*PIGEONS (*COLUMBA LIVIA*) APPROACH NASH EQUILIBRIUM IN EXPERIMENTAL MATCHING PENNIES COMPETITIONS*

FEDERICO SANABRIA[1] AND ERIC THRAILKILL[1,2]

[1]ARIZONA STATE UNIVERSITY
[2]UTAH STATE UNIVERSITY

The game of Matching Pennies (MP), a simplified version of the more popular Rock, Papers, Scissors, schematically represents competitions between organisms with incentives to predict each other's behavior. Optimal performance in iterated MP competitions involves the production of random choice patterns and the detection of nonrandomness in the opponent's choices. The purpose of this study was to replicate systematic deviations from optimal choice observed in humans when playing MP, and to establish whether suboptimal performance was better described by a modified linear learning model or by a more cognitively sophisticated reinforcement-tracking model. Two pairs of pigeons played iterated MP competitions; payoffs for successful choices (e.g., ''Rock'' vs. ''Scissors'') varied within experimental sessions and across experimental conditions, and were signaled by visual stimuli. Pigeons' behavior adjusted to payoff matrices; divergences from optimal play were analogous to those usually demonstrated by humans, except for the tendency of pigeons to persist on prior choices. Suboptimal play was well characterized by a linear learning model of the kind widely used to describe human performance. This linear learning model may thus serve as default account of competitive performance against which the imputation of cognitively sophisticated processes can be evaluated.

Key words: choice, competition, mixed-strategy equilibrium, behavioral economics, model, key peck, pigeons

Psychology, economics, and ethology are concerned with how human and nonhuman organisms choose between alternatives in the context of scarce resources. Choices made by one organism are often linked to those made by others, such that obtaining a resource depends on what others do. In competitive scenarios—where one's gain is another's loss—the chances of obtaining a preferred resource are greatly enhanced if an opponent's choices are anticipated. Consequently, natural selection condemns predictable behavior and favors the detection of nonrandomness among competitors (Miller, 1997). In predator–prey encounters, for instance, unpredictable motor behavior is afforded by inheritable morphological symmetries and sensory-motor systems that allow fast changes of speed and direction (Driver & Humphries, 1988). It is unclear, however, the extent to which individual animals may learn unpredictable behaviors by repeated exposure to predator/prey-associated stimuli, and if such a learning process is general or niche specific.

Mixed strategy games generically describe the kind of competitive situations considered here, where unpredictability is encouraged (Camerer, 2003). This class of interaction is characteristic of the relation between cheetahs and gazelles, goalkeepers and penalty-kickers, cops and robbers, and any other pair of agents where a pursuer must predict the actions of an evasive opponent. Representing competitions as mixed strategy games permits the specification of normative (Nash) equilibria in the distribution of choices made by two opponents. Consider the *Rock, Paper, Scissors* game. If a player chose to always play one alternative (rock, paper, or scissors), she would easily be defeated by an opponent who always made the complementary move (paper, scissors, or rock). Thus, it is not optimal to pick a single alternative unless the opponent picks a single alternative. It is easy to see how, if both players understand the rules of the game and maximize their chances to win, their choice strategy will converge at selecting any alternative

This research was supported by NIMH R01 MH066860 to Peter R. Killeen, to whom we are indebted for his support and advice. We also thank Howard Rachlin for commentaries on early drafts, Weihua Chen for data analysis, and James Mazur, Armando Machado, and an anonymous reviewer for their helpful comments. This study was presented at the 2006 Annual Meeting of the Psychonomic Society, Houston, TX.

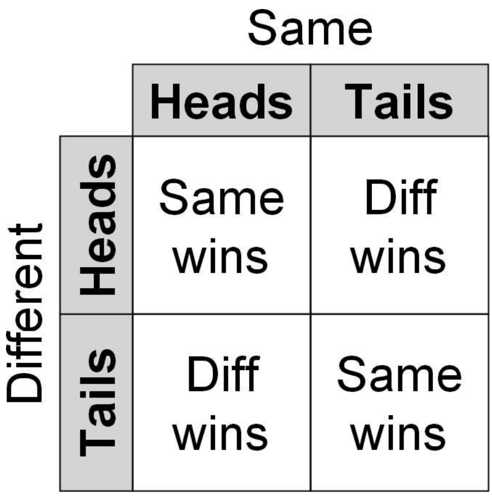
Correspondence concerning this article should be addressed to Federico Sanabria, Department of Psychology, Arizona State University, P.O. Box 871104, Tempe, Arizona 85287-1104, most expeditiously at Federico.Sanabria@asu.edu.

doi: 10.1901/jeab.2009.91-169

randomly with $p = 1/3$; any deviation from this strategy could be exploited by the opponent and reduce one's chances to win. When each player chooses between rock, paper, and scissors with $p = 1/3$ the Nash equilibrium of the game is achieved.

When investigating actual behavior in mixed strategy games, there are two important considerations regarding Nash equilibria. First, perfect randomization is not necessary for optimal performance—as long as the opponent cannot predict a player's move, the player's chances to win are as high as they can be. Although randomization is optimal regardless of the opponent's competence (and thus will serve here as criterion of optimality), mixed strategy games encourage only unpredictable behavior. Second, Nash equilibria specify *what* players would learn if they were maximizing payoffs, but not *how* they would learn it. In this paper, we are concerned with whether Nash equilibria are attained by real players in a mixed strategy game, and how they learn to attain it.

Experimental studies on how individual living organisms actually learn mixed strategies have been focused on human players (for a review, see Camerer, 2003), with only a few studies on other species, mostly primates (Dorris & Glimcher, 2004; Flood, Lendenmann & Rapoport, 1983; Lee, Conroy, McGreevy & Barraclough, 2004; Lee, McGreevy, & Barraclough, 2005). Research on human randomization suggests systematic suboptimal biases, but quasirandom sequences may be learned (Neuringer, 1986). Nonhuman organisms have also shown learned randomization (Machado, 1989). Although the demands of repeated mixed strategy games may enhance randomization in humans (Rapoport & Budescu, 1992), subjects with extensive training systematically fail to produce serially independent choice sequences: Humans tend to overalternate (Brown & Rosenthal, 1990; Towse & Maclachlan, 1999), whereas rhesus monkeys (*Macaca mulatta*) tend to perseverate (Lee et al., 2004). The capacity to detect nonrandom patterns in humans appears to be limited by memory and other computational constraints (Falk & Konold, 1997); comparable data are not yet available from nonhuman subjects. Without a wider base of studies involving nonhuman subjects, it is difficult to establish whether sophisticated cognitive mechanisms, such as top-down executive functions, are necessary for the acquisition of novel and optimal competitive behavior.



		Same	
		Heads	Tails
Different	Heads	Same wins	Diff wins
	Tails	Diff wins	Same wins

Fig. 1. Schematic representation of MP competition. Each of two players (*Same* and *Different*) have a choice between heads and tails. *Same* wins if both players make the same choice; *Different* wins if players make different choices.

We examined how pigeons (*Columba livia*) learn to compete against a conspecific in a mixed strategy game known as *Matching Pennies* (MP), a two-choice version of Rock, Paper, Scissors. Matching Pennies involves two players, each with a penny that can be played *heads* or *tails* and an assigned role as *Same* or *Different*. If both players play the same side of the coin, *Same* keeps both coins; if each player plays a different side of the coin, *Different* keeps both coins (Figure 1). As in Rock, Paper, Scissors, optimal play in iterated MP competitions involves producing unpredictable sequences of choices and detecting nonrandomness in the opponent's choices. This is also the Nash equilibrium of MP: Against a choice-randomizing opponent, the best strategy is to randomize one's own choices.

We evaluated performance of pigeons in MP against two predictions derived from optimality: (1) choices are serially independent, and (2) in each game, choices stabilize at the Nash equilibrium. Then, we compared two learning models as accounts for the performance of pigeons and their deviations from optimality. Finally, we validated the best learning model by using its parameter estimates to simulate MP performance.

METHOD

Subjects

Four adult pigeons (*Columba livia*) were housed individually in a room with a 12:12-hr day: night cycle, with dawn at 0600 hr. They had free access to water and grit in their home cages. The pigeons' running weights were based on 80% of their free-feeding weights. Each pigeon was weighed immediately prior to an experimental session and was excluded from a session if its weight exceeded 8% of its running weight. When required, a supplementary feeding of *ACE-HI* pigeon pellets (Star Milling Co.) was given at the end of each day, at least 12 hr before experimental sessions were conducted. Supplementary feeding amounts were equal to 50% of the average amount fed over the last 15 days, plus 50% of the current deviation from target running weight.

Apparatus

Experimental sessions were conducted in four modular test chambers (305 mm long, 241 mm wide, and 292 mm high), each enclosed in a sound- and light-attenuating box equipped with a ventilating fan. The floor consisted of thin metal bars positioned above a catch pan. The front and rear walls and the ceiling of the experimental chambers were made of clear plastic, and the front wall was hinged and functioned as a door to the chamber. One of the two aluminum side panels served as a test panel. The test panel contained three plastic transparent response keys (25 mm in diameter) aligned horizontally, 70 mm from the ceiling. The keys could be illuminated by white, green and red light emitted from two diodes located behind the keys. A rectangular opening (52 mm wide, 57 mm high) located 20 mm above the floor and centered on the test panel could provide access to milo (grain sorghum) when a grain hopper behind the panel was activated. A house light was mounted 12 mm from the ceiling on the side wall opposite the test panel. The ventilation fan mounted on the rear wall of the sound-attenuating chamber provided masking noise of 60 dB. Experimental events were arranged via a Med-PC® interface connected to a PC controlled by Med-PC IV® software.

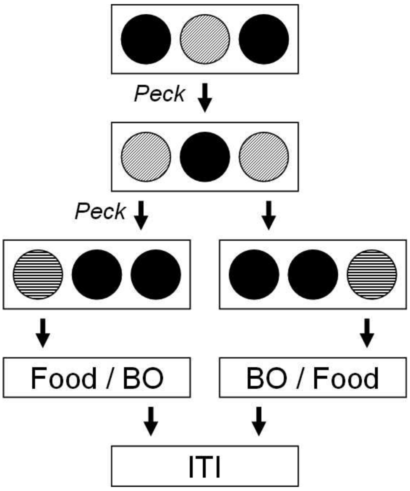


Fig. 2. Choice procedure for unbiasing protocol and MP game. A peck on the center key (illuminated according to block or role) illuminated both side keys. A peck on a side key constituted a choice (left = heads, right = tails); it changed key color to red and extinguished the opposite side key before delivering food or a blackout.

Procedure

Unbiasing protocol. Before starting the experiment proper, we sought to reduce any bias toward pecking either choice key when illuminated with colors used during the experiment. Each of six daily sessions consisted of one block of consecutive ''green'' trials, and one block of consecutive ''white'' trials, presented in random order. Each trial—diagrammed in Figure 2—began with the darkening of the house light and the illumination of the middle key with the assigned color (green or white). A key peck extinguished the middle key and illuminated both choice keys with the assigned color. A key peck on either choice key changed its color to red and extinguished the opposite key. After a 2-s delay, the red-colored key was also extinguished and food—which served as reinforcer—was presented probabilistically for 2.5 s. The probability of reinforcement of each choice was approximately the fraction of total pecks made on the opposite key. A 2.5-s blackout was presented when reinforcement was absent. The house light was illuminated between trials for 2.5 s. ''Color'' blocks alternated after 30 feedings; key colors were not related to reinforcement probability. Sessions ended after 60 feedings or 90 min, whichever happened first.

Table 1

Number of sessions and duration of reinforcers in each game.

	Number of Sessions		Duration of Reinforcers			
			Pigeon A		Pigeon B	
Game	Pigeons A1, B1	Pigeons A2, B2	Heads	Tails	Heads	Tails
1	21	20	*short*	*short*	*short*	*short*
2	32	35	*short*	*LONG*	*short*	*short*
3	33	47	*short*	*LONG*	*short*	*LONG*
4	21	29	*LONG*	*short*	*LONG*	*short*

Note. ''Short'' reinforcers were 2 or 2.5 s of access to food. ''LONG'' reinforcers were 7.5 s of access to food. Unreinforced choices were followed by 2.5-s blackout.

Matching Pennies (MP) game. After completing the unbiasing protocol, pigeons were paired without regard to preexperimental performance, and played a series of MP games exclusively against each other. Choices in these games were arranged similarly to those in the unbiasing protocol (Figure 2). Sessions were divided in two parts. During the first half of each session, one pigeon played the role of *Same*, where reinforcers were delivered only after the subject pecked on the same side as its opponent. Meanwhile, the other pigeon played *Different*, and reinforcers were delivered only after it pecked on the opposite side than its opponent did. Roles were assigned randomly at the beginning of the session and were signaled by the color of both keys. The size of each reinforcer—the *payoff*—was constant within sessions, but varied across experimental conditons, ranging between 2 and 7 s of access to food. The first half of each session ended once any 1 of the 2 pigeons had accumulated at least 80 s of access to food over the session. During the second half of the session, roles were reversed. The second half of the session ended when 1 of the 2 pigeons had accumulated at least 160 s of food access over the session.

Before the beginning of each trial, the house light was turned on for 10 to 12 s. The variability in intertrial interval was used to synchronize the beginning of each trial across pigeons. Each trial began with the extinction of the house light simultaneously with the illumination of the middle key on the response panel. The middle key was illuminated green, when playing *Same*, or white, when playing *Different*. A peck on the middle key extinguished it and illuminated both left and right choice keys (*heads* and *tails*, respectively) according to role. A peck on either choice key turned the pecked key red and extinguished the key not pecked. The chosen key remained red until 2 s after the opponent made its choice, then it was turned off and the reinforcer was delivered if it was due; if it was not due, a 2.5-s blackout was presented instead. If a choice was not made within 10 s from the beginning of the trial, the key lights were turned off, the choice for that round was marked as a forfeit, and a 2.5-s blackout ensued. When the opponent forfeited, either choice was reinforced. Forfeits, which represented 1% of the trials, were excluded from analysis and thus *heads* and *tails* choices added to 100%. Reinforcement and blackouts marked the end of each trial.

Payoffs were varied from game to game by changing the duration of reinforcers. Table 1 indicates the number of sessions played and the duration of reinforcers in each game. In Game 1, the durations of heads and tails reinforcers were identical; in Game 2, the duration of tails reinforcers was increased only for one pigeon (A) in each pair; in Game 3, the duration of tails reinforcers was increased also for pigeon B in each pair; in Game 4, the duration of tails and heads reinforcers were switched for both pigeons in each pair. Unreinforced choices were always followed by a 2.5-s chamber blackout. Note that assignment of pigeons as A or B was not related to role assignment: In each session, both pigeons, A and B, played *Same* and *Different*.

RESULTS

Serial Independence and Choice Perseveration

Optimal choices in iterated Matching Pennies games are unpredictable, and thus should

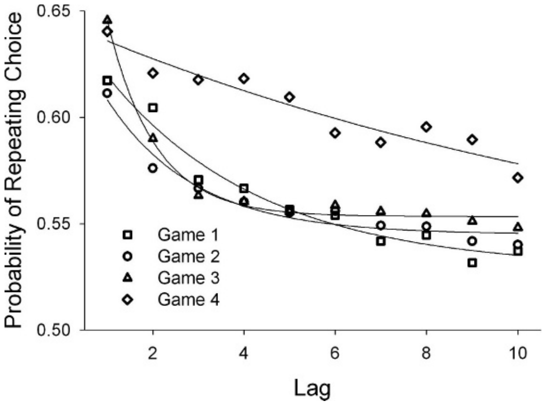


Fig. 3. Probability of choosing the same alternative 1 to 10 trials in the future, in each of the four games. Probability of repeating a choice was calculated as $y = p$(heads in lag 0|heads in lag x) + p(tails in lag 0|tails in lag x). Fitted curves trace exponential decay functions of the form $y = y_0 + ae^{-x}$, with y_0 and a varying across games. Fitting was conducted using SigmaPlot 2004 for Windows 9.01.

be independent of prior choices. To evaluate this expectation, we examined the probability of repeating a choice in future trials. In an unpredictable sequence of choices, the probability of repeating the same choice after any number of trials should be .5. To the extent that this probability is above .5, it demonstrates perseveration in choice; to the extent that it is below .5, it demonstrates overalternation. This measure is insensitive to preference towards an alternative: If a player prefers playing heads over tails, the probability of repeating a heads choice would be over .5, but the probability of repeating a tails choice would be below .5, averaging at .5 if there are no perserveration or overalternation trends.

Figure 3 shows the probability that choice in any trial t would be repeated in trial $t + x$, with x ranging between 1 and 10. The probability of repeating a choice was above .5 in all games, regardless of lag, which indicates a systematic positive influence of past choices on current choices in pigeons. This influence appears to decay exponentially over time, faster in Games 1, 2, and 3 than in Game 4, as shown by the fitted curves in Figure 3. We may therefore represent choice on any given trial as a function of an exponentially weighted moving average (EWMA) of prior choices,

$$p'_{t+1} = \alpha C_t + (1 - \alpha)p'_t, \quad 0 < \alpha < 1 \qquad (1)$$

where the p'_t is the predicted probability of a choice in trial t. For clarity, it will be assumed throughout this paper that the variable of interest is the probability of heads choices; our inferences would apply also to tails choices, because their probability complements the probability of heads choices. Thus, p'_t is the probability of heads in trial t; C_t is 1 if heads was chosen in t, or zero if tails was chosen in t; α, the only free parameter, is the rate of decay of the moving average (1 = no decay, 0 = immediate decay). To illustrate how α was estimated, consider the following example: Suppose a choice of heads in trial 99 was expected with $p'_{99} = .8$, but the actual choice was tails ($C_{99} = 0$). If $\alpha = .5$, the expected choice of heads for trial 100 would be $p'_{100} = .5 \times 0 + .5 \times .8 = .4$; if $\alpha = .1$, $p'_{100} = .1 \times 0 + .9 \times .8 = .72$. Higher values of α indicate that current choices are more predictive of subsequent choices. Fitting α to each pigeon's choices across all four games, using the method of maximum likelihood[1], yielded a median of .074, ranging between .039 and .103 (see EWMA in Table 2). Thus, the typical influence of a choice on the subsequent choice was systematic but relatively small, and it decayed to about half (3.7%) between $t + 1$ and $t + 10$ (i.e., $\alpha(1 - \alpha)^{10-1} \approx 0.5\alpha(1 - \alpha)^{1-1}$).

Nash Equilibrium and Sensitivity to Own Payoffs

The second prediction from optimal iterated play of Matching Pennies is that choices stabilize near the Nash equilibrium. The Nash equilibrium is the probability of choosing heads such that neither player has an incentive to change how its own choices are allocated. The incentive to change choice distribution is absent only when the expected utility (*EU*) obtained from choosing heads or tails is the same. To estimate the stable allocation of choices predicted by Nash equilibrium, it may be assumed that the *EU* of choosing an alternative is the product of the utility of reinforcement (which we expected to covary positively with duration of reinforcers) and the

[1] Maximum likelihood was computed by varying free parameters to fit predictions (p'_t) to data (C_t) on each trial. Fitting was conducted by maximizing

$$\sum \log[C_t p'_t + (1 - C_t)(1 - p'_t)], \qquad (F1)$$

which yields the log-likelihood of the model, the log probability of the data given the best estimates of model parameters.

Table 2

Parameters of EWMA (Eq. 1), SLLp, and WRM.

Parameter	A1	A2	B1	B2
EWMA				
α	.039	.103	.079	.069
SLLp				
α_L	.110	.241	.175	.041
γ	.032	.099	.020	.010
λ_E	.077	.201	.090	.011
LLR*	111.1	265.4	130.6	137.7
WRM				
α_M	.126	.119	.005	.017
s	.308	.404	.632	.666
β_{HEADS}	0.748	1.16	1.12	1.18
LLR*	101.6	135.0	75.3	133.2

* LLRs (log-likelihood ratios) are computed as L(model) − L(EWMA), where $L(Y)$ is the log-likelihood of model Y.

probability of reinforcement. Probability of reinforcement depended on the opponents' choices and on role being played: For example, the probability that a heads choice will be reinforced while playing *Same* is the probability that the opponent will also choose heads. Thus, we can express the *EU* of choosing heads while playing *Same* as $EU_S(H) = p_D U_S(H)$, where p_D is the probability that the opponent (who is playing *Different*) will choose heads, and $U_S(H)$ is the utility of a reinforcer obtained for choosing heads. The subscripts in this equation indicate the player to whom each variable is ascribed: The utility of reinforcement and the expected utility correspond to one player (S in our example), whereas the probability of choosing heads correspond to the opponent (D in our example). At the Nash equilibrium, when *EU*s of playing heads and tails are equal, we may solve for the opponent's p, and this value may be contrasted against data to verify that choices converged at the Nash equilibrium. To solve for p in each role, a more general algebraic statement of the Nash equilibria must be made:

$$\begin{aligned} EU_D(H)&=EU_D(T)=(1-p_S)U_D(H)=p_S U_D(T) \\ EU_S(H)&=EU_S(T)=p_D U_S(H)=(1-p_D)U_S(T) \end{aligned} . \quad (2)$$

The first line of Equation 2 may be read as "while playing *Different*, the expected utility of choosing heads or tails is the same, which means that the probability that the opponent will choose tails ($1 - p_S$), times the utility of a reinforced heads choice, is equal to the probability that the opponent will choose heads times the utility of a reinforced tails choice." The second line may be read analogously. From these equations p_S and p_D may be solved:

$$\begin{aligned} p_S &= \frac{U_D(H)}{U_D(T)+U_D(H)} \\ p_D &= 1 - \frac{U_S(H)}{U_S(T)+U_S(H)} \end{aligned} . \quad (3)$$

To determine whether choices converged on Nash equilibria, or if there were any systematic deviations from this normative prediction, p_S, p_D and utilities for each pigeon were fitted to the proportions of heads choices during the last 5 sessions of each game (i.e., after the pigeons had at least 15 sessions of experience). To the extent that p_S and p_D successfully tracked changes in heads choices, Nash equilibrium would be validated as a behavioral predictor. There were two constraints to the estimation of utilities in the Nash equilibrium model: (1) Utilities did not change across roles or games within the same pigeon—for example, the utility of a 2.5-s reinforcer did not change for a pigeon regardless of whether it was playing *Same*, *Different*, Game 1, or Game 4; and (2) utilities did not covary negatively with duration of reinforcement within the same pigeon—for example, the utility of a 2.5-s reinforcer was equal to or less than the utility of a 7.5-s reinforcer. This implied two free parameters in the Nash equilibrium model fitted to each pigeon, one for the utility of each of the two reinforcer durations used. A third parameter was included to account for payoff-independent biases in choice—for example, $U_S(H)$ was the utility of the reinforcer of successful heads choices while playing *Same*, plus the "utility" of choosing heads.

Figure 4 shows the proportion of heads choices in each role (thin and dotted lines) and estimates of p_S (thick gray bars) and p_D (thick black bars) across games. From simply looking at the thick bars and comparing them to the terminal distribution of choices in each game, it is difficult to determine if the Nash equilibria correctly predicted where choices would converge—compare, for instance, data from pigeon A1 in Game 1, where the Nash equilibria fared well, with data from pigeon B2 in Game 4, where they did not. In some games

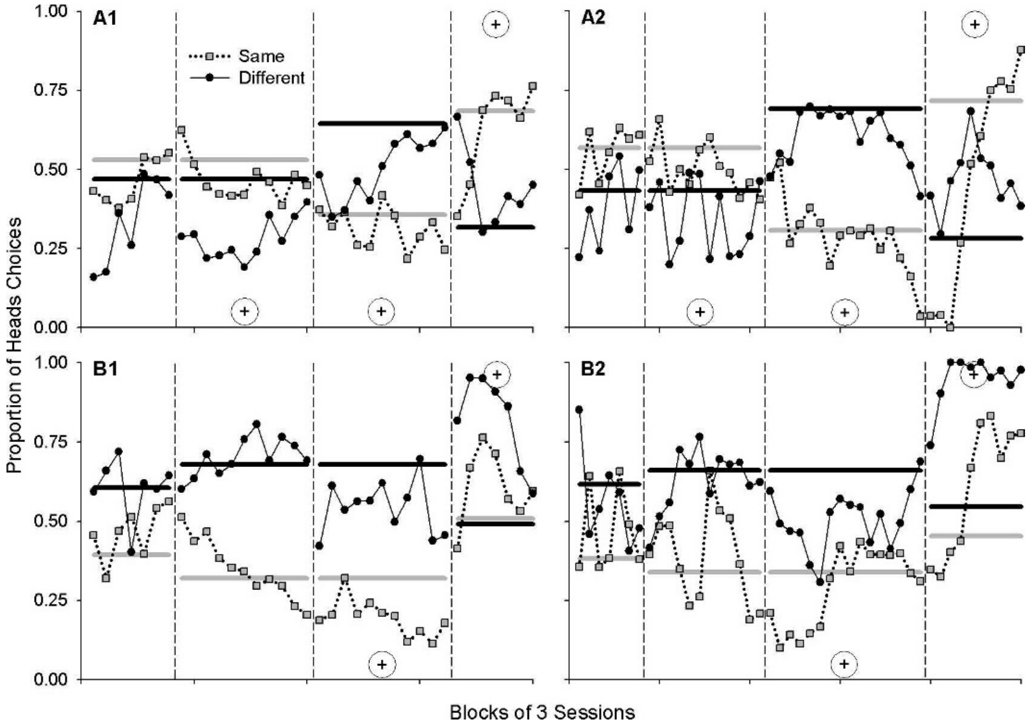


Fig. 4. Proportion of heads choices across MP games. Pigeon A1 played against B1, and pigeon A2 played against B2. Horizontal bars indicate expected proportion of heads choices according to a fitted Nash Equilibrium model (Equation 3). The ⊕ symbol represents the successful choice with larger reward (top = heads, bottom = tails). Deviations from horizontal bars towards ⊕ are indicative of own-payoffs effect.

it is not even clear that choices stabilized after training (e.g., pigeon A2 in Game 4). In almost all cases, however, whether heads was chosen more often while playing one role or another coincided with predictions from the Nash equilibria. A more stringent assessment may be based on two important features of choices predicted by the Nash equilibria that are revealed by Equation 3.

First, note the difference in subscripts between the left and right hand sides of Equation 3: they indicate that optimal choices are not determined by one's own payoffs, but by the opponent's. This implies that if payoffs are larger for tails relative to heads, it is the opponent who should adjust her choices, choosing heads less often if playing *Same*, or more often if playing *Different*. Here is an intuitive account of why this should happen: If pigeon A1 plays *Different* and its payoff for successful tails choices increases (as in Game 2), it would be motivated to choose heads less often—but so would be its opponent, B1, who is playing *Same*. This would discourage A1 from choosing heads less often than tails, but only while B1 makes it less likely that A1 would win if it chose tails. Thus, a change in A1's payoffs should be reflected in B1's behavior. Knowledge of the opponent's payoff, which our pigeons could not have, should not be necessary—feedback via reinforcement should suffice: If A1 chooses tails more often, B1 would increase its rate of reinforcement by choosing tails too, which would discourage A1 from playing tails. Furthermore, unequal payoffs should affect the opponent's choice in opposite directions depending on whether it is playing *Same* or *Different*—tracking the richer alternative in the former, avoiding it in the latter.

The effect of payoff changes on an opponent's choices expected from the Nash equilibrium is visible in the thick bars of Figure 4: Nash equilibria changed between games only when payoffs were changed for the opponent. In Game 2, when reinforcer duration for tails choices was increased for pigeons A1 and A2, the Nash equilibria predicted a decrease in p_S

and an increase in p_D in pigeons B1 and B2. A similar change was predicted for pigeons A1 and A2 in Game 3, when reinforcer duration for tails choices was increased for pigeons B1 and B2. In Game 4, when reinforcer duration for heads choices was increased and for successful tails choices was decreased, p_S was predicted to increase and p_D to decrease relative to prior games. Actual choices roughly followed these predictions. In Game 2, choices diverged less across roles for pigeons that experienced unequal payoffs (A1 and A2) than for their opponents (B1 and B2); once unequal payoffs were presented to B1 and B2 in Game 3, their opponents' choices also diverged across roles. In Game 4, the simultaneous reversal of reinforcer duration for both pigeons either reversed choices across roles (pigeons A1 and A2) or at least reduced their difference (pigeons B1 and B2). Thus, consistent with the Nash equilibria, changes in choices between games were sensitive to changes in opponent's payoffs. Moreover, rapid changes in choice with changes in role demonstrated that game strategy was controlled by visual stimuli that cued role.

The second feature of choice at the Nash equilibrium revealed by Equation 3 is that, because utilities did not change across roles within the same pigeon (our first constraint for estimating utilities), $p_S = 1 - p_D$ for any pigeon. This is visible in Figure 4: Nash equilibria for both roles within each game were always equidistant to .5 and thus added to 1. Relative to these predictions, however, choices in each role at the end of game appear to be biased towards the richer alternative (''⊕'' symbols in Figure 4) in most games with unequal payoffs; the opposite bias was never observed. The sensitivity of pigeons' choices to their own payoffs, however intuitive, was not expected from the Nash equilibria.

The Nash equilibria provided normative predictions about asymptotic choice distribution in MP competitions. These predictions were partially validated by the pigeons' behavior. Consistent with predictions, choices were sensitive to the payoffs of the opponent; however, in conflict with predictions, choices were also sensitive to each pigeon's own payoffs. These molar characteristics of observed choice may now serve to evaluate the global output of models of trial-by-trial performance.

Learning Models

Learning processes may be characterized by algorithms that specify trial-by-trial changes in performance as a function of various inputs from the organism and its environment (Sutton & Barto, 1998). An algorithm that accurately represents the learning process of pigeons in MP must be consistent with the data presented here: It must approximate the Nash equilibrium and be sensitive to role cues, but it must also display choice perseveration and oversensitivity to one's own payoffs. We considered two candidate algorithms to describe MP performance: A *Stochastic Linear Learning* model that incorporates choice perseveration (SLLp), and a *Weighted Reinforcement Matching* (WRM) model. Choice perseveration alone (EWMA, Equation 1) served as a default nonlearning algorithm against which learning models were evaluated.

Stochastic Linear Learning Model with Perseveration (SLLp)

Stochastic linear learning (SLL) models assume that learning agents maintain a probability distribution of choices: If a choice is reinforced, the probability of that choice increases; if a choice is not reinforced, the probability of that choice either decreases or remains constant. This algorithm has been used often as the backbone of models of conditioning (Bush & Mosteller, 1951; Rescorla & Wagner, 1972), and has also provided accurate descriptions of human and nonhuman primate learning of mixed strategies (Erev & Roth, 1998; Lee et al., 2004; 2005; Mookherjee & Sopher, 1994). In the only report of a MP game between laboratory animals, however, neither the Nash equilibrium nor an SLL model captured the performance of pairs of rats (Flood et al., 1983). This result may have been due to poorly discriminable rewards and strong biases generated by visual contact between opponents. Nevertheless, the demonstrable operation of SLL mechanisms in nonhuman learning suggests that a modified version of SLL may provide a good account of nonhuman performance in game environments.

For any given alternative, the SLL model may be formulated as:

$$p_{t+1} = p_t + f(C_t, \pi_t). \qquad (4)$$

Table 3

Probability change function in Eq. 4 for choosing heads.

	Payoff (π_t)	
Prior Choice (C_t)	0	> 0
Heads	$\lambda_E\ (0 - p_t)$	$\lambda_R\ (1 - p_t)$
Tails	$\lambda_E\ (1 - p_t)$	$\lambda_R\ (0 - p_t)$

Equation 4 simply states that a change in probability of choosing an alternative from one trial to the next is a function of the last choice made (C_t) and the payoff for that choice (π_t). Table 3 specifies the change function $f\,(C_t,\pi_t)$ evaluated in this paper. The two-way table assumes that p_t is the probability of choosing heads; an analogous table may be constructed for choosing tails by flipping the top and bottom cells. Two free parameters modulate the impact of a trial on the probability of choosing an alternative: Rate of extinction λ_E, which operates when reinforcement is absent ($\pi_t = 0$), and rate of learning λ_R, which operates when reinforcement is present ($\pi_t > 0$). The left cells of Table 3 indicate that the probability of choosing an alternative, heads in this case, should decrease after each trial by a factor of λ_E if that choice is not reinforced, and increases by the same factor if the opposite choice is not reinforced. The right cells of Table 3 indicate that the probability of choosing an alternative should increase after each trial by a factor of λ_R if that choice is reinforced, and decrease by the same factor if the opposite choice is reinforced.

Whereas it may suffice to use a single parameter across all games to define rate of extinction λ_E, the use of two different positive payoffs, small and large, implies a potential use of two free parameters to define rate of learning λ_R. We relied on prior empirical research to minimize the number of free parameters—rate of learning λ_R was specified as a function of immediately preceding positive payoffs (Börgers & Sarin, 1997), using Killeen's (1985) value function for food duration in animals,

$$\lambda_R = 1 - e^{-\gamma\pi}. \tag{5}$$

Parameter γ is the curvature of the utility function for food rewards—higher values of γ yield more concave functions.

The SLL model is completed with the incorporation of choice perseveration into Equation 4, forming what will be called SLLp. Perseveration was incorporated by substituting p_t in Equation 4 and Table 3 with p'_{t+1} from EWMA (Equation 1). Thus, in determining choice, the model assumes that perseveration operates first, transforming the just-prior estimate of choice probability $p'_t = p_t$ into p'_{t+1} following EWMA (Equation 1); reinforcement then adds or substracts to the momentum-driven propensity following Equations 4 and 5 and Table 3.

The nesting of EWMA into SLLp yields two equations that represent the change in probability of choosing an alternative following reinforced and nonreinforced trials, respectively,

$$\text{If } \pi_t > 0,\ p_{t+1} = \alpha_L C_t + (1 - \alpha_L)[p_t + (1 - e^{-\gamma\pi})(C_t - p_t)], \tag{6a}$$

$$\text{If } \pi_t = 0,\ p_{t+1} = \lambda_E(1 - C_t) + (1 - \lambda_E)[\alpha_L C_t + (1 - \alpha_L)p_t]. \tag{6b}$$

Whereas EWMA has a single free parameter, α, SLLp model has three free parameters: Persistence (α_L; the subscript distinguishes it from EWMA's parameter), concavity of utility function (γ), and rate of extinction (λ_E). Note that Equation 6a can be reduced to EWMA by setting $\gamma = 0$, and Equation 6b by setting $\lambda_E = 0$. That is, if choice is insensitive to reinforcement and extinction, only persistence operates. Equation 6a implies that persistence and reinforcement determine choice following a reinforced trial: The first term on the right-hand side indicates that as $\alpha_L \rightarrow 1$ (strong persistence), choice depends more on the just-prior choice; the second term indicates that as $\alpha_L \rightarrow 0$, choice changes in the direction of the reinforced choice at a rate determined by the utility of the reinforcer. Similarly, Equation 6b implies that extinction and persistence determine choice following a nonreinforced trial: The first term on the right-hand side indicates that as $\lambda_E \rightarrow 1$ (high sensitivity to nonreinforcement), choice alternation is more likely; the second term indicates that as $\lambda_E \rightarrow 0$, choice is determined more by its own momentum.

To account for stimulus control, the model assumed separate values of p_t for each role.

When roles changed between t and $t + 1$, p_{t+1} was predicted by the last p' estimated for the new role. However, the values of C_t and π_t used to predict p_{t+1} were those of the preceding trial, regardless of role.

The SLLp model was evaluated against EWMA (Equation 1). As had been done already with EWMA, SLLp's free parameters were estimated by fitting predicted choice probabilities to actual choices, using the method of maximum likelihood. The relative merit of each model was determined by log-likelihood ratios (LLRs). Log-likelihood ratios were computed for each pigeon as the log of the ratio of the likelihood of SLLp over the likelihood of EWMA (cf. footnote 1). Parameter estimates and LLRs are shown in Table 2. According to LLRs, the data were substantially (e^{111} to e^{265}) more probable given SLLp than given EWMA. This difference cannot be accounted for only by the difference in the number of free parameters (one in EWMA, three in SLLp). On the basis of the Akaike Information Criterion (Burnham & Anderson, 2002), two additional free parameters would cast doubt on inferences based on LLRs of about 4, but not on 3-digit LLRs.

It should not be suprising that SLLp predicted data better than EWMA. The large LLRs indicate that choices in MP are not driven just by perseveration alone, and that SLLp captures at least part of the unaccounted variance by attributing it to an obvious suspect, reinforcement. This does not mean that perseveration is unimportant; in fact, the incorporation of reinforcement effects revealed stronger choice perseveration than EWMA estimated for most pigeons. Nonetheless, to evaluate SLLp's account of reinforcement in MP it is necessary to compare it against another dynamic model of choice.

Weighted Reinforcement Matching (WRM)

A well established regularity in choice behavior is that the proportion of choices of an alternative matches the proportion of reinforcement obtained from that alternative (Herrnstein, 1961); this regularity is known as the matching law (Herrnstein, 1974). This is a global aspect of behavior for which local mechanisms are still uncertain (Lefebvre & Sanabria, 2008; MacDonall, 1999). Weighted Reinforcement Matching (WRM) is proposed here as a local mechanism that, consistent with melioration theory (Herrnstein & Vaughan, 1980), assumes that the matching law operates locally. More specifically, WRM postulates that, when choosing between alternatives, the proportion of recent payoffs obtained from each alternative is compared against the proportion of recent opportunities in which that alternative was chosen. If the former were larger than the latter, the alternative would be likely chosen; if the former were smaller than the latter, the alternative would not be likely chosen. This implies a significant contrast with the low cognitive sophistication of SLLp: Whereas the mnemonic demands of SLLp are restricted to the agent's own behavior (i.e., the propensity towards each choice), local matching also demands separate counts of rewards obtained from each alternative.

Although choice distribution generally covaries with reward distribution, there are two well known systematic deviations from strict matching: Bias and sensitivity to reward size (Baum, 1974). Bias towards an alternative is the reward-independent tendency to choose that alternative; a high/low sensitivity to reward size means that larger rewards are chosen more/less than expected from strict matching (Alsop & Porritt, 2006). For binary choices, these deviations are accounted for by the general form of the matching law (Baum, 1974; Davison & McCarthy, 1988),

$$p_A = \frac{\beta_A (R_A)^s}{\beta_A (R_A)^s + (R_B)^s}, \tag{7}$$

where p_A is the long-term probability of choosing alternative A instead of B, R_A is the cumulative reinforcement obtained from choosing A divided by the total number of trials, β_A is bias towards choosing A, and s is sensitivity to reward size expressed as a power function[2].

The probability p of choosing an alternative may be dynamically updated by substituting R with a EWMA of reinforcement of that choice,

$$R_{t+1} = \alpha_M \pi_t C_t + (1 - \alpha_M) R_t. \tag{8}$$

[2] Equation 7 is mathematically equivalent to the more conventional ratio expression

$$\frac{\sum C_A}{\sum C_B} = \beta_A \left(\frac{R_A}{R_B} \right)^s, \tag{F2}$$

where the left-hand side of the equation is the ratio of choices.

According to Equation 8, if a choice is not reinforced ($\pi_t = 0$) or not made ($C_t = 0$), average reinforcement of that choice declines at a rate of $1 - \alpha_M$. If $\alpha_M = 0$, R_t is constant and equivalent to R_A in Equation 7; values of α_M closer to 1 weigh recent payoffs more heavily in updating average reinforcement. Parameter α_M may be regarded as the rate of persistence of memory for past choices and reinforcement—as memory of old items fades, their weight in the computation of average reinforcement is reduced. The WRM model incorporates this memory into the matching function of Equation 7, thus involving three free parameters: memory persistence (α_M), sensitivity to reward size (s), and bias toward choosing *heads* (β_{HEADS}; $\beta_{TAILS} = 1 / \beta_{HEADS}$).

To implement WRM, Equation 8 was computed for each alternative, on each trial; R_{t+1} for heads and tails were entered as R_A and R_B, respectively, in Equation 7. A separate register of R_t was kept for each role such that, if roles changed between t and $t + 1$, R_{t+1} depended on the last R computed for the new role. However, the values of C_t and π_t used to compute R_{t+1} were those of the preceding trial, regardless of role.

Free parameters were estimated using the maximum likelihood method. According to these estimates (shown in Table 2), all pigeons were relatively insensitive to reward size ($s < 1$), and most of them were biased toward choosing heads ($\beta_{HEADS} > 1$). This account was more likely than EWMA to yield the observed data, as shown by LLRs that ranged between 75.3 and 135. These large LLRs preclude the possibility that the advantage of WRM over EWMA was due to additional free parameters.

SLLp versus WRM

Despite the superior performance of WRM relative to EWMA, it did not surpass SLLp in predicting the obtained data. Log-likelihood ratios were larger for SLLp than for WRM for every pigeon. Differences between LLRs, which evaluate each learning model against EWMA, are in fact the LLRs between SLLp and WRM. These differences ranged from 4.5 (pigeon B2) to 130.4 (pigeon A2), favoring SLLp. The superiority of SLLp over WRM could not depend on the number of free parameters because there were three free parameters for both models. Thus, SLLp provided a better account of MP performance than WRM did.

The superiority of SLLp relative to WRM suggests that choices probably did not depend on separate counts of rate of reinforcement for each alternative action. Instead, SLLp implies a simpler behavioral machinery, a ''stream'' of choices that carried its own momentum and that was directed by reinforcement. But, could such a simple mechanism yield the global patterns of behavior depicted in Fig. 4? Would its output be sensitive to Nash equilibrium in each role? Would it produce own-payoffs biases? To answer these questions, we resorted to simulations of SLLp.

Simulation of SLLp

We attempted to reproduce the observed performance of pigeons by feeding an SLLp simulator with the best fitting parameters (Table 2). The simulator was run 1000 times on the same sequence of games experienced by the pigeons. Results are shown in Figure 5. Like the actual pigeons, performance of simulated pigeons that learned using SLLp approached the Nash equilibria (note the separation of the lines in SimB1 and SimB2 during Game 2) and were oversensitive to their own payoffs (note changes in performance from Game 1 to Game 2 in SimA1 and A2, and from Game 2 to Game 3 in SimB1 and SimB2).

Despite the approximate reproduction of the pigeons' performances, SLLp simulations showed two significant departures from the empirical data. First, pigeons did not acquire asymptotic performance as fast as the mean SLLp simulation performance suggests. Second, the asymptotic sequential variance of choice proportions across experimental sessions was systematically lower than in its simulated counterpart (the seemingly stable curves shown in Figure 5 are due to the averaging of 1000 simulator runs). Both inconsistencies point in the same direction: SLLp parameters are mostly dependent on fast within-session changes in choice, and thus misrepresent substantially slower between-session changes. By assuming that fast changes in choice carried over from one session to the next, the simulations yielded steeper acquisition curves that were less stable at asymptote than those observed in pigeons.

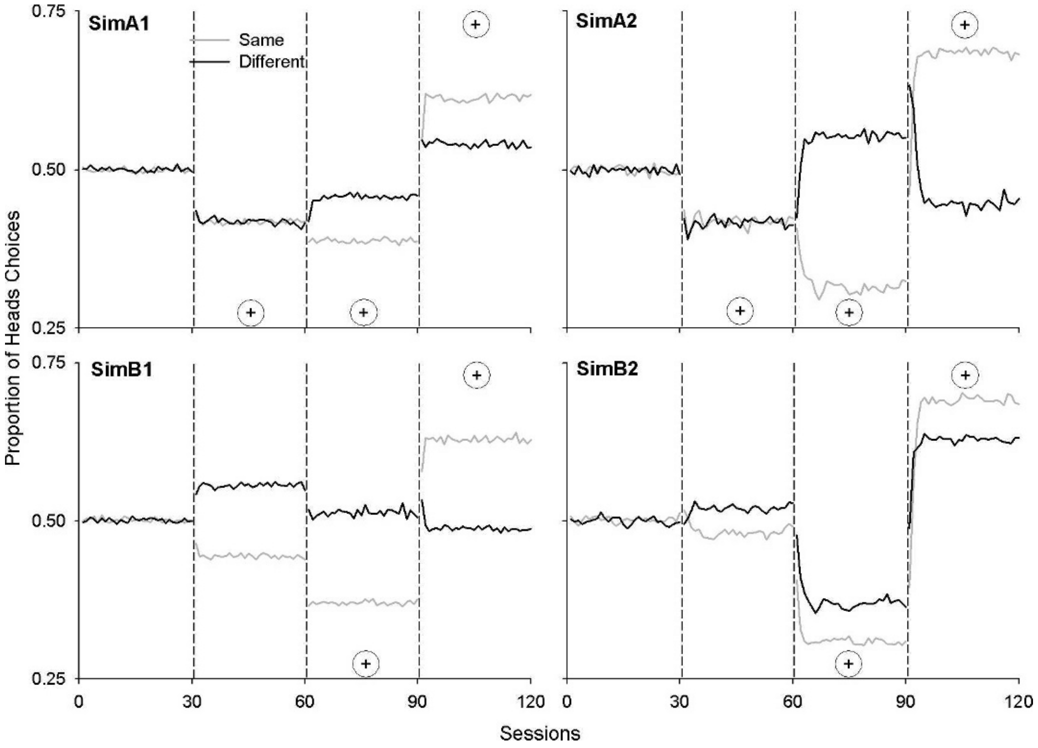


Fig. 5. Mean performance of 1000 simulations of SLLp, with parameters extracted from pigeons' performance. The simulations reproduce reversals in preference across roles predicted by Nash equilibrium, as well as own-payoff effects. Notation is as in Figure 4.

Differences in within- and between-session learning rates are regularly observed in laboratory animals in various experimental preparations, and are best illustrated by the phenomenon of spontaneous recovery (Bouton, 2002): Animals learn to stop responding when reinforcement is discontinued, but responding is likely to recover spontaneously after an interruption of experimental conditions, such as the interval between experimental sessions. Important associative processes appear to underlie spontaneous recovery (Robbins, 1990). The net result is that, at the beginning of each experimental session, learning does not pick up where it left off at the end of the last session, but is somewhat backtracked.

Choice acquisition by pigeons also appeared to backtrack at the beginning of each MP session. This effect is illustrated by the performance of pigeon B2 in Game 4, while playing *Same* (dotted curve in the bottom-right panel of Figure 4). In Game 4, B2 transitioned from choosing heads with $p = .35$ to $p = .78$. Backtracking is revealed when p is plotted for the first and second half of each session, as in Figure 6. Before transition (first four sessions), p repeatedly declined between the first and second half of each session, suggesting a fast within-session learning process that backtracked between sessions. During and after transition, downward within-session trends in p became rare; substantial within-session changes in p were typically upward, suggesting a reversal in the within-session learning process, often interfered by prior ''downward'' learning. The reversal of within-session trends is more clearly depicted in the inset of Figure 6: Downward trends in p during the first four sessions (negative change) were followed by a mixture of upwards (positive) and flat trends in subsequent sessions. Similar backtracking was also visible in other pigeons while substantial changes in choice were in progress (Pigeon A1, Game 4; Pigeon B1, Game 2, playing *Same*; Pigeon A2, Game 4, playing *Same*).

Within-session changes in choice around transitions suggest that a more precise account

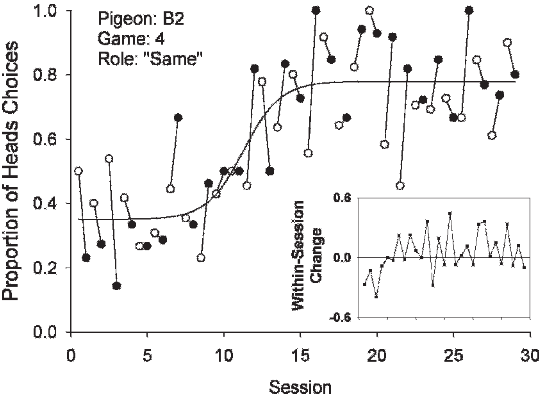


Fig. 6. Proportion of heads choices by pigeon B2 in Game 4 while playing *Same*, in the first and second half of each session (open and closed circles, respectively). The general trend in change of choice over the game is depicted by a continuous sigmoidal function, $y = y_0 + (a - y_0) / (1 + e^{-(x - b) / c})$, where y_0, a, b, and c were fitted to choice proportions. The inset shows the difference in choice proportions between the second and first half of each session.

of multisession game behavior would be attained if the processes underlying spontaneous recovery were incorporated to the SLLp model. The specifications of this module could be laid out in future developments of SLLp.

DISCUSSION

Like humans (Rapoport & Budescu, 1992) and other primates (Lee et al., 2004), pigeons can learn to compete efficiently in MP games. Such efficiency, however, is somewhat abated by a slight yet consistent tendency to persist in prior choices and an excessive responsiveness to their own payoffs. Choice persistence in pigeons contrasts with over-alternation in adult humans playing mixed strategy games (Brown & Rosenthal, 1990). Response persistence, however, is not an isolated phenomenon: It has been reported in other species (Lee et al., 2004) and in other experimental contexts (Baum & Davison, 2004; Killeen, 2003; Reboreda & Kacelnik, 1993; but see Dember & Fowler, 1958). Moreover, persistence cannot be reduced to the automatic repetition of a motoric response because, in the experiment reported here, choices were separated by intertrial intervals of 10 s or more. This divergence from optimality in the opposite direction of humans suggests that overalternation depends on cognitive idiosyncrasies of *Homo sapiens*; perhaps it is a verbally acquired misrepresentation of randomness, thus absent in preverbal children and nonhumans. This hypothesis has yet to be evaluated.

Suboptimal sensitivity to one's own payoffs in mixed strategy games has been observed in human players (Binmore, Swierzbinski, & Proulx, 2001; Ochs, 1995). Prior to this experiment, however, the manipulation of payoffs necessary to evaluate this effect had not been conducted in nonhuman subjects. Our results indicate an important invariance across species that further research should confirm; it implies that accounts of own-payoffs effects (Goeree, Holt, & Palfrey, 2003) may be modeling basic behavioral processes that humans share with other species.

A stochastic linear learning (SLL) algorithm, which has successfully described human performance in mixed strategy games (Erev & Roth, 1998; Mookherjee & Sopher, 1994), approximated optimal performance and reproduced own-payoffs effects. With the addition of a persistence module, this algorithm (SLLp) described pigeon performance better than the more sophisticated reward-tracking WRM model. This result, however, does not rule out reward tracking in other species, or other forms of reward tracking different from dynamic matching. Research in this direction has been conducted by Lau and Glimcher (2005) and Corrado, Sugrue, Seung, and Newsome (2005). Their studies suggest reward-tracking mechanisms that account for the performance of rhesus monkeys in concurrent schedules of reinforcement without relying on dynamic matching. Ideally, learning models should be developed to account for performance across schedules of reinforcement and, to the extent that it is possible, across species. This study and those by Lau and Glimcher and Corrado and colleagues are formulating some candidate models. The challenge, now, is to identify common ground and devise critical tests.

An important implication of SLLp is that it does not require players to learn anything about their opponent, but only about their own outcomes. The approximation of pigeons' choices to the Nash Equilibrium did not (and could not) depend on them ''knowing'' their opponent's payoffs, as it is presumed in game-theoretical agents. Similarly, the formation of evolutionarily stable strategies does not depend on a population ''knowing'' the fitness payoffs of another population (Maynard Smith, 1974). Whether over trials or over generations, a rudimentary feedback loop suffices for the formation of stable stochastic strategies. Choices in mixed strategy games thus appear to be susceptible to the often invoked parallelism between instrumental learning and natural selection (cf. Skinner, 1981).

It is possible that sophisticated cognitive and social processes, such as hypothesis testing and considerations of fairness, play a role in mixed strategy game playing, particularly in humans. To invoke such processes, they must account for variance that is not accounted for by more parsimonious learning mechanisms, such as those described here. Critical tests for the involvement of complex processes are, thus, defined in part by how simpler processes operate in game-theoretical scenarios. Because these simpler processes are more efficiently studied using animal models, we hope to see the field of behavioral game theory extend its domain into the acquired behavior of nonhuman species.

REFERENCES

Alsop, B., & Porritt, M. (2006). Discriminability and sensitivity to reinforcer magnitude in a detection task. *Journal of the Experimental Analysis of Behavior, 85,* 41–56.

Baum, W. M., & Davison, M. (2004). Choice in a variable environment: Visit patterns in the dynamics of choice. *Journal of the Experimental Analysis of Behavior, 81,* 85–127.

Baum, W. M. (1974). On two types of deviation from the matching law: Bias and undermatching. *Journal of the Experimental Analysis of Behavior, 22,* 231–242.

Binmore, K., Swierzbinski, J., & Proulx, C. (2001). Does minimax work? An experimental study. *Economic Journal, 111,* 445–464.

Börgers, T., & Sarin, R. (1997). Learning through reinforcement and replicator dynamics. *Journal of Economic Theory, 77,* 1–14.

Bouton, M. E. (2002). Context, ambiguity, and unlearning: Sources of relapse after behavioral extinction. *Biological Psychology, 52,* 976–986.

Brown, J. N., & Rosenthal, R. W. (1990). Testing the minimax hypothesis: A reexamination of O'Neill's game experiment. *Econometrica, 38,* 1065–1081.

Burnham, K. P., & Anderson, D. R. (2002). *Model selection and multimodel inference: A practical information-theoretic approach.* New York: Springer-Verlag.

Bush, R. R., & Mosteller, F. M. (1951). A mathematical model for simple learning. *Psychological Review, 58,* 313–323.

Camerer, C. F. (2003). *Behavioral Game Theory, Experiments in Strategic Interaction.* New York/New Jersey: Princeton/Russell Sage.

Corrado, G. S., Sugrue, L. P., Seung, H. S., & Newsome, W. T. (2005). Linear-Nonlinear-Poisson models of primate choice dynamics. *Journal of the Experimental Analysis of Behavior, 84,* 581–617.

Davison, M., & McCarthy, D. (1988). *The matching law: A research review.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Dember, W. N., & Fowler, F. (1958). Spontaneous alternation behavior. *Psychological Bulletin, 55,* 412–428.

Dorris, M. C., & Glimcher, P. W. (2004). Activity in posterior parietal cortex is correlated with the relative subjective desirability of action. *Neuron, 44,* 365–378.

Driver, P. M., & Humphries, D. A. (1988). *Protean behaviour: The biology of unpredictability.* Oxford: Oxford University Press.

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.

Falk, R., & Konold, C. (1997). Making sense of randomness: Implicit encoding as a basis for judgment. *Psychological Review, 104,* 301–318.

Flood, M., Lendenmann, K., & Rapoport, A. (1983). 2×2 games played by rats: Different delays of reinforcement as payoffs. *Behavioral Science, 28,* 65–78.

Goeree, J. K., Holt, C. A., & Palfrey, T. R. (2003). Risk averse behavior in generalized matching pennies games. *Games and Economic Behavior, 45,* 97–113.

Herrnstein, R. J. (1961). Relative and absolute strength of response as a function of frequency of reinforcement. *Journal of the Experimental Analysis of Behavior, 4,* 267–272.

Herrnstein, R. J. (1974). Formal properties of the matching law. *Journal of the Experimental Analysis of Behavior, 21,* 159–164.

Herrnstein, R. J., & Vaughan, W. (1980). Melioration and behavioral allocation. In J. E. R. Staddon (Ed.), *Limits to action, the allocation of individual behavior* (pp. 143–176). New York: Academic Press.

Killeen, P. R. (1985). Incentive theory: IV. Magnitude of reward. *Journal of the Experimental Analysis of Behavior, 43,* 407–417.

Killeen, P. R. (2003). Complex dynamic processes in sign tracking with an omission contingency (negative automaintenance). *Journal of Experimental Psychology: Animal Behavior Processes, 29,* 49–61.

Lau, B., & Glimcher, P. W. (2005). Dynamic response-by-response models of matching behavior in rhesus monkeys. *Journal of the Experimental Analysis of Behavior, 84,* 555–579.

Lee, D., Conroy, M. L., McGreevy, B. P., & Barraclough, D. J. (2004). Reinforcement learning and decision making in monkeys during a competitive game. *Cognitive Brain Research, 22,* 45–58.

Lee, D., McGreevy, B. P., & Barraclough, D. J. (2005). Learning and decision making in monkeys during a rock-paper-scissors game. *Cognitive Brain Research, 25,* 416–430.

Lefebvre, V. A., & Sanabria, F. (2008). Matching by fixing and sampling: A local model based on internality. *Behavioural Processes, 78*, 204–209.

MacDonall, J. S. (1999). A local model of concurrent performance. *Journal of the Experimental Analysis of Behavior, 71*, 57–74.

Machado, A. (1989). Operant conditioning of behavioral variability using a percentile reinforcement schedule. *Journal of the Experimental Analysis of Behavior, 52*, 155–166.

Maynard Smith, J. (1974). The theory of games and the evolution of animal conflicts. *Journal of Theoretical Biology, 47*, 209–221.

Miller, G. F. (1997). Protean primates: The evolution of adaptive unpredictability in competition and courtship. In: A. Whiten, & R. W. Byrne (Eds.), *Machiavellian Intelligence II: Extensions and Evaluations* (pp. 313–340). Cambridge, MA: Cambridge University Press.

Mookherjee, D., & Sopher, B. (1994). Learning behavior in an experimental matching pennies game. *Games and Economic Behavior, 7*, 62–91.

Neuringer, A. (1986). Can people behave ''randomly?'': The role of feedback. *Journal of Experimental Psychology: General, 115*, 62–75.

Ochs, J. (1995). Games with unique, mixed strategy equilibria: An experimental study. *Games and Economic Behavior, 10*, 202–217.

Rapoport, A., & Budescu, D. V. (1992). Generation of random series in 2-person strictly competitive games. *Journal of Experimental Psychology: General, 121*, 352–363.

Reboreda, J. C., & Kacelnik, A. (1993). The role of autoshaping in cooperative two-player games between starlings. *Journal of the Experimental Analysis of Behavior, 60*, 67–83.

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black, & W. F. Prokasy (Eds.), *Classical conditioning II: Current research and theory* (pp. 64–99). New York: Appleton-Century-Crofts.

Robbins, S. J. (1990). Mechanisms underlying spontaneous recovery in autoshaping. *Journal of Experimental Psychology: Animal Behavior Processes, 16*, 235–249.

Skinner, B. F. (1981). Selection by consequences. *Science, 213*, 501–504.

Sutton, R. S., & Barto, A. G. (1998). *Reinforcement Learning: An Introduction*. Cambridge, MA: MIT.

Towse, J. N., & Maclachlan, A. (1999). An exploration of random generation among children. *British Journal of Developmental Psychology, 17*, 363–380.

Received: October 6, 2008
Final Acceptance: November 17, 2008