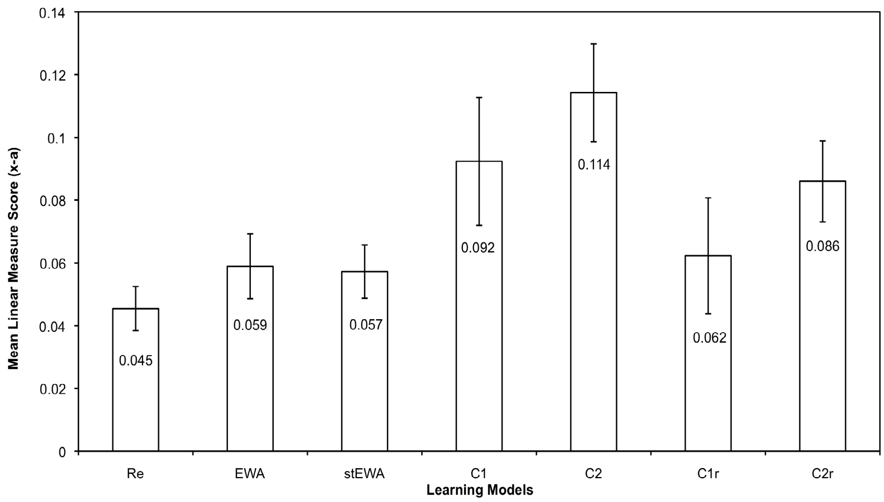

Figure 3. Mean Linear Measure Scores for Different Learning Models

Appendices

Daniel T. Knoepfle, Joseph Tao-yi Wang, and Colin F. Camerer

September 1, 2009

Links to Main Text

References in the main text to material in these appendices are listed below in the order they appear in the paper:

- "Subjects play four asymmetric non-zero-sum two-player 4x4 normal form games (see KWC A1 for the game matrices and A2 for design details)." **See page 3 for section A1 and page 5 for section A2.**
- "Another screen (not shown here) appears at the conclusion of each period, displaying the subject's choice, his opponent's choice, and his received payoff for the period; we discuss a possible bias related to this screen in KWC A6.2.2." **See page 26 for section A6.2.2.**
- "As learning is the topic of interest, all of the results we report herein exclude these games and use only the "learning games" in which players do not make the same choice every time (see KWC A5.3 and A6.3 for results computed using all games)." **See page 17 for section A5.3 and page 28 for A6.3.**
- "Choices in the games appear to converge to Nash equilibrium relatively rapidly; the frequency of Nash play tends to increase over the ten periods of each game and averages 79.7% in the final period (see KWC 5.1)." **See page 11 for section A5.1.**
- "Parameters are estimated at the subject level (see KWC A5.2 for details on the behavioral estimation)." **See page 15 for section A5.2.**
- "Out of sample goodness-of-fit measures are reported in Figure 2 separately for eye-tracked subjects and for all subjects. The best models put predicted probability of about 0.5 on the actual choices, well above a random guessing benchmark of 0.25. The results are similar across the four games (see KWC A5.3)." **See page 17 for section A5.3.**
- "We find that our qualitative results are generally robust to different pre-processing and analysis choices; discussion of these pre-processing methods and analysis choices can be found in the online appendix (KWC A6)." **See page 23 for section A6.**

- "A notable analysis choice is our use of fixation counts rather than durations as the atoms of analysis; see KWC A6.4.5 for discussion of this issue." **See page 44 for section A6.4.5.**

- "We consider only payoff matrix lookups here, and discuss observed history lookup behavior in KWC A6.5." **See page 46 for section A6.5.**

- "There is a sharp decline in the average number of lookups from the first period to the second and a more gradual decline across the following nine periods of each game (see KWC A6.3)." **See page 28 for section A6.3.**

- "An often-suggested alternative measure is the ratio measure x/a. This ratio averages around 2 for Re and stEWA and 1.4 for all other theories. Selten (1991) notes that the ratio measure favors point theories (given a set of equal-sized areas, the ratio measure is maximized by a model including only the area(s) with the highest hit-rate). We consider it a poor summary of lookups (see KWC A6.2 for discussion)." **See page 25 for section A6.2.**

- "We impose a simple order restriction requiring at least one lookup in a stage's lookup area before lookups in the next stage's area count as hits, and adjust the model areas accordingly to maintain a zero expected LM score under uniform looking (see KWC 6.2.1 for additional details)." **See page 25 for section A6.2.1.**

- "Given that C1 and C2 perform worse than the adaptive models in predicting behavior, one wonders why they have superior lookup scores. The true learning process may have a similar or identical set of relevant information but map this information into choices differently. We explore this possibility in KWC A5, varying the behavioral specifications corresponding to the Ck models." **See page 11 for section A5.**

A1 Games

The games used in the experiment are asymmetric two-player games and are not zero-sum. They are shown in Table 1; Nash equilibria are bolded and bracketed, and the Camerer-Ho-Chong cognitive hierarcy (CH) predicted frequencies are given on the margins (computed with $\tau = 1.5$). All four games have unique pure strategy Nash equilibria. Game 1 has a dominant strategy for the column player (column 2) and was designed as a warm-up. In all four games, the Nash equilibria can be found by iterated deletion of dominated strategies.

All four games are 4x4 normal-form games. We chose this size because it represented a reasonable tradeoff between the desires for type separation and complexity and the need to work within display constraints. Using smaller games with pure strategy equilibria would have reduced the type-separation power and would likely have produced faster Nash-convergence. We would have liked to have used games with a larger number of strategies; however, increasing the number of strategies increases the number of payoff boxes that must be fit onto the screen (see Figures 1 and 2) and necessarily decreases the display separation between different pieces of information, increasing the error in associating gaze with information lookups.

The games were designed to have CH predictions that differ to some extent from the Nash equilibria. In each game, the CH model's predicted frequency for a non-Nash strategy is at least 0.391 for at least one of the players. As the intent was to study learning behavior in games, and since CH often provides a good indication of one-shot play or initial play in repeated games, we supposed that players of these games would move from the non-Nash CH strategies to the Nash strategy as they learned over time. Note that the row and column players had different payoff multipliers mapping game payoffs to cash payoffs and that these multipliers were not common knowledge; this feature was included to prevent direct summation of the payoffs and thereby help disrupt altruistic or joint-maximizing behavior. In our analysis, we ignore altruism and social preferences and assume risk neutrality with respect to game payoffs.

#1	0.056	0.832	0.056	0.056
0.056	20,5	-20,9	-40,7	110,-1
0.497	-40,-3	**[100,13]**	40,3	-60,1
0.391	40,13	20,29	130,11	100,7
0.056	120,5	0,7	40,-3	-20,-1

#2	0.768	0.056	0.056	0.121
0.832	160,10	-20,11	140,-6	120,16
0.056	-40,36	-30,22	50,12	110,-8
0.056	0,14	160,-4	60,14	**[150,15]**
0.056	-30,30	-20,5	60,5	120,-6

#3	0.391	0.056	0.056	0.497
0.056	35,16	75,8	-80,3	**[120,17]**
0.832	150,13	80,15	70,12	60,24
0.056	0,-2	65,16	60,-4	-30,13
0.056	-10,30	55,-14	50,20	-40,-20

#4	0.391	0.056	0.246	0.307
0.497	150,-1	0,-3	45,5	-5,-8
0.391	95,-9	5,-19	60,9	**[50,14]**
0.056	25,27	-5,29	-95,-19	-55,1
0.056	65,15	15,-1	-85,11	-15,3

Table 1: *The games used in the experiment. Subjects play each game for ten consecutive periods; the order of the games was fixed and corresponds to the above number labels. Nash equilibria are bolded and bracketed and CH-predicted frequencies (for* $\tau = 1.5$*) are given along the margins.*

A2 Experiment Details

The experiments were conducted at the Social Science Experimental Laboratory (SSEL) at Caltech. Subjects were recruited via email from the SSEL subject pool. They participated in experiment sessions in groups of six. As described above, subjects played four two-player 4x4 normal form games. Subjects played the four games in a set order (corresponding to their numerical labelings), playing 10 consecutive periods of each game. The subjects were randomly paired with another opponent from their session each period, and if possible, the pairing was performed so that they would not see the same subject in the immediate next period. Each subject maintained their role as row or column player throughout the experiment.

At the beginning of the session, subjects played three practice periods on a different 4x4 game to familiarize themselves with the computer interface. They were not paid for the practice rounds. In the final stage of the experiment, subjects were asked to fill out a questionnaire and to answer the free form question, "What is your strategy?" At the conclusion of the session, they were paid, in cash, the total amount they earned in the games plus a show-up fee of \$5 or \$10 dollars.

The experimental task for non-eye-tracked subjects was coded and run using the Zurich Toolbox for Readymade Economic Experiments (z-Tree) developed by Fischbacher (2007)[4]. See Figure 1 for an example of the screen display during the non-eye-tracked experimental task, and see below for discussion of the interface design.

After each round of choice, subjects were shown a result screen informing them the other players' actions and the payoffs they earn (in this round and overall). The history of past outcomes (pairs of own/other choices and past earnings) is also displayed at the bottom of the screen during each round, though subjects seldom look at them (presumably, they can easily recall them from their short-term memory).

A2.1 Eye-tracked Subjects

We designed the experiment to have two subjects eye-tracked in each session; in practice, hardware calibration issues sometimes interfered. The eye-tracked subjects were taken to a separate location on campus and mounted with Eyelink II head-mounted eye-tracking systems (SR Research, Ontario, CA). The eye-tracked subjects were instructed the same way as the non-eye-tracked subjects, but were given additional instructions related to the eye-tracking procedure.

Prior to the start of the experimental task, eye-tracked subjects were mounted with the eye-tracker and underwent a calibration procedure performed in accordance with the manufacturer's technical instructions. This calibration allows the eye-tracker to estimate subjects' gaze locations in screen coordinates on the basis of raw measurements from head-mounted cameras, two focused on their eyes and one focused on infrared markers at the corners of the screen. The calibration was performed by asking the subject to fixate on nine different points on the screen and recording the corresponding eye and head positions. After the calibration, a nine point validation was performed (similar to the calibration process) to make sure the calibration was accurate. Generally, we accepted calibrations and allowed subjects to proceed only if their average error of measurement was less than one degree of viewing angle; recalibrations were performed if needed, and eye-tracking was halted if these were unsuccessful. At the beginning of each period, a drift correction was performed to correct for drifts in the calibration (and a recalibration was performed if necessary).

The eye-tracked subjects used a different task program than the other subjects; this task was coded and run with Matlab, using the Psychophysical Toolbox version 2.54[1, 7] for task display and subject input and the Eyelink Toolbox[2] to interface with the eye-tracker hardware.[1] The experimental task was displayed on a Iiyama HM204DT 22-inch monitor with a resolution of 1600x1200 and a refresh rate of 85 Hz. Subjects were seated approximately 24 inches away from the screen; this distance implies that a 1° change in gaze equals an approximately 40 pixel change in screen location.

On the screen, payoffs of oneself are placed on the left and the opponent's payoffs on the right (to make it easier for the eye-tracker to identify which payoff cell the subject is viewing). For the column players, we transposed the payoff matrix so they were visually choosing rows (to make it comparable to the tracked row players). Subjects were asked to make their choices by fixating on one of the four boxes on the left of the screen for 0.8 seconds.[2]

[1] To avoid the complications of integrating different software packages, experimenters served as the interface between each eye-tracked subject and the rest of the group, communicating strategy choices back and forth between a laptop running z-Tree and a response pad connected to the Psychophysical Toolbox task.

[2] This input method likely produced some spurious lookups; see A6.1 and A6.4.3 for relevant discussion. We chose this input method because of one nice feature: when the quality of the eye-tracker calibration declines beyond a certain point, the subject is unable to proceed. Thus, it is easier to identify when recalibration is needed. On the whole, however, this input method is probably undesirable.

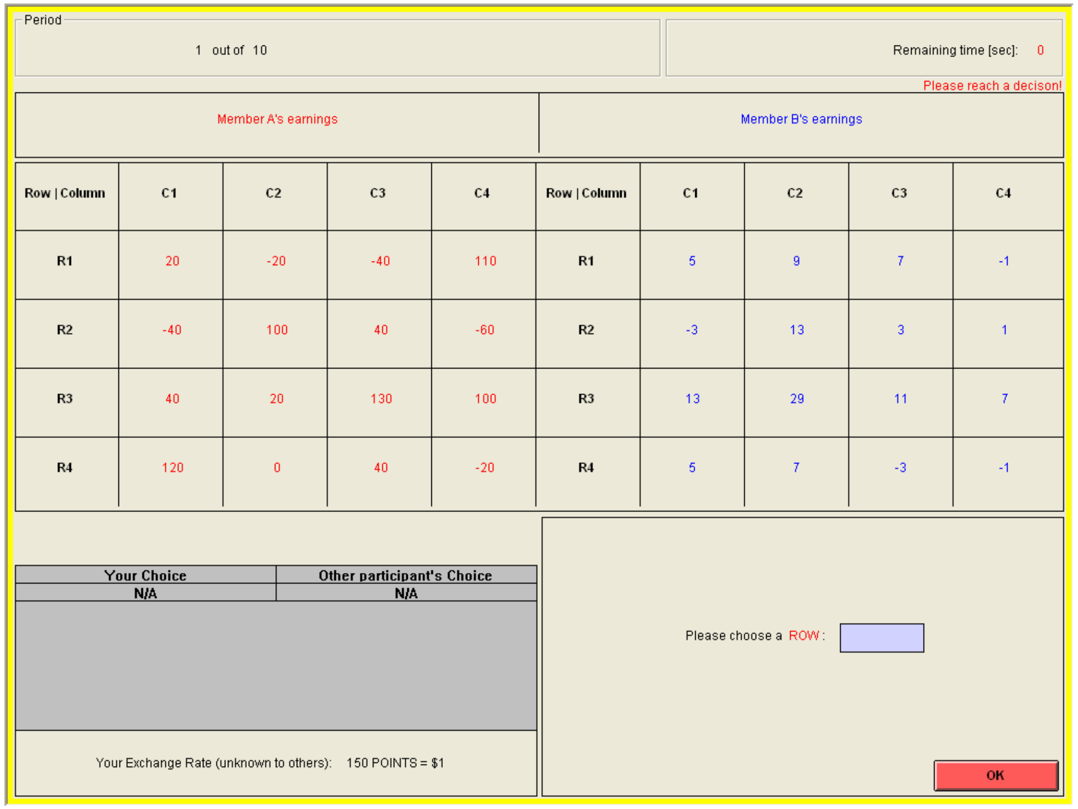

Figure 1: *Example screen display for a non-eye-tracked subject (here, a row player in period one of game one). Note the separation of own payoff and other payoff matrices, the keyboard input method, and the history of strategy choices and own payoffs at the bottom left of the screen.*

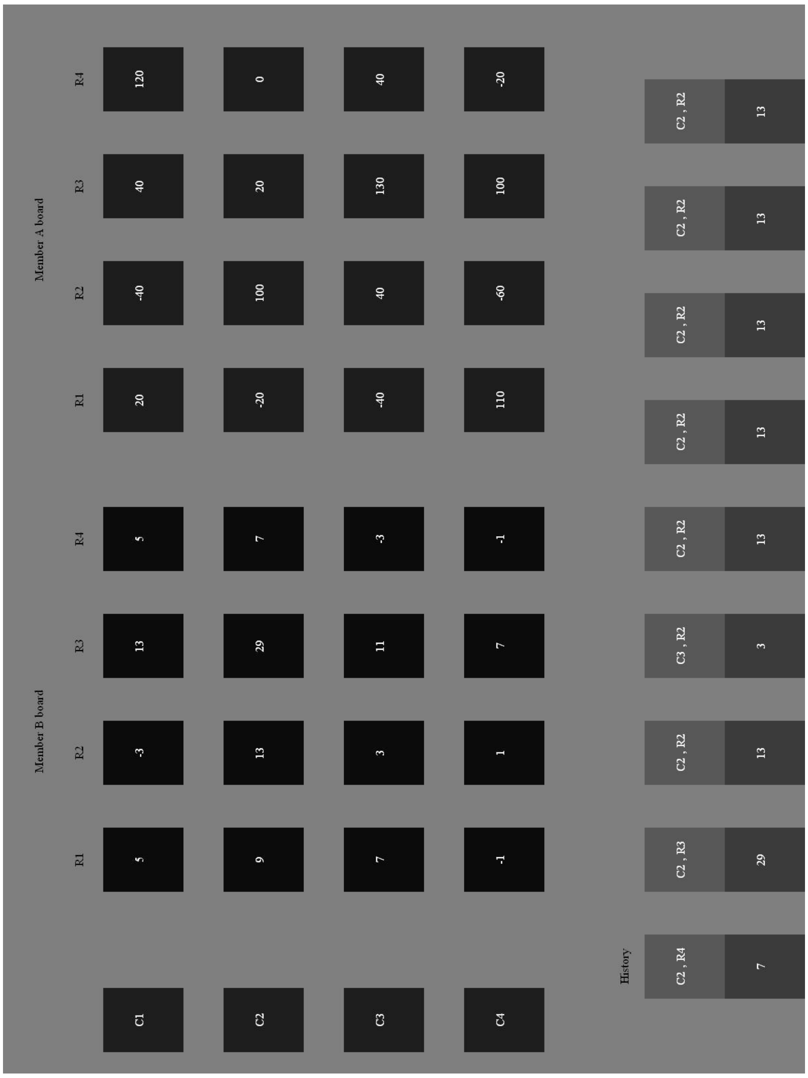

Figure 2: *Example screen display for an eye-tracked subject (here, a column player in period ten of game one). Note the separation of own payoff and other payoff matrices, the choice boxes on the left side of the screen, and the history of strategy choices and own payoffs at the bottom of the screen.*

A3 Summary of Data Collected

A total of eight sessions were conducted. Five sessions were conducted during August 12-20, 2005 (Two sessions each on August 12 and 13, and one session on August 20), resulting in twenty-four non-eye-tracked subjects and six eye-tracked row player subjects. Due to technical difficulties, we had to stop eye-tracking two other eye-tracked subjects during the experiment; these subjects continued with the experimental task and only their behavioral choices are included in the analysis. Another three sessions were conducted on September 11, 2006, resulting in eight non-eye-tracked subjects and six eye-tracked column player subjects. Due to a low show-up rate, two of the sessions were conducted with only four subjects in the group, two of them eye-tracked. Also, due to experimenter error, a subject's opponent's choice was transferred between z-Tree and the Psychophysics Toolbox incorrectly in one of the periods for two subjects. We consider it unlikely that this would affect any of our analyses, and therefore do not exclude these subjects on the basis of these minor errors. In total, we collected behavioral data for forty-four subjects and lookup data for twelve of these subjects (six row players and six column players).

A4 Models

We consider five models of learning: reinforcement (Re), experience-weighted attraction learning (EWA), self-tuning EWA (stEWA), and Cournot-type anticipatory response level one (C1) and two (C2).

A4.1 Experience-Weighted Attraction (EWA)

EWA is a general adaptive model that nests types of reinforcement and belief-based learning as special cases. In EWA, strategies have numerical attractions which begin with subject-specific initial attractions and are updated based on received and foregone payoffs. In its general form, EWA has three psychologically-inspired parameters: ϕ, δ, and κ.

Using notation similar to Ho, Camerer, and Chong (2007), we have players i, games g, and trials t; each player chooses strategy j from $1, \ldots, m_i$. Player i's strategy choice in period t of game g is $s_{ig}(t)$; his opponent's strategy choice is $s_{-ig}(t)$. Likewise, s_{ig}^j denotes player i's j-th strategy in period t of game g. Players have subject-specific initial attractions $A_{ig}(0) = (A_{ig}^1(0), \ldots, A_{ig}^{m_i}(0))$ and initial weight $N_{ig}(0)$. Attractions update after the choices and receipt of information in a trial, according to the following equation:

$$\begin{aligned} A_{ig}^j(t) = & \frac{\phi_i * N_{ig}(t-1) * A_{ig}^j(t-1)}{N_{ig}(t)} \\ & + \frac{\Big(\delta_i + (1-\delta_i)\,\mathrm{I}\big(s_{ig}(t), s_{ig}^j\big)\Big)\pi_{ig}\big(s_{ig}^j, s_{-ig}(t)\big)}{N_{ig}(t)}, \end{aligned} \tag{A4.1}$$

and $N_{ig}(t)$ updates according to

$$N_{ig}(t) = \phi_i * (1 - \kappa_i) * N_{ig}(t-1) + 1. \tag{A4.2}$$

In (A4.1), the first term is the attraction from the previous period discounted by some factor and the second term is the weighted reinforcement value corresponding to the strategy. Chosen strategies have a reinforcement weight of 1 whereas unchosen strategies have a reinforcement weight of δ.

These attractions map to strategy choice probabilities via some function; following Camerer and Ho (1999) and Ho, Camerer, and Chong (2007), we choose the "logit" form which uses a sigmoid function with a subject-specific response parameter λ. The probability of choosing strategy s_{ig}^j in period t is the following function of period $t-1$ attractions:

$$p_{ig}^j(t) = p_{ig}^j(A_{ig}^j(t-1); \lambda_i) = \frac{\exp\big(\lambda_i * A_{ig}^j(t-1)\big)}{\sum_{j=1}^{m_i} \exp\big(\lambda_i * A_{ig}^j(t-1)\big)} \tag{A4.3}$$

The lookup area associated with EWA is the own payoff matrix column corresponding to $s_{-i}(t)$.

A4.2 Reinforcement (Re)

As noted above, in the EWA model, the parameter δ is the reinforcement weight placed on foregone payoffs. In addition to unconstrained EWA, we consider a reinforcement-type model obtained by fixing $\delta = 0$, which implies payoffs from unchosen strategies do not enter into the calculation of attractions and are therefore irrelevant. Reinforcement-type models have the own payoff matrix cell corresponding to the chosen strategies of the period $\pi_{ig}(s_{ig}(t), s_{-ig}(t))$ as their lookup area.

A4.3 Self-tuning EWA (stEWA)

Ho, Camerer and Chong (2007) created a "self-tuning" form of EWA (stEWA) in which two psychologically uninteresting parameters are removed and the remaining free parameters are replaced by functions[5]. The reinforcement weight $\delta + (1-\delta)I(s_i^j, s_i(t))$ is replaced by a function $\delta_{ig}^j(t)$ that equals 1 for strategies which are weakly better responses than the chosen strategy (i.e., $\pi_{ig}(s_{ig}^j, s_{-ig}(t)) \geq \pi_{ig}(s_{ig}(t), s_{-ig}(t))$) and 0 for worse responses. In addition, the decay weight ϕ on old attractions is specified as a function of experience $\phi_{ig}(t)$. The stEWA model constrains $\kappa = 0$ and $N(0) = 1$. See Ho, Camerer and Chong (2007) for discussion of the inspiration behind these choices.

As in the vanilla EWA model, attractions update after the choices and receipt of information in a trial, according to the following equation:

$$A_{ig}^j(t) = \frac{\phi_{ig}(t) * N_{ig}(t-1) * A_{ig}^j(t-1)}{N_{ig}(t)} + \frac{\delta_{ig}^j(t) * \pi_{ig}\big(s_{ig}^j, s_{-ig}(t)\big)}{N_{ig}(t)}, \tag{A4.4}$$

and $N_{ig}(t)$ updates according to

$$N_{ig}(t) = \phi_{ig}(t) * N(t-1) + 1. \tag{A4.5}$$

The function $\phi_{ig}(t)$ is defined to be

$$\phi_{ig}(t) = 1 - \frac{1}{2}\sum_{k=1}^{m_{-i}}\Big(h_i^k(t) - \mathrm{I}\big(s_{-i}^k, s_{-i}(t)\big)\Big)^2, \tag{A4.6}$$

where

$$h_i^k(t) = \frac{1}{t}\sum_{\tau=1}^{t}\mathrm{I}\big(s_{-i}^k, s_{-i}(\tau)\big); \tag{A4.7}$$

and the function $\delta_{ig}(t)$ is defined to be

$$\delta_i^j(t) = \begin{cases} 1 & \text{if } \pi_i\big(s_i^j, s_{-i}(t)\big) \geq \pi_i(t) \\ 0 & \text{otherwise.} \end{cases} \tag{A4.8}$$

As before, attractions map to probabilities according to (A4.3).

The lookup area for stEWA is the set of cells with positive reinforcement weight in the own payoff matrix column corresponding to $s_{-i}(t)$.

A4.4 Cournot-type Models (C1 and C2)

The two "sophisticated" models C1 and C2 anticipate opponents' behavior on the basis of the observed choice in the previous period. In one model (C1), players expect others will choose a best response to their own last choice (i.e., they expect others use a Cournot rule). Note that this rule requires player i to look at payoffs $\pi_{-i}(s_{-i}^j, s_i(t))$ (i.e., the other player's payoffs given what player i just did), compute $s_{-i}^* = \{s_{-i}^j : s_{-i}^j = \arg\max \pi_{-i}(s_i(t), s_{-i}^j)\}$ and then look at $\pi_i(s_i^j, s_{-i}^*)$.

An iteration of this rule (C2) assumes that row players think column players think they (the row players) will follow a Cournot rule. This requires looking at $\pi_i(s_i^j, s_{-i}(t))$, computing $s_i^* = \{s_i^j : s_i^j = \arg\max \pi_i(s_i^j, s_{-i}(t))\}$, then looking at $\pi_{-i}(s_i^*, s_{-i}^j)$ to compute $s_{-i}^* = \{s_{-i}^k : s_{-i}^k = \arg\max \pi_{-i}(s_i^*, s_{-i}^k)\}$, then looking at $\pi_i(s_i^j, s_{-i}^*)$ to choose a best response. (These models obviously could be iterated further than two steps.)

A5 Behavior

A5.1 Summary of Observed Behavior

Twenty-eight of the forty-four subjects have at least one game in which they chose the same strategy across all ten periods (sixteen subjects have one, five subjects have two, six subjects have three, and one subject has four). These "non-learning games" make up 27.3% of the observed games and 29.2% of the eye-tracked games. Since we conjecture that lookups and behavior in these "non-learning games" are less likely to be the result of a learning process, we focus on the subset of the data which excludes them (we call this the "learning games" subset).

Graphs of the proportion of subjects playing their Nash strategy in each period of the four games are Figures 3 and 4, computed for all games and for the "learning games" subset, respectively. Two features are immediately evident: there is a substantial amount of non-Nash play in the first period of each game, and the proportion playing Nash tends to increase over the ten periods. Furthermore, the proportion of Nash play is reduced in the "learning games" subset, as would be expected.

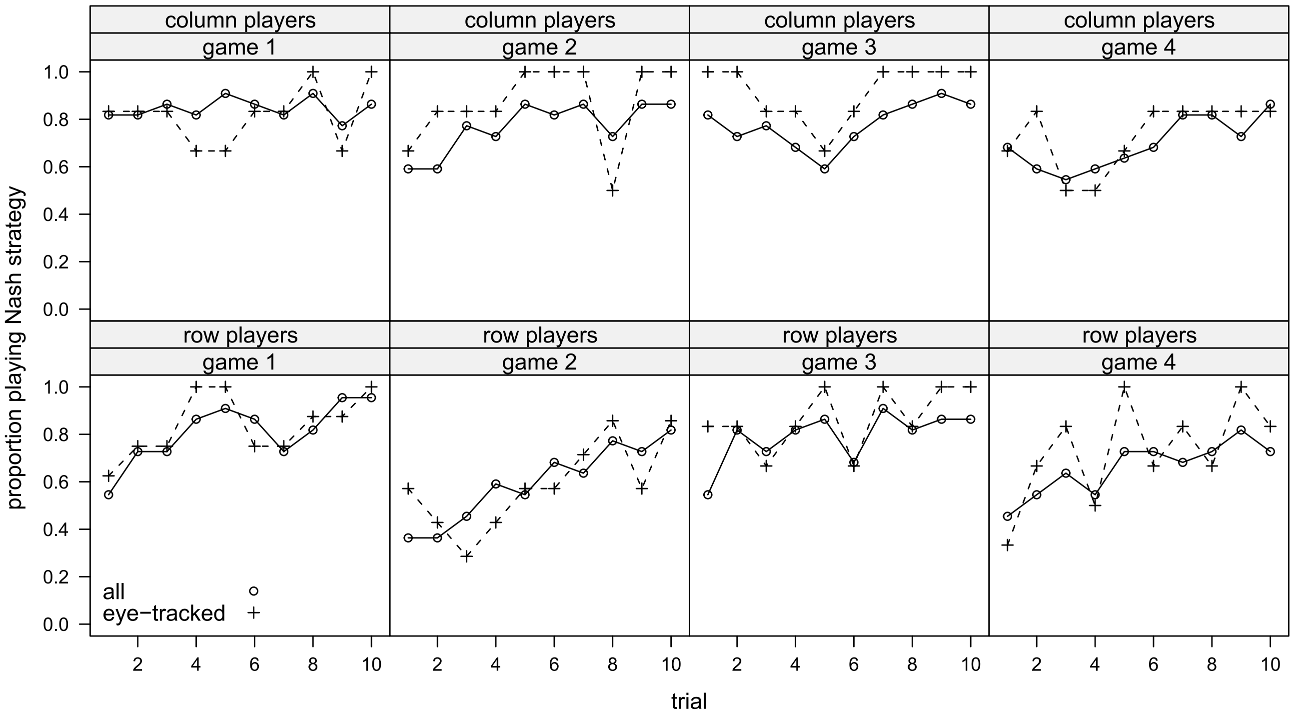

Figure 3: *Proportion of subjects playing their Nash strategy, computed using all trials of all games. Here, we do not drop trials from games in which the subject played the same strategy in every trial (the "non-learning games"). Results computed using "learning games" only, as in the main paper, are in Figure 4 below.*

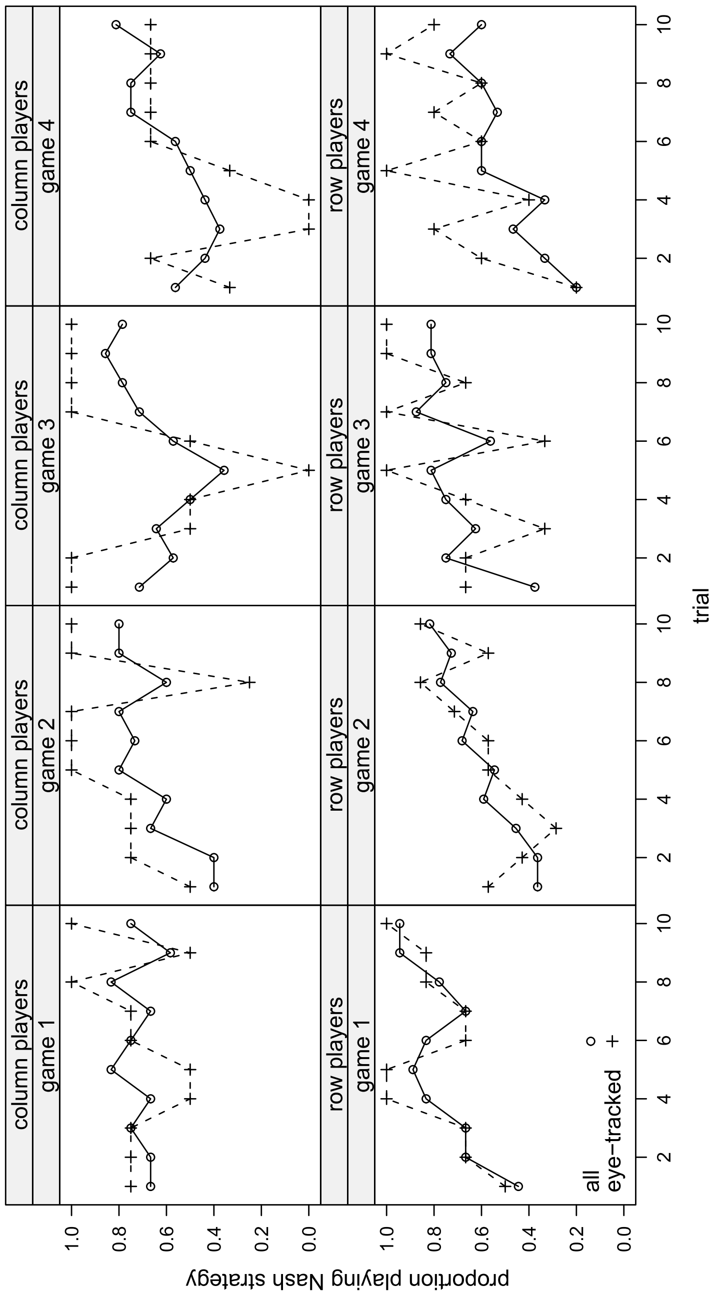

Figure 4: *Proportion of subjects playing their Nash strategy, computed using the "learning games" subset. Note that the proportion of Nash play is reduced relative to Figure 3.*

A5.2 Methods for Behavioral Estimation

The basic features to note are:

- Initial attractions are burned-in using first period choices.
- Parameters are subject-specific but constrained to be equal across games.
- Choice probabilities are defined by the logit-type map of (A4.3).
- Models are estimated using maximum likelihood.

A5.2.1 Maximum Likelihood Estimation

For each player i and model M, we numerically maximize the likelihood function with respect to a vector of parameters $\theta_i^{\mathrm{M}} = (\mu_i^{\mathrm{M}}, \sigma_i^{\mathrm{M}}, \lambda_i^{\mathrm{M}})$ subject to relevant constraints, where μ_i^{M} is the (possibly length-0) vector of model parameters, σ_i^{M} is the initial attractions scale parameter (included for adaptive models only), and λ_i^{M} is the response parameter in the probability map of (A4.3). Omitting the model superscript, this likelihood is

$$L(\theta_i | s_i, s_{-i}) = \prod_g \prod_t p_{ig}^j(A_{ig}^j(t-1; \mu_i, \sigma_i); \lambda_i). \tag{A5.1}$$

A5.2.2 Initial Attractions

For the adaptive models (EWA and its variants), we need to provide initial attractions for every player, for each strategy in each game. Instead of fitting these as free parameters, we "burn-in" the initial attractions using the first-period choice, and give the initial attractions an unrestricted scale parameter σ_i. Specifically, $A_{ig}^j(0) = \sigma_i * \mathrm{I}(s_{ig}^j = s_{ig}(0))$. We also fix $N(0) = 1$ in all adaptive models. This is a constraint in stEWA that we apply to Re and EWA as well; the modulation of initial attraction weight by having $N_i(0)$ a free parameter is accomplished here indirectly by the scale parameter for the initial attractions.

A5.2.3 Nash-type Benchmark

As a non-learning benchmark, we also estimate a model in which players assume their opponent will choose their unique Nash strategy with certainty (we call this model "Nash"). In this model, the base attractions are expected payoffs under this assumption: $A_{ig}^j(t) = \pi_{ig}(s_{ig}^j, s_{-ig}^{\mathrm{Nash}})$.

A5.2.4 Ck Model Variants

We consider several variants of the Ck models to examine whether minor changes to our behavioral specifications affect our qualitative conclusions. These variants, described below, alter the myopic best response of the basic Ck models, reducing the responsiveness to opponents' play in the previous period.

Our base Ck model, as described above, assigns attractions equal to expected payoffs assuming the opponent is a hard-maximizing C(k − 1) type; that is, $A^j_{ig}(t) = \pi_{ig}(s^j_{ig}, s^{\mathrm{C(k-1)}}_{-ig})$.

A5.2.5 Geometric Ck Variant

The geometric variants of the Ck models assign attractions equal to a geometrically-discounted average of the current Ck attractions and attractions in the previous periods. If the base Ck-model attractions for $\tau = 1, \ldots, t$ are $B^j_{ig}(\tau)$, the geometric attractions for period t are

$$A^j_{ig}(t) = \sum_{\tau=1}^{t} \gamma^{t-\tau} B^j_{ig}(\tau),$$

where γ is a free parameter.

A5.2.6 *p*-switch Ck Variant

The p-switch variants of the Ck models simply add a quantity γ to the attraction for the strategy chosen by the player in the previous period. Essentially, players are predicted to play the same strategy as the previous period with some probability and to switch to the myopically best-responding Ck strategy otherwise.

A5.2.7 "Hard" and "Soft" Maximization

The theoretical Ck-type and Nash-type models assume that the opponent will choose a certain unique strategy with certainty. In these models, the base attractions are expected payoffs under this assumption; for model M, we have $A^j_{ig}(t) = \pi_{ig}(s^j_{ig}, s^{\mathrm{M}}_{-ig}(t))$.

We can map these attractions into probabilities using either "hard" or "soft" maximization, depending on our assumptions about the error structure. Essentially, in "soft" maximization, when subjects make "errors" and do not choose optimal strategies as specified by the model, they are more likely to choose strategies with higher expected payoffs under the assumptions of the model; in contrast, with "hard" maximization, these errors are uniform across strategies. Using "hard" maximization is in a sense the more straightforward choice (and is the choice we made for results reported in the main paper)—it assumes that subjects choose the model-specified strategies with probability p and randomize uniformly over the other strategies.

With "soft" maximization, the expected payoff-based attractions are fed directly into (A4.3), and the free parameter in the map from attractions to choice probabilities determines the degree of responsiveness to the model-implied expected payoffs.

With "hard" maximization, the expected payoff-based attractions are transformed thusly

$$A^j_{ig}(t) = \frac{\mathrm{I}\big(A^j_{ig}(t) = \max_j A^j_{ig}(t)\big)}{\sum_{j=1}^{m_i} \mathrm{I}\big(A^j_{ig}(t) = \max_j A^j_{ig}(t)\big)} \tag{A5.2}$$

and then fed into (A4.3). That is, strategies which are best responses under the assumptions of the model are given an attraction of one and all others are given an attraction of zero, and the free parameter in (A4.3) determines the error probability.

A5.2.8 Out of Sample Validation

We report goodness of fit for periods outside the sample used for fitting the models. In these out of sample "validation" fits, we estimate parameters using choices from periods 2–7 in each game and report the corresponding likelihoods for the data from periods 8–10.

A5.3 Behavioral Estimation Results

Table 2 reports the validation goodness of fit (as described above), recapitulating the results reported in the main paper alongside results including all trials. The goodness of fit is given as geometric mean likelihoods, averaged across subjects. As can be seen in the table, there are no apparent differences in behavioral fit for the eye-tracked subjects. Tables 3 and 4 report uncorrected p-values from paired Mann-Whitney-Wilcoxon tests of equality between each of the behavioral model specifications, for the "learning games" subset and for all games, respectively. The "soft" maximization variants have been dropped. Tables 5 and 6 report results of the same tests, correcting for multiple comparisons by the method of Holm (1979)[6].

Figure 5 plots the EWA parameter estimates for each subject, fit using the "learning games" subset and periods 2–7. This plot can be compared with similar plots in the EWA literature.

		all subjects		eye-tracked subjects	
		all trials	learning games	all trials	learning games
Re		0.555	0.474	0.497	0.405
EWA		0.561	0.499	0.559	0.454
stEWA		0.588	0.505	0.601	0.460
C1	hard	0.535	0.420	0.557	0.400
	soft	0.496	0.409	0.532	0.411
geometric C1	hard	0.541	0.433	0.566	0.399
	soft	0.530	0.450	0.556	0.438
p-switch C1	hard	0.582	0.481	0.556	0.450
	soft	0.573	0.464	0.603	0.471
C2	hard	0.538	0.425	0.576	0.429
	soft	0.511	0.429	0.551	0.433
geometric C2	hard	0.540	0.425	0.586	0.437
	soft	0.519	0.442	0.557	0.441
p-switch C2	hard	0.605	0.509	0.602	0.491
	soft	0.587	0.493	0.577	0.470
Nash	hard	0.549	0.444	0.573	0.425
	soft	0.512	0.431	0.560	0.437

Table 2: *Validation fits by subset. Results reported in the main paper correspond to columns 2 and 4. Qualitatively, it can be seen that validation fit is uniformly worse when considering the "learning games" subset instead of all the trials. In addition, there are no significant differences in validation fit between eye-tracked subjects and untracked subjects.*

	Re	EWA	stEWA	C1	geo C1	p-sw C1	C2	geo C2	p-sw C2
EWA	0.05501								
stEWA	0.06518	0.88776							
C1	0.02403	0.00290	0.00001						
geo C1	0.12154	0.01155	0.00101	0.13274					
p-sw C1	0.69760	0.21823	0.13698	0.00069	0.08993				
C2	0.06701	0.00374	0.00015	0.30656	0.73074	0.00769			
geo C2	0.12154	0.00746	0.00064	0.08752	0.48735	0.01967	0.68583		
p-sw C2	0.11018	0.77880	0.41886	0.00009	0.00065	0.02468	0.00015	0.00025	
Nash	0.19636	0.00831	0.00084	0.01242	0.41502	0.10747	0.14148	0.90867	0.00390

Table 3: *Pairwise validation fit Mann-Whitney-Wilcoxon tests using all subjects and the "learning games" subset. The Ck and Nash models use "hard" maximization. In this table, p-values are uncorrected for multiple comparisons. Table 5 reports results from these tests corrected for multiple comparisons by the method of Holm (1979).*

	Re	EWA	stEWA	C1	geo C1	p-sw C1	C2	geo C2	p-sw C2
EWA	0.56646								
stEWA	0.24392	0.67200							
C1	0.03743	0.04054	0.00449						
geo C1	0.06735	0.10406	0.00897	0.04924					
p-sw C1	0.70368	0.98005	0.57366	0.00180	0.05261				
C2	0.05738	0.05200	0.00402	0.43202	0.90247	0.00740			
geo C2	0.16312	0.10949	0.01075	0.05342	0.62156	0.01000	0.06103		
p-sw C2	0.19845	0.41637	0.82193	0.00033	0.00269	0.04209	0.00047	0.00049	
Nash	0.20763	0.11380	0.02467	0.00711	0.43769	0.05200	0.12628	0.44458	0.00483

Table 4: *Pairwise validation fit Mann-Whitney-Wilcoxon tests using all subjects and all games. The Ck and Nash models use "hard" maximization. In this table, p-values are uncorrected for multiple comparisons. Table 6 reports results from these tests corrected for multiple comparisons by the method of Holm (1979).*

	Re	EWA	stEWA	C1	C2
EWA	0.38506				
stEWA	0.39109	0.88776			
C1	0.19226	0.03474	0.00026		
C2	0.39109	0.04109	0.00204	0.61312	
Nash	0.58907	0.08306	0.01088	0.11182	0.56593

Table 5: *Pairwise validation fit Mann-Whitney-Wilcoxon tests using all subjects and the "learning games" subset. The Ck and Nash models use "hard" maximization, and other Ck variants have been dropped. In this table, p-values are corrected for multiple comparisons by the method of Holm (1979).*

	Re	EWA	stEWA	C1	C2
EWA	1.00000				
stEWA	1.00000	1.00000			
C1	0.41177	0.41177	0.06282		
C2	0.46801	0.46801	0.06027	1.00000	
Nash	1.00000	0.79661	0.29604	0.09237	0.79661

Table 6: *Pairwise validation fit Mann-Whitney-Wilcoxon tests using all subjects and all games. The Ck and Nash models use "hard" maximization, and other Ck variants have been dropped. In this table, p-values are corrected for multiple comparisons by the method of Holm (1979).*

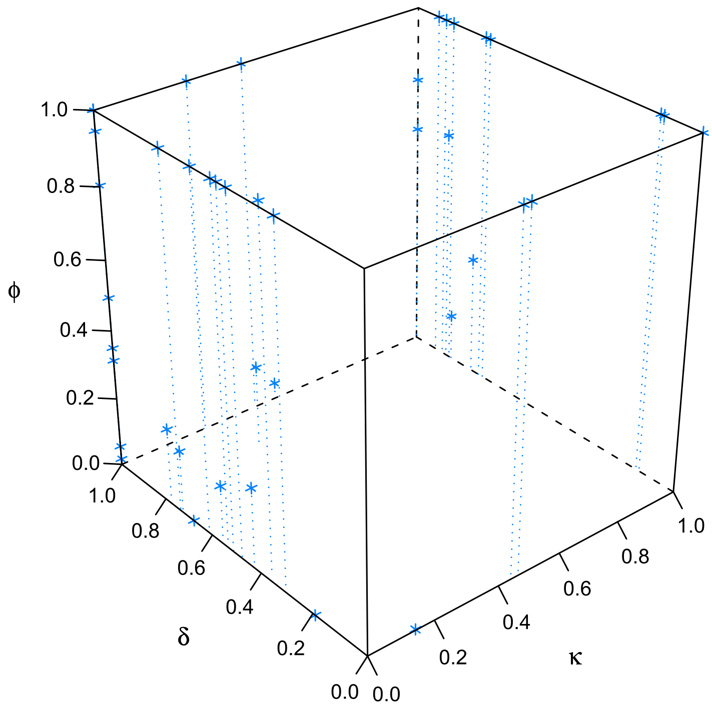

Figure 5: *EWA parameter cube plotting the estimated parameters for each subject, computed using the "learning games" subset, periods 2–7.*

A6 Eye-tracking

A6.1 Preprocessing and Filtering

Gaze samples were parsed into fixations and saccades by the EyeLink II on-line parser, using standard parameters for cognitive studies[9]. Fixations map to lookups via defined interest areas, which are boxes partitioning the screen area on and around the displayed payoff matrices into corresponding cells. A fixation in a payoff matrix cell's interest area counts as a lookup of that payoff matrix cell. Consecutive fixations in the same cell were combined and considered as one lookup. We use fixation counts for our analysis and not fixation durations (with the exception of the analysis in A6.4.5); thus, combining consecutive fixations prevents us from double-counting lookups. See Figure 6 for an example of the preprocessed output from the eye-tracker, showing interest areas, fixations, and saccades prior to combining consecutive fixations or filtering.

The figure should also convey how noisy eye-tracking data can be and how difficult it is to unambiguously map fixations to information lookups. Note that we chose a relatively clean trial; nonetheless, there are fixations outside defined interest areas that might reasonably be considered lookups, depending on the (unknown) discrepancy between the calibrated estimated screen locations of fixations and the true gaze location, and also on the angle from the gaze location within which subjects can read our displayed payoff numbers. Furthermore, a number of fixations are located near edges or corners of interest areas and might very well represent lookups of several different payoff matrix cells. In this study, we tend to take our lookups at face value and hope that noise in our data translates to noise, but not bias, in our analysis results (and especially in relative comparisons across theories). Our main results derived from eye-tracking data aggregate observations across trials, games, and eventually subjects; thus, these results should be reasonable estimates of the average relative weight of visual attention on different pieces of information within subjects and within our experimental population. Other analyses, like those described in A6.2.1, involve assumptions or interpretations of the data that are less robust to noise and calibration error.

We perform minor filtering of the fixations after combining consecutive fixations in a cell but prior to any analysis. We always drop the first fixation in a trial, as it represents the gaze location at the moment when the payoff matrices are displayed, and should not represent an information lookup. We also always drop the last fixation (not counting the fixation that chooses a strategy and ends the trial), as this fixation often appeared to represent a spurious lookup produced while the subject was attempting to make a strategy choice (recall that we used a "choose-with-eyes" input method; see A2.1). After this filtering, all lookups outside payoff matrix or history interest areas are dropped; history lookups are dropped for the majority of our analysis and discussed in A6.5.

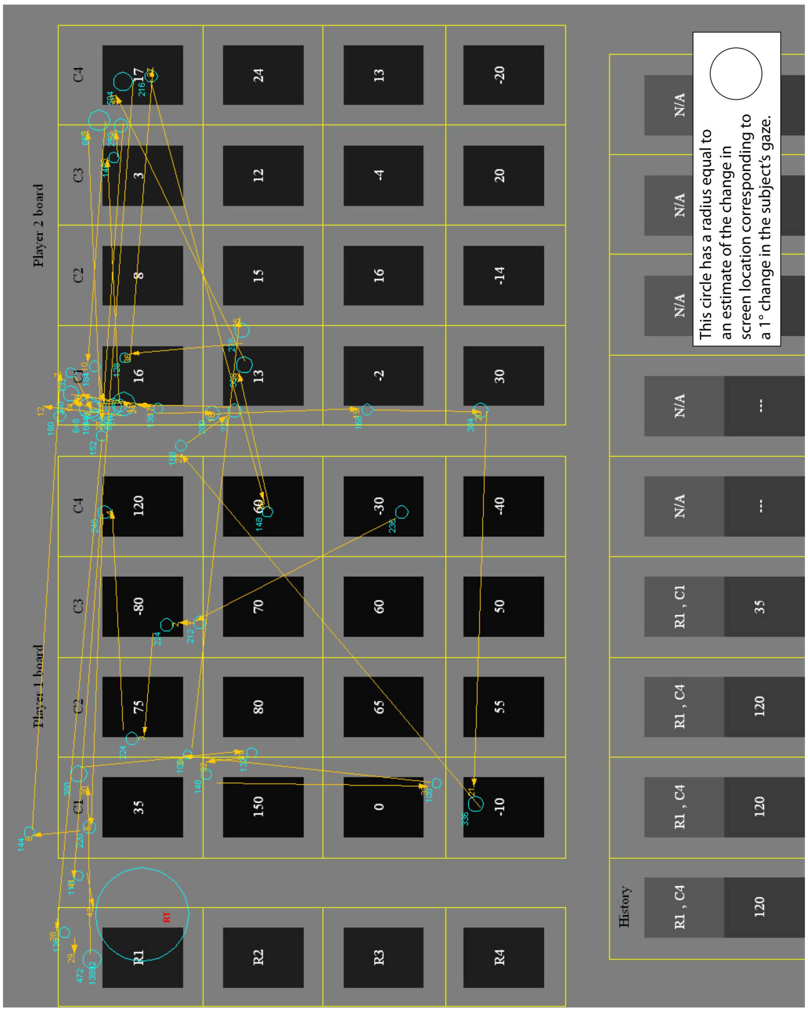

Figure 6: *Graphical representation of example preprocessed output from the eye-tracker. The yellow boxes are interest areas. Fixations are displayed as teal circles and saccades as yellow arrows.*

A6.2 Model Lookup Predictions and Measures

We assume that each learning model implies an area theory corresponding to the cells containing the minimal set of information necessary to implement the learning model. Define the following measures as functions of the hit-rate x and the proportional area a:

- Linear measure:
$$LM(x, a) = x - a \tag{A6.1}$$
- Ratio measure:
$$RM(x, a) = \frac{x}{a} \tag{A6.2}$$
- Outer measure:
$$OM(x, a) = \frac{x - a}{1 - a} \tag{A6.3}$$

The hit-rate x can be taken as the proportion of time fixated, the proportion of fixations, or the proportion of observed cells. Arguments made by Selten (1991) suggest that the linear measure is in some sense optimal [8]. The ratio measure is an oft-suggested alternative, but, as Selten (1991) notes, it favors point theories. The outer measure is another alternative; noting $x - a = (1 - x) - (1 - a)$, we can write

$$OM(x, a) = \frac{(1 - x) - (1 - a)}{1 - a}, \tag{A6.4}$$

which is simply the ratio measure for the area model's complement. The outer measure favors large area theories.

A6.2.1 Time Course-Restricted Ck Models

In addition to area predictions, the level-k "anticipatory" models suggest certain restrictions on the order of lookups. Specifically, the C1 model suggests that subjects undergo the following stages:

1. Look at the row in their opponent's payoff matrix corresponding to their previous choice and find their opponent's best response strategy $\mathrm{BR}_{-i}(s_i(t-1))$.
2. Look at the column in their own payoff matrix corresponding to this best response and find their best response to it $\mathrm{BR}_i(\mathrm{BR}_{-i}(s_i(t-1)))$.

Likewise, the C2 model suggests that subjects undergo the following stages:

1. Look at the column in their own payoff matrix corresponding to their opponent's previous choice and find their best response strategy $\mathrm{BR}_i(s_{-i}(t-1))$.
2. Look at the row in their opponent's payoff matrix corresponding to this best response and find their opponent's best response to it $\mathrm{BR}_{-i}(\mathrm{BR}_i(s_{-i}(t-1)))$.

3. Look at the column in their own payoff matrix corresponding to this best response and find their best response to it $\text{BR}_i(\text{BR}_{-i}(\text{BR}_i(s_{-i}(t-1))))$[3].

We examine lookup scores for "restricted Ck" models (denoted "C1r" and "C2r"); for these models, we implement a simple form of order restrictions by requiring that subjects hit $r = 1$ of the four cells in the first stage area before counting lookups in the second stage area as hits for the model, and likewise, for three-stage models, require that subjects hit $r = 1$ of the four cells in the second stage area (subject to the first stage restriction) before counting lookups in the third stage area as hits for the model.[4]

In order for the linear measure scores for these restricted models to have zero expectations under random looking, we must adjust the areas a to account for the restrictions.

The relevant distribution is the negative binomial, which, for a series of i.i.d. Bernoulli trials with success probability p, gives the probability of receiving the r-th success on the $(k+r)$-th trial:

$$\text{NegBin}(k; r, p) = \binom{k+r-1}{k} p^r (1-p)^k$$

For a trial with n total fixations, the expected count-wise hit rates under uniform random lookup for the three stages can be written as:

$$\begin{aligned}
a_1 &= p \\
a_2 &= \frac{1}{n} \sum_{u=1}^{n} \text{NegBin}(u-r; r, p) \left(\frac{p(n-u)}{n} \right) \\
a_3 &= \frac{1}{n} \sum_{u=1}^{n} \text{NegBin}(u-r; r, p) \left(\sum_{v=1}^{n-u} \text{NegBin}(v-r; r, p) \left(\frac{p(n-u-v)}{n} \right) \right)
\end{aligned}$$

Thus, the adjusted two-stage model area is $a = a_1 + a_2$ and the adjusted three-stage model area is $a = a_1 + a_2 + a_3$.

A6.2.2 Post-Period Feedback and "Reinforcement-enhanced" Scores

A notable feature of our experimental task for eye-tracked subjects is the inclusion of a post-period feedback screen displaying the subject's strategy choice, his opponent's strategy choice, and the subject's received payoff for the period. Importantly, this information on the subject's received payoff is exactly the information contained in the "reinforcement cell" of the payoff matrices. Given this redundancy, one wonders if the reinforcement-type model has well-defined lookup predictions in our experiment.

[3] Note that in the C2 model the third stage area and first stage area can coincide in certain trials (in equilibrium, for instance). When they do, C2 is effectively a two-stage model.

[4] Lookups in earlier stage areas continue to count as hits after later stage areas become operative.

One strategy for dealing with this issue attempts to integrate this feedback screen information into our lookup counts by adding some number of hypothetical automatic lookups of the reinforcement cell in every trial. For instance, we could suppose that subjects' viewing of the post-period feedback screen translates into one automatic lookup of the reinforcement cell in the subsequent trial. This strategy implicitly assumes that ignoring the post-period feedback screen undercounts the true lookups of the reinforcement cell information (the subject's received payoff). We'll call the lookup scores produced using this strategy "Reinforcement-enhanced scores". In subsequent sections, we include figures, tables, and analysis of these Reinforcement-enhanced scores. There is further discussion of the Reinforcement-enhanced scores interacting with our assumption that the number of lookups is exogenous in A6.4.4.

It's important to remember, however, that all of the models we consider, with the exception of C1, have lookup area predictions that include the reinforcement cell. Specifically, EWA's lookup area is the entire column in the subject's own payoff matrix corresponding to his opponent's strategy choice in the previous period; stEWA's lookup area is some subset of this EWA column but always includes the reinforcement cell. Furthermore, since the C2-type player first finds his best response to his opponent's strategy choice in the previous period, C2's lookup area always contains the EWA column. Recall that, as we have defined the hierarchy of models, the Ck-type player assumes his opponent is C(k − 1); the C1-type player assumes his opponent uses a Cournot rule and simply best-responds to *his* opponent's previous strategy. Thus, any Ck model with k even will have a lookup area that includes the EWA column and thus the reinforcement cell. C1's lookup area does not necessarily contain the reinforcement cell, but it contains the EWA column (and thus the reinforcement cell) if the opponent's previous strategy choice was a best response to the subject's previous strategy choice. In our data (eye-tracked subjects, "learning games" only), C1's lookup area contained the EWA column 64.9% of the time.[5]

Thus, while our lookup counts may be undercounting the true attention to the reinforcement cell information, "correcting" this bias using the above strategy only affects comparisons between C1 and the other models.[6] The "Reinforcement-enhancement" affects the magnitudes of the lookup scores for the other models, but these magnitudes aren't particularly interesting, beyond the fact that every model scores above the uniform looking benchmark.

It is debatable whether the post-period feedback would affect all of the types in the same way, as hypothesized above, by providing the information contained in the reinforcement cell. Practically, subjects implementing something like a Reinforcement-type model might use the reinforcement cell information on its own to update attractions; those implementing something like the other

[5] The same statistic computed using all trials from eye-tracked subjects, including the "non-learning games", is 74.0%; it is of course higher, as these "non-learning games" tend to have steady Nash play.

[6] As our lookup scores use the linear measure $LM(x, a) = x - a$, which is linear in both the hit-rate x and the proportional area a, comparisons between models depend only on the hit-rates in areas that differ between the models.

models may wish to have that information available at the same time as other payoff information. Certainly, we might expect this to be the case with subjects who, like those implementing a C2-type model, would use the information in the reinforcement cell when comparing payoffs in that column to find a best response. It could even be argued that lookups of the reinforcement cell of the payoff matrix represent evidence against Reinforcement-type reasoning, as subjects implementing a simple Reinforcement-type learning model receive the only information they consider relevant on the post-period feedback screen; they should have no motivation to look for it in the payoff matrix, and it would surely stick in their short-term memory long enough for them to make their strategy choice.

A6.3 Eye-tracking Results

Table 7 contains mean LM scores, both raw and "Reinforcement-enhanced" (as described above in A6.2.2), computed using all trials or using only "learning games". The results reported in our main paper are the mean raw LM scores computed using only "learning games". Table 8 contains *p*-values for pairwise Mann-Whitney-Wilcoxon tests between models' LM scores (uncorrected for multiple comparisons), using "learning games" only; these are the tests reported in the main paper. Table 9 reports results of the same tests, corrected for multiple comparisons using the method of Holm (1979)[6].

Figures 7 and 8 display mean numbers of lookups across periods of the games, broken down by their location (own payoff matrix, opponent payoff matrix, and total), for the "learning games" subset (Figure 7) and for all trials (Figure 8). There is an apparent downward trend in all three statistics across periods of each game, but no apparent trend in the relative share of lookups on one's own payoff matrix versus one's opponent's payoff matrix.

		Re	EWA	stEWA	C1	C2	C1r	C2r
raw	all trials	0.0458	0.0562	0.0553	0.1016	0.1153	0.0714	0.0843
		(0.0084)	(0.0093)	(0.0091)	(0.0169)	(0.0179)	(0.0179)	(0.0120)
	learning games	0.0454	0.0589	0.0572	0.0924	0.1143	0.0622	0.0860
		(0.0070)	(0.0104)	(0.0086)	(0.0203)	(0.0156)	(0.0184)	(0.0129)
Re+1	all trials	0.1570	0.1572	0.1649	0.1567	0.1886	0.0495	0.1865
		(0.0338)	(0.0316)	(0.0338)	(0.0231)	(0.0290)	(0.0145)	(0.0323)
	learning games	0.1188	0.1257	0.1289	0.1222	0.1643	0.0457	0.1502
		(0.0218)	(0.0213)	(0.0225)	(0.0197)	(0.0195)	(0.0167)	(0.0191)

Table 7: *Mean LM scores, raw and Reinforcement-enhanced ("Re+1"), by subset (all games or "learning games"). Standard errors are in parentheses. The mean raw LM scores for the "learning games" are the scores reported in the main paper.*

	Re	EWA	stEWA	C1	C2	C1r
EWA	0.09229					
stEWA	0.00098	0.90967				
C1	0.07715	0.12939	0.12939			
C2	0.00146	0.01611	0.00244	0.05225		
C1r	0.10986	0.26611	0.30127	0.00684	0.00049	
C2r	0.00342	0.07715	0.03418	0.46973	0.01221	0.23340

Table 8: *Pairwise LM score Mann-Whitney-Wilcoxon tests, computed using "learning games" only. These are the tests reported in the main paper. In this table, as in the main paper, p-values are uncorrected for multiple comparisons. Table 9 reports results from these tests corrected for multiple comparisons by the method of Holm (1979)*

	Re	EWA	stEWA	C1	C2	C1r
EWA	0.849					
stEWA	0.020	1.000				
C1	0.849	0.906	0.906			
C2	0.028	0.226	0.044	0.627		
C1r	0.879	1.000	1.000	0.109	0.010	
C2r	0.058	0.849	0.444	1.000	0.183	1.000

Table 9: *Pairwise LM score Mann-Whitney-Wilcoxon tests, computed using "learning games" only. In this table, p-values are corrected for multiple comparisons by the method of Holm (1979).*

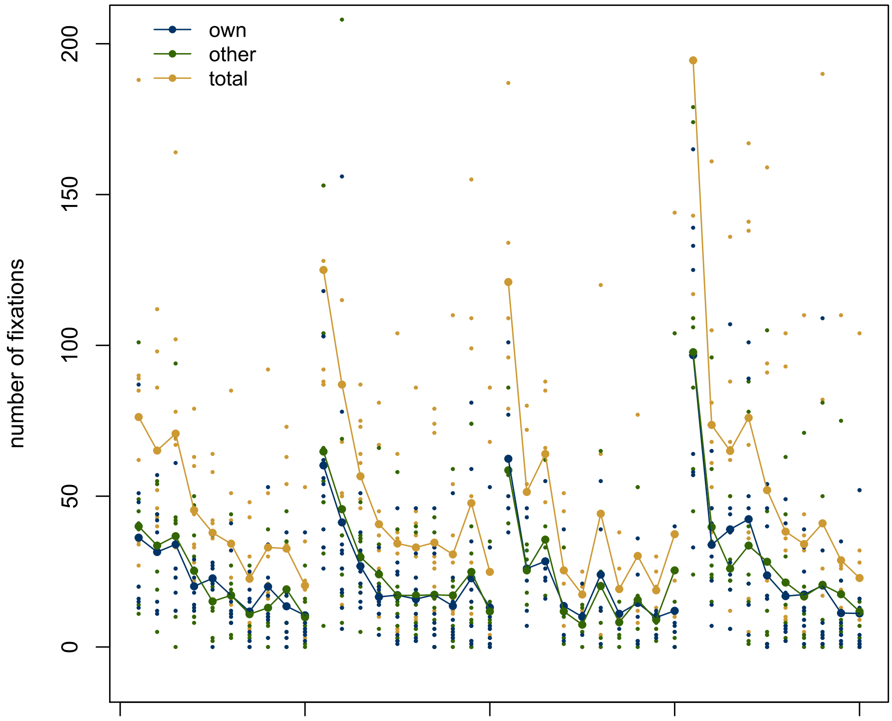

Figure 7: *Number of own payoff matrix lookups, other payoff matrix lookups, and total payoff matrix lookups by game and period, computed using only "learning games".*

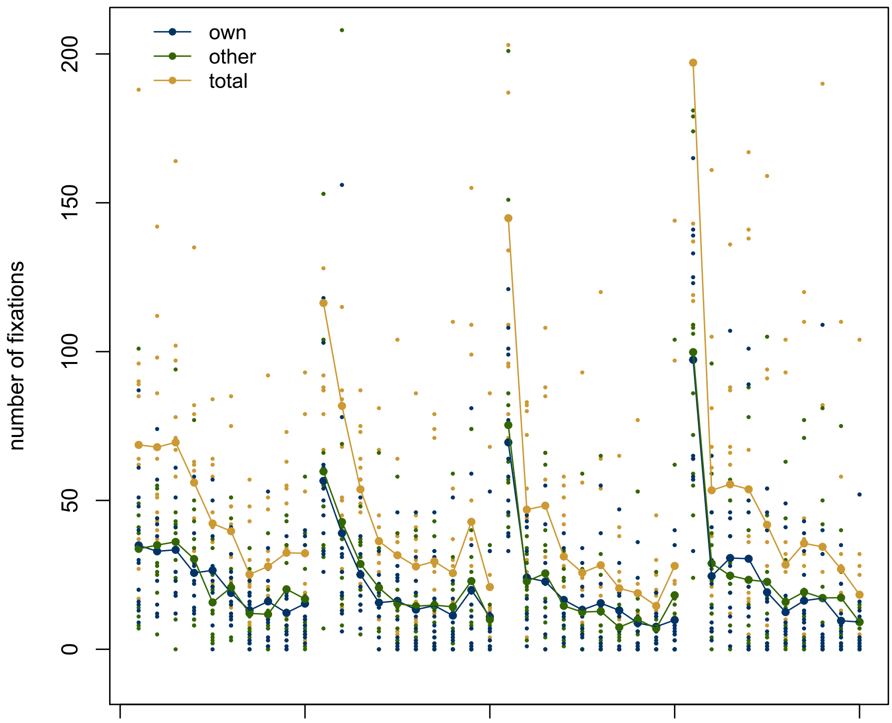

Figure 8: *Number of own payoff matrix lookups, other payoff matrix lookups, and total payoff matrix lookups by game and period, computed using all trials (both "learning games" and "non-learning games").*

A6.4 Eye-tracking Data Breakdown

A6.4.1 Variation by Game and Period

Figure 9 plots the mean LM scores, computed using only "learning games", broken down by game, along with the mean number of lookups for each game. While there is some variation in mean LM scores across games, it does not appear substantial. There is no apparent trend in mean number of lookups across games. Figure 10 plots the mean LM scores, computed using only "learning games", broken down by period, along with the mean number of lookups for each period. There is substantial variation in mean LM scores across periods, but no clear trends in relative scores. As seen above, there is an apparent downward trend in the number of lookups across periods of a game. Figure 11 shows the LM score and number of lookups in each period for a single subject.

A6.4.2 Subjectwise Heterogeneity

An anonymous referee noted the importance of examining subjectwise heterogeneity. To this end, Table 10 contains the mean LM scores and share of lookups on opponents' payoffs for each subject, including only "learning games". Note that four of the twelve subjects have opponent payoffs lookup shares above 50%, four have shares between 40% and 50%, and four have shares below 40%; the minimum share is 27.5%. Thus, all subjects have substantial numbers of lookups in their opponents' payoff matrices, and most look at their opponents' payoffs nearly as much as at their own.

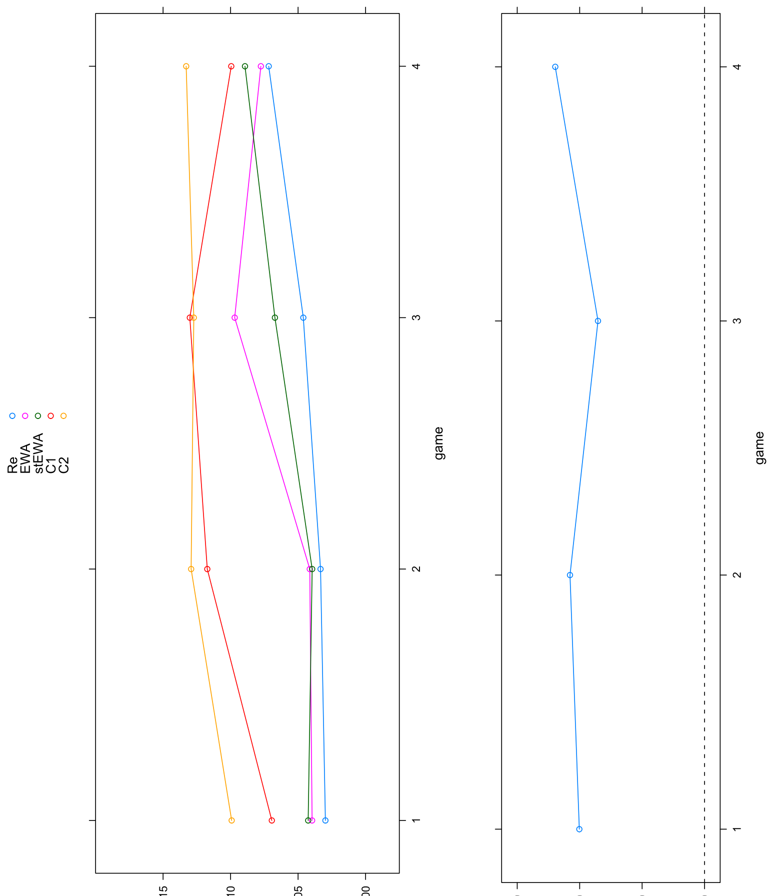

Figure 9: *Mean LM scores and number of lookups by game, computed using "learning games" only. Note that the ordering of the scores is essentially stable across all four games, as is the average number of lookups.*

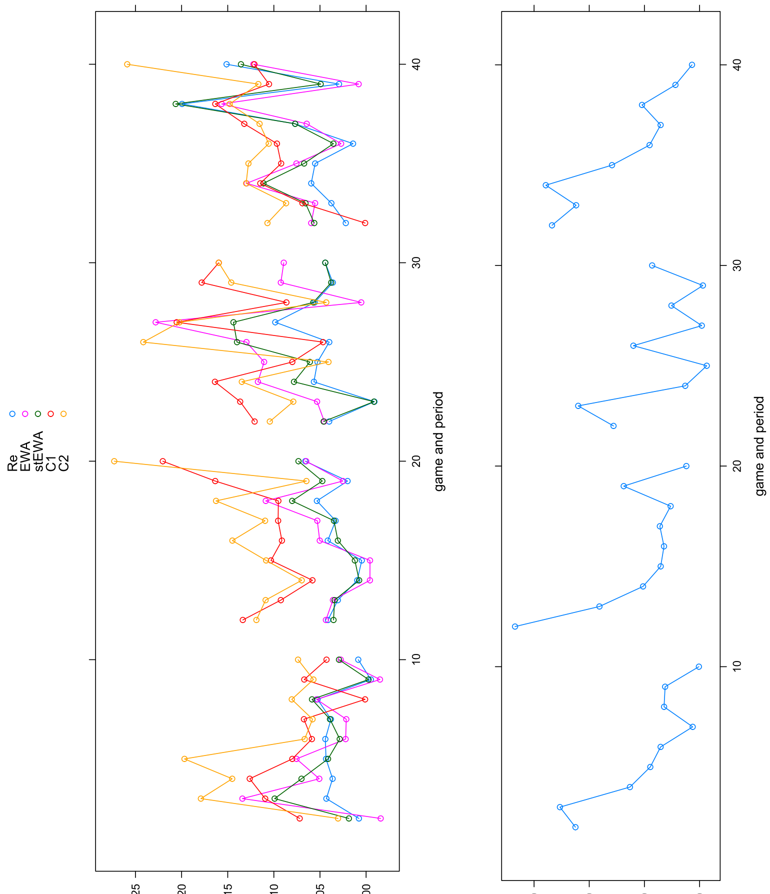

Figure 10: *Mean LM scores and number of lookups by game and period, computed using "learning games" only. Note that there is significant trial-by-trial variability in both LM scores and average number of lookups owing the limited sample size.*

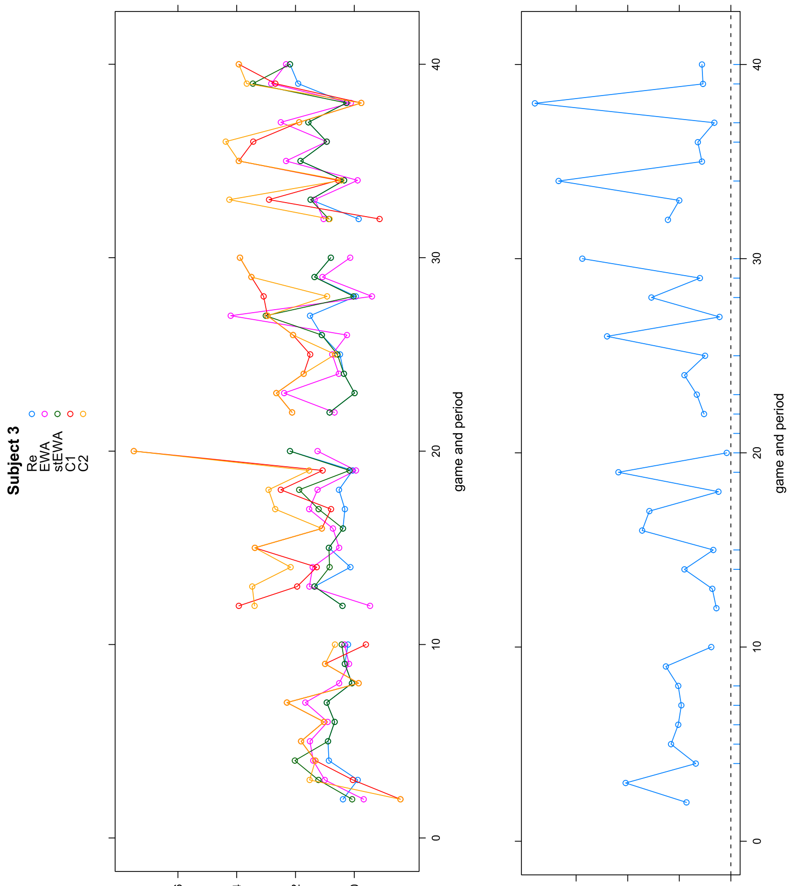

Figure 11: *Raw single-subject LM scores and number of lookups, by game and period, provided as an example of an individual subject's results. Compare with Figure 10 above.*

subject	Re	EWA	stEWA	C1	C2	C1r	C2r	opponent payoffs looking share
2	0.030	0.036	0.041	0.084	0.103	0.077	0.072	46.1%
3	0.076	0.104	0.103	0.196	0.221	0.165	0.198	46.4%
5	0.022	0.029	0.038	0.046	0.073	0.030	0.035	46.1%
6	0.009	0.034	0.017	0.075	0.102	0.076	0.080	56.4%
7	0.027	-0.005	0.025	0.093	0.101	0.047	0.101	57.3%
8	0.019	0.036	0.022	0.076	0.067	0.062	0.050	59.4%
9	0.043	0.096	0.068	0.101	0.126	0.070	0.082	36.5%
10	0.062	0.114	0.070	-0.041	0.044	-0.100	0.057	27.5%
11	0.080	0.091	0.099	0.018	0.094	0.037	0.111	31.8%
12	0.043	0.058	0.047	0.090	0.099	0.062	0.088	46.3%
13	0.068	0.057	0.080	0.167	0.217	0.134	0.121	61.9%
14	0.067	0.058	0.077	0.202	0.124	0.085	0.037	36.3%
mean	0.045	0.059	0.057	0.092	0.114	0.062	0.086	46.0%

Table 10: *Mean LM scores and share of lookups on opponents' payoffs, by subject, including only "learning games".*

A6.4.3 Breakdown by Time Within a Trial

We examined the variation in lookup scores within the time course of a trial by computing the scores for moving windows of fixations within the trials.

Specifically, for subject i, game g, and trial t, we have a string of K fixations $f_1^{igt}, \ldots, f_k^{igt}, \ldots, f_K^{igt}$; for each fixation compute its order percentile $q_k^{igt} = 100 * k/K$. We define a moving window of size s as $W_s = [0, s]$. Then, taking samples $U = \{1, \ldots, 100 - s\}$, we compute the LM scores for each $u \in U$ taking only fixations for which $q_k^{igt} \in u + W_s$. The scores are aggregated within subjects by taking an unweighted mean, and aggregated between subjects by taking another unweighted mean.[7] We also estimate the standard error of this latter mean.

The moving window LM scores (±1 standard error) for the five models are graphed in Figure 12 with window size 20%. Figure 13 shows the corresponding LM scores for two area masks of interest: the own payoff matrix (labeled "own") and the own payoff matrix row corresponding to the period's choice (labeled "own choice, own matrix"). This latter figure shows an important characteristic of the lookup data: near the end of each trial, there is a substantial tendency for lookups to be in the own payoff matrix; furthermore, there is a strong bias towards the row in the own payoff matrix corresponding to the strategy chosen in the period. This was most likely produced by the task design—since eye-tracked subjects chose their strategies by fixating on choice boxes in line with the strategy rows to the left of the own payoff matrix on the screen, their gaze must move leftwards prior to their choice. Furthermore, any difficulty in triggering the gaze-contingent selection of their strategy (due to poor calibration of the eye-tracker, for instance) is likely to produce fixations nearby the choice box; with enough error in the gaze location, these fixations may overlap the own payoff matrix (most likely in the row corresponding to their choice).

Seminar participants at the 2007 North American meeting of the Economic Science Association suggested that subjects implementing adaptive-type learning might produce lookup scores unfavorable to these models if they make the relevant lookups quickly, early in the trial, and then make lookups randomly or driven by "curiosity" for the remainder of the period. Examining Figure 12 in light of the above issues, we find that the order of the models' LM scores is generally maintained within the time course of a trial; the drop in C1 (and slight drop in C2) towards the end can be explained by the discussion above. There appears to be a downward trend in adaptive scores in the first quarter of a trial, but this trend is matched in the C1 and C2 scores. The stability of our ordinal conclusions across the time course within a trial provides some slight evidence against the suggestion that the "true" types are obscured by non-learning looking during the later parts of a trial.

There are other arguments against this suggestion, of course. Since each trial ends when all subjects have entered their strategy choice, there is a disincentive to spend extra time on irrelevant looking. Furthermore, uniform random looking (or looking that is uncorrelated with the locations of the model-relevant

[7] This does not weight by the number of observations (trials) each subject has in the included subset.

information) should merely decrease the difference in observed LM scores. Also, one wonders why a subject would be "curious" about other payoff information but not use this payoff information in choosing their strategy.

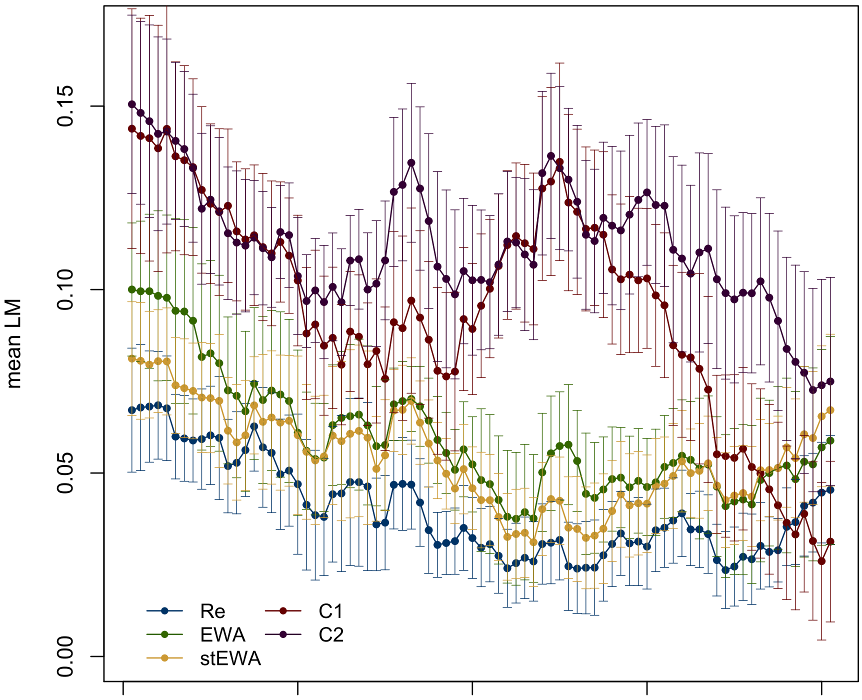

Figure 12: *Moving window analysis of LM scores within each trial by fixation index %, with window size 20%. The bars are ±1 standard error.*

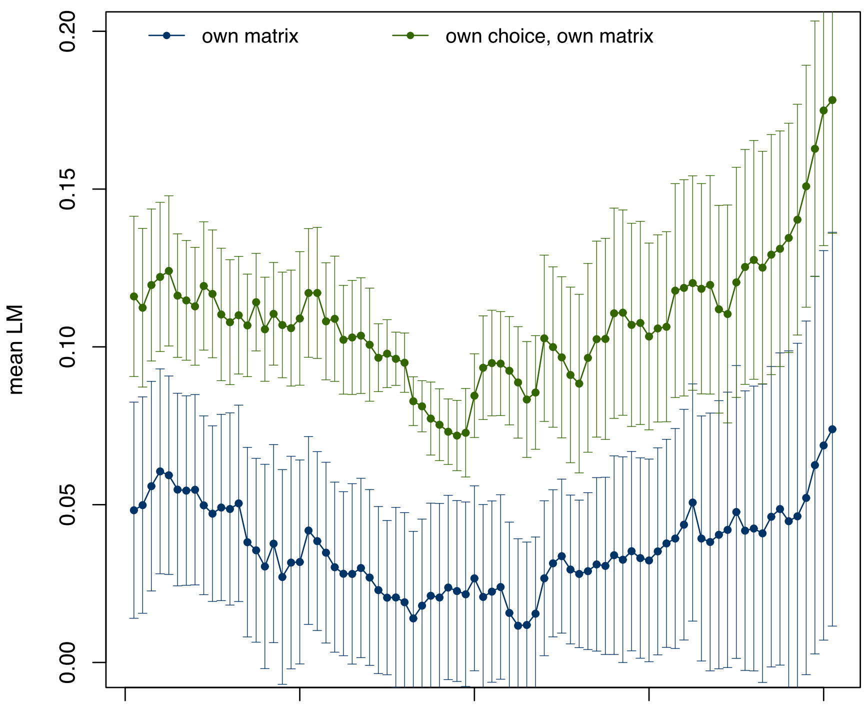

Figure 13: *Moving window analysis of the LM scores corresponding to the own payoff matrix ("own") and the own payoff matrix row corresponding to the period's choice ("own choice, own matrix") within each trial by fixation index %, with window size 20%. The bars are* ± 1 *standard error.*

A6.4.4 Breakdown by Number of Lookups

It must be emphasized that our analysis takes the number of lookups to be exogenous. We lack a framework for determining the number of lookups implied by various models or player-types, especially in the context of eye-tracking. While we might suppose that implementing a given model *requires* looking up a certain set of information, this assumption would oblige us to place a great deal of confidence in our ability to identify information lookups from our eye-tracking data, and especially our ability to infer that the absence of an information lookup of a payoff matrix cell from the absence of a recorded fixation in a payoff matrix cell. Instead, we consider our eye-tracking data as representing a rough measure of the weight of visual attention on different pieces of information, and believe that the data provides useful results when it is combined across trials.

We can, however, examine whether our results change if we look at subsets of the data with different numbers of lookups. Figures 14 and 15 display the mean LM scores for the models as a function of a minimum lookups cutoff. For value n on the x-axis, the scores are computing excluding trials with number of payoff matrix lookups less than or equal to n. Figure 14 shows limited variation in mean raw LM scores and rankings with varying n, especially for reasonable cutoff choices (as can be seen in the bottom graph, if we were to increase n to 40 and above, we would be discarding more than half of our observed periods).

Figure 15 shows important trends for the Reinforcement-enhanced LM scores discussed above in A6.2.2. Ignoring the restricted-Ck scores for a moment, note that all of the [unrestricted] models' scores decline steeply as n increases from zero. This is because of two facts: all of the models gain from the automatically-added reinforcement cell lookup (with C1 gaining in 64.9% of trials, and the other models gaining in every trial), and the contribution of this lookup is approximately proportional to the reciprocal of the number of actual lookups in a period. This first fact is especially notable—recall that Re, stEWA, EWA, and C2 all predict relevance of the reinforcement cell $\pi_{ig}(s_{ig}(t-1), s_{-ig}(t-1))$ in every trial. Furthermore, note that C1 predicts relevance of this cell when the previously-chosen strategy is a best response to their opponent's best response to their previous-chosen strategy, e.g., when both the subject and their opponent played their Nash strategy in the previous period. Now, considering the restricted-Ck models, note that the reinforcement cell is predicted to be relevant in the first stage of a C2 player's reasoning but is never predicted to be relevant for a C1 player until the second stage of his or her C1 reasoning. Thus, applying the simple order restriction discussed in 6.2.1 (and ignoring the automatic reinforcement cell lookup with regards to satisfying order restrictions), the automatic reinforcement cell lookup continues to count as a hit for C2r but never counts as a hit for C1r. As a result, the score for C1r for small n is hurt by the automatic reinforcement cell lookup.

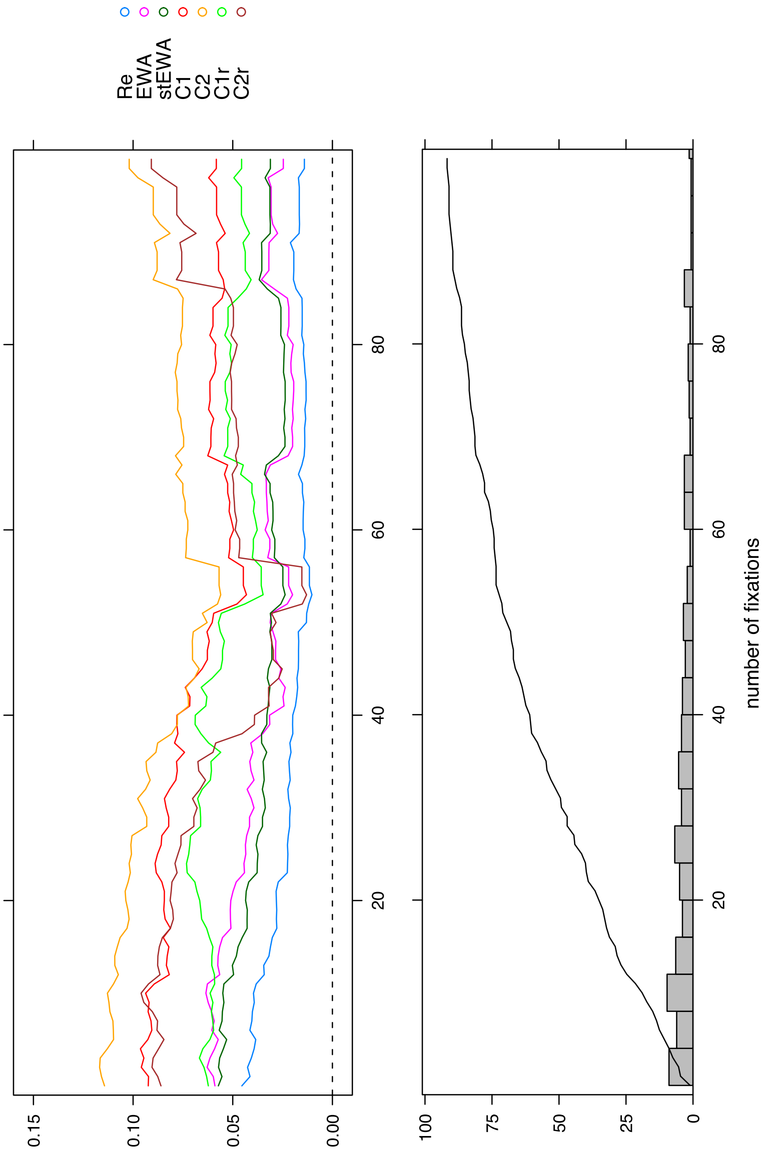

Figure 14: *The top graph displays mean LM scores by minimum number of lookups cutoff; below, the corresponding histogram and empirical CDF for the number of lookups.*

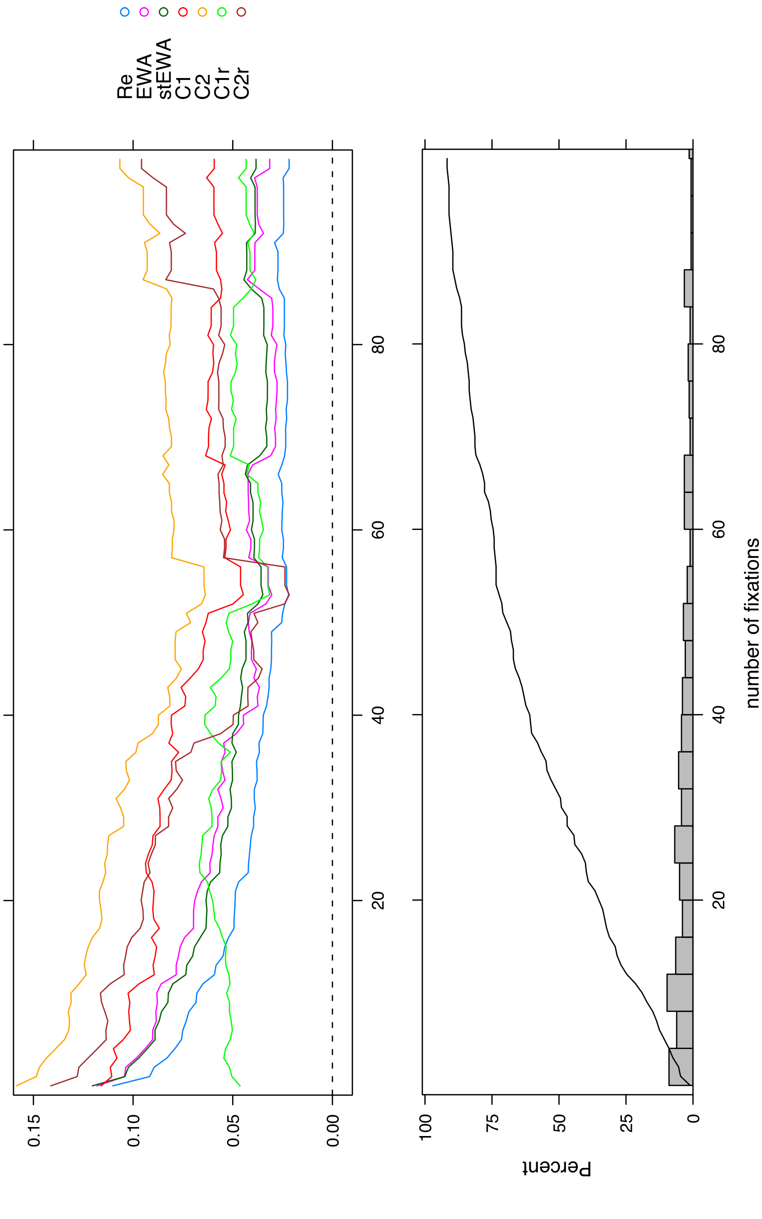

Figure 15: *The top graph displays mean "Reinforcement-enhanced" LM scores by minimum number of lookups cutoff; below, the corresponding histogram and empirical CDF for the number of lookups.*

A6.4.5 Breakdown by Fixation Duration Cutoff

We considered the possibility that fixation duration might carry some useful information about whether a given fixation represents an information lookup. Crawford (2007) notes the exclusion of short lookups (less than 180 ms) in previous mouse-tracking studies of information search in games, suggesting that these lookups are "too short for comprehension"[3]. A comparable threshold for fixations in eye-tracking experiments should be lower, but there is no commonly agreed-upon threshold for these kinds of experiments. Figure 16 displays the model scores as a function of a minimum fixation duration cutoff. The fixations (and thus their durations) are combined when consecutive fixations lie in the same cell (see A6.1). One fact readily apparent in the figure is the absence of any fixations with duration less than 50 ms. This is a consequence of the parameters used in preprocessing of gaze location samples into fixations. Fixations of less than 100 ms in duration could reasonably be considered "short" for the purposes of cognitive studies and are often discarded; here, after preprocessing, they represent only 4.4% of fixations[9]. The graph shows no substantial changes in scores for the range of minimum durations from 0 ms to 250 ms. As a result, we chose to report results without a minimum fixation duration, and believe that results would be nearly identical for any reasonable choice of minimum fixation duration.

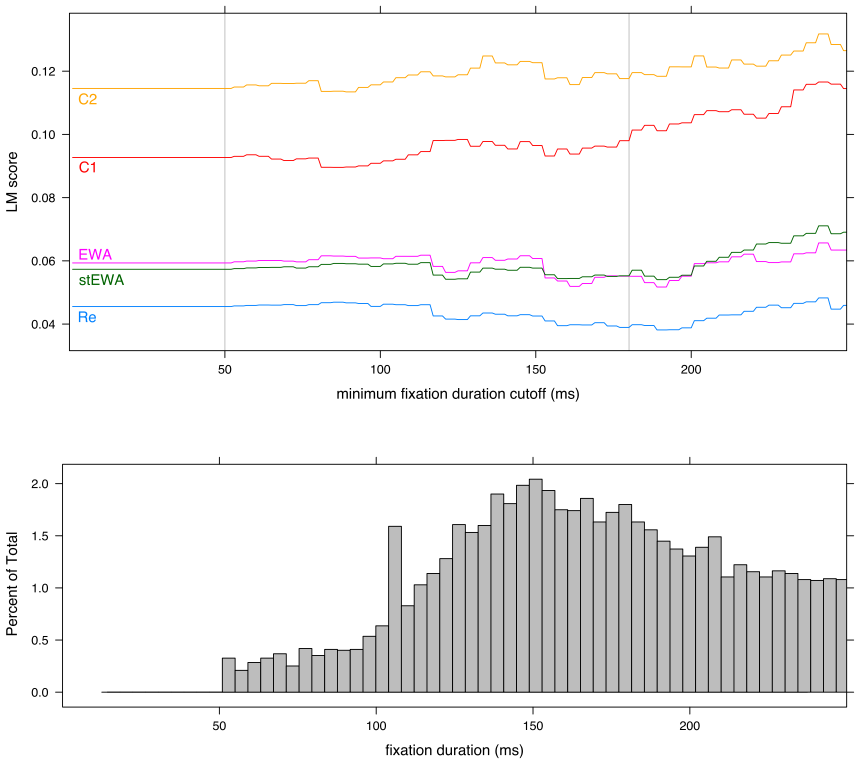

Figure 16: *The top graph displays mean LM scores by minimum fixation duration cutoff; below, the corresponding histogram for fixation durations.*

A6.5 History Looking

As can be seen in Figure 2, the task program for eye-tracked subjects displays history information in boxes at the bottom of the screen. Each box corresponds to a period of the game and contains the subject's chosen strategy, his opponent's chosen strategy, and his own payoff for that period. Unfortunately, the limited display area implies any increase in the space used for history information must decrease the size of the displayed payoff matrices. Thus, by design, the different pieces of history information are not visually separated. As a result, observed patterns of history information acquisition cannot contribute to identification of learning models or features thereof without imposing strong assumptions about the cognitive processes implied by the models. A player observed to be doing frequent history lookups might be implementing belief learning, computing his opponent's choice frequencies; he might be implementing reinforcement learning, recalculating his strategy attractions on the basis of his received payoffs; he might instead be an anticipatory or sophisticated player, examining his opponents' choices and weighting his Ck model prediction accordingly (perhaps along the lines of the geometric or p-switch Ck behavioral variants), or using their observed behavior as hints for finding dominance relationships or equilibria.

While we might have gained additional insights by emphasizing visual separation of history lookup components in our design, some basic analyses of history looking in our data suggests that it is fairly meager. The amount of observed history box looking is limited relative to payoff matrix looking. There are about 17.6 payoff matrix lookups for every lookup of a history box, even including lookups in empty history boxes corresponding to future periods; if we consider only non-empty history boxes (prior period boxes), this ratio increases to about 26.5 payoff matrix lookups per non-empty history box lookup. In addition, one subject *never* looked at any history boxes, empty or non-empty. Furthermore, if we ignore this subject and consider only history box lookups and calculate the subject-wise mean LM scores for the non-empty box area in each period, four of the eleven subjects have negative scores, implying that they looked at empty boxes more often than informative boxes.

A7 Experiment Instructions

EXPERIMENT INSTRUCTIONS

The experiment you are participating in consists of 4 sessions, each having 10 rounds. At the end of the last session, you will be asked to fill out a questionnaire and paid the total amount you have accumulated during the course of the sessions in addition to a $5 show-up fee. Everybody will be paid in private after showing the record sheet. You are under no obligation to tell others how much you earned.

During the experiment all the earnings are denominated in POINTS. Your dollar earnings are determined by a private POINT/$ exchange rate, which will be displayed on the computer screen and is only known to you.

In each round, you will be paired with another participant to form a group. In each group, one participant will be member A, and the other member B. The matching is done randomly and so that you will NOT see the same participant again in the next round. Member A and member B will simultaneously choose an action. Member A will choose R1, R2, R3, or R4, while member B will choose C1, C2, C3, or C4. Each member's earnings depend on the two actions chosen, as shown in the tables displayed on the screen. Note that your POINT/$ exchange rate is different from the other participant, and hence, even if you and the other participant earn the same amount of points, you will NOT earn the same amount of dollars.

Practice Session: 3 Rounds

Session 1: 10 Rounds

Session 2: 10 Rounds

Session 3: 10 Rounds

Session 4: 10 Rounds

Additional Information: Two participants will be wearing eye-tracking device and will participate from another location in campus. Below are special instructions for these participants:

You will wear an eye-tracking device which will track your eye movements. As a result, you will receive an additional $5 beside the normal show up fee. Please make sure you are not wearing contact lenses. You will be seated in front of the computer screen, showing the earnings tables, and make your choice by looking at the R1, R2, R3, R4 boxes on the side of the screen. When looking at a box, it will light up, and will become your choice of action if you keep looking at it for 0.8 seconds.

At the beginning of each session, the experimenter will adjust and calibrate the eye-tracker if needed. The calibration is done by looking at the black dot at the center of the screen, and tracing it around as it appears at different locations. This procedure is done twice to validate your calibration. At the start of each round, you will perform a drift-correction by looking at the center of the screen (black dot) and hit the space bar. Please tell the experimenter if you have any concerns.

NOTE: **Your payments will be rounded up to the next dollar.** Thank you for your participation.

References

[1] David H. Brainard. The psychophysics toolbox. *Spatial Vision*, 10(4):433–436, 1997.

[2] Frans W. Cornelissen, Enno M. Peters, and John Palmer. The eyelink toolbox: Eye tracking with matlab and the psychophysics toolbox. *Behavior Research Methods, Instruments & Computers*, 34(4):613–617, November 2002.

[3] Vincent P. Crawford. Look-ups as the windows of the strategic soul: Studying cognition via information search in game experiments. Working Paper, January 2007.

[4] Urs Fischbacher. z-tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics*, 10(2):171–178, June 2007.

[5] Teck H. Ho, Colin F. Camerer, and Juin-Kuan Chong. Self-tuning experience weighted attraction learning in games. *Journal of Economic Theory*, 133(1):177–198, March 2007.

[6] Sture Holm. A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, 6(2):65–70, 1979.

[7] Denis G. Pelli. The videotoolbox software for visual psychophysics: transforming numbers into movies. *Spatial Vision*, 10(6):437–442, 1997.

[8] Reinhard Selten. Properties of a measure of predictive success. *Mathematical Social Sciences*, 21(2):153–167, April 1991.

[9] SR Research, Ltd. EyeLink II user manual, version 2.12, July 2006.