A702

ENCYCLOPEDIA OF COGNITIVE SCIENCE

2000

Learning in Economics Experiments

Reinforcement learning, belief learning, experiments, probability matching, market price-choice games, computer simulations.

Goeree, Jacob K

Jacob K Goeree

University of Virginia, Charlottesville, Virginia USA

Holt, Charles A

Charles A Holt

University of Virginia, Charlottesville, Virginia USA

This paper explains how simple psychological models of reinforcement and belief learning can be used to explain dynamic patterns of adjustment in economics experiments.

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 behaviour 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 by psychologists extensively over thirty years ago. This paper 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 behaviour that is known in the psychology literature as "probability matching."

The second part of the paper uses these learning models to analyze behaviour in an economic market where firms choose prices. Markets and games are more complex than individual decision tasks in the sense 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 toward 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 learning basic learning models in the contest of a binary prediction task that has been familiar to psychologists for over fifty years. This task is also of special interest, since 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 each correct choice, and penalties may be deducted for each incorrect choice.

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 illuminates three-fourths of the time, then subjects would come to learn this by experience and then would tend to predict Left three-fourths of the time. This behavior is not rational, since predicting the more likely event will be correct three-fourths of the time. Matching behavior will only generate a correct prediction with a probability of (3/4)(3/4) + (1/4)(1/4), where the first term corresponds to predicting the more likely event with probability 3/4 and being correct with this prediction three-fourths of the time, and the second term is analogous. The probability of being correct under probability matching, therefore, is: $10/16 = 5/8 < 3/4$.

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 behaviour is perplexing given that non-humans are quite adept at optimal behaviour 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 forty years ago provides a high standard of careful reporting and procedures, appropriate statistical techniques, and the use of financial incentives where appropriate. His experiments on probability matching are a good example of this work. In one experiment, 36 male Penn State students were allowed to make predictions for 100 trials, and then 12 of these 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 graphed in Figure 1, with each point being the 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 averages are plotted as the heavy dashed line, which begins 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 the level of 0.75 (shown by a

horizontal line on the right) predicted by probability matching, with a leveling off at about trial 100.

In the "pay-loss" treatment, 12 participants received 5 cents for each correct prediction, and they lost 5 cents for each incorrect decision. The 20-trial averages are plotted as the dark solid line in the figure. Notice that the line converges to a level of about 0.9, as shown by the upper horizontal line on the right. A third "pay" treatment offered a 5-cent reward but no loss for an incorrect prediction, and the results (not shown) are in between the other two treatments, and clearly above 0.75. Clearly, incentives matter, and probability matching is not observed with incentives in this context.

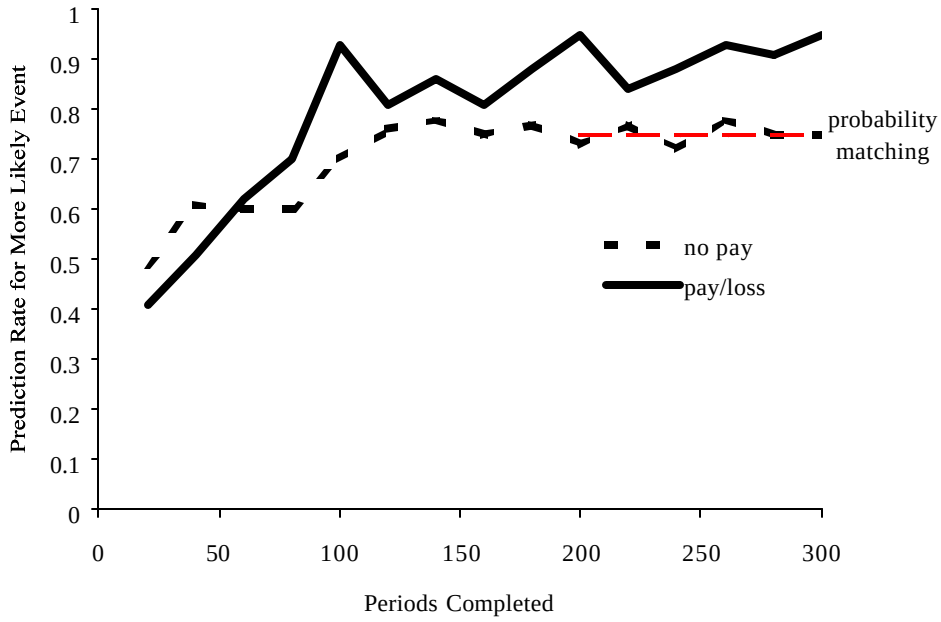

Figure 1. Prediction Proportions for the Event with Frequency 0.75
Source: Siegel et al. (1964)

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" and 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 fifty years of probability matching experiments, Vulcan (1998) separates those studies that used real incentives from those that did not, and he concluded that probability matching is generally not observed with real payoffs, although humans can be surprisingly slow learners in this simple setting. For this reason, probability matching data are particularly interesting for evaluating alternative learning theories.

Reinforcement Learning

One prominent theory of learning associates changes in behaviour 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. We will model this in terms of a positive parameter, α, which will be explained below.

$$(1) \qquad \Pr(\mathrm{L}) = \frac{a}{a+a} \quad \text{and} \quad \Pr(\mathrm{R}) = \frac{a}{a+a} \qquad \text{(priors)},$$

Of course, at this point α plays no role since the above probabilities are both 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. One way to model this is to let the choice probability be:

$$(2) \qquad \Pr(\text{choose L}) = \frac{a+x}{a+x+a} \quad \text{and} \quad \Pr(\text{choose R}) = \frac{a}{a+x+a} \ .$$

Reinforcements for the Right choice are defined similarly. Notice that the α parameters determine how quickly learning responds to the reinforcements.

As additional reinforcements are received they are added into the relevant numerator, and to both denominators to ensure that the probabilities add to 1. 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 e_L, would be less than xN_L. Similarly, let e_R be the total earnings from the correct R predictions. The choice probabilities would then be:

$$(3) \qquad \Pr(\text{choose L}) = \frac{a+e_L}{2a+e_L+e_R} \quad \text{and} \quad \Pr(\text{choose R}) = \frac{a+e_R}{2a+e_L+e_R} \ .$$

This kind of model might explain some aspects of behaviour 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 percent of the time, and the resulting asymmetries in reinforcement would tend to raise prediction probabilities 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 as e_L gets larger. Thus the probability of choosing L in (3) would tend to converge to 1.

The learning model in (3) can be simulated by using past accumulated earnings to compute choice probabilities. Then a random-number generator determines the actual choices. To be comparable we simulate decisions of a cohort of twelve individuals for 300 periods and calculate the twenty-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 behaviour, which will disappear after forty or fifty periods. It is apparent from (3) that an α equal to 5 is analogous to having had

each decision reinforced five times. The dashed line in Figure 2 shows simulated choice averages together with Siegel's original data. The simulated data are smoother and start somewhat higher, but the general pattern and final tendencies are quite similar. Erev and Roth (1997) have used reinforcement learning to explain behavior in simple matrix games.

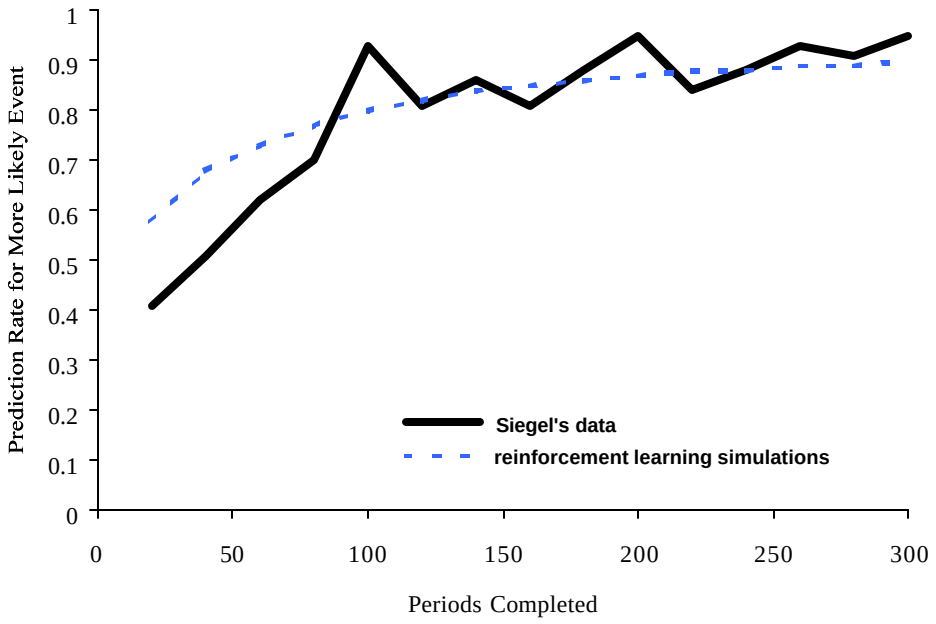

Figure 2. Data and Simulations for Siegel's Probability Matching Experiment

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. As before, we model initial beliefs for the probability of events L and R in terms of parameter β:

(4) $$\Pr(\mathrm{L}) = \frac{b}{b+b} \quad \text{and} \quad \Pr(\mathrm{R}) = \frac{b}{b+b} \qquad \text{(priors)},$$

If event L is observed, then Pr(L) should increase, so let us add 1 to the numerator for Pr(L). To make the two probabilities sum to 1, we must add 1 to the denominators for each probability expression:

$$(5) \qquad \Pr(\mathrm{L}) = \frac{b+1}{b+1+b} \quad \text{and} \quad \Pr(\mathrm{R}) = \frac{b}{b+1+b} \qquad \text{(after observing L).}$$

Note that β determines how quickly the probabilities respond to the new information; a large value of β will keep these probabilities close to 1/2. Continuing to add 1 to the numerator of the probability for the event just observed, and to add 1 to the denominators, we have a formula for the probabilities after N_L observations of event L and N_R observations of event R. Let N be the total number of observations to date. Then the resulting probabilities are:

$$(6) \qquad \Pr(\mathrm{L}) = \frac{b+N_L}{2b+N} \quad \text{and} \quad \Pr(\mathrm{R}) = \frac{b+N_R}{2b+N} \qquad \text{(after } N \text{ observations).}$$

where $N = N_L + N_R$. This formula for calculating probabilities can be derived from Bayesian statistical principles (see DeGroot, 1970, p. 160). In the early periods, the totals, N_L and N_R, might switch in terms of which one is higher, but the more likely event will soon dominate, and therefore Pr(L) will be greater than 1/2.

The beliefs in (6) determine the expected payoffs (or utilities) for each decision, which in turn determine the decisions made. In theory, the decision with the highest expected payoff is selected with certainty. The prediction of the belief-learning model is, therefore, that all people will eventually start to predict the more likely event every time.

In an experiment, however, some randomness in decision-making might be expected if the expected payoffs for the two decisions are not too different. 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 positive but not perfectly related to expected payoffs. Let π_L and π_R denote the expected payoffs from choosing Left and Right respectively. Luce provided a set of axioms under which the choice probability is calculated as:

$$(7) \quad \Pr(\text{choose L}) = \frac{(p_L)^{1/m}}{(p_L)^{1/m} + (p_R)^{1/m}} \quad \text{and} \quad \Pr(\text{choose R}) = \frac{(p_R)^{1/m}}{(p_L)^{1/m} + (p_R)^{1/m}}.$$

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

In the probability matching experiment, the expected payoff of choosing Left is the reward of 1 times the probability of Left that represents the person's beliefs. Thus the expected payoff of Left is Pr(L) and, similarly, the expected payoff of Right is Pr(R). It follows from (7) that the probability of choosing Left 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 in (6) 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 true probability of the more likely event (0.75). The dashed lines show the simulated

average choice frequencies for three different levels of the error parameter. With high error ($\mu = 1$), the dashed line choice frequencies bounce around the belief line, which would correspond to probability matching. This result can be expected from (7), since expected payoffs are equal to belief probabilities. Therefore the denominator on the right side of (7) 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 dashed lines representing simulated choice frequencies move upward toward the optimal level of 1. The top dashed line with $\mu = 1/3$ converges to the level of about 0.9, which is close to the choice frequency observed by Siegel.

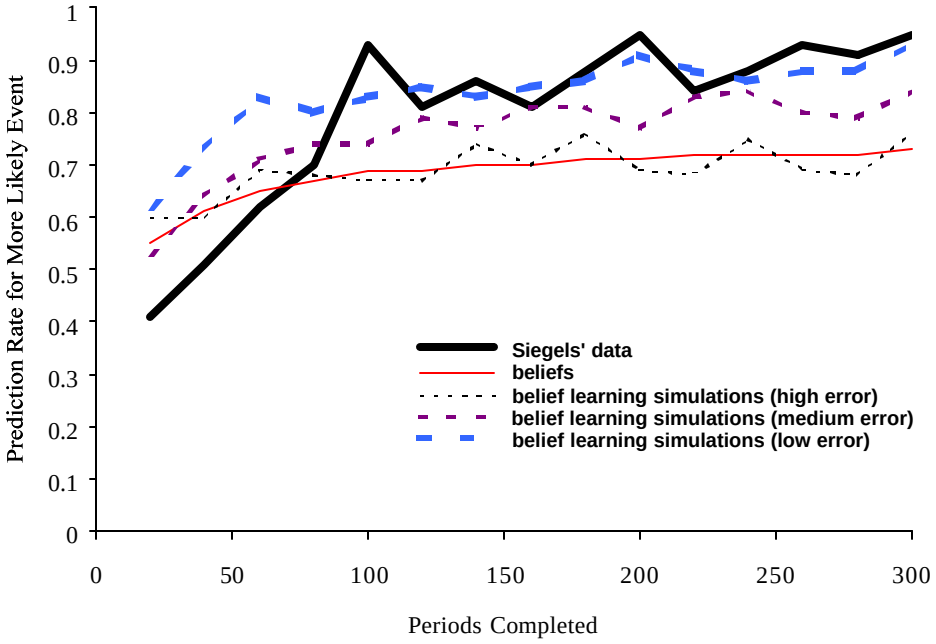

Figure 3. Data and Simulations for Siegel's Probability Matching Experiment

Generalizations

Both of the learning models discussed here are somewhat simple, which is part of their appeal. The reinforcement model builds in some randomness in behaviour 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, which seems like an unreasonable prediction. Obviously, people learn something in the absence of previously received 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 received partial reinforcement even if nothing is earned.

These learning models can be enriched in other ways to obtain better predictions of behaviour. 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 like L in the most recent trail 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 (6) 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, etc. This type of "recency" effect will be discussed in the next section in the context of an interactive market game.

Finally, the "Luce probabilistic choice rule" in (7) is often replaced with the "logit rule:"

$$\text{Pr(choose L)} = \frac{\exp(p_L/m)}{\exp(p_L/m)+\exp(p_R/m)}, \qquad (8)$$
$$\text{Pr(choose R)} = \frac{\exp(p_R/m)}{\exp(p_L/m)+\exp(p_R/m)},$$

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 payoffs are increased by a constant, and the Luce probabilities are unchanged when all payoffs 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]. Demand is assumed to be a fixed total quantity ("perfectly inelastic"). The sales quantity of the firm with the low 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 degree to 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 due to 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 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, firms have an incentive to undercut the other by a small amount to increase 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 simple economic intuition that the degree of buyer inertia will affect the behaviour 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 $R \geq 1$), 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 group of ten subjects was randomly paired with new partners in each of ten 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. Payoffs 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 done 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 for the other treatment (the dashed lines are simulation results explained below). 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 impact on price, even though the Nash equilibrium remains unchanged.

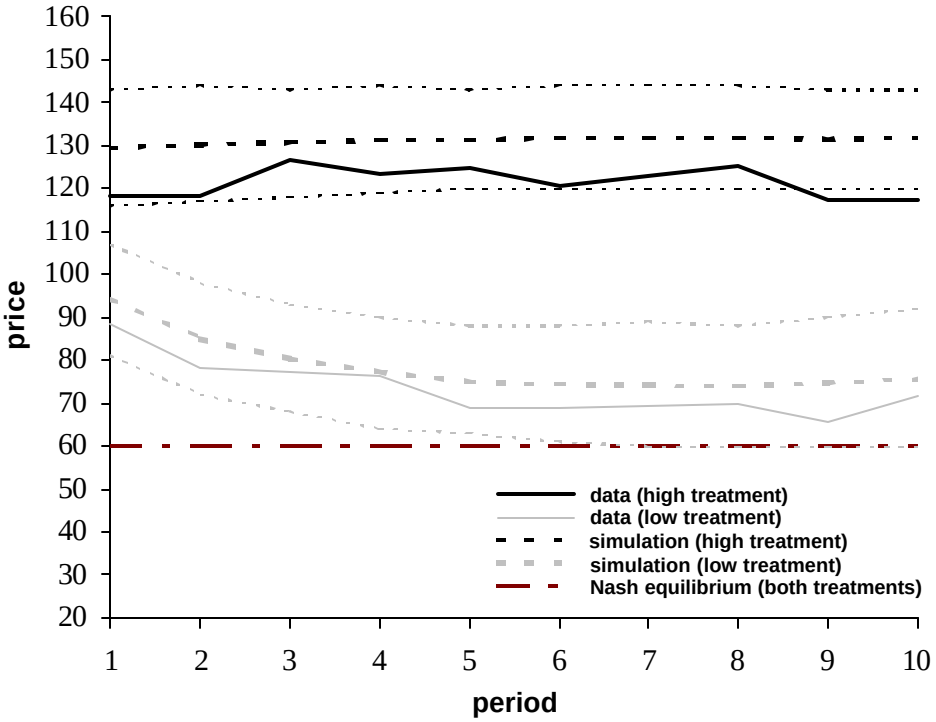

Figure 4. Data and Simulations (plus or minus two standard deviations)

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 payoffs based on these

beliefs determine players' choice probabilities via a logit rule. This model was used to simulate behaviour 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 rival's 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 $w \leftarrow \rho w + 1$, whereas the other weights are discounted by ρ: $w \leftarrow \rho w$. The belief probabilities in each period are obtained by dividing the weight of each category by the sum of all weights. Hence, the model is one in which 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). 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 that observation. For intermediate values of ρ, the weight given to past observations declines geometrically over time.

The expected payoff for player i choosing a price in category j is denoted by $\pi_i^e(j|\rho)$, which determines player i's decision probabilities via the logit rule in (8):

$$(9) \qquad P_i(j \mid r) = \frac{\exp(p_i^e(j \mid r)/m)}{\sum_{k=1}^{101} \exp(p_i^e(k \mid r)/m)}, \quad j = 1,..,101.$$

The ρ notation indicates the dependence of choice probabilities and expected payoffs 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 ana lyzed with simulation techniques.

The structure of the computer simulation program matches that of the experiment to be 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 payoffs for each price, which in turn, determine the choice probabilities via the logit rule in (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 payoffs for priors that are uniform on [60, 160]. The simulated players are randomly divided into five pairs, and each one "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. Next, the process is repeated, with the period 2 priors determining expected payoffs, 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 (thick dashed lines) obtained from 1,000 simulations together with plus or minus two standard deviations of the

average (thin dashed lines). 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. To summarize, 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 thereby forming new histories. Due to 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 exist when there are a finite number of decisions and players have finite, but possibly long, memories.

Given a specific learning rule, it is possible to solve for the stochastic learning equilibrium. To illustrate, consider the market price game under to extreme cases, 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. (Incidentally, 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 adds 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 traces out the price distribution for the case of one-period memory ($\rho = 0$). Both of these distributions are hump-shaped with means near the observed price average in the experiment.

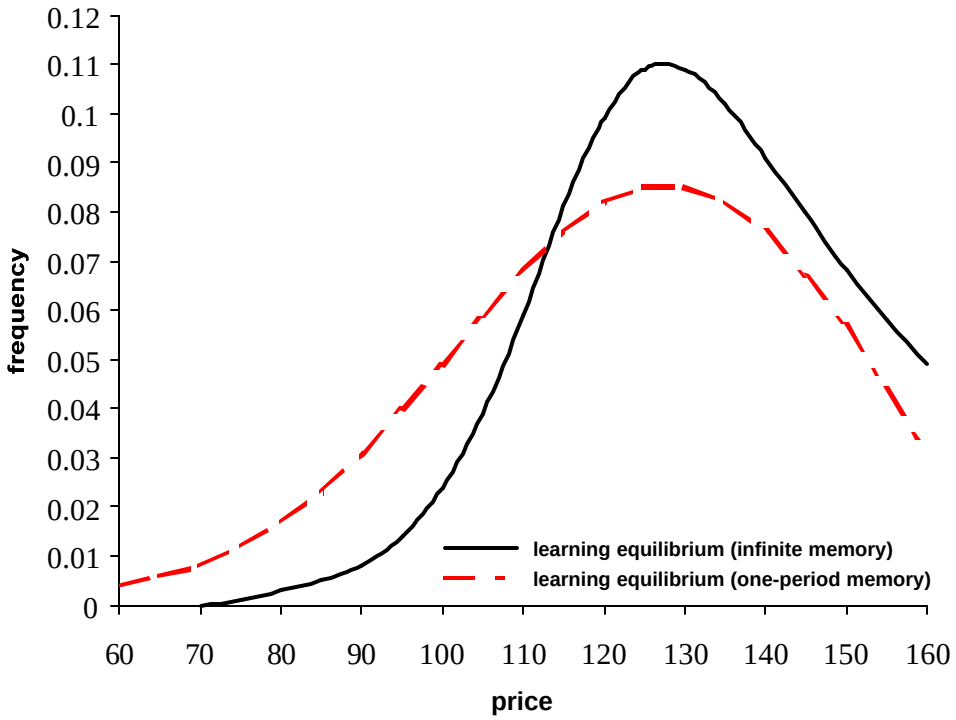

Figure 5. Stochastic learning equilibrium distribution of prices for $R = 0.8$.

Summary

The main ideas can be recapped by reviewing the five figures. The probability matching data in Figure 1 display a relatively slow learning process that never really settles down, as the jagged patterns in 20 period averages continue for many periods. Simulations of learning in Figure 2 are too smooth. This is because these simulations were done with reinforcement learning rules that do not allow anything to be forgotten, so the effects of new draws are rapidly overwhelmed by the weight of all past history. One way to introduce recency effects is to let past observations receive diminishing weights, as is the case for the geometrically declining weights used in the belief learning simulations described above. These effects will cause choices to bounce around as beliefs continue to me moved by the more recent observations.

Recency effects are even more likely in interactive systems where the observations being predicted (e.g. others' prices) are not exogenous, but rather, are themselves generated by learning mechanisms and stochastic choice. For example, consider an extreme case where a person can only remember the two most recent observations. There are four possible remembered histories in the probability matching experiment: LL, LR, RL, and RR, with exogenously determined probabilities of (3/4)(3/4), (3/4)(1/4), (1/4)(3/4), and (1/4)(1/4) respectively. In an interactive market or game, histories are generated by players' decisions, so they will depend on additional factors such as the payoffs and error parameters from the stochastic choice rules.

A *stochastic learning equilibrium* (Goeree and Holt, 2002) is a steady-state probability distribution over all possible histories. The formulation of this model in terms of histories (instead of single-period choice distributions) allows the possibility

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.

Figure 5 shows the implications of two special cases of the stochastic learning equilibrium for the market price game: the (limiting) case of an infinite history ($\rho = 1$) and the case of a one-period history ($\rho = 0$). The implied distribution of price choices is flatter and more dispersed for the latter case, since beliefs are being moved around by recent observations, which introduces extra randomness. Both of these extreme cases, however capture the salient feature of the prices observed in the high *R* treatment of the market experiment, i.e. that 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 products of 0.72 decline very quickly, and the equilibrium price distribution is quite close to the flatter ($\rho = 0$) case, as we confirmed with computer simulations. Simulations of individual cohorts of ten subjects (not shown) also 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 toward the Nash prediction in the other. Thus computer simulations of learning models can explain the 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., 1999). The simulations helped us select the two values of the treatment parameter, *R*, which would ensure that there would be a strong treatment effect that is not predicted by the Nash equilibrium.

The learning models used 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 on theories of learning, biases, and heuristics that have a richer cognitive content. We wish to stress that the classical learning and stochastic choice techniques have proved to yield important insights in explaining economics experiments where the anonymity and repetitiveness of market interactions dominate, although the incorporation of insights from the heuristics and biases literature may also prove to be valuable in the future.

References

Bush R and Mosteller F (1955) *Stochastic Models for Learning*, New York: 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 (2002) Learning and Noisy Equilibrium Behavior in an Experimental Study of Imperfect Price Competition. *International Economic Review*, forthcoming.

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

Erev I and Roth AE (1997) Predicting How People Play Games: Reinforcement Learning in Experimental Games with Unique, Mixed Strategy Equilibria. *American Economic Review* **88**(4): 848-881.

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: John Wiley and Sons.

McKelvey RD and Palfrey TR (1995) Quantal Response Equilibria for Normal Form Games. *Games and Economic Behavior* **10**: 6-38.

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

Further Reading

Capra CM, Goeree JK, Gomez R and Holt CA (1999) Anomalous Behavior in a Traveler's Dilemma? *American Economic Review* **89**(3): 678-690.

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.