Hoffman RR, Bringmann W, Bamberg M and Klein R (1987) Some historical observations on Ebbinghaus. In: Gorfein DS and Hoffman RR (eds), *Memory and Learning: The Ebbinghaus Centennial Conference*, pp. 57–76. Hillsdale, NJ: Lawrence Erlbaum.

Hothersall D (1984) *A History of Psychology*. New York, NY: Random House.

Jaensch ER (1909) Hermann Ebbinghaus. *Zeitschrift für Psychologie* **51**: 3–8.

James W (1885) Experiments in memory. *Science* **6**: 198–199.

Peterson J (1925) *Early Conceptions and Tests of Intelligence*. Chicago, IL: World Book Co.

Stigler SM (1978) Some forgotten work on memory. *Journal of Experimental Psychology: Human Learning and Memory* **4**: 1–4.

Titchener EB (1910) The past decade in experimental psychology. *American Journal of Psychology* **21**: 404–421.

Woodworth RS (1909) Hermann Ebbinghaus. *Journal of Philosophy and Scientific Methods* **6**: 253–256.

Woodworth RS (1938) *Experimental Psychology*. New York, NY: Holt.

Economics Experiments, Learning in

Advanced article

Jacob K Goeree, University of Amsterdam, Amsterdam, The Netherlands
Charles A Holt, University of Virginia, Charlottesville, Virginia, USA

CONTENTS

Models of learning in economics are used to explain how people use information about past prices and other signals to make good decisions. These models can be tested with economics experiments in which decisions are made in a sequence of rounds or trading periods.

INTRODUCTION

The main focus of economic analysis is on equilibrium steady states, e.g. on prices determined by the intersection of supply and demand. The preoccupation with equilibrium is perhaps due to the fact that many markets operate for protracted periods of time under fairly stationary conditions. The awareness that there may be multiple equilibria, some of which are bad for all concerned, has raised interest in why behavior might converge to one equilibrium and not to another. As a result, there is renewed interest among economists in mathematical models of learning that were studied extensively by psychologists in the 1960s. This article will describe two of those models, 'reinforcement learning' and Bayesian 'belief learning'. These models and their generalizations will be discussed in the context of a binary prediction task, which may generate the kind of behavior that is known in the psychology literature as 'probability matching'.

We will then use these learning models to analyze behavior in an economic market where firms choose prices. Markets and games are more complex than individual decision tasks, in that people's choices affect others' beliefs. One role of learning models in such situations is to provide an explanation of the dynamic paths of prices, which can shed light on the nature of adjustment towards equilibrium. The equilibrium is characterized by an unchanging (steady-state) distribution of beliefs across individuals, which we call a 'stochastic learning equilibrium'.

TYPES OF LEARNING MODELS

We will introduce the basic learning models in the context of a binary prediction task that has been familiar to psychologists since the 1950s. This task is of special interest because humans are thought to be slow learners in this context. The typical setup involves two lights, each with a corresponding lever (or computer key). A signal light indicates that a decision can be made, and then one of the levers is pressed. Finally, one of the lights is illuminated. Animal subjects like rats and chicks are reinforced with food pellets when the prediction is correct. Human subjects are sometimes told to 'do your best' to predict accurately or to 'maximize the number of correct choices'. In other studies, humans are paid small amounts, typically pennies, for correct choices, and penalties may be deducted for incorrect choices.

The general result seems to be that humans are subject to 'probability matching', predicting each event with a frequency that approximately matches the frequency with which it actually occurs. For example, if the left light is illuminated three-fourths of the time, then subjects would come to learn this by experience and would tend to predict 'left' three-fourths of the time.

This behavior is not rational. A consistent prediction of the more likely event will be correct three-fourths of the time. Matching behavior will only generate a correct prediction with a probability of $\frac{3}{4} \times \frac{3}{4} + \frac{1}{4} \times \frac{1}{4}$ (the first term corresponds to predicting the more likely event with probability $\frac{3}{4}$ and being correct with this prediction three-fourths of the time, and similarly for the second term). Thus, the probability of being correct under probability matching is $\frac{5}{8} < \frac{3}{4}$.

In a recent summary of the probability-matching literature, the psychologist Fantino (1998, pp. 360–361) concludes: 'Human subjects do not behave optimally. Instead they match the proportion of their choices to the probability of reinforcement.… This behavior is perplexing given that non-humans are quite adept at optimal behavior in this situation.' For example, Mackintosh (1996) conducted probability-matching experiments with chicks and rats, and the choice frequencies were well above the probability-matching predictions in most treatments.

The resolution of this paradox may be found in the work of Sidney Siegel, who is perhaps the psychologist who has had the largest impact on experiments in economics. His early work from the 1960s exhibits a high standard of careful reporting and procedures, appropriate statistical techniques, and the use of financial incentives where appropriate. His work on probability matching is a good example. In one experiment, 36 male Penn State students were allowed to make predictions for 100 trials, and then 12 of these students were brought back on a later day to make predictions in 200 more trials (Siegel *et al.*, 1964). The proportions of predictions for the more likely event are shown in Figure 1, in which each point is an average over 20 trials.

The 12 subjects in the 'no pay' treatment were simply told to 'do your best' to predict which light bulb would be illuminated. These prediction rates begin at about 0.5, as would be expected in early trials with no information about which event is more likely. Notice that the proportion of predictions for the more likely event converges to 0.75, as predicted by probability matching, with very close matching from about trial 100.

In the 'pay/loss' treatment, 12 participants received 5 ¢ for each correct prediction, and they lost 5 ¢ for each incorrect prediction. The rate seems to converge to about 0.9.

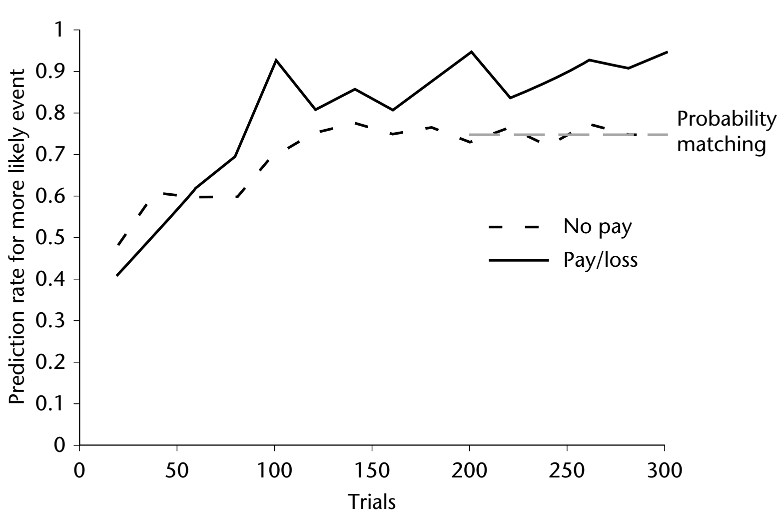

Figure 1. Prediction frequencies for an event that occurs with probability $\frac{3}{4}$ (Siegel *et al.*, 1964). Each point plotted represents an average over 20 trials. The 'no pay' group were simply told to 'do your best'. The 'pay/loss' group received 5¢ for each correct prediction and forfeited 5¢ for each incorrect prediction.

A third 'pay' treatment offered a 5 ¢ reward for a correct prediction but no loss for an incorrect prediction. The results (not shown) are in between those of the other two treatments, and clearly above 0.75.

Clearly, then, incentives matter. Probability matching is not observed with incentives in this context.

Siegel's findings suggest a resolution to the paradox that rats are smarter than humans in binary prediction tasks. You cannot tell a rat to 'do your best': incentives such as food pellets must be used. Consequently, the choice proportions are closer to those observed with financially motivated human subjects. In a recent survey of over 50 years of probability-matching experiments, Vulkan (2000) separates those studies that used real incentives from those that did not, and he concluded that probability matching is generally not observed with real pay-offs. However, humans can still be surprisingly slow learners in this simple setting. For this reason, probability-matching data are particularly interesting for valuating alternative learning theories.

Reinforcement Learning

One prominent theory of learning associates changes in behavior to the reinforcements actually received. Initially, when no reinforcements have been received, it is natural to assume that the choice probabilities for each decision are equal to one-half. Suppose that in the experiment there is a reinforcement of x for each correct prediction, and nothing otherwise. So if one predicts event L and is correct, then the probability of choosing L should increase. The extent of the behavioral change is assumed to depend on the size of the reinforcement. As the total earnings received for a particular decision increase, the probability of making that decision is assumed to increase. Suppose that event L has been predicted N_L times and that the predictions have sometimes been correct and sometimes not. Then the total earnings for predicting L, denoted by e_L, would be less than xN_L. Similarly, let e_R be the total earnings from the correct R predictions. One way to model the effect of total earnings for each decision on choice probabilities is to choose L or R with the following probabilities:

$$P(\text{choose L}) = \frac{\alpha + e_L}{2\alpha + e_L + e_R} \quad (1)$$

$$P(\text{choose R}) = \frac{\alpha + e_R}{2\alpha + e_L + e_R} \quad (2)$$

The parameter α determines how quickly learning responds to the reinforcements. Note that, as additional reinforcements are received, they are added into the relevant numerator, and to both denominators, to ensure that the probabilities sum to 1.

This kind of model might explain some aspects of behavior in probability-matching experiments with financial incentives. The choice probabilities would be equal initially, but a prediction of the more likely event will be correct 75% of the time, and the resulting asymmetries in reinforcement would tend to raise the prediction probability for that event, and the total earnings for this event would tend to be much larger than for the other event. If L is the more likely event, then e_L would be growing faster, so that e_R/e_L would tend to get smaller. Thus the probability of choosing L would tend to converge to 1.

This learning model can be simulated by using past accumulated earnings to compute choice probabilities. Then a random-number generator determines the actual choices. To make our data comparable with Siegel's data, we simulate decisions of a cohort of 1000 individuals for 300 periods, and calculate the 20-period choice averages for the more likely event.

The simulations were done for $\alpha = 5$ and $x = 1$. The value of α was chosen to create some initial inertia in behavior, which will tend to disappear after 40 or 50 periods. Setting α equal to 5 is analogous to having had each decision reinforced five times initially. Figure 2 shows simulated choice averages together with Siegel's original data. The simulated data are smoother, and start somewhat higher to start with, but the general pattern and final tendencies are similar. Erev and Roth (1998) have used reinforcement learning to explain behavior in simple matrix games.

A Simple Model of Belief Learning

With reinforcement learning, beliefs are not explicitly modeled. An alternative approach that is more natural to economists is to model learning in terms of (Bayesian) updating of beliefs. Given the symmetry of Siegel's experimental setup, a person's initial beliefs ought to be that each event is equally likely, but the first observation should raise the probability associated with the event that was just observed. Moreover, the probability of event L should be an increasing function of the number N_L of times that this event has been observed, and a decreasing function of the number N_R of times that event R has been observed. Let N be the total

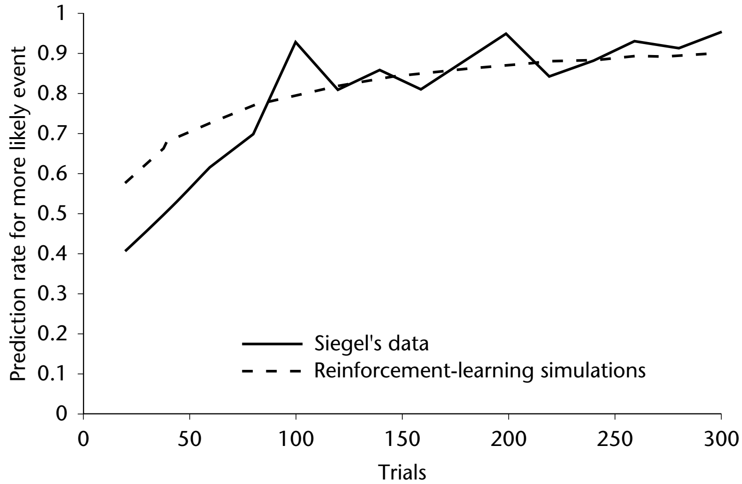

Figure 2. Data for Siegel's probability-matching experiment ('pay/loss' condition), with reinforcement-learning simulation data superimposed ($\alpha = 5$).

number of observations to date. Then a standard belief-learning model is:

$$P(\mathrm{L}) = \frac{\beta + N_{\mathrm{L}}}{2\beta + N} \tag{3}$$

$$P(\mathrm{R}) = \frac{\beta + N_{\mathrm{R}}}{2\beta + N} \tag{4}$$

where $N = N_{\mathrm{L}} + N_{\mathrm{R}}$. Note that β determines how quickly the probabilities respond to the new information; a large value of β will keep these probabilities close to $\frac{1}{2}$. These formulae for calculating probabilities can be derived from Bayesian statistical principles (DeGroot, 1970, p. 160). In the early periods, the totals N_{L} and N_{R} might sometimes switch in terms of which one is higher, but the more likely event will soon dominate, and therefore $P(\mathrm{L})$ will be greater than $\frac{1}{2}$.

The beliefs determine the expected pay-offs (or utilities) for each decision, which in turn determine the decisions made. In theory, the decision with the highest expected pay-off should be selected with certainty. The prediction of the belief-learning model is, therefore, that all people will eventually predict the more likely event every time.

In an experiment, however, some randomness in decision-making might be observed if the expected pay-offs for the two decisions are similar. This randomness may be due to changes in emotions, calculation errors, selective forgetting of past experience, etc. Following Luce (1959), we introduce some 'noise' via a probabilistic choice model, where decision probabilities are positively but not perfectly related to expected pay-offs. Let π_{L} and π_{R} denote the expected pay-offs from choosing 'left' and 'right' respectively. Luce provided a set of axioms under which the choice probability is calculated as

$$P(\text{choose L}) = \frac{(\pi_{\mathrm{L}})^{1/\mu}}{(\pi_{\mathrm{L}})^{1/\mu} + (\pi_{\mathrm{R}})^{1/\mu}} \tag{5}$$

$$P(\text{choose R}) = \frac{(\pi_{\mathrm{R}})^{1/\mu}}{(\pi_{\mathrm{L}})^{1/\mu} + (\pi_{\mathrm{R}})^{1/\mu}} \tag{6}$$

The parameter μ is an 'error' parameter. It determines the sensitivity of choice probabilities to differences in expected pay-offs. In the limit when μ tends to zero, the decision with the higher expected pay-off is selected with probability 1. In the other extreme as μ gets large, behavior is random and independent of pay-offs.

In the probability-matching experiment, the expected pay-off of choosing 'left' is the reward (of 1, say) times the probability of 'left' that represents the person's beliefs. Thus the expected pay-off of 'left' is $P(\mathrm{L})$, and similarly the expected pay-off of 'right' is $P(\mathrm{R})$. $P(\mathrm{L})$ is greater than one-half if 'left' is more likely, and the error parameter μ determines how close the choice probability for the more likely event is to 1.

Figure 3 shows a simulation of the belief-learning model for $\beta = 20$. The thin solid line represents the average of the belief probabilities for the 12 simulated subjects. Notice that beliefs start close to one-half and converge to the true probability of the more likely event $\left(\frac{3}{4}\right)$. These beliefs determine expected pay-offs, and hence choice probabilities, via equations 5 and 6.

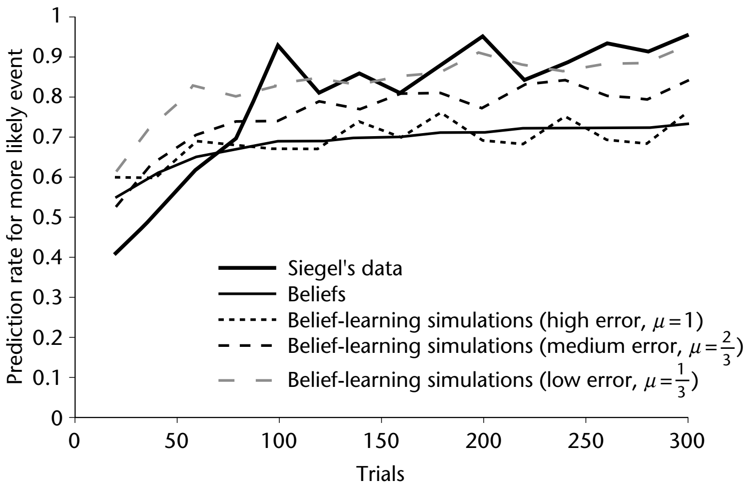

Figure 3. Data for Siegel's probability-matching experiment ('pay/loss' condition), with belief-learning simulation data superimposed ($\beta = 20$). The error parameter μ ranges from 1 (high error) to $\frac{1}{3}$ (low error).

The dashed lines show the simulated average choice frequencies for three different levels of the error parameter. With high error ($\mu = 1$), the choice frequencies are close to the belief line, which would correspond to probability matching. This result is to be expected, since expected pay-offs are equal to belief probabilities. The denominator on the right-hand side of equations 5 and 6 is 1 when $\mu = 1$, and hence the probability of choosing 'left' equals π_L, which is equal to the belief probability.

As the error is reduced, the simulated choice frequencies move upwards towards the optimal level of 1. The top line, with $\mu = \frac{1}{3}$, converges to the level of about 0.9, which is close to the choice frequency observed by Siegel.

The simulations in Figure 3 were done for a cohort of size 12, to be comparable with Siegel's experiment. This allows us to see the degree of variation in the simulated data with a small group. In order to predict the average over a large number of individuals, we ran the simulation 1000 times, and the average proportions of choices for the more likely event were: 0.76 for $\mu = 1$, 0.80 for $\mu = 0.67$, and 0.87 for $\mu = 0.33$.

Generalizations

Both of the learning models discussed above are somewhat simple, and this is part of their appeal. The reinforcement model builds in some randomness in behavior, and has the appealing feature that incentives matter. But it has less of a cognitive element. There is no reinforcement for decisions not made. For example, suppose that a person chooses L three times in a row (by chance) and is wrong each time. Since no reinforcement is received, the choice probabilities stay at 0.5 even after three incorrect predictions. This seems unreasonable. People do learn something in the absence of previous reinforcement, since they realize that making a good decision may result in higher earnings in the next round. Camerer and Ho (1999) have developed a generalization of reinforcement learning that contains some elements of belief learning. Roughly speaking, outcomes that are observed receive partial reinforcement even if nothing is earned.

These learning models can be enriched in other ways to obtain better predictions of behavior. For example, the sums of event observations in the belief-learning model weigh each observation equally. It may be reasonable to allow for 'forgetting' in some contexts, so that the observation of an event in the most recent trial may carry more weight than something observed a long time ago. This can be done by replacing sums with weighted sums. For example, if event L was observed three times, N_L in equation 3 would be 3, which can be thought of as $1+1+1$. If the most recent observation (listed on the right in this sum) is twice as prominent as the one before it, then the prior event would get a weight of one-half, and the one before that would get a weight of one-fourth, and so on. This type of 'recency' effect will be discussed below in the context of an interactive market game.

Finally, 'Luce's probabilistic-choice rule' (equations 5 and 6) is often replaced with the 'logit rule':

$$P(\text{choose L}) = \frac{\exp(\pi_L/\mu)}{\exp(\pi_L/\mu) + \exp(\pi_R/\mu)} \qquad (7)$$

$$P(\text{choose R}) = \frac{\exp(\pi_R/\mu)}{\exp(\pi_L/\mu) + \exp(\pi_R/\mu)} \quad (8)$$

where μ is an error parameter as before. The Luce and logit rules are often similar in effect, and both are commonly used. The logit probabilities are unchanged when all pay-offs are increased by a constant, and the Luce probabilities are unchanged when all pay-offs are multiplied by a positive constant.

LEARNING AND PRICE DYNAMICS IN A MARKET GAME

We use a simple price competition example from Capra *et al*. (2002) to illustrate the effects of learning in an interactive setting. Consider a market game in which firms 1 and 2 simultaneously choose prices p_1 and p_2 in the range [60, 160] (units are cents). Demand is assumed to be a fixed total quantity ('perfectly inelastic'). The sales quantity of the firm with the lower price $p_{\min}$ is normalized to be one, so the low-price firm earns an amount equal to its price. The high-price firm sells a 'residual' amount R, which is less than 1. The amount by which this residual is less than 1 indicates the degree of buyer responsiveness to price. The high-price firm has to match the lower price in order to make any sales, but some sales are lost because of the initially higher price. We assume that the high-price firm only earns $Rp_{\min}$, where $R<1$. In the event of a tie, the $1+R$ sales units are shared equally, so that each seller earns $(1+R)\ p_{\min}$.

As long as the high-price firm obtains less than half the market ($R<1$), the Nash equilibrium prediction is for both firms to set the lowest possible price of 60. To see this, note that at any common price, each firm has an incentive to undercut the other by a small amount to increase its market share. Therefore, the unique Nash equilibrium involves both firms charging the lowest possible price. The harsh competitive nature of the Nash prediction seems to go against the simple economic intuition that the degree of buyer inertia will affect the behavior of firms. When $R=0.8$, say, the loss from having the higher price is relatively small, and firms should be more likely to set prices above 60 when there is a small chance that rivals will do the same. Indeed, in the extreme case when $R=1$ it becomes a dominant strategy for both firms to choose the highest possible price of 160. While a standard Nash analysis predicts no change as long as $R<1$ (and then an abrupt change when $R=1$), it seems plausible that prices will gradually rise with R.

We ran an experiment based on this market game, using six cohorts of 10 subjects each. Each group of 10 subjects was randomly paired, with new partners in each of 10 periods. A period began with all subjects selecting a price in the interval [60, 160]. After these prices were recorded, subjects were matched, and each person was informed about the other's price choice. Pay-offs were calculated as described above: the low-price firm earned an amount equal to its price, and the high-price firm earned R times the lowest price. Three sessions were conducted with $R=0.2$ and three with $R=0.8$. Figure 4 shows the period-by-period average price choices. The upper solid line shows the average prices when buyers were relatively unresponsive ($R=0.8$), and the lower solid line shows average prices when buyers were relatively responsive ($R=0.2$). Recall that the Nash equilibrium was 60 for both treatments, as shown by the horizontal dashed line at 60. As intuition suggests, changes in the buyers' responsiveness has a large effect on price, even though the Nash equilibrium remains unchanged.

Notice that prices start high and stay high in the $R=0.8$ treatment, while prices decline before leveling off in the $R=0.2$ treatment. Standard economic models cannot explain either the levels or the patterns of adjustment. Our approach is to consider a naive learning model in which players use observations of rivals' past prices to update their beliefs about others' future actions. In turn, the expected pay-offs based on these beliefs determine players' choice probabilities via a logit rule. This model was used to simulate behavior in the experiment.

To obtain a tractable model, the price range [60, 160] is divided into 101 one-cent categories. Players assign weights to each category and use observations of their rivals' choices to update these weights as follows: each period, all weights are 'discounted' by a factor ρ and the discounted weight of the observed category is increased by 1. In other words, the weight w of an observed category is updated as $w \rightarrow \rho\, w+1$, whereas the other weights are updated as $w \rightarrow \rho w$. The belief probabilities in each period are obtained by dividing the weight of each category by the sum of all the weights. Hence, the learning parameter ρ determines the importance of new observations relative to previous information. Since the most recent observation gets a weight of 1, a lower value of ρ reduces the importance of prior history and increases recency effects.

Generally ρ will be between 0 and 1. When $\rho=0$, the observations prior to the most recent one are ignored, and the model is one of best response to the previously observed price (Cournot dynamics).

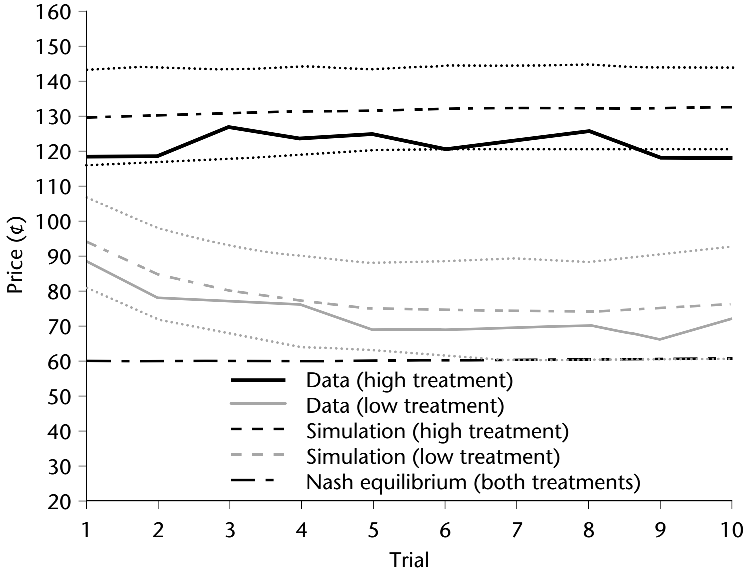

Figure 4. Data and simulations (plus or minus two standard deviations indicated by dotted lines) for a market game. In the 'high' treatment, buyers were relatively unresponsive to differences in price; in the 'low' treatment, buyers were relatively responsive. The simulations are based on a simple belief-learning model using a logit rule to determine probabilities.

At the other extreme, when $\rho = 1$, the model reduces to 'fictitious play', in which each observation is given equal weight, regardless of the number of periods that have elapsed since. For intermediate values of ρ, the weight given to past observations declines geometrically over time.

The expected pay-off for player i choosing a price in category j is denoted by $\pi_i^e(j|\rho)$. This determines player i's decision probabilities via the logit rule

$$P_i(j|p) = \frac{\exp\left(\pi_i^e(j|\rho)/\mu\right)}{\sum_{k=1}^{101} \exp\left(\pi_i^e(k|\rho)/\mu\right)}, \; j = 1, \ldots, 101 \qquad (9)$$

Choice probabilities and expected pay-offs depend on the learning parameter. In this dynamic model, beliefs, and hence choices, depend on the history of what has been observed up to that point. Since individual histories are realizations of a stochastic process, the predictions of this model will be stochastic and can be analyzed with simulation techniques.

The structure of the computer simulation program matches that of the experiment reported below: for each session, or 'run', there are 10 simulated subjects who are randomly matched in a sequence of 10 periods. We specify initial prior beliefs for each subject to be uniform on the integers in the set [60, 160]. These priors determine expected pay-offs for each price, which in turn determine the choice probabilities via the logit rule in equation 9. The simulation begins by determining each simulated player's actual price choice for period 1 as a draw from the logit probabilistic response to the pay-offs for priors that are uniform on [60, 160]. The simulated players are randomly divided into five pairs, and each player 'sees' the other's actual price choice. These price observations are used to update players' beliefs using the naive learning rule explained above, with a learning parameter $\rho = 0.72$ (which was estimated from the data). The updated beliefs, which become the priors for period 2, will not all be the same if the simulated subjects encountered different price choices in period 1. The process is repeated, with the period-2 priors determining expected pay-offs, which in turn determine the logit choice probabilities, and hence the observed price realizations for that period. The whole process is repeated for 10 periods.

Figure 4 shows the sequences of average prices obtained from 1000 simulations, together with dotted lines indicating two standard deviations of the average. These simulation results predict that average prices decline in the $R = 0.2$ treatment and stay the same in the $R = 0.8$ treatment, as observed in the data. Thus, the learning model explains the salient features of the experimental data: both the directions of adjustment and the steady-state levels.

STOCHASTIC LEARNING EQUILIBRIUM

Next we consider what the learning model implies about the long-run steady-state distribution of price decisions. In particular, will learning generate a price distribution that corresponds to some equilibrium?

At any point in time, different people will have different experiences or histories. These differences may be due to the randomness in individuals' decisions or to randomness in the random matching. For each person, the history of what they have seen will determine a probability distribution over their decisions. This mapping of histories to decision probabilities may be direct, as in reinforcement learning. Alternatively, histories may generate beliefs, which in turn produce decisions via a probabilistic choice rule. The decisions made are then appended to the existing histories, forming new histories. Because of the randomness in decision-making, there will be a probability distribution over all possible histories.

In a steady state of the learning model, the probability distribution over histories remains unchanged over time. The 'stochastic learning equilibrium' is defined as the steady-state probability distribution over histories. This formulation is general and includes many learning models as special cases. Goeree and Holt (2002) show that this equilibrium always exists when there is a finite number of decisions and players have finite (but possibly long) memories.

For example, consider the extreme case where a person can only remember the two most recent observations in the probability-matching experiment. There are four possible remembered histories: LL, LR, RL, and RR, with exogenously determined probabilities of $\frac{3}{4} \times \frac{3}{4}$, $\frac{3}{4} \times \frac{1}{4}$, $\frac{1}{4} \times \frac{3}{4}$, and $\frac{1}{4} \times \frac{1}{4}$, respectively. A stochastic learning equilibrium in this context would be a vector of transition probabilities between these states. The formulation of this model in terms of histories (instead of single-period choice distributions) allows the possibility of dynamic effects such as cycles and endogenous learning rules. The focus on histories (sequences of vectors of players' decisions) also facilitates the proof that a stochastic learning exists under fairly general conditions.

Given a specific learning rule, it is possible to determine the stochastic learning equilibrium. To illustrate, consider the market price game under two extreme conditions, fictitious play ($\rho = 1$) and Cournot best response ($\rho = 0$). Since there is no 'forgetfulness' in fictitious play, any steady-state distribution of decisions will eventually be fully learned by all players, i.e. the empirical frequencies of price draws from the distribution will converge to that distribution. In this case, each player is making a logit probabilistic best response to the empirical distribution, and these best responses match the empirical distribution. Notice that all players must have identical beliefs in this equilibrium. This is known as a 'quantal response equilibrium' as defined by McKelvey and Palfrey (1995.)

When $\rho = 0$, a player's history is simply the most recent observation, and beliefs are necessarily different across players. These differences in individuals' beliefs add extra randomness into the steady state. Figure 5 illustrates these observations for the high-*R* treatment of the price-choice game. The solid line represents the stochastic learning equilibrium with an infinite memory ($\rho = 1$), and the dashed line represents the price distribution for the case of one-period memory ($\rho = 0$). Both of these distributions are hump-shaped, with means near the price average observed in the experiment. The implied distribution of price choices is flatter and more dispersed for the case of one-period memory, since beliefs are being moved around by recent observations, which introduces extra randomness. Both cases, however, capture the salient feature of the prices observed in the high-*R* treatment of the market experiment. In particular, price averages are more than twice as high as the unique Nash equilibrium prediction.

When maximum-likelihood techniques are used to estimate the learning parameter from the choices made by the human subjects, the resulting estimate ($\rho = 0.72$) is intermediate between the extreme cases shown in Figure 5, and the resulting steady-state price distribution will also be intermediate. In fact, the weights determined by powers of 0.72 decline very quickly, and the equilibrium price distribution is rather close to the flatter ($\rho = 0$) case, as can be confirmed with computer simulations. Simulations of individual cohorts of 10 subjects (not shown) show the same up-and-down patterns exhibited by comparably sized cohorts of humans. The simulation averages shown in Figure 4 track the main features of the human data: prices start high and stay high in one treatment, and they start high and decline towards the Nash prediction in the other. Thus, computer simulations of learning models can explain data patterns that are not predicted with standard equilibrium techniques. In fact, we ran the computer simulations before we ran the experiments with human subjects, using the learning and error parameter estimates from a previous experiment (Capra *et al.*,

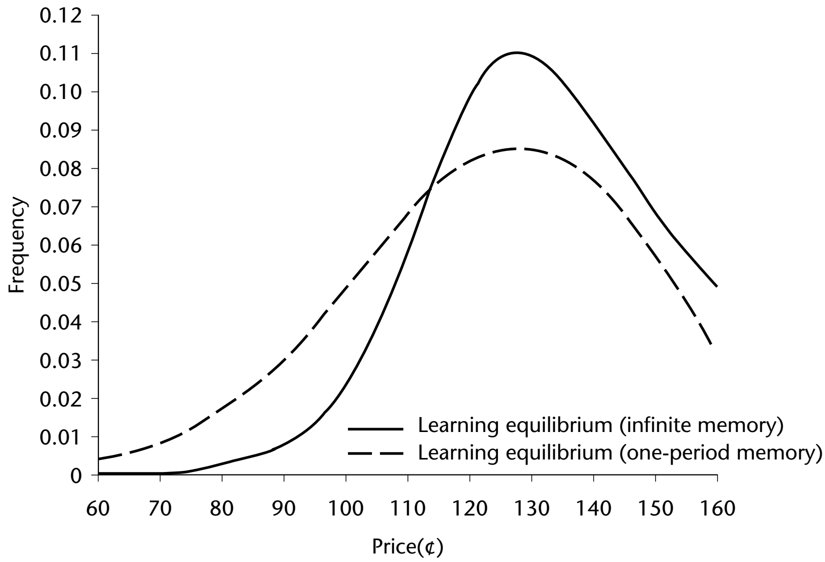

Figure 5. Stochastic learning equilibrium distribution of prices in the price-choice game for $R = 0.8$.

1999). The simulations helped us select two values of the treatment parameter R that would ensure that there would be a strong treatment effect that is not predicted by the Nash equilibrium.

SUMMARY

The learning models presented here were pioneered by Bush and Mosteller (1955), and the stochastic choice models were introduced by the mathematical psychologist Luce (1959) and others. These techniques no longer receive much attention in the psychology literature, where the main interest is in theories of learning, biases, and heuristics that have a richer cognitive content. Yet they have yielded important insights in explaining economics experiments where the anonymity and repetitiveness of market interactions dominate. The incorporation of insights from the literature on heuristics and biases may also prove to be valuable in the future.

The belief- and reinforcement-based learning models depend on past history in a somewhat mechanical manner. In contrast, some laboratory experiments provide situations in which learning seems to proceed in response to cognitive insights. For example, if one subject observes that another has earned more money, the first person may decide to try to imitate the other. Consider a market in which subjects choose 'production quantities' simultaneously, with the advance knowledge that the price at which all production is sold will be a decreasing function of the total quantity produced. Since all output is sold at the same price, the person with the highest quantity will have the highest sales revenue. To the extent that high revenues translate into high profits, the sellers with low quantities may be tempted to imitate the high-output strategies of those who have higher earnings. In this context, the process of imitating high-production sellers can cause the total production to be high. The implications of imitation learning have been studied in a series of recent economics experiments. Offerman and Sonnemans (1998), for example, find evidence that learning is induced to some extent both by imitation of others and by one's own experience.

When some people are following predictable learning patterns, it may be advantageous to try to manipulate others' beliefs via 'strategic teaching'. For example, a dominant seller may punish new entrants by expanding production quantity and thereby driving prices down. This behavior may be intended to 'teach' potential rivals not to enter. This is an important area for future research, and it is complicated by the fact that the person doing the teaching should have a mental model of the others' learning processes.

In some economic situations, learning may occur as a sudden realization that some different decision will provide higher earnings or will avoid losses. A first step in the study of this type of learning may be to measure biological indicators of mental activity for economic tasks that may involve sharp changes in behavior or attempts to anticipate others' decisions (McCabe *et al.*, 2000).

References

Bush R and Mosteller F (1955) *Stochastic Models for Learning*. New York, NY: Wiley.

Camerer C and Ho T-H (1999) Experience weighted attraction learning in normal-form games. *Econometrica* **67**: 827–874.

Capra CM, Goeree JK, Gomez R and Holt CA (1999) Anomalous behavior in a traveler's dilemma? *American Economic Review* **89**(3): 678–690.

Capra CM, Goeree JK, Gomez R and Holt CA (2002) Learning and noisy equilibrium behavior in an experimental study of imperfect price competition. *International Economic Review* **43**(3): 613–636.

DeGroot MH (1970) *Optimal Statistical Decisions*. New York, NY: McGraw-Hill.

Erev I and Roth AE (1998) Predicting how people play games: reinforcement learning in experimental games with unique, mixed strategy equilibria. *American Economic Review* **88**(4): 848–881.

Fantino E (1998) Behavior analysis and decision making. *Journal of the Experimental Analysis of Behavior* **69**: 355–364.

Goeree JK and Holt CA (2002) Stochastic learning equilibrium. Working paper, University of Virginia. [Presented at the Economic Science Association Meetings in New York City, June 2000.]

Luce D (1959) *Individual Choice Behavior*. New York, NY: Wiley.

Mackintosh NJ (1969) Comparative psychology of serial reversal and probability learning: rats, birds, and fish. In: Gilbert R and Sutherland NS (eds) *Animal Discrimination Learning*, pp. 137–167. London: Academic Press.

McCabe K, Coricelli G, Houser D, Ryan L and Smith VL (2000) Other minds in the brain: a functional imaging study of 'theory of mind' in two-person exchange. Working paper, University of Arizona.

McKelvey RD and Palfrey TR (1995) Quantal response equilibria for normal form games. *Games and Economic Behavior* **10**: 6–38.

Offerman T and Sonnemans J (1998) Learning by experience and learning by imitating successful others. *Journal of Economic Behavior and Organization* **34**(4): 559–575.

Siegel S, Siegel A and Andrews J (1964) *Choice, Strategy, and Utility*. New York, NY: McGraw-Hill.

Vulkan N (2000) An economist's perspective on probability matching. *Journal of Economic Surveys* **14**(1): 101–118.

Further Reading

Chen Y and Tang FF (1998) Learning and incentive compatible mechanisms for public goods provision: an experimental study. *Journal of Political Economy* **106**: 633–662.

Cooper DJ, Garvin S and Kagel JH (1994) Adaptive learning vs. equilibrium refinements in an entry limit pricing game. *RAND Journal of Economics* **106**(3): 662–683.

Fudenberg D and Levine DK (1998) *Learning in Games*. Cambridge, MA: MIT Press.

Goeree JK and Holt CA (1999) Stochastic game theory: for playing games, not just for doing theory. *Proceedings of the National Academy of Sciences* **96**: 10564–10567.