Journal of Economic Dynamics & Control ▮ (▮▮▮▮) ▮▮▮–▮▮▮

Contents lists available at SciVerse ScienceDirect

Journal of Economic Dynamics & Control

journal homepage: www.elsevier.com/locate/jedc

Learning about learning in games through experimental control of strategic interdependence

Jason Shachat [a], J. Todd Swarthout [b,*]

[a] Wang Yanan Institute for Studies in Economics (WISE), The MOE Key Laboratory in Econometerics, Xiamen University, China
[b] Department of Economics, Georgia State University, Atlanta, GA 30303, USA

ARTICLE INFO

Article history:
Received 21 January 2010
Received in revised form
27 August 2011
Accepted 12 September 2011

JEL classification:
C72
C92
C81

Keywords:
Learning
Repeated games
Experiments
Simulation

ABSTRACT

We report results from an experiment in which humans repeatedly play one of two games against a computer program that follows either a reinforcement or an experience weighted attraction learning algorithm. Our experiment shows these learning algorithms detect exploitable opportunities more sensitively than humans. Also, learning algorithms respond to detected payoff-increasing opportunities systematically; however, the responses are too weak to improve the algorithms' payoffs. Human play against various decision maker types does not vary significantly. These factors lead to a strong linear relationship between the humans' and algorithms' action choice proportions that is suggestive of the algorithms' best response correspondences.

1. Introduction

Researchers have exerted considerable efforts identifying how individuals adjust behavior in repeated decision making tasks. A sizable literature has approached this question by developing hypotheses about the learning process, then embedding the hypotheses in a parametric model, and subsequently estimating the model's parameters from experiments in which human subjects repeatedly play some stage game. The learning-in-games models within this literature have consolidated to a formulation with two main components. The first component is a rule that assigns a value to each of a player's actions conditional upon the history of play, and the second component converts these values into a probability distribution that governs the player's action choice—in effect, a mixed strategy. Three examples of such models are cautious fictitious play (Fudenberg and Levine, 1995), reinforcement learning (Erev and Roth, 1998), and experienced weighted attraction (Camerer and Ho, 1999).

We uncover two important properties of this class of models. First, the models are more sensitive than human subjects at detecting exploitable trends in opponent play. Moreover, the models' action frequencies exhibit strikingly linear adjustments toward best response. Second, these adjustments are too weak to generate any significant gains in payoffs. The probabilistic choice components generate responses that are less aggressive than those made by human subjects in other studies.

* Corresponding author.
E-mail addresses: jason.shachat@gmail.com (J. Shachat), swarthout@gsu.edu (J.T. Swarthout).

0165-1889/$ - see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.jedc.2011.09.007

We adopt a novel research methodology that allows us to identify these properties. As in previous studies, we start with treatments in which each human plays against another human and use these data to estimate parameters of alternative learning models. Then we conduct hybrid treatments in which each human repeatedly plays against a computer implementation of one of the estimated learning models. Finally we generate parallel simulations in which an estimated version of a model plays against itself. Through our hybrid treatments we control for the strategic interdependence of the adaptive behavior of subjects—a main source of the econometric difficulties when estimating these models (Salmon, 2001).

Our methodological approach and analysis is congruent with some of the research in agent-based computational economics. Frequently in agent-based studies, simulation results are sensitive to the behavioral rules assigned to the agents. In an effort to develop appropriate simulation models, agent-based studies often adopt behavioral rules estimated from human experiments.[1] Conducting hybrid studies with humans and agents interacting is a natural extension that is suggested by influential survey articles such as Duffy (2006) and Richiardi et al. (2006). The notion that an appropriate model should generate data in both simulations and human–computer interaction similar to data from pure human interaction is the core idea of our research approach.

A simple ideal motivates our approach: a model is considered the "true" model if human versus computer play is indistinguishable from human versus human play. When differences in play can be identified (which will almost surely happen given the lofty benchmark), these differences should suggest how one can refine the model. This hybrid procedure can also help in developing new learning models: by systematically varying the strategies against which people play, we can obtain descriptions of how humans play against a variety of models (Shachat and Swarthout, 2004; Spiliopoulos, 2008). In addition to identifying how closely a model mimics human behavior, we can also assess the effectiveness of a model by comparing its earnings to those of humans in the same settings.[2]

In our study we consider reinforcement (RE hereafter) and experience-weighted attraction (EWA hereafter) learning models.[3] Since the introduction of these two models over a decade ago, researchers have used variations of the models in a wide assortment of settings. Within the behavioral game theory literature, these models have been successfully applied to behavior in a variety of scenarios: games with unique mixed strategy Nash equilibrium (Mookherjee and Sopher, 1994, 1997), bargaining games (Slonim and Roth, 1998; Grosskopf, 2003), public goods games (Chen and Tang, 1998; Bo and Frechette, 2011), games of asymmetric information (Feltovich, 2000), and games uniquely solved by rationalizability (Ho et al., 1998; Weber, 2003).[4] Beyond this literature, RE and EWA have been used to analyze applied economic problems such as the performance of organ donor matching mechanisms (Zhang, 2010), portfolio allocation and savings decisions (Malmendier and Nagel, 2011; Choi et al., 2009), formation of market price expectations (Heemeijer et al., 2009; Hommes, 2011), and inventory management in supply chains (Bostian et al., 2008; Bolton and Katok, 2008). The widespread success of the RE and EWA models within the behavioral game literature and their adoption by researchers in areas beyond make them natural candidates for our investigation.

We also adopt a pair of 2×2 normal forms games, each having a unique Nash equilibrium that is in mixed strategies. We select these games because of specific properties which benefit our purposes. First, each player has only two strategies, minimizing the number of parameters we need to estimate, and correspondingly increasing the power of our statistical inferences and tests. Next, with a low number of action profiles, we aim to avoid games for which play quickly converges to a pure Nash equilibrium strategy profile, and consequently limits the ability of the two learning models to generate distinct predictions. Hence we choose games with a mixed strategy Nash equilibrium. Further, we chose to avoid games with multiple Nash equilibria, as we want to avoid the confound of equilibrium selection. Finally, we select our games because in pure human play, many cohorts do not converge to the Nash equilibrium within 200 repetitions, and further exhibit interesting non-constant dynamics in choice proportions. This presents an opportunity for the learning algorithms to provide a better fit to the data than static equilibrium notions, such as Nash or Quantal Response Equilibrium (McKelvey and Palfrey, 1995; Selten and Chmura, 2008), which is not often the case with low dimension normal form games with a unique, mixed strategy solution.

We believe this choice of games proved prudent, as we find a number of significant results, summarized as follows:

1. In the human versus human treatments we find that, as is found uniformly across studies, the estimated adaptive rules have significant memory. As a consequence, the impact of recent outcomes on action values (and thus on mixed strategies) is very small. This leads to viscous adjustment rules in simulations. By viscous, we simply mean that the mixed strategies exhibit little change from period to period.

[1] For example, see Andreoni and Miller (1995), Duffy (2001), Markose et al. (2007), and Houser and Kurzban (2002).

[2] This is similar in spirit to the notion of a model's economic value as discussed by Camerer et al. (2002, 2003).

[3] There are many other similarly structured models worthy of studying with our technique, but models in this class tend to generate similar play (Salmon, 2001).

[4] But there are caveats as well: the models fail to capture the heterogeneous behavior exhibited across subjects (Cheung and Friedman, 1997); humans have an ability to detect and exploit intertemporal choice patterns by their opponents which the models do not (Sonsino and Sirota, 2003; Mukherji and Runkle, 2000); and humans are more successful at establishing reciprocal behavior in the repeated game environments (Andreoni and Miller, 1993; Cooper et al., 1996).

Please cite this article as: Shachat, J., Swarthout, J.T., Learning about learning in games through experimental control of strategic interdependence. Journal of Economic Dynamics and Control (2011), doi:10.1016/j.jedc.2011.09.007

2. In the simulations, models generate behavior that correspond to average human play but display less variation. Specifically, we find that, as is found in other studies, the average joint choice frequencies generated by the model simulations correspond to those generated by human pairs. However, the variance of these joint frequencies in the simulations is much lower than we observe in the human treatments.
3. In both simulations and human treatments, there is no correlation in the joint action frequencies across fixed pairs of opponents within the same game treatment.
4. In stark contrast to the previous result, there exists a highly correlated relationship between the joint action frequencies of computer–human pairs in the hybrid treatments. This *linear best response* is a strong linear relationship between the humans' and algorithms' action choice proportions that are suggestive of the algorithms' best response correspondence. Accordingly, when a human's action frequency deviates from the Nash equilibrium proportion, the algorithm's action frequency proportionally adjusts toward its pure strategy best response. This allows us to conclude that the learning models are more sensitive than humans at detecting payoff-increasing opportunities and correspondingly adjusting play whenever a human opponent's action frequency deviates from the Nash equilibrium proportion.
5. While the adjustments by the models are remarkably systematic when pitted against humans, the magnitude of the adjustments is quite weak. Indeed, too weak to result in statistically significant gains in payoffs. In all six computer–human treatments, the average earnings of the models are not statistically significantly more than those of humans in equivalent roles. Moreover, in two of these cases the humans earn statistically significantly more. However, in both of these cases, we show that the probabilistic best response component is the main source of model's lower earnings.

Our results explain and unify some previous findings counter to the implications of the considered learning models. First, the viscous property of the learning models' mixed strategy sequences suggest that a more appropriate model should generate more dramatic changes in period-to-period mixed strategy formulation. One such example is presented by Nyarko and Schotter (2002) (NS hereafter). NS elicit subjects' beliefs of opponents' actions in a repeated game similar to what we use, and observe that subjects wildly revise their elicited beliefs from period to period. NS formulate a model in which the expected values of actions, calculated using the stated beliefs, are incorporated into a probabilistic logit choice rule. After conducting a series of goodness-of-fit exercises, NS conclude this model outperforms both the RE and EWA models. Our results suggest that the NS model gains its effectiveness from its ability to allow for greater period-to-period changes in the likelihoods of action choices. The implied volatile mixed strategies of the NS model are supported when subjects are asked to explicitly choose mixed strategies in studies such as Shachat (2002) and Noussair and Willinger (2003).

Second, our demonstration that probabilistic learning models accurately detect but only weakly best respond to nonequilibrium play is in direct contrast with what humans do. In studies such as Shachat and Swarthout (2004), Fox (1972), and Lieberman (1962), subjects play against non-optimal but unknown mixed strategies in repeated zero-sum games. Subjects only detect these non-optimal strategies if they are far enough removed from the minimax strategies. However, once detected, subjects move decisively toward best response and increase their payoffs.

We next proceed with a discussion of several past studies that incorporate human versus computer game play. We then present the two learning models used in our study. In the fourth section we discuss the games used in our experiment and our experimental procedures. The fifth section covers our experiment results, findings, and interpretations. In conclusion, we integrate our results with other experimental results to provide a summary of human play in games, and suggest future directions for formulation and parameter selection in learning models.

2. Man versus the machine

Human players and computerized decision makers have interacted in a number of previous studies. This technique has been used to identify social preferences in strategic settings (Houser and Kurzban, 2002; McCabe et al., 2001), to establish experimental control over player expectations in games (Roth and Schoumaker, 1983; Winter and Zamir, 2005), and to identify how humans play against particular strategies in games (Walker et al., 1987). Especially noteworthy, Duersch et al. (2010) investigated how human subjects perform against a variety of learning algorithms in Cornout duopoly games, and found that humans regularly increased their earning against most adaptive type algorithms. The source of this exploitation is the vulnerability of adaptive algorithms to be coaxed into predictable choice patterns.

For the remainder of this section, we summarize prior results on how humans play against unique minimax solutions, non-optimal stationary mixed strategies, and variants of the fictitious play dynamic (with deterministic choice rules) in repeated constant-sum games with unique minimax solutions in mixed strategies. Many of the studies we discuss used fixed human–computer pairs playing repetitions of one of the zero-sum games presented in Fig. 1.[5] Studies by Lieberman (1962), Messick (1967), Fox (1972), and Shachat et al. (2011) all contain treatments with humans playing against an experimenter-implemented minimax strategy. In these studies, the human participants were not informed of the explicit

[5] In some of these studies the experimenters implemented stationary mixed strategies by using pre-selected computer generated random sequences in their non-computerized experiments.

Please cite this article as: Shachat, J., Swarthout, J.T., Learning about learning in games through experimental control of strategic interdependence. Journal of Economic Dynamics and Control (2011), doi:10.1016/j.jedc.2011.09.007

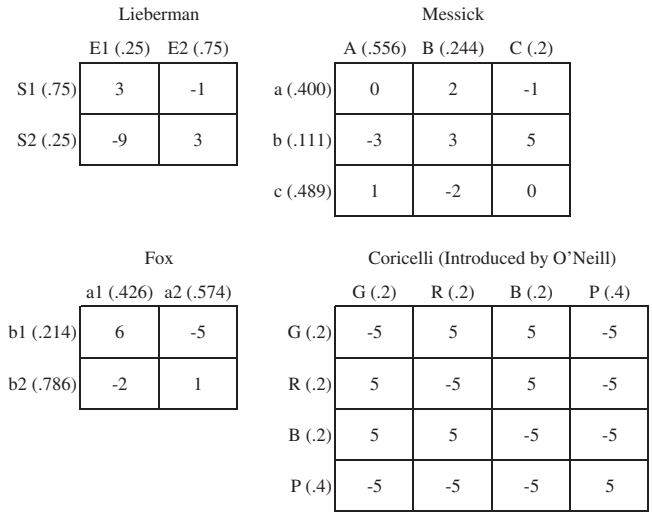


Fig. 1. Zero-sum games used in previous studies (payoffs are for row player, minimax strategy proportions are next to action names).

mixed strategy adopted by their computerized counterparts.[6] All four studies reach the same conclusion: human play does not correspond to the minimax prediction, and only in the Fox study does the human play adjust – albeit weakly – toward the minimax prediction. These results are not surprising: when a "computer" adopts its minimax strategy, the expected payoffs of a human player's actions are all equal.

This indifference is not present when the computer adopts non-minimax mixed strategies. Lieberman (1962) and Fox (1972) also studied human play against non-optimal stationary mixed strategies and discovered that human players do significantly adjust their play (although not to the extent of exclusively playing the pure strategy best response) and also significantly increase – in a statistical sense – their payoffs above minimax value levels. In the relevant Lieberman treatment, subjects played against the experimenter for a total of 200 periods. In the first 100 periods, the experimenter played his minimax strategy of (0.25, 0.75) and then in the final 100 periods the experimenter played a non-minimax strategy of (0.5, 0.5). Human players were not informed that their opponent had adjusted his strategy. Human play adjusted from best responding approximately 20% of the time immediately after the experimenter began non-minimax play, to best responding approximately 70% of the time by the end of the session. This shift toward the best response was also a shift toward the human's minimax strategy, making it difficult to differentiate between the attractiveness of the minimax strategy and the best response.

In one of Fox's treatments, each human participant played 200 periods against a computer which played the non-minimax mixed strategy (0.6, 0.4) for the entire session. This design placed the human's best response of (1, 0) on the opposite side of (0.5, 0.5) from the human's minimax strategy of (0.214, 0.786). Initial human play of their first action was slightly above 50%, and then slowly adjusted toward the pure strategy best response over the course of the experiment. Specifically, human players approached the best response 75% of the time. These experiments demonstrate that human participants will adjust their behavior (but not as much as possible) to take advantage of exploitable stationary mixed strategies. Furthermore, the human subjects in both studies statistically improved their payoffs.

In a study designed to ascertain how far a mixed strategy must deviate from the minimax strategy before humans exploit it, Shachat and Swarthout (2004) systematically vary the fixed mixed strategy across subjects and observe that when a computerized strategy deviates by more than 15% from the minimax strategy of two-thirds, human play begins converging to best response.[7] Furthermore, many subjects in this study adjusted to exclusive play of the best response action – a behavior which was not apparent in the aggregate data presented in the previous studies.

Messick (1967), Coricelli (2001), and Spiliopoulos (2008) all conducted experiments to evaluate how human players respond when playing against variations of fictitious play.[8] These experiments are notable in that the computer's strategy was responsive to the actions selected by its opponent. Messick studied human subjects matched against two fictitious play algorithms: one with unlimited memory and the other with only a five period memory. Against unlimited memory fictitious play, human players earned substantially more than their minimax payoff level. Human players earned an even greater average payoff against limited memory fictitious play. In the study by Coricelli, there are two treatments – both utilizing the game form introduced by O'Neill (1987) – in which human participants play against unlimited memory

[6] When reported, human participants were instructed something similar to: "The computer has been programmed to play so as to make as much money as possible. Its goal in the game is to minimize the amount of money you win and to maximize its own winnings," Messick (1967, p. 35).

[7] This study used the Pursue-Evade game adopted herein and presented in Fig. 2.

[8] In the original formulations of fictitious play (Brown, 1951; Robinson, 1951), a player uses the empirical distribution of the entire history of his opponent's action choices as his belief of the opponent's current mixed strategy and then chooses a best response to this belief.

Please cite this article as: Shachat, J., Swarthout, J.T., Learning about learning in games through experimental control of strategic interdependence. Journal of Economic Dynamics and Control (2011), doi:10.1016/j.jedc.2011.09.007

fictitious play with and without a belief bias. This bias holds that human subjects tend not to repeat their "P" action. In both treatments, human participants win significantly more often against the algorithms than they do against human opponents.[9] Spiliopoulos studied human play against a variety of fictitious play algorithms, including those with varying memory, pattern detection, and aggressive probabilistic choice rules.[10] The main findings were that subjects generally could exploit the algorithms and that subjects play differently conditional on the algorithm they faced. Establishing that humans can "outgame" these algorithms is significant, and has theoretical basis. It is well known that in games with a unique mixed strategy equilibrium, the fictitious play algorithm can generate strong positively serially correlated action choices that are easily exploited (Jordan, 1993; Gjerstad, 1996). It was this speculated vulnerability that partially motivated game theorists to propose and study adaptive learning models which incorporated probabilistic choice as a key component.

To summarize, prior experiments pairing human subjects against algorithms in constant sum games with strictly mixed strategy solutions have taught us that: (1) human players do not tend to play their minimax strategy in response to opponents playing their minimax strategy; (2) human players exploit opponents who play mixed strategies significantly different from their minimax strategy; and (3) human players exploit adaptive algorithms which generate highly serially correlated action choices.

3. Response algorithms

In this section we describe the reinforcement learning model of Erev and Roth (1998) and the experience Weighted Attraction model of Camerer and Ho (1999). Our descriptions of the model formulations and estimation techniques follow the original presentations as close as possible. Nonetheless, we only consider 2×2 games and in some instances we simplify notation without changing the models.

3.1. Reinforcement learning

Erev and Roth's model (hereafter RE) is motivated by the reinforcement hypothesis from psychology: an action's score is incremented by a greater amount when it results in a "positive" outcome rather than a "negative" outcome. More formally, let $R_{ij}(t)$ denote player i's score for his jth action prior to the game at iteration t; let $\sigma_{ij}(t)$ denote the probability that i chooses j at iteration t; and let X_i denote the set of player i's possible stage-game payoffs. The two initial conditions of the dynamic system are: (1) at the initial iteration, each of a player's actions has the same probability of being selected (in the 2×2 case each action is chosen with probability one-half); and (2) the initial score of each actions is

$$R_{ij}(1)=\sigma_{ij}(1)S(1)\overline{X_i},$$

where $S(1)$ is an unobservable strength parameter, which influences the player's sensitivity to subsequent experience, and $\overline{X_i}$ is the absolute value of player i's payoff averaged across all action profiles.

After an iteration, each action's score is updated as follows:

$$R_{ij}(t+1)=(1-\phi)R_{ij}(t)+((1-2\varepsilon)I_{(a_i(t)=j)}+\varepsilon)(\pi_i(j,k)-\min\{X_i\}),$$

where ϕ is an unobservable parameter that discounts past scores, $I_{(a_i(t)=j)}$ is an indicator function for the event that player i selected action j in period t, ε is an unobservable parameter determining the relative impacts on the scores of the selected versus the unselected action, and $\pi_i(j,k)$ is i's payoff when he plays action j against the opponent's action k. Also player i's minimum possible payoff for any action profile, $\min\{X_i\}$, is subtracted from $\pi_i(j,a_{-i}(t))$ as a normalization to avoid negative scores. The second component of the model, a probabilistic choice rule, is specified as

$$\sigma_{ij}(t)=\frac{R_{ij}(t)}{\sum_k R_{ik}(t)}.$$

For each game we consider, parameters of the model are estimated along the lines suggested by Erev and Roth. We estimate the values of $S(1)$, ϕ, and ε by minimizing the mean square error of the predicted proportions of Left play in 20-period trial blocks for the human versus human treatments. More specifically, for each fixed triple of parameter values from a discrete grid we proceed as follows: we simulate the play of 500 fixed pairs engaging in 200 iterations, and then we calculate separately the frequency of Left play by the 500 Row players and by the 500 Column players in each 20-period block. These frequencies are the model's predictions for that triple of parameter values. The grid is then searched for the optimal parameters.

3.2. Experience-weighted attraction

We use the version of EWA developed by Camerer and Ho (1999). While the structure of the EWA formulation is similar to the RE learning model, it adopts a different parametric form of probabilistic choice and it updates actions' scores

[9] Human versus human data for this conclusion are taken from O'Neill (1987) and Shachat (2002).

[10] The computerized decision makers followed the algorithm prescription 80% of the time and played the minimax strategy 20% of the time.

Please cite this article as: Shachat, J., Swarthout, J.T., Learning about learning in games through experimental control of strategic interdependence. Journal of Economic Dynamics and Control (2011), doi:10.1016/j.jedc.2011.09.007

according to what actions actually earned in past play, as well as what actions hypothetically would have earned if they had been played.

According to EWA, subjects choose stage-game actions probabilistically according to the logistic distribution

$$\sigma_{ij}(t) = \frac{e^{\lambda R_{ij}(t)}}{\sum_k e^{\lambda R_{ik}(t)}},$$

where at stage t player i chooses action j with probability $\sigma_{ij}(t)$, where λ is the inverse precision (variance) parameter, and where $R_{ij}(t)$ is a scoring function, as in the RE model, albeit defined (i.e., updated) differently. The updating of $R_{ij}(t)$ involves a discounting factor $N(t)$, which is updated according to $N(t+1) = \rho N(t)+1$ for $t \geqq 1$, where ρ is an unobservable discount parameter and $N(1)$ is an unobservable parameter, interpreted as the strength of experience prior to the beginning of play. The score $R_{ij}(t)$ is updated as follows:

$$R_{ij}(t+1) = \frac{N(t)\phi R_{ij}(t) + ((1-\varepsilon)I_{(a_i(t)=j)} + \varepsilon)\pi_i(j,k)}{N(t+1)},$$

where $\pi_i(j,k)$, ϕ, and ε are interpreted as in the Erev and Roth model. Initial scores $R_{ij}(1)$ for each i and j are additional unobservable parameters.

Parameters of the EWA model are estimated via maximum likelihood. EWA is a flexible specification that includes several other models as special cases. For example, a simple reinforcement learning model, which has a different parametric form than RE, is generated when $N(1)=0$, $\varepsilon = 0$, and $\rho = 0$; and probabilistic fictitious play is generated when $\varepsilon = \rho = \phi = 1$.[11]

4. Experimental procedure

We have three basic steps in our experimental methodology. First, we collect baseline data samples consisting of fixed human versus human pairs that play 100 or 200 rounds of one of two 2×2 games. Second, we estimate parameters for the two learning models separately for each of the two games. In the third step, we conduct new sessions with inexperienced subjects. Each subject is paired with a computer clone programmed with one of the estimated algorithms for his game role, and this pair plays against another human–clone pair. We proceed by describing the two games we use and then present more details on the outlined steps.

4.1. The two games

The first game we consider is a zero-sum asymmetric game called Pursue-Evade. This game was introduced by Rosenthal et al. (2002) (hereafter RSW). The normal form representation of the game is given in Fig. 2. The minimax solution (and Nash equilibrium) of this game is symmetric with each player choosing Left with probability of two-thirds.

There are several reasons why this game is a strong candidate to use in our study. First, zero-sum games eliminate social utility concerns often found in experimental studies of games, thereby mitigating some behavioral effects that might arise if a human suspects he is playing against a computer rather than another human. Second, with some standard behavioral assumptions, the repeated game has a unique Nash equilibrium path which calls for repeated play of the stage game Nash equilibrium. This eliminates potential repeated game effects that the algorithms are not designed to address. Third, Pursue-Evade is a simple game in which the Nash equilibrium predictions differ from equiprobable choice. This provides a powerful test against the alternative hypothesis of equiprobable play.

Our second game poses a more difficult challenge to the learning algorithms. We refer to our second game, presented in Fig. 3, as Gamble-Safe. Each player has a Gamble action (Left for each player) from which he receives a payoff of either two or zero, and a Safe action (Right for each player) which guarantees a payoff of one. This game has a unique mixed strategy in which each player chooses his Left action with probability one-half, and his expected Nash equilibrium payoff is one. Notice that this game is not constant-sum; therefore the minimax solution need not coincide with the Nash equilibrium. In this game, Right is a pure minimax strategy for both players that guarantees a payoff of one. A game for which minimax and Nash equilibrium solutions differ but generate the same expected payoff is called an unprofitable game.[12] The potential attraction of the minimax strategy can (and does) prove to be difficult for the learning algorithms which, loosely speaking, probabilistically best respond.

4.2. Protocols

The laboratory sessions reported in this study were conducted in October and November of 2000 at the dedicated experimental economics laboratory at the IBM TJ Watson research center in Yorktown Heights, NY, and at the University of California, San Diego EconLab in the department of economics. All subjects were undergraduate students at either Pace

[11] We refer the reader to Camerer and Ho (1999) for more discussion of how EWA can emulate various models and for a more complete interpretation of the parameters.

[12] Morgan and Sefton (2002) present a noteworthy study of human play in unprofitable games.

Please cite this article as: Shachat, J., Swarthout, J.T., Learning about learning in games through experimental control of strategic interdependence. Journal of Economic Dynamics and Control (2011), doi:10.1016/j.jedc.2011.09.007

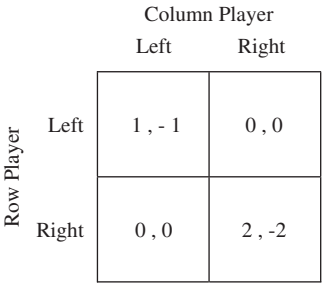


Fig. 2. The Pursue-Evade game.

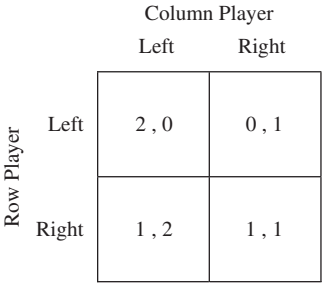


Fig. 3. The Gamble-Safe game.

Table 1
Description of laboratory sessions.

Game	Opponent type	Location	# periods	# subjects
Gamble-Safe	Human	IBM	100	10
Gamble-Safe	Human	IBM	200	14
Gamble-Safe	Human	UCSD	200	6
Gamble-Safe	RE	UCSD	200	24
Pursue-Evade	EWA	IBM	200	6
Pursue-Evade	EWA	UCSD	200	24
Pursue-Evade	RE	UCSD	200	30

Note: Pursue-Evade Human data are from RSW and thus not listed above.

University or UCSD, recruited through flyers posted on campus or from visiting classroom lectures within the economics department or business school. Subjects were paid a show-up fee independent of, and in addition to, their performance in the experiment. Each session was completed in under an hour. At UCSD, the exchange rate for experimental currency to US dollars was ten to one. At IBM, a more generous exchange rate typically of four to one was used. In this case, the compensation was larger due to the inconvenience of traveling to the TJ Watson research facility. As in Rosenthal et al. (2002), subjects in the Evader role of the Pursue-Evade game received an endowment of 300 units of experimental currency. The breakdown of sample sizes based upon game, opponent, location, and number of periods is given in Table 1.[13]

4.2.1. Human versus human baselines

For the human versus human baseline play in the Pursue-Evade game we use the data from RSW. In their hand-run experiment, a pair of subjects was seated on the same side of a table with an opaque screen between the subjects. Each player was given two index cards: one labeled Left and the other labeled Right. At each iteration the players slid their chosen cards face down to the experimenter seated across the table. Then the experimenter simultaneously turned over the cards, executed the payoffs, and recorded the actions. The exchange rate of experimental currency to US dollars was six to one. Twenty pairs of human subjects played this treatment: fourteen for 100 periods and six for 200 periods. Subjects in the Evader role were given endowments of 150 and 300 units of experimental currency for 100 and 200 period sessions, respectively.[14]

[13] We explain in Section 5 why we have no observations for the EWA Gamble-Safe treatment.
[14] While bankruptcy was theoretically possible, no subject came remotely close to a balance of zero in the Evader role.

Please cite this article as: Shachat, J., Swarthout, J.T., Learning about learning in games through experimental control of strategic interdependence. Journal of Economic Dynamics and Control (2011), doi:10.1016/j.jedc.2011.09.007

The human versus human baseline treatment for the Gamble-Safe game was executed via computerized interaction. Each subject was seated at a separate computer terminal such that no subject could observe the screen of any other subject. Subjects began by reading interactive instructions on their computer.[15] Within each fixed pair, each subject played either the Row or Column role for the entire session. Fifteen pairs of subjects participated in this treatment, with five pairs each playing 100 periods and ten pairs each playing 200 periods. At the beginning of each repetition, a subject saw a graphical representation of the game on the screen. A Column player's display of the game was transformed so that he appeared to be a Row player. Thus, each subject selected an action by clicking on a row, and then confirmed his selection. Each subject was free to change his row selection before confirmation. Once an action was confirmed, a subject waited until his opponent also confirmed an action. Then each subject saw the outcome highlighted on his game display, as well as a text message stating both players' actions and his own earnings for that repetition. Finally, at all times a history of past play was displayed to the subject. This history consisted of an ordered list with each row displaying the number of the iteration, the actions selected by both players, and the subject's earnings.

4.2.2. Human versus algorithm treatments

In creating the procedures for these treatments, we faced the difficult decision of what to tell subjects about their opponents. The basic tenet of our study is to evaluate whether a model will generate indistinguishable data between human versus human interactions and human versus model interactions. For experimental control we desire that a subject, regardless whether he is playing against a human or algorithm, have the same homegrown beliefs regarding the process generating his opponent's play, and also constant preferences over the joint payoffs of game outcomes. Some studies, for example Eckel and Grossman (1996), show that subjects' preferences change significantly when the other player's payoffs go to a third party. Further, Fehr and Tyran (2007) find treatment effects when subjects are told they are playing against a computer implemented strategy rather than another person. And perhaps most striking, McCabe et al. (2001) find that some subjects, in addition to playing differently, exhibit activity in different regions of the brain depending on whether an opponent is known to be a computer or another person. Consequently, given our specific research objectives, we believe it is imperative for us to maintain constant beliefs about opponents across treatments, and so we refer to the human–computer opponent pair using the neutral term "opponent." Specifically, the first screen of the computerized instructions reads,

> Today you will play 200 rounds of a simple game. You have been matched to play against one other opponent. You will play all rounds of this game against this same opponent.

We conducted our hybrid treatments using the same instructions, experimental software, and protocol used for the Gamble-Safe game baseline. In these treatments, each human–computer pair was matched with another human–computer pair at the beginning of the session, with no rematching in subsequent periods. The two human members within a given matching played against each other for the first 23 repetitions of the game.[16] Then, beginning in repetition 24, the two humans stopped playing against each other and for the remainder of the session they each played against the computer member of their opponent pair that implemented either the EWA or RE learning algorithm. We acknowledge that some may disagree with not informing subjects of this switch. However, given our aim to maintain homegrown beliefs, and the findings of the prior literature reviewed above, we concluded at the time that this approach was our best option.

We used an initial phase of human versus human play to minimize the impact of estimated initial score values of actions and focus our evaluation on the dynamics of the algorithm. During the first 23 repetitions, we allowed the action value scores to "prime" themselves with the play generated by the subjects. (Although updating of scores was determined by the parameter estimates obtained from the baseline treatments). That is, even though the response algorithms were not selecting actions during the first 23 repetitions, the scores were still being updated according to the specifications of the previous section. For example, consider the 24th repetition of a game. The human Row player now faces the computer member of the opposing Column pair. Moreover, during the first 23 repetitions, the computer Column player updated the scores associated with Column's actions based on the chosen action profiles of both its human counterpart and opponent pair's human counterpart.

We took steps to ensure that each subject proceeds through the periods with the same natural timing whether he was facing the human or computer member of his opponent pair. We adopted a simple technique to make the "split" seamless from the subjects' perspectives. From period 24 on, the two humans within a matching had no interaction except for the timing of when stage game results were revealed. Specifically, although the computer opponents generated their action choices instantly, stage game results were not revealed until both humans had selected their actions. This protocol preserved the natural timing rhythm established by the humans in the first 23 stage games.

[15] Screenshots of the instructions are available at http://excen.gsu.edu/swarthout/learning. The experimental software is also available upon request from the authors.

[16] The number 23 was chosen arbitrarily.

Please cite this article as: Shachat, J., Swarthout, J.T., Learning about learning in games through experimental control of strategic interdependence. Journal of Economic Dynamics and Control (2011), doi:10.1016/j.jedc.2011.09.007

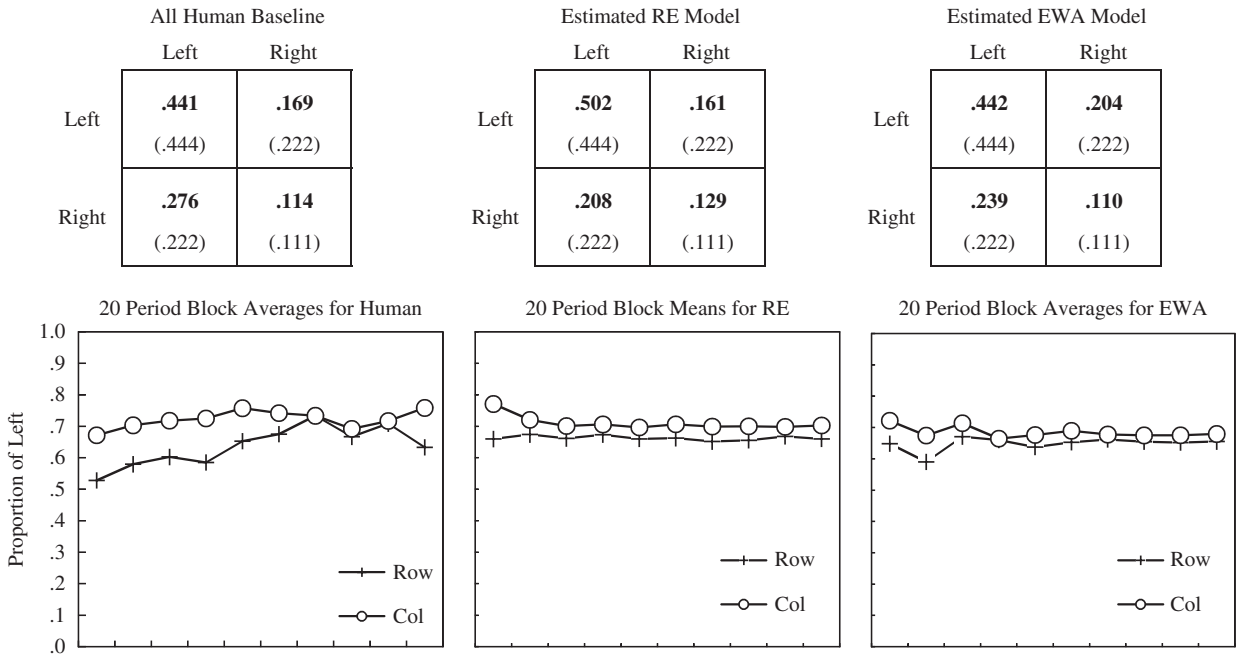


Fig. 4. Baseline data and estimated model summary for Pursue-Evade game.

5. Experimental results

5.1. Baseline experiments, model estimation, and simulation

Our experimental baselines are the human versus human play in each of the two games. Inspection of the aggregate data reveals that play in the two games departs from the Nash equilibrium and the dynamic features of the data suggest non-stationarity of play. After estimating the unobserved parameters of the learning models, we simulated large numbers of sessions based upon these estimated versions of the models. Simulations reveal that the learning models generate aggregate choice frequencies similar to the experimental data, but only weakly mimic the experimental data time series. Furthermore, the simulations do not reveal striking differences between the two learning models.

We use the data from RSW as the Pursue-Evade game baseline data set. Fig. 4 shows contingency tables for the data aggregated across subject pairs and stage games. A graph of the time series of the average proportion of Left play for the Row and Column players is shown below each contingency table. Each observation in a series is the average across a 20 period time block. As noted by RSW, the contingency table is distinctly different from the Nash equilibrium predictions (the numbers in parentheses) and Column subjects play Left significantly more often than the Row subjects.[17] In the block average time series, we see that the Column series almost always lies above the Row series and that both series exhibit an increasing trend. Furthermore, as noted by RSW, there is no convergence to the Nash equilibrium by either player role.

Using these data, RSW estimated the parameters of both the RE and EWA models which are reported in Tables 2 and 3. The estimation results are quite distinct. The estimated RE model reflects a weak strength of initial attraction levels, moderate discounting resulting in long memory, and a very low rate of updating the scores of unselected actions. The estimated EWA model is very different in nature, and nearly coincides with probabilistic fictitious play. Notice that the discount rates (ϕ and ρ) are nearly one – indicating long memory – and both strategies are fully updated by their payoff against the opponent's action, $\lambda = 1$. Also of interest is how the initial estimated attraction levels reflect loss averse preferences. We normalize the initial attractions for the Row and Column players by setting the $R_{RL}(1) = R_{CL}(1) = 0$. Recall the sets of possible payoffs for the Right action are {2,0} and {−2,0} for Row and Column respectively. For Row, this lies in the gain space and the estimated initial attraction is 0.657, but for Column this is in the loss domain and the estimated attraction is −1.863—almost triple the magnitude.

We provide a corresponding analysis for the Gamble-Safe game in Fig. 5. In the contingency table for the baseline data we observe that the Row subjects play Right significantly more than Left, while Column subjects played Left more often. This result partly comes from two pairs in which the Row and Column subjects' action profile sequence eventually converged to the profile (Safe, Gamble). This is evident around the midpoint of the times series for the baseline treatment, where we see the Column and Row subjects' series diverge.

[17] Moreover, the Column subject plays Left more frequently than his Row counterpart in almost all pairs.

Please cite this article as: Shachat, J., Swarthout, J.T., Learning about learning in games through experimental control of strategic interdependence. Journal of Economic Dynamics and Control (2011), doi:10.1016/j.jedc.2011.09.007

Table 2
RE parameter estimates for human vs. human treatments.

	Pursue-Evade	Gamble-Safe
$S(1)$	3	34
ϕ	0.47	0.78
ε	0.045	0.45
Avg. sum squared error	0.205	0.201

Table 3
EWA parameter estimates for Pursue-Evade game.

Parameter	$N(1)$	$R_{RL}(1)$	$R_{RR}(1)$	$R_{CL}(1)$	$R_{CR}(1)$	ρ	ϕ	ε	λ
Estimate	0.833	0.000	0.657	0.000	−1.863	0.993	0.998	1.000	0.578
Std. error	0.071	–	0.075	–	0.054	0.047	0.004	[0.997,1][a]	0.032
Ln Likelihood	−16045								

[a] Estimated using GAUSS and its constrained maximum likelihood package. Note, the constraint $\varepsilon \leq 1$ is binding and thus the normal standard errors invalid. Hence, we report a bootstrapped 99% confidence interval.

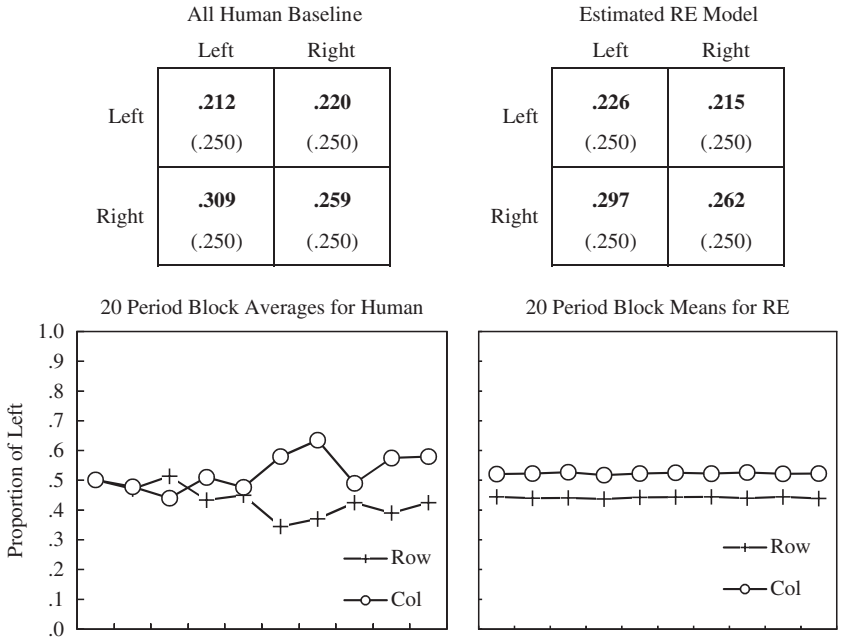


Fig. 5. Baseline data and estimated model summary for Gamble-Safe game.

This convergence to minimax play by the Row subjects in these two pairs is problematic for the maximum likelihood estimation used in the EWA model. Specifically, the long strings of Left by Column leads the EWA model to assign a near zero probability to Right (Safe) by Row for any possible parameter values. However, since Row is repeatedly choosing Right in these instances there is a zero likelihood problem in estimating the EWA parameters. Rather than violate the maximum likelihood criterion for parameter selection specified by Camerer and Ho (1999) we chose not to conduct a Human versus EWA treatment for this game.

Since the parameter selection of the RE model does not rely upon maximum likelihood estimation, we obtain estimates which generate the best fit for the baseline data, which are presented in Table 2. Here the estimated parameters are quite different from those for the Pursue-Evade game. Most notably, the estimated strength of initial reinforcements is much higher, as is the updating of reinforcements for the action not chosen. Also, the decay for action scores is lower, indicating even stronger memory than in the Pursue-Evade game. Further, the estimated RE models perform practically the same under the minimum squared error criteria for the Pursue-Evade and Gamble-Safe games.

Next, we conduct simulations to evaluate model performance. First, we simulate an experiment with 20 pairs playing 200 iterations for each estimated model and game. Second, we calculate the sample proportion and variance of Left play in the simulated and baseline data. These are presented in Table 4. We evaluate the similarity of the first two moments of the

Please cite this article as: Shachat, J., Swarthout, J.T., Learning about learning in games through experimental control of strategic interdependence. Journal of Economic Dynamics and Control (2011), doi:10.1016/j.jedc.2011.09.007

Table 4
Sample moment comparison of simulation results with human baseline data.

	Pursue-Evade RE		Pursue-Evade EWA		Gamble-Safe RE	
	Row	Column	Row	Column	Row	Column
% Left human	0.611	0.718	0.611	0.718	0.444	0.510
Sample variance	0.013	0.010	0.013	0.010	0.013	0.021
Observations	15	15	20	20	20	20
% Left simulation	0.683	0.715	0.643	0.688	0.434	0.524
Sample variance	0.003	0.001	0.001	0.001	0.001	0.001
Observations	20	20	20	20	20	20
t-Statistic	−2.555	0.111	−1.224	1.257	0.330	−0.361
p-Value	0.008	0.456	0.117	0.111	0.373	0.362
F-Statistic	10.693	9.029	4.488	8.506	20.068	39.932
p-Value	0.000	0.000	0.001	0.000	0.000	0.000

simulation and experimental data. In rows seven and eight, we report the results of *t*-tests that the sample proportions of Left play are the same. The hypothesis that the proportion for the human data is the same as the simulation is rejected in only one out six cases: the Pursue-Evade Row player RE simulation. Next, we conduct *F*-tests that the sample variance of the Left proportions are the same and report results in the last two rows of Table 4. The hypothesis that the variances are the same is resoundingly rejected in all cases, and clearly the variance in the simulations is lower. This suggests that the learning models accurately track average play but not the variability.

We now consider how well the estimated models track the dynamic features of experimental game data. Using the estimated models, we simulate 10,000 experiments of 20 pairs playing the corresponding game for 200 iterations. Averages from the 10,000 simulated experiments were used to construct contingency tables and time series in the same format as those presented for the baseline data. These results are presented alongside the baseline results in Figs. 4 and 5. For the Pursue-Evade game casual observation suggests that the EWA model generates an expected contingency very close to the human baseline and the RE model more accurately mimics dynamics in the times series. This is not surprising given the respective objective functions used to select model parameters and estimated values. For the Gamble Safe game, we see that the RE contingency table is remarkably similar to the baseline table. However, the predicted RE dynamics are excessively smooth and do not resemble the baseline time series.

Comparison of the experimental data to simulations based upon estimated versions of the learning models suggests that the learning models successfully capture some features of the humans' disequilibrium behavior. However, time series views of the simulation data exhibit less variable dynamics than the experiment data, which suggest that learning models are not as responsive as humans and tend to simply fit aggregate human choice frequencies. We will see that this conclusion could not be further from the truth. In the human–algorithm treatments the learning algorithms demonstrate an acute ability to detect exploitable opportunities and distinctly adjust play. However, these adjustments are too timid to be profitable.

5.2. Analysis of learning algorithm response to opponents' play

Inspection of the pair-level data from human–algorithm treatments reveals that the learning algorithms generate choice frequencies that are linear better responses to the choice frequencies of their human opponents. Each of Figs. 6–8 is a 2×2 array of scatterplot panels. The rows of each panel array correspond to the decision maker type of the Row player: the top row corresponds to the human decision maker and the bottom row corresponds to the computer decision maker. Similarly the columns of each panel array correspond to the decision maker type of the Column player: the left column for human and the right column for computer. Hence the upper left panel is from the human–human baselines, the lower right panel is from the algorithm–algorithm simulations, and the off-diagonal panels are from the human–algorithm treatments. The scatterplots show the proportions of Left play by the Row and Column players in each pair after the first 23 iterations. In the simulation panel we only use the data from the single simulated experiment used to calculate the values in Table 4. Also, each off-diagonal scatterplot displays a trend line, which is obtained by regressing the Computers' proportions of Left on the Humans' proportions of Left.

Inspecting the two main diagonal panels of each figure reveals that both human–human play and pure simulations of model interactions generate uncorrelated "clouds" of joint Left frequencies with the simulations' clouds exhibiting much less dispersion, consistent with the statistics reported in Table 4. The scatterplots of human–algorithm play are dramatically different. In most of the off-diagonal panels the joint frequencies exhibit strong linear correlations. Moreover, the linear relationship suggests that the algorithms' frequencies adjust toward best responding to the frequencies of their human opponents.

These notions are quantitatively expressed in Table 5. In this table we report for each scatterplot the correlation coefficient and a hypothesis test of whether the coefficient is different than zero. For four of the five pure human treatment

Please cite this article as: Shachat, J., Swarthout, J.T., Learning about learning in games through experimental control of strategic interdependence. Journal of Economic Dynamics and Control (2011), doi:10.1016/j.jedc.2011.09.007

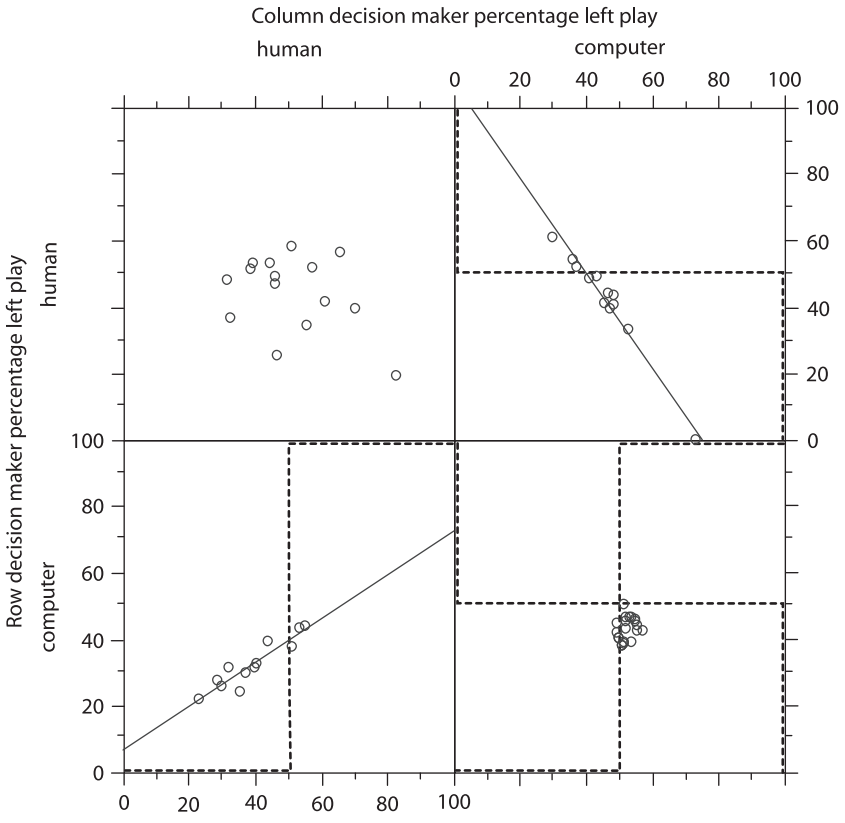


Fig. 6. Gamble-Safe joint densities of proportion Left; RE interactions.

or pure simulation panels we fail to reject zero correlation at the 5% level of significance. However, we get a near opposite result in the human–algorithm treatments. We do reject zero correlation for four out of six of the human–algorithm cases. Moreover, for each of the human–algorithm treatments, the sign of the correlation coefficient is consistent with linear better response by the algorithm.

One example really highlights the result that the algorithms linearly better respond – rather than best respond – to human play. Consider the upper right scatter plot of Fig. 6. In this scatterplot, Column RE players face human Row players in the Gamble-Safe game. One of the human players chose his Minimax strategy, Right, exclusively and his computer RE opponent best responded to this only about 70% of the time. More striking is how all the observations in this panel, including this extreme observation, very closely align with the fitted line.

We further explore the idea that the learning algorithms are better than the human subjects at detecting and adjusting to exploitable play by presenting the OLS results of regressing the learning algorithms' Left frequencies on their human counterparts' Left frequencies.[18] A learning algorithm that is highly sensitive and adjusts systematically to opponents' play should generate regressions that explain a high percentage of the variance of the algorithm's Left frequencies, and the estimated slope coefficient should be consistent with the best response correspondence. These features are found in Table 6 regressions: the slope of each regression has the correct sign, three of the regressions have exceedingly large adjusted R^2 statistics, and a fourth is still quite large considering the data is cross sectional. These adjusted R^2 results reflect the tight clustering to the fitted regression line observed in the scatterplots. Correspondingly, *F*-tests for these four regressions do not reject the significance of the regressions at the 5% level of significance. Interestingly, the two cases where *F*-tests reject the regressions are when the EWA and RE algorithms assume the Column role in the Pursue-Evade game. We do not see a reason for the differential performance, but do note that the mean of the computers' data is close to their minimax strategy in this case.

To summarize, we see that the frequency of Left play by the learning algorithms moves toward (but not all the way to) best response, and the magnitude of these responses by the algorithms is described by a surprisingly predictable linear relationship. So can we conclude that the learning algorithms out play humans?

[18] When running these regression we are now assuming that there is a causality which we did not assume in the correlation analysis. Given the consistency of the correlations with the direction of best response for the computer we feel that the assumption is not too egregious.

Please cite this article as: Shachat, J., Swarthout, J.T., Learning about learning in games through experimental control of strategic interdependence. Journal of Economic Dynamics and Control (2011), doi:10.1016/j.jedc.2011.09.007

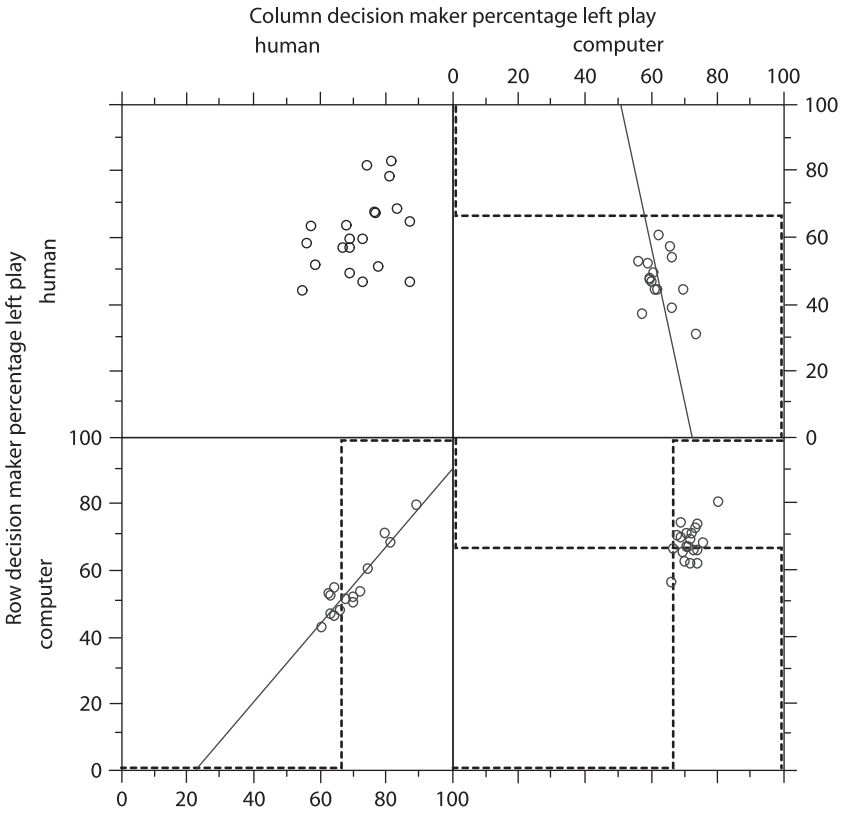


Fig. 7. Pursue-Evade joint densities of proportion Left; RE interactions.

5.3. *Learning algorithms' lack of effective exploitation*

Previous arguments established that the learning algorithms sensitively detect human opponents' exploitable action choice frequencies and then the algorithms respond with linear better responses. However, we will now see that these statistically significant responses are too weak in magnitude to generate statistically significant payoff gains. Table 7 presents the average stage game winnings for all decision maker types when pitted against a human for each role and game. If the learning algorithms successfully exploit human decision makers, we would expect the algorithms in each game and role to have greater winnings than a human when playing against a human in the competing role. The average stage game winnings in Table 7 do not exhibit this trait.

The reported average stage game payoff statistics are calculated by first taking the total session payoffs for each decision maker who plays against a human, and dividing by the number of stage games played. Then we partition these decision makers according to the game played, role played, and decision maker type. Finally, we report the average stage game payoffs across decision makers in each partition. For each game and player role we conduct both a *t*-test, with assumed different variances, and a Mann–Whitney *U*-test for the null hypothesis that on average a computer decision maker earns the same as a human when the opponent is a human. At a 5% level of significance we fail to reject the null hypothesis in four out of the six cases for both tests. In the two rejections, the human average actually exceeds the algorithm average.

These two rejections merit closer inspection because it is tempting to conclude that human subjects are able to "outsmart" the algorithms, even though the learning algorithms appear to be linearly better responding to the opponents more often over the course of play. In fact, this is one of the noteworthy results presented by Duersch et al. (2010). We show this is not the case here, as the rejections arise from the combination of two other factors: (1) the timidness of the algorithm to best respond, as dictated by the probabilistic choice rule; and (2) these are the two cases where Human play differs depending on whether the opponent is an algorithm or another human.

Let us first consider the Gamble-Safe sessions with Human Column players, where we see Human Row players on average earn more than RE Row players. In Fig. 9 we first graph a family of iso-expected payoff curves for the Row player as a function of the joint frequency of Left play by the Row and Column player. Notice that the Row player expects a payoff of one whenever he plays Left with probability zero (i.e., he chooses the safe action) or the Column player plays his Nash equilibrium Left frequency of fifty percent. More importantly, whenever Column plays Left more than 50% of the time

Please cite this article as: Shachat, J., Swarthout, J.T., Learning about learning in games through experimental control of strategic interdependence. Journal of Economic Dynamics and Control (2011), doi:10.1016/j.jedc.2011.09.007

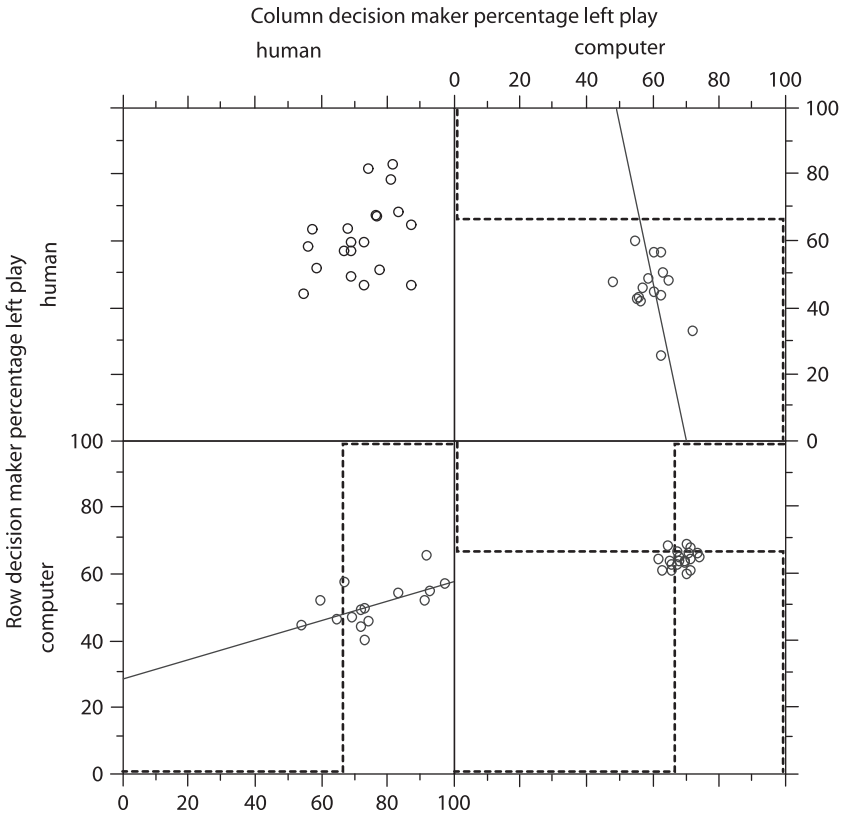


Fig. 8. Pursue-Evade joint densities of proportion Left; EWA interactions.

Table 5
Correlation coefficients by game and interaction factors.

Game	Row player	Column player	Correlation coefficient	*t*-statistic	d.o.f.	*p*-value
Gamble-Safe	Human	Human	−0.379	−1.533	14	0.148
Gamble-Safe	Human	RE	−0.982	−17.441	11	0.000
Gamble-Safe	RE	Human	0.928	8.285	11	0.000
Gamble-Safe	RE	RE	0.270	1.220	19	0.237
Pursue-Evade	Human	Human	0.383	1.805	19	0.087
Pursue-Evade	Human	RE	−0.338	−1.345	14	0.200
Pursue-Evade	RE	Human	0.930	9.450	14	0.000
Pursue-Evade	RE	RE	0.490	2.452	19	0.024
Pursue-Evade	Human	EWA	−0.314	−1.237	15	0.237
Pursue-Evade	EWA	Human	0.581	2.674	14	0.018
Pursue-Evade	EWA	EWA	0.200	0.891	19	0.384

Row's expected payoff is bounded below by one, and whenever Column plays Left less than 50% of the time Row's expected payoff is bounded above by one.

Next we plot the joint frequencies of Left play for Human Row versus Human Column pairs (denoted with open circle marks) and the RE Row versus Human Column pairs (denoted with solid triangle marks). When facing Human Row players, 7 of the 15 Human Column players choose Left more than half the time and thus ensuring that their opponents expected payoff is greater than one. However, the RE Row players typically face less favorable opportunities. When facing RE Row players, 9 out of the 12 Human Column players choose Left less than 50% of the time. In these nine instances, the only way the RE Row player can expect to achieve a payoff equal to the average Human Row payoff of 0.99 would be to exclusively best respond and always choose right. Of course, the probabilistic choice rule prevents consistent best responding.

We next consider the Pursue-Evade sessions with Human Row players, where we see Human Column players on average earn more than EWA Column players. In Fig. 10 we first graph a family of iso-expected payoff curves for the Column player as a function of the joint frequency of Left play by the Row and Column players. We see that the Column player expects to earn their minimax payoff of −0.67 whenever either the Row or Column player plays their minimax Left frequency of two-thirds. We have already observed that the EWA Column players earn less than the minimax payoff,

Please cite this article as: Shachat, J., Swarthout, J.T., Learning about learning in games through experimental control of strategic interdependence. Journal of Economic Dynamics and Control (2011), doi:10.1016/j.jedc.2011.09.007

Table 6
OLS regression results, computer left frequency $= \alpha + \beta *$subject left frequency.

Game	Algorithm role	Human role	α t-stat	β t-stat	Adjusted R-square	F-stat	F-stat p-value
Gamble-Safe	RE Row	Column	0.07 (2.11)	0.66 (7.90)	0.85	62.40	0.00
Gamble-Safe	RE Column	Row	0.75 (40.03)	−0.69 (−16.63)	0.96	276.54	0.00
Pursue-Evade	RE Row	Column	−0.26 (−2.89)	1.16 (9.11)	0.85	82.92	0.00
Pursue-Evade	RE Column	Row	0.72 (9.40)	−0.21 (−1.30)	0.05	1.68	0.22
Pursue-Evade	EWA Row	Column	0.28 (3.24)	0.29 (2.58)	0.29	6.64	0.02
Pursue-Evade	EWA Column	Row	0.69 (8.85)	−0.20 (−1.19)	0.03	1.42	0.25

Table 7
Average stage game payoffs for decision makers when facing a human opponent.

Game	Human role	Human's opponent	Decision maker avg. payoff	t-stat	Approx. d.o.f.	p-value	Mann–Whitney U-stat	p-value
Gamble-Safe	Row	Human Column	1.0776	N/A	N/A	N/A	N/A	N/A
Gamble-Safe	Row	RE Column	1.0786	−0.012	23	0.990	−0.224	0.808
Gamble-Safe	Column	Human Row	0.9888	N/A	N/A	N/A	N/A	N/A
Gamble-Safe	Column	RE Row	0.8983	2.187	25	0.038	2.123	0.034
Pursue-Evade	Row	Human Column	−0.6709	N/A	N/A	N/A	N/A	N/A
Pursue-Evade	Row	RE Column	−0.6829	0.498	32	0.622	0.567	0.571
Pursue-Evade	Row	EWA Column	−0.7205	2.312	33	0.027	2.00	0.046
Pursue-Evade	Column	Human Row	0.6709	N/A	N/A	N/A	N/A	N/A
Pursue-Evade	Column	RE Row	0.6395	1.285	31	0.208	1.533	0.125
Pursue-Evade	Column	EWA Row	0.6395	1.557	32	0.129	1.400	0.161

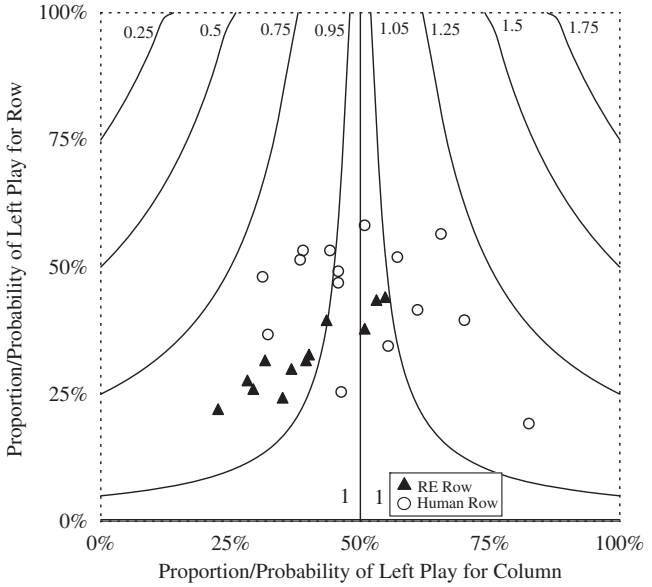


Fig. 9. Row payoff contours and joint frequencies for RE and human Row players versus human Column Payers in Gamble-Safe game.

which is roughly what the Human Column players earn when facing Human Row players. This is also evident in Fig. 10 when we look at the scatterplot of the joint frequencies of Left play for EWA Column versus Human Row pairs (denoted with solid triangle marks). Notice that for all of these pairs the Human Row player chooses Left less than two-thirds of the time, and in all but one case the EWA opponent best responds, by choosing Left, more than half the time.

So why are EWA decision makers earning less than their minimax payoff? From the formulation of the algorithm one sees that if the opponent plays his minimax strategy then both Left and Right will tend to have similar values overtime. Accordingly, the probabilistic choice rule leads the EWA algorithm to play Left and Right with equal frequency. Now as the

Please cite this article as: Shachat, J., Swarthout, J.T., Learning about learning in games through experimental control of strategic interdependence. Journal of Economic Dynamics and Control (2011), doi:10.1016/j.jedc.2011.09.007

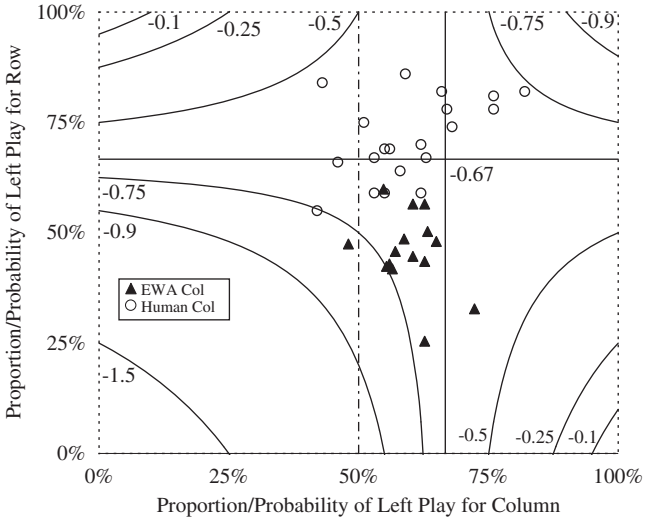


Fig. 10. Column payoff contours and joint frequencies for EWA and human Column players versus human Row payers in Pursue-Evade game.

opponent deviates away from his minimax strategy, the EWA algorithm will deviate from equiprobable play toward the best response. This is what we observe, but in this instance the magnitude of the algorithm's choice frequency adjustment is so small that it fails to approach its minimax strategy of two-thirds. As we see from the location of the level curves, by failing to best respond at least two-thirds of the time the EWA algorithm earns less than the minimax payoff. This surprising result is due to the fact that EWA does not assess payoffs relative to what it can achieve via its minimax strategy and that its probabilistic choice rule leads to weak adjustments that are based relative to the equiprobable mixed strategy.

At this point, the human–algorithm treatments have provided us with the identification of two previously unknown properties of the learning-in-games models: the models adjust linearly toward their best responses to human play, and the adjustments are extremely weak. Now we turn our attention to the question of what the human–algorithm treatments tell us about human subject play.

5.4. Human play conditional on opponent decision maker type

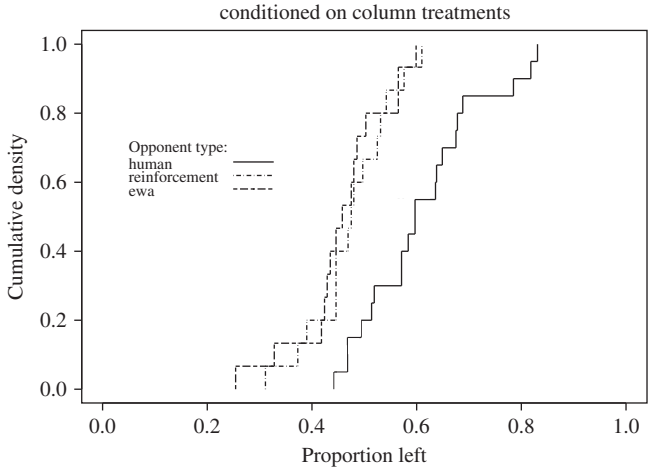
Past studies have demonstrated that humans play differently against Nash equilibrium strategies than they do against other humans. However, we have also presented evidence that play by learning algorithms is more responsive to opponents' decisions than human play is. A natural question to ask is: when unaware of opponent type, do humans play differently against learning algorithms than they do against other humans? To answer this question we compare the empirical distributions of the proportions of Left play by humans when facing the different decision-making types as presented in the scatterplots of Figs. 6–8. We report a series of Kolmogorov–Smirnov two-sample goodness-of-fit tests (hereafter denoted KS) comparing the distributions of Human Left play proportions when facing human opponents to Human Left play proportions when facing the alternative algorithms. The main result is that we do not observe differences in human play except in two cases: when the human is the Row player in the Pursue-Evade game and when the human is the Column player in the Gamble-Safe game. These are the same two cases we just discussed for which subjects out-earned algorithms. Of course, informing the subjects of opponent type may indeed change these results.

Fig. 11 shows the empirical CDFs of proportion of Left play by human Row players as they face human, RE, and EWA Column decision maker types in the Pursue-Evade game. Additionally, the figure reports the results of Kolmogorov–Smirnov tests of whether the Humans' distribution of Left play frequencies differs when facing an algorithm opponent as opposed to a human opponent. Previously we have observed that the learning algorithms performed differently in the Column role of the Pursue-Evade game than in any other situation. This trend continues as the proportions of Left by humans in the Row role are significantly different when facing each learning algorithm than when facing another human.

Next we consider the CDFs generated by human Column players when playing against Human, RE, and EWA Row decision maker types in the Pursue-Evade game. We see in Fig. 12 that play against human opponents is statistically indistinguishable from play against both EWA and RE opponents.

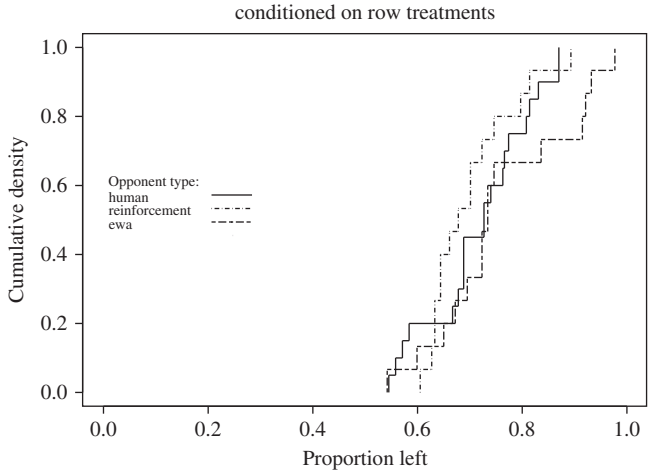
Next, we turn our attention to human play in the Gamble-Safe game. Fig. 13 shows that human Row players' CDFs of proportion of Left play are not statistically different as they face Human and RE Column decision maker types. Finally, the CDFs and associated KS tests generated by human Column players in the Gamble-Safe game are shown in Fig. 14. We see that play against human opponents differs from play against RE opponents at the six-percent level of significance.

Please cite this article as: Shachat, J., Swarthout, J.T., Learning about learning in games through experimental control of strategic interdependence. Journal of Economic Dynamics and Control (2011), doi:10.1016/j.jedc.2011.09.007



Dist. Left when facing Human tested against dist. Left when facing:	KS statistic	P-value
RE	0.567	0.005
EWA	0.633	0.001

Fig. 11. Distributions of left by human Row players in Pursue-Evade.



Dist. Left when facing Human tested against dist. Left when facing:	KS statistic	P-value
RE	0.283	0.435
EWA	0.267	0.507

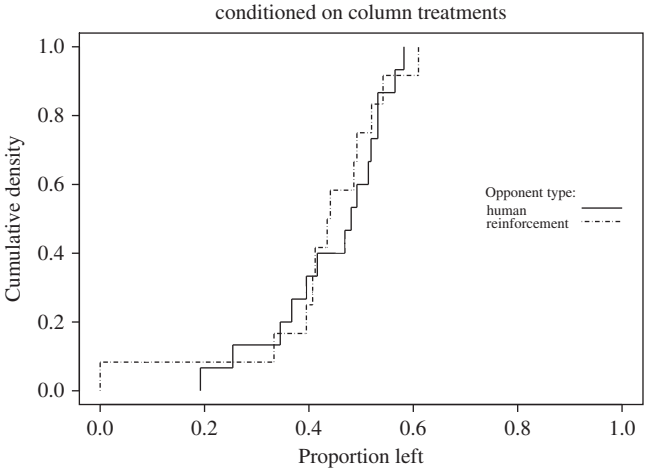
Fig. 12. Distributions of left by human Column players in Pursue-Evade.

6. Discussion

We present an experiment in which humans play games against computer-implemented learning algorithms, and establish that humans neither detect nor exploit the non-stationary, but rather viscous, mixed strategy processes of the RE and EWA algorithms. Our experiment also establishes that the learning models are more sensitive than humans in detecting exploitable opponent play. Furthermore, we show that the learning algorithms' action choice frequencies respond uniformly and linearly to opponents' non-equilibrium action choice frequencies. However, the corresponding mixed strategy adjustments of the learning models in response to detected exploitable play are too weak to increase their payoffs.

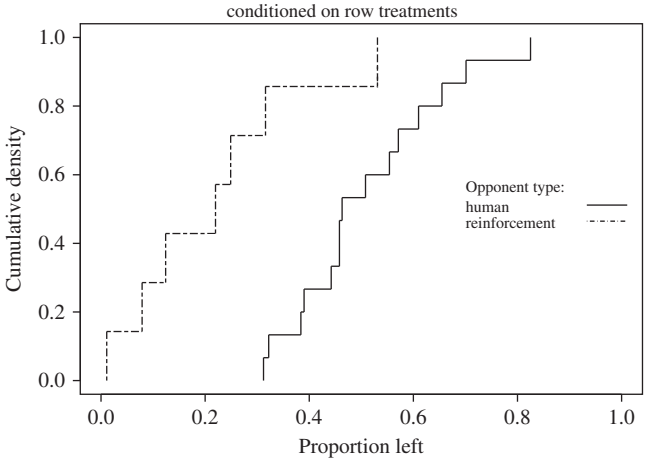
Our results, in conjunction with those of other studies, reveal a different depiction of human learning in games than those suggested by currently proposed models of adaptive behavior. First, through the technique of pitting humans against

Please cite this article as: Shachat, J., Swarthout, J.T., Learning about learning in games through experimental control of strategic interdependence. Journal of Economic Dynamics and Control (2011), doi:10.1016/j.jedc.2011.09.007



Dist. Left when facing Human tested against dist. Left when facing:	KS statistic	P-value
RE	0.183	0.952

Fig. 13. Distributions of left by human Row players in Gamble-Safe.



Dist. Left when facing Human tested against dist. Left when facing:	KS statistic	P-value
RE	0.483	0.061

Fig. 14. Distributions of left by human Column players in Gamble-Safe.

algorithms we know that humans successfully increase their payoffs (but not as much as possible) against non-optimal but stationary mixed strategy play and against adaptive play that generates highly serially correlated action sequences. On the other hand humans do not exploit the subtle dynamic mixed strategy processes of the learning models examined in this paper.

Some sources of behavioral departure between learning models and humans are identified in experiments that elicit subjects' beliefs (Nyarko and Schotter, 2002) or subjects' mixed strategies (Shachat, 2002). Elicited beliefs are highly volatile and often times correspond to a belief that one action will be chosen with certainty. Similarly, elicited mixed strategies show erratic adjustments and a significant amount of pure strategy play.

Despite the deficiencies uncovered here, the RE and EWA models have many applications in the literature, as there are few alternative models which allow for such rich dynamic adjustments of the relative assessments between strategies and corresponding dynamics of proportional choices. So a productive way to proceed would be to ask how can one adjust the formulations of these models to address the issues of muted response to payoff increasing opportunities and the

Please cite this article as: Shachat, J., Swarthout, J.T., Learning about learning in games through experimental control of strategic interdependence. Journal of Economic Dynamics and Control (2011), doi:10.1016/j.jedc.2011.09.007

oversensitivity, relative to human decision makers, of payoff assessments? To this end we suggest four possible directions in formulation and model selection:

1. *Adopt coarser assessment functions*: With a continuous probabilistic choice function, increasing the variance of the random component of payoff assessments will lead to more switching of the most likely chosen action choice, but pushes away from best response to toward equiprobable choice. However, while reducing this variance leads to best response, it also leads to highly positive serially correlated action choices as long as the "observed" component payoff assessments adjust smoothly. Thus, under the probabilistic choice formulation one can try to use assessment functions that are coarse and discontinuous. This is a key element of the success of Nyarko and Schotter (2002) model in which they assess payoffs according to the subjects' elicited beliefs—which display high variance. Of course, this model has the drawback that it is quite unusual to have the needed individual forecast information available.
2. *Change what is probabilistic*: One could formulate a model in which the payoff assessments are random but the choice rule is to best reply with probability one. For example, Shachat and Swarthout (2004) consider a belief based learning model in which each player's stage game beliefs regarding the opponent's action are random variables with a hierarchal probability structure for which the hyper-parameters are functions of the game history. With respect to action choice, each player simply best responds to his realized belief. This model's shortcoming is that it does not make use of one of the lessons taught by RE and EWA that assessment adjustments differ for actions played versus those that were not played.
3. *Incorporate richer individual heterogeneity*: One productive direction in the learning-in-games literature is the incorporation of individual level heterogeneity either exogenously (Camerer and Ho, 1998; Ho et al., 2008) or endogenously (Ho et al., 2007; Spiliopoulos, 2008). These significant advances are difficult to assess with the methodology of this study. In order to apply the techniques here, one would need to estimate and sample from the distribution of the models' random coefficients. This would require a much grander scale experiment than we conducted. We do suggest an alternative way to model heterogeneity. Consider a set of learning rules and treat them as a state space in which a subject's adoption is a latent variable. Then, one can model the dynamics of their adoption as a Markov process. This type of hidden Markov modeling is common in biostatistics and speech recognition, but seldom used in behavioral game theory.
4. *Consider alternative model selection criteria*: The maximum likelihood and minimum square forecast error criteria for estimating EWA and RE parameters lead to strong fits of the mean but not the other moments of experimental data, as we observe when we compare pure model simulation to human experimental outcomes. One alternative would be to estimate parameters that give the best fit of the first and second moments of joint play in either the aggregate or time series analysis. A second alternative would be to consider the same approach but to fit the moments between pure human treatments and the hybrid treatments we consider. Of course, this would be very costly in terms of time and money, and how does one adjust parameter values after each iteration of hybrid play?

In conclusion, this study establishes benchmarks which new learning models should explain. Furthermore, the use of human–algorithm interactions can play an important role in future efforts to identify how humans adapt in strategic environments. First, the technique brings increased power in evaluating proposed models and overcomes some current econometric and numerical limitations. Second, this technique can be used to identify human learning behavior through the adoption of carefully selected algorithms and the subsequent measurement of human responses to these algorithms. In doing so, the algorithms are not being directly evaluated but rather used as carefully chosen stimuli to control for strategic interdependence and produce informative measurements of human behavior.

Acknowledgments

We thank Steven Gjerstad, John Wooders, and Burkhard Schipper for useful discussions. We also thank the Economics Laboratory at the University of California—San Diego for the use of their facility to conduct some of our sessions. We acknowledge support from the IBM TJ Waston Research Center. Jason Shachat also thanks Xiamen University and the Fujian Provincial Government for financial support.

References

Andreoni, J., Miller, J.H., 1995. Auctions with artificial adaptive agents. Games and Economic Behavior 10 (1), 39–64.

Andreoni, J.A., Miller, J.H., 1993. Rational cooperation in the finitely repeated prisoner's dilemma: experimental evidence. Economic Journal 103, 570–585.

Bo, P.D., Frechette, G.R., 2011. The evolution of cooperation in infinitely repeated games: experimental evidence. American Economic Review 101 (February), 411–429.

Bolton, G.E., Katok, E., 2008. Learning by doing in the newsvendor problem: a laboratory investigation of the role of experience and feedback. Manufacturing & Service Operations Management 10 (July), 519–538.

Bostian, A.A., Holt, C.A., Smith, A.M., 2008. Newsvendor "pull-to-center" effect: adaptive learning in a laboratory experiment. Manufacturing & Service Operations Management 10 (4), 590–608.

Brown, G.W., 1951. Iterative solutions of games by fictitious play. in: Koopmans, T.C. (Ed.), Activity Analysis of Production and Allocation, John Wiley.

Please cite this article as: Shachat, J., Swarthout, J.T., Learning about learning in games through experimental control of strategic interdependence. Journal of Economic Dynamics and Control (2011), doi:10.1016/j.jedc.2011.09.007

Camerer, C., Ho, T., Chong, J.-K., 2002. Sophisticated experience-weighted attraction learning and strategic teaching in repeated games. Journal of Economic Theory 104, 137–188.
Camerer, C., Ho, T., Chong, J.-K., 2003. Models of thinking, learning, and teaching in games. American Economic Review 93, 192–195.
Camerer, C., Ho, T.-H., 1998. Experience-weighted attraction learning in coordination games: probability rules, heterogeneity, and time-variation. Journal of Mathematical Psychology 42 (June), 305–326.
Camerer, C.F., Ho, T.-H., 1999. Experience-weighted attraction in games. Econometrica 67, 827–874.
Chen, Y., Tang, F.-F., 1998. Learning and incentive-compatible mechanisms for public goods provision: an experimental study. Journal of Political Economy 106 (3), 633–662.
Cheung, Y.-W., Friedman, D., 1997. Individual learning in normal form games: some laboratory results. Games and Economic Behavior 19, 46–76.
Choi, J.J., Laibson, D., Madrian, B.C., Metrick, A., 2009. Reinforcement learning and savings behavior. Journal of Finance 64 (6), 2515–2534.
Cooper, R., DeJong, D.V., Forsythe, R., Ross, T.W., 1996. Cooperation without reputation: experimental evidence from prisoner's dilemma games. Games and Economic Behavior 12, 187–218.
Coricelli, G., 2001. Strategic interaction in iterated zero-sum games. Technical Report 01–07, University of Arizona.
Duersch, P., Kolb, A., Oechssler, J., Schipper, B.C., 2010. Rage against the machines: how subjects learn to play against computers. Economic Theory 43 (3), 407–430.
Duffy, J., 2001. Learning to speculate: experiments with artificial and real agents. Journal of Economic Dynamics and Control 24 (3), 295–319.
Duffy, J., 2006. Agent-based models and human-subject experiments. In: Tesfatsion, L., Judd, K.L. (Eds.), Handbook of Computational Economics, Agent-Based Computational Economics. Handbooks in Economics Series, vol. 2. , North-Holland/Elsevier, Amsterdam.
Eckel, C.C., Grossman, P.J., 1996. Altruism in anonymous dictator games. Games and Economic Behavior 16, 181–191.
Erev, I., Roth, A.E., 1998. Predicting how people play games: reinforcement learning in experimental games with unique, mixed strategy equilibria. American Economic Review 88, 848–881.
Fehr, E., Tyran, J.-R., 2007. Money illusion and coordination failure. Games and Economic Behavior 58 (2), 246–268.
Feltovich, N., 2000. Reinforcement-based vs. beliefs-based learning models in experimental asymmetric-information games. Econometrica 68, 605–642.
Fox, J., 1972. The learning of strategies in a simple, two-person zero-sum game without saddlepoint. Behavioral Science 17, 300–308.
Fudenberg, D., Levine, D., 1995. Consistency and cautious fictitious play. Journal of Economic Dynamics and Control 19, 1065–1089.
Gjerstad, S., 1996. The rate of convergence of continuous fictitious play. Economic Theory 7, 161–178.
Grosskopf, B., 2003. Reinforcement and directional learning in the ultimatum game with responder competition. Experimental Economics 6 (2), 141–158.
Heemeijer, P., Hommes, C., Sonnemans, J., Tuinstra, J., 2009. Price stability and volatility in markets with positive and negative expectations feedback: an experimental investigation. Journal of Economic Dynamics and Control 33 (5), 1052–1072.
Ho, T., Wang, X., Camerer, C., 2008. Individual differences in EWA learning with partial payoff information. Economic Journal 118 (525), 37–59.
Ho, T.-H., Camerer, C., Weigelt, K., 1998. Iterated dominance and iterated best response in experimental "p-beauty contests". American Economic Review 88 (4), 947–969.
Ho, T.H., Camerer, C.F., Chong, J.-K., 2007. Self-tuning experience weighted attraction learning in games. Journal of Economic Theory 133 (1), 177–198.
Hommes, C., 2011. The heterogeneous expectations hypothesis: some evidence from the lab. Journal of Economic Dynamics and Control 35 (1), 1–24.
Houser, D., Kurzban, R., 2002. Revisiting confusion in public good experiments. American Economic Review 92, 1062–1069.
Jordan, J.S., 1993. Three problems in learning mixed-strategy Nash equilibria. Games and Economic Behavior 5, 368–386.
Lieberman, B., 1962. Experimental studies of conflict in some two-person and three-person games. in: Criswell, J.H., Solomon, H., Suppes, P. (Eds.), Mathematical Methods in Small Group Processes, Stanford University Press, pp. 203–220.
Malmendier, U., Nagel, S., 2011. Depression babies: do macroeconomic experiences affect risk taking? Quarterly Journal of Economics 126 (1), 373–416.
Markose, S., Arifovic, J., Sunder, S., 2007. Advances in experimental and agent-based modelling: asset markets, economic networks, computational mechanism design and evolutionary game dynamics. Journal of Economic Dynamics and Control 31 (6), 1801–1807.
McCabe, K., Houser, D., Ryan, L., Smith, V., Trouard, T., 2001. A functional imaging study of cooperation in two-person reciprocal exchange. Proceedings of the National Academy of Sciences, USA 98 (20), 11832–11835.
McKelvey, R.D., Palfrey, T.R., 1995. Quantal response equilibria for normal form games. Games and Economic Behavior 10 (1), 6–38.
Messick, D.M., 1967. Interdependent decision strategies in zero-sum games: a computer-controlled study. Behavioral Science 12, 33–48.
Mookherjee, D., Sopher, B., 1994. Learning behavior in an experimental matching pennies game. Games and Economic Behavior 7, 62–91.
Mookherjee, D., Sopher, B., 1997. Learning and decision costs in experimental constant sum games. Games and Economic Behavior 19, 97–132.
Morgan, J., Sefton, M., 2002. An experimental investigation of unprofitable games. Games and Economic Behavior 40, 123–146.
Mukherji, A., Runkle, D.E., 2000. Learning to be unpredictable: an experimental study. Federal Reserve Bank of Minneapolis Quarterly Review 24, 14–20.
Noussair, C., Willinger, M., 2003. Efficient mixing and unpredictability in an experimental game. Technical Report, Emory University, unpublished manuscript.
Nyarko, Y., Schotter, A., 2002. An experimental study of belief learning using elicited beliefs. Econometrica 70, 971–1005.
O'Neill, B., 1987. Nonmetric test of the minimax theory of two-person zerosum games. Proceedings of the National Academy of Sciences, USA 84, 2106–2109.
Richiardi, M.G., Leombruni, R., Contini, B., 2006. Exploring a new ExpAce: the complementarities between experimental economics and agent-based computational economics. Journal of Social Complexity 3 (1), 1–9.
Robinson, J., 1951. An iterative method of solving a game. Annals of Mathematics 54, 296–301.
Rosenthal, R.W., Shachat, J., Walker, M., 2002. Hide and seek in Arizona. International Journal of Game Theory 32, 273–293.
Roth, A.E., Schoumaker, F., 1983. Expectations and reputations in bargaining: an experimental study. American Economic Review 73, 362–372.
Salmon, T.C., 2001. An evaluation of econometric models of adaptive learning. Econometrica 69, 1597–1628.
Selten, R., Chmura, T., 2008. Stationary concepts for experimental 2×2-games. American Economic Review 98, 938–966 (29).
Shachat, J., 2002. Mixed strategy play and the minimax hypothesis. Journal of Economic Theory 104, 189–226.
Shachat, J., Swarthout, J.T., 2004. Do we detect and exploit mixed strategy play by opponents? Mathematical Methods of Operations Research 59, 359–373.
Shachat, J., Swarthout, J.T., Wei, L., 2011. Man versus Nash: an experiment on the self-enforcing nature of mixed strategy equilibrium. Working Paper 2011-4, Experimental Economics Center, Georgia State University. Available from: 〈http://excen.gsu.edu/workingpapers/GSU_EXCEN_WP_2011-04.pdf〉.
Slonim, R., Roth, A.E., 1998. Learning in high stakes ultimatum games: an experiment in the Slovak Republic. Econometrica 66 (3), 569–596.
Sonsino, D., Sirota, J., 2003. Strategic pattern recognition—experimental evidence. Games and Economic Behavior 44, 390–411.
Spiliopoulos, L., January 2008. Humans Versus Computer Algorithms in Repeated Mixed Strategy Games. Technical Report 6672. Munich Personal RePEc Archive.
Walker, J.W., Smith, V.L., Cox, J.C., 1987. Bidding behavior in first price sealed bid auctions: use of computerized Nash competitors. Economics Letters 23, 239–244.
Weber, R.A., 2003. 'Learning' with no feedback in a competitive guessing game. Games and Economic Behavior 44 (1), 134–144.
Winter, E., Zamir, S., 2005. An experiment with ultimatum bargaining in a changing environment. Japanese Economic Review 56 (3), 363–385.
Zhang, J., 2010. The sound of silence: observational learning in the U.S. kidney market. Marketing Science 29 (2), 315–335.

Please cite this article as: Shachat, J., Swarthout, J.T., Learning about learning in games through experimental control of strategic interdependence. Journal of Economic Dynamics and Control (2011), doi:10.1016/j.jedc.2011.09.007