Eastern Economic Journal, **2011**, **37**, (109–133)
© 2011 EEA 0094-5056/11
www.palgrave-journals.com/eej/

Reinforcement Learning in Experimental Asset Markets

Shu-Heng Chen and Yi-Lin Hsieh
AI-ECON Research Center, Department of Economics, National Chengchi University, Taipei 116, Taiwan.
E-mails: chchen@nccu.edu.tw; littleyam1982@yahoo.com.tw

In this paper, we study the learning behavior possibly emerging in six series of prediction market experiments. We first find, from the experimental outcomes, that there is a general positive correlation between subjects' earning performance and their reliance on using limit orders to trade. We therefore focus on the subjects' learning behavior in terms of their use of limit orders or market orders by estimating a three-parameter Roth–Erev reinforcement learning model. The results of the estimated parameters show not just their great heterogeneity, but also the sharp contrasts among subjects, which in turn impact the subjects' earning performance.

Eastern Economic Journal (2011) **37,** 109–133. doi:10.1057/eej.2010.57

Keywords: reinforcement learning; prediction market experiments; limit order; market order; zero intelligence

JEL: C51; C91; D83

MOTIVATION

Learning is a multi-disciplinary area that continuously draws the attention of economists, psychologists, cognitive scientists, computer scientists, mathematicians, and neuroscientists. Among the learning models proposed, *reinforcement learning* seems to be one that is most commonly shared by all these disciplines. In psychology, Bush and Mosteller [1955] proposed the first mathematical model of reinforcement learning; this Bush–Mosteller model has since been adapted and generalized in economics by Cross [1973], Arthur [1991], Arthur [1993], Roth and Erev [1995], and Erev and Roth [1998]. On the basis of this development, the relevance of reinforcement learning in economic experiments has become increasingly apparent. In particular, given the rapid growth of the bounded-rationality literature, reinforcement learning stands for an important class of learning. It is a kind of *reflexive learning*, that is, learning involving little reasoning [Bossaerts et al. 2008]. Owing to this unique position, reinforcement learning has been frequently compared with other more sophisticated learning models in either experimental economics or agent-based computational economics [Duffy 2006].

The mathematical development of reinforcement models has constantly added some interesting cognitive or psychological parameters to the model. For example, *memory capacity* and *generalization capacity* have been added by Roth and Erev in the Roth–Erev three-parameter reinforcement learning (RL) models [Roth and Erev 1995]. These cognitive psychological parameters, in general, are expected to differ among agents, as we have learned from various results from psychometric tests [Borghans et al. 2008]. However, this heterogeneity has been largely unexploited in the empirical studies of RL models in experiments involving human subjects. By contrast, since the pioneering work conducted by Arthur [1991; 1993], most empirical applications of RL models have focused only on aggregate behavior, and hence only aggregate RL models have been built. Doing so, implicitly assumes that all subjects share the same cognitive psychological parameters, even though it is known that there were evident variations in subjects' behavior in the experiments.[1] Therefore, two issues are not sufficiently addressed in the literature. First, if the agents' learning can be realistically represented by reinforcement learning, then how heterogeneous can they possibly be? Second, if they are heterogeneous to some extent, do those cognitive psychological parameters matter for agents' performance? The second issue corresponds to the recent surge in interest in exploring the economic consequences of personal psychology traits, for example, Borghans et al. [2008].

This paper attempts to address the above-mentioned two issues. In the context of *experimental asset markets*, we study subjects' trading behavior in terms of the two types of order, namely the *market order vs limit order*. Market microstructure theory informs us of different risk exposures associated with these two types of order: the former to the *price risk*, and the latter to the *execution risk*.[2] In this paper, we design six different experimental asset markets and observe 120 subjects' choices regarding these two types of orders. By assuming that these observable choices are *adaptive* and can be described by a simple learning process, we apply the three-parameter Roth–Erev model to each subject and estimate the parameters of the model using maximum likelihood estimation (MLE).

Since each of the three parameters can be given a psychological consideration, what we infer from the data can be regarded as the psychological traits of each individual. We first find that subjects differ quite clearly in terms of the estimated parameter values. Our statistical analysis further shows that this inferred heterogeneity is not due to the experimental designs, and suggests that it mainly reflects the diversity of subjects' personality traits. Finally, we then examine whether these personality traits may have economic consequences, such as an effect on earning capacity. Statistical tests indicate that two out of three parameters may be important, which we identify as *attention control* and *responsiveness*. Subjects with stronger attention control and greater responsiveness tend to earn more than otherwise. Hence, to be winners, not only do subjects need to learn, but they also need to learn with right psychological traits.

The rest of the paper is organized as follows. The next section highlights the key features of the experimental design. The third section discusses the likely learning behavior of subjects in the experiments. The fourth section introduces the three-parameter RL model, which we employ to study the learning behavior of subjects. The fifth section shows the estimation of the three-parameter RL model, and presents the great heterogeneity of our subjects. The sixth section addresses the issue regarding the causes of the observed heterogeneity. The penultimate section discusses the possible economic consequences of the observed heterogeneity. The final section concludes the paper.

EXPERIMENTAL DESIGNS

To run experiments, we used the AI-ECON Prediction Market [Chen and Wu 2009], which was originally designed as a virtual political future market. As we are using political future markets, agents in these experiments can trade based on their

Table 1 Experimental designs

(1) *Markets*	*(2)* *Perfect insiders*	*(3)* *Imperfect insiders*	*(4)* *Insiders*	*(5)* *Outsiders*	*(6)* *Market participants*	*(7)* *Composition information*
A	0 (0%)	0 (0%)	0 (0%)	20 (100%)	20	Y
B	5 (25%)	5 (25%)	10 (50%)	10 (50%)	20	Y
C	8 (40%)	8 (40%)	16 (80%)	4 (20%)	20	Y
D	6 (30%)	10 (50%)	16 (80%)	4 (20%)	20	Y
E	10 (50%)	6 (30%)	16 (80%)	4 (20%)	20	Y
F	5 (25%)	5 (25%)	10 (50%)	10 (50%)	20	N

Note: Inside the parentheses are the percentages of the respective types of traders in the market. The last column indicates whether the exact composition of traders is made known to market participants. "Y" indicates that the composition is public information, and "N" refers to the opposite.

expectations of the "true value" of an asset at the expiration date, also called the *expiration price*. In the political future market, the expiration price is not known by any subjects. However, in our experiments, we introduce *insiders* to markets, and release some information regarding the expiration price to these insiders. Motivated by Ang and Schwarz [1985], we further distinguish *perfect insiders* from *imperfect insiders*. Perfect insiders know the exact asset price at the expiration date, whereas imperfect insiders do not know the exact expiration price, but are aware of a specific range covering the true price.

In total, we have six different market designs. Each market experiment is composed of 20 traders. Each trader is initially endowed with 10,000 in cash and 20 shares of an asset. Each experiment comprises 12-14 rounds. Each round lasts for 7 min. In each market round, traders can trade either by submitting *limit orders* or *market orders*. Short sales and margin buying are not permitted. Table 1 provides a summary of the experimental designs.

The six market experiments are correlated in the following way. Markets A, B, and C are closely related because from A to C we increase (decrease) the number of insiders (outsiders). Markets C, D, and E are closely related because from D to C and to E we increase (decrease) the number of perfect insiders (imperfect insiders) while keeping the number of outsiders unchanged. Markets B and F are also closely related because they are exactly same in terms of the composition of traders, except for the availability of this composition information. In each of our six experiments, all traders know of the existence of insiders, but in Market F they do not know the exact numbers of various types of traders (see column (7) of Table 1). Details are given in the appendix.

WHAT CAN WE LEARN?

In a lab with human subjects, it is sometimes difficult to be precise as to what agents learned or were trying to learn. In principle, they should have an incentive to learn everything that they consider to be relevant to their gains or profits. However, deciding what is relevant and what is not is itself a part of learning. In this study, we assume that the main subject for agents to learn is *the use of limit orders*, more exactly the *intensity* of using limit orders, as opposed to the alternative, market orders. We shall first define what the intensity of the limit order (ILO) is (the first subsection) and then justify its choice as a focus to learn (the second subsection). However, focusing only on one dimension of learning and ignoring others can be problematic; we, therefore, also briefly discuss this issue (the last subsection).

Intensity of the limit order

The intensity of using limit orders (ILO) is defined as in equation (1). For each market, at the end of each round, say, the *t*th round, we can observe the total submission of each subject, say, *j*. This total submission is the sum of the limit order submission (LOS) and the market order submission (MOS). The intensity of using the limit order by the *j*th subject in the *t*th round, $ILO_{j,t}$, can be defined as follows:

$$ILO_{j,t} = \frac{LOS_{j,t}}{LOS_{j,t} + MOS_{j,t}} \tag{1}$$

ILO as a target of learning

Needless to say, choosing ILO as the learning target is certainly a simplification of the potentially more complex and multifaceted learning. Nonetheless, statistics from our experimental results (Table 2) show that the intensity of using the limit order is positively correlated with the realized profits.

Let $\pi_{j,t}$ be the profit earned by subject *j* at the end of the *t*th round of one market experiment. By summing up the submission and the associated profit over all rounds, one can have

$$ILO_j = \frac{\sum_{t=1}^{T} LOS_{j,t}}{\sum_{t=1}^{T} LOS_{j,t} + \sum_{t=1}^{T} MOS_{j,t}}$$

and

$$\pi_j = \sum_{t=1}^{T} \pi_{j,t}$$

where *T* is the total number of rounds in that market experiment. Table 2 gives the correlation between ILO_j and π_j. To take into account the possible divergence, both the Kendall and Spearman correlations are provided, while, qualitatively, the result is not much different. They both clearly show the significant positive correlation that exists, which indicates that individuals who use the LOS more intensively also tend to earn a higher profit.[3]

Furthermore, Figure 1 gives the dynamics of the ILO of the best-performing five subjects and the worst-performing five subjects in Markets A, C, and E.[4] From these

Table 2 Rank correlation between the intensity of the limit order submission and profits

Market	*Kendall correlation*	*Spearman correlation*
A	0.568*	0.721*
B	0.720*	0.865*
C	0.717*	0.872*
D	0.727*	0.894*
E	0.765*	0.898*
F	0.377†	0.566*

Note: The symbol * refers to the statistical significance at a significance level of 1 percent, and † refers to the statistical significance at a significance level of 5 percent.

figures, we can see that the ILO dynamics vary among subjects. Both fluctuating patterns and converging patterns exist. The general feature is that those subjects whose ILO converges to or fluctuates around a higher value of ILO tend to be among the few best performers, and those subjects whose ILO converges to or fluctuates around a lower value tend to be among the few worst performers. This feature, which is shown in all markets, justifies ILO as a learning target.

However, one reservation for ILO as a learning target is that the result of Table 2 and Figure 1 may be superficial, or is *ex ante* obvious. This concern comes from the design of our experiments involving perfect (and imperfect) insiders, who know the expiration price and hence they tend to use a limit order when the market price deviates from the expiration price and hence can make profits from this information. The concern includes two hypotheses. The first hypothesis is that insiders would trade using limit orders more intensively than outsiders; as a consequence, the

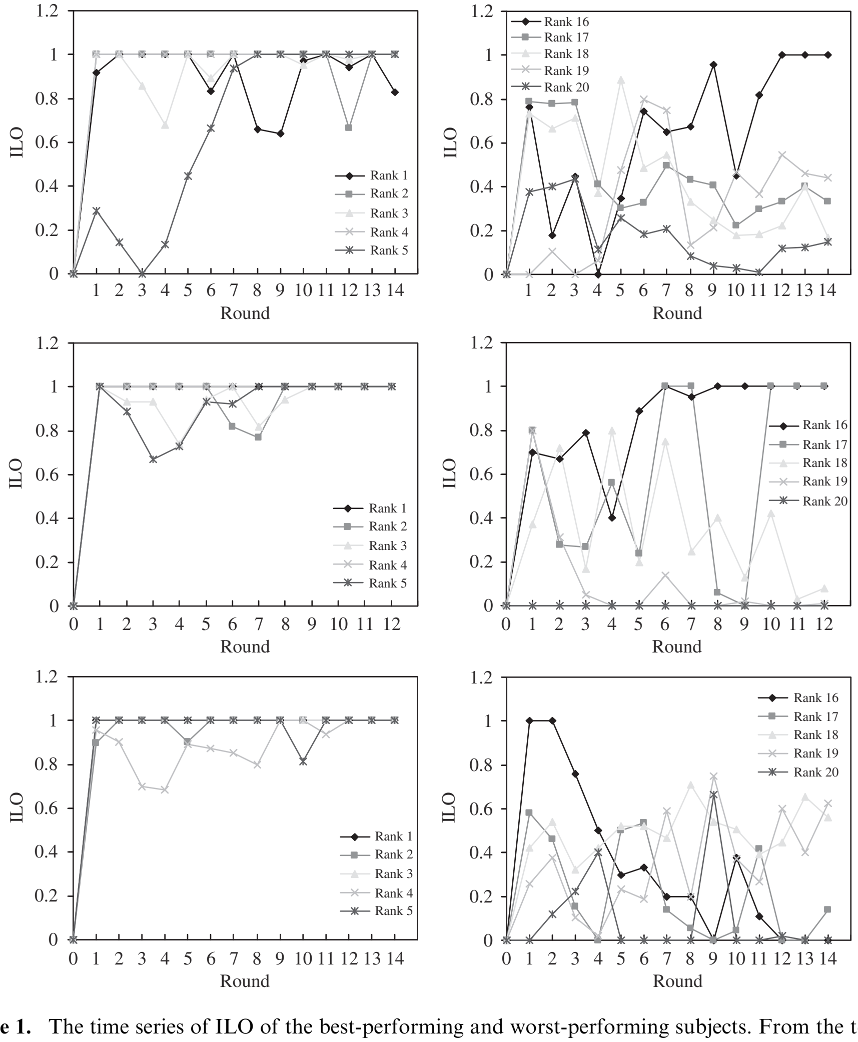


Figure 1. The time series of ILO of the best-performing and worst-performing subjects. From the top to the bottom are the results of Markets A, C, and E, respectively. For each market, the ILO time series of the five best-performing subjects are presented in the left-hand side, and the ILO time series of the five worst-performing subjects are presented in the right-hand side.

second hypothesis is that the former profited more than the latter. Hence, putting them together, the positive relationship between the ILO and the profits earned is only superficial. We did run the Wilcoxon rank sum test for these two hypotheses, round to round and pair by pair from Markets B to F, where both insiders and outsiders appeared. The two hypotheses were rejected in almost all rounds (see Tables A1 and A2 in the appendix). First, the insiders did not perform better than the outsiders; second, they did not trade more intensively using limit orders than the outsiders.[5] In sum, our institutional design did not directly prompt the insiders to follow any specific way to trade and to profit. As a result, we have reason to believe that some agents, if not all, were trying (learning) the ILO during the experiment.

To reduce the possible dimensions of learning, we further assume that the subjects have tried to learn whether they should increase their reliance on the use of limit orders, decrease it, or keep it unchanged. This can, of course, be a rough approximation since these three options are not made explicitly for subjects to choose from, and not explicitly in terms of the extent of the *multi-armed bandit experiments* in psychology and economics [Arthur 1993; Gabaix et al. 2006]. Nonetheless, from the screens that subjects are able to see during the trading process, the distinction between the option “market order” and the alternative “limit order” is very noticeable (see Appendix). In each period, we can see the change in their ILO_j index, as shown in Figure 1. We, therefore, assume that they know these options implicitly: “Look at that! Market orders made me suffer a great loss, so I will change my order type to the limit order,” or “Gee! It seems difficult to have my limit order matched and I just missed a good trading opportunity. Let me change to the market order next time.” After all, as we shall see in the fifth section, the subjects’ choice among these three actions is not random; therefore, how their decision was made deserves a closer look.

Other dimensions of learning

In addition to the choice of trading scheme (ILO), other explicit decisions that need to be made include when to trade and how much to trade (as we can see in the appendix), which may also involve price expectations, and hence learning. We do not intend to ignore them, but assume that they can be separated from our focus on ILO. In other words, we assume a decomposable structure of learning. Hence, the learning behavior observed in complex economic experiments can be studied by taking advantage of this structure and by applying a different learning model to each constituent. For example, in our case decomposability could mean that decisions on ordering schemes and price expectations are independent, so that their learning can be modeled separately. If this assumption is not true, then a multi-dimensional learning model that simultaneously incorporates multiple decisions is more appropriate. Regardless of what this model can be, it would be much harder than the one we pursue here; hence, we leave it as a direction for further study.

THE THREE-PARAMETER RL MODEL

Why reinforcement learning?

In this section, we shall introduce the RL model used in this paper. Before proceeding further, we notice that in the learning literature reinforcement learning has been frequently used in horse racing with other alternatives: some are popular in normal-form games such as *belief-based learning*, and some are popular in agent-based simulations such as *genetic algorithms* (GAs). To the best of our knowledge, GAs have not been applied to fit the behavior of individual subjects observed in experiments, while they have been extensively used to construct artificial agents to mimic human subject experiments. One of the possible reasons is that GAs have many parameters to be determined, but there are only limited observations available from human subject experiments (humans can get fatigued quite quickly). In addition, given our discussion in the third section, using multidimensional models such as GAs is not necessary.

We did not consider belief-based learning and several more sophisticated and general extensions, such as exponentially weighed attraction [Camerer and Ho 1999], because the applicability of these learning models to areas outside normal form-games is yet to be seen.[6] In particular, these learning models demand data on opponents' histories of play, which in our experiment was largely unavailable to subjects (see Appendix). One virtue of using the RL model is its simplicity in the sense that it only requires the information of the subject's own payoff and his own chosen action. As Duffy [2006] has pointed out, if subjects' information is restricted to their own histories of play, then RL models are more adequate.

Updating of propensities

The intuitive idea of the RL model is that *choices that have led to good outcomes in the past are more likely to be repeated in the future*. To make it operational, the RL model assigns each possible action a probability of being activated, chosen, or taken. The entire probability function over the action space is based on the *propensity* of each action. The propensity of an action is its accumulated received rewards (utilities) over the past. The propensity and the activation probability of each action are constantly updated by taking into account the rewards received most recently.

In the literature, several different versions of RL models have been proposed. They differ in how the propensity is updated and how it is mapped to the activation probability. The specific model that we consider in this paper is a version of the *Roth–Erev model* [Roth and Erev 1995, Erev and Roth 1998].

At the beginning, the propensity of each action is treated equally and is given by q_1 ($q_{i,1}=q_1$, $\forall i=1, \ldots, N$). Then how it is updated depends on whether it has been activated. If i is activated (chosen) in the previous period, then its propensity at time t will be updated by adding the payoff received from the activation, that is, Π_{t-1}; otherwise, it is the same as $q_{i,t-1}$:

$$q_{i,t} = \begin{cases} q_{i,t-1} + \Pi_{t-1} & \text{if } i \text{ is chosen in period } t-1 \\ q_{i,t-1} & \text{otherwise} \end{cases} \tag{2}$$

From the psychological consideration/a psychological perspective, Roth and Erev [1995] further introduced two additional elements on top of this basic model (2). These two "weak psychological assumptions," as referred to in Erev and Roth [1998], are the *recency effect* and *experimentation*.

Recency effect

The first element is to take the influence of *time* into account, and assume that propensity will decay with time. Denote the constant decaying rate by φ. Then equation (2) can be developed into equation (3).

$$q_{i,t} = \begin{cases} (1-\varphi) * q_{i,t-1} + \Pi_{t-1} & \text{if } i \text{ is chosen in period } t-1 \\ (1-\varphi) * q_{i,t-1} & \text{otherwise} \end{cases} \tag{3}$$

A reasonable range of the parameter φ is between 0 and 1. If $\varphi = 1$, the propensity of the last period $q_{i,t-1}$ will be completely ignored (forgotten), and the strength-updating only depends on the most recently received payoffs. On the other hand, if $\varphi = 0$, the past propensity will not decay at all, and it in its entirety will roll over to this period. Hence, φ is also known as the *forgetting parameter*.

Experimentation

The second added element had its original motivation known as *experimentation* or *generalization* when it was first introduced in Roth and Erev [1995]. The idea is not only to update the propensity of the activated action with its received payoff, but also to use this payoff to update the propensity of *similar* actions. As interpreted by Erev and Roth [1998], "Not only are choices which were successful in the past more likely to be employed in the future, but *similar choices* will be employed more often as well, and players will not (quickly) become locked in to one choice in exclusion of all others" (*ibid*, p. 863), or "players will generalize their most recent experience in a way that leads to experimentation among the most similar strategies" (*ibid*, p. 863) To achieve this purpose, the experimentation parameter, ε, is introduced to the model, and equation (2) is expanded to (4):

$$q_{i,t} = \begin{cases} q_{i,t-1} + (1-\varepsilon) * \Pi_{t-1} & \text{if } i = i^*_{t-1} \\ q_{i,t-1} + \dfrac{\varepsilon}{(N_{i,t-1}-1)} * \Pi_{t-1} & \text{if } i \sim i^*_{t-1} \\ q_{i,t-1} & \text{otherwise} \end{cases} \tag{4}$$

Here, $N_{i,t-1}$ refers to the size of the neighborhood $\{i|i{\sim}{}^{*}_{t-1}\}$, that is, the collection of non-chosen actions that, however, are similar to the chosen one i^*_{t-1}. $1-\varepsilon$ is the reserve of the own payoff to the chosen action. The part that is not reserved is then equally shared by its neighbors (similar actions), $\varepsilon/(N_{i,t-1}-1)$. A reasonable range of ε also lies between 0 and 1. When it is 1, there is no reserve; the current received payoff is completely shared with others, while when it is 0, there is no sharing.

By joining these two psychological elements, the recency effect (3) and experimentation (4), the basic model (2) is then augmented to become (5):

$$q_{i,t} = \begin{cases} (1-\varphi) * q_{i,t-1} + (1-\varepsilon) * \Pi_{t-1} & \text{if } i = i^*_{t-1} \\ (1-\varphi) * q_{i,t-1} + \dfrac{\varepsilon}{(N_{i,t-1}-1)} * \Pi_{t-1} & \text{if } i \sim i^*_{t-1} \\ (1-\varphi) * q_{i,t-1} & \text{otherwise} \end{cases} \tag{5}$$

The local experimentation restricted to similar actions is, however, not very useful for our setting where there are only a total of three actions being considered. We, therefore, modify the above propensity-updating rule by replacing the local experimentation with global experimentation. That is, given the very small action space, each action is a "neighbor" of other actions.[7] This leads to a slight modification of equations (5) to (6):

$$(6) \qquad q_{i,t} = \begin{cases} (1-\varphi) * q_{i,t-1} + (1-\varepsilon) * \Pi_{t-1} & \text{if } i \text{ is chosen in period } t-1 \\ (1-\varphi) * q_{i,t-1} + \dfrac{\varepsilon}{(N-1)} * \Pi_{t-1} & \text{otherwise} \end{cases}$$

Attention effect

This change also makes the original motivation of the behavioral parameter ε no longer appropriate. When experimentation becomes global, it is not the kind of generalization normally used in the learning literature. We, therefore, propose a new interpretation of ε, namely *attention control*, which is also frequently discussed in modern psychology. $\varepsilon = 0$ means that the subject has a quite sharp focus on the cause of success, and hence only the activated action is updated, whereas when ε increases the subject's attention is distracted, and the cause of success is not in sharp focus. What is particularly interesting is that when $\varepsilon = 2/3$, there is no discrimination between chosen actions and non-chosen actions. In this case, the subject may not even consciously know what action he/she is taking, leading to the received payoffs. Therefore, there is no attention being directed to the choice behavior at all.[8]

Stochastic choice

One essence of the RL model is that *choice behavior is stochastic and the stochastic rule must be consistent with the law of effect*. This idea was not only shared by many earlier mathematical psychologists who studied learning behavior, but also generally holds in the latest interdisciplinary models of learning, such as evolutionary computation. The original probabilistic choice rule considered in the Roth-Erev model is linear (7):

$$(7) \qquad p_{i,t} = \frac{q_{i,t}}{\sum_i q_{i.t}}$$

where $p_{i,t}$ is the probability of choosing action i at time t. This linear rule is consistent with the law of effect and the power law of practice [Roth and Erev 1995]. However, to make $p_{i,t}$ have a probability meaning, the propensity of each action must be positive, which in turn requires that the payoff be positive. To guarantee a positive payoff, what Roth and Erev did was to take the following truncation of the raw payoff:

$$(8) \qquad \Pi_{i,t} = \pi_{i,t} - \pi_{min}$$

where π_{min} is smallest possible payoff. Nonetheless, this design does not fit our experimental asset market, since π_{min} can change with the market price, and is difficult to figure out and fix at the beginning of the experiment. Therefore, we do not continue to follow their setting in equations (7) and (8). Instead, we follow the device popularly used in the agent-based financial model by replacing the linear probabilistic rule with the Gibbs–Boltzmann distribution:

$$(9) \qquad p_{i,t} = \frac{\exp(\lambda * q_{i,t})}{\sum_i \exp(\lambda * q_{i.t})}$$

The parameter λ introduced above matches the parameter known as the *speed of learning* in the Roth-Erev model, while for the latter it is introduced through the initial propensity equally assigned to each action, $q_{i,1}=q_1$, $\forall i$. Therefore, the parameter λ, replacing q_1, becomes the third parameter of our RL model.[9]

While the Gibbs–Boltzmann distribution does not require a positive payoff, the scale of our payoff in the experimental asset market can vary over a very wide range. Therefore, it is desirable to take a logarithmic transformation of the raw payoff, as shown in equation (10). When the raw payoff is positive ($\pi(t)>0$), its log transformation is straightforward; when it is negative ($\pi(t)<0$), an absolute value of $\pi(t)$ is taken first before the log transformation, and the sign is changed from positive to negative after the transformation. However, a complication may arise due to our integer unit of account. Since $\pi(t)$ can be $\ldots,-1, 0, 1,\ldots$, and nothing in between, to make our monotone transformation work for this integer unit of account, an artificial discontinuity is introduced in equation (10) to keep the transformation of -1, 0 and 1 in its original order:

$$\Pi_t = \begin{cases} ln\ (\pi_t) + 1 & \text{if } \pi(t)>0 \\ 0 & \text{if } \pi(t) = 0 \\ -ln\ |\pi_t| - 1 & \text{if } \pi(t)<0 \end{cases} \tag{10}$$

The raw payoff is then defined by equation (11):

$$\pi_t = m_t + P_t\times v_t - (m_0 + P_t\times v_0) \tag{11}$$

where m_0 and m_t refer to the cash held by the subject at the beginning and the end of the trading period t. v_0 and v_t are the shares owned by the subject at the beginning and the end of that period, and P_t is the closing price. With this measure, if the subject was idle and did not engage in any trade, then $m_0=m_t$ and $v_0=v_t$, and his end-of-period profit, π_t, will simply be zero. If he did engage in transactions, π_t can be positive and negative.

A remark

After endowing these three parameters with psychological interpretations, one might question whether there are some redundancies among them, in particular, the last two referred to as attention control and speed of adjustment.[10] In fact, we have conducted the test for the necessity of including the third parameter, the speed of adjustment, given that we have the other two. The result, which will not be shown here, supports the inclusion of λ. In addition, we also run the correlation of the estimated parameter values (to be discussed more in the next section), $\hat{\varepsilon}_j$ and $\hat{\lambda}_j$, over 120 subjects, and the correlation coefficient is only about 0.062. Hence, these two behavioral parameters have little redundancy in relation to each other.

BEHAVIOR OF THE THREE PARAMETERS

The three parameters to be estimated are the recency effect (φ), the experimentation parameter (attention effect) (ε), and the intensity of choice (speed of learning) (λ). The parameter spaces for the three are $\varphi\in[0, 1]$, $\varepsilon\in[0, 1]$, and $\lambda\in[0, \infty]$. However, based on what we have discussed above, there are specific ranges to be expected for ε and λ. For example, we may be interested in a low value of ε (subjects with a sharp focus) and a high value of λ (subjects with a fast response). Therefore, some interesting values of these coefficients are $\varepsilon\approx 0$, $\varepsilon=2/3$ (normal case) or 1/2 (corner case), and $\mathscr{H}_0{:}\lambda\gg 0$.

Maximum likelihood estimation

To estimate the three parameters $(\varphi, \varepsilon, \lambda)$, the method of MLE will be applied. For each subject *j*, we can observe a sequence of his choices among the three options, say $i = 1, 2, 3$, denoting, respectively, increasing ILO_j, decreasing it and keeping it unchanged. Denote this sequence of actions by $\{a_{i,t}\}(i = 1, 2, 3)$, where $a_{i,t} = 1$, if action *i* is taken in period *t*, and $a_{i,t} = 0$ otherwise. Then the likelihood function for $(\varphi, \varepsilon, \lambda)$ can be written as follows.

$$\mathbf{L}(\varphi, \varepsilon, \lambda) = \prod_{t=1}^{T}\left(\prod_{i=1}^{3} p_{i,t}^{a_{i,t}}\right) = \prod_{t=1}^{T} p_{1,t}^{a_{1,t}} p_{2,t}^{a_{2,t}} p_{3,t}^{a_{3,t}}$$

$$= \prod_{t=1}^{T} p_{i_t^*,t}(\varphi, \varepsilon, \lambda), \quad a_{i,t} \in \{0, 1\} \tag{12}$$

where i_t^* is the action taken at period *t*. $p^*_{i_t,t}$ is derived from the Gibbs–Boltzman distribution with the RL model (equations (11), (10), (6), and (9)) and using the observations $\{\pi_{i_t^*,t}\}_{t=1}^{T}$. Following the convention of taking the logarithm of **L**, we have the log likelihood function:

$$\mathbf{l}(\varphi, \varepsilon, \lambda) = \sum_{t=1}^{T} ln\, p_{i_t^*,t}(\varphi, \varepsilon, \lambda) \tag{13}$$

and the MLE estimation of φ, ε, and λ as the solution of

$$\max_{\varphi,\varepsilon,\lambda}\ \mathbf{l}(\varphi, \varepsilon, \lambda) = \max_{\varphi,\varepsilon,\lambda} \sum_{t=1}^{T} ln\, p_{i_t^*,t}(\varphi, \varepsilon, \lambda) \tag{14}$$

However, equation (14) is not analytically solvable. We, therefore, take a numerical algorithm, known as *differential evolution*, to solve this optimization problem.[11]

Notice that there are 20 subjects in each experiment, and there are six experiments. Therefore, we have a total of 120 subjects, and each of their observations $\left\{\pi_{i_t^*,t}, a_{i_t^*,t}\right\}_j$ is used to derive the MLE estimates $\hat{\varphi}_j$, $\hat{\varepsilon}_j$, and $\hat{\lambda}_j$. The results are presented in Figure 2 as a histogram for each estimated parameter.

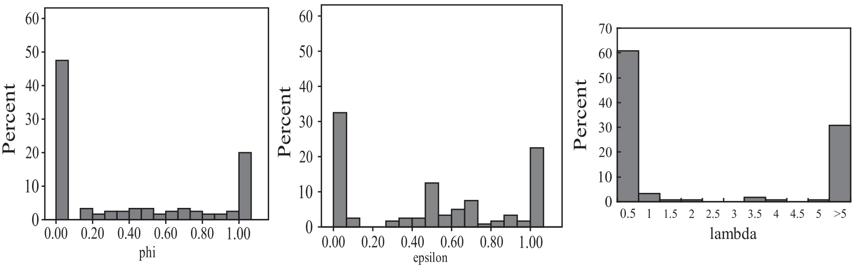


Figure 2. Distribution of the three estimated parameters ($\hat{\varphi}$, $\hat{\varepsilon}$, $\hat{\lambda}$) of all subjects.
Both the parameter φ and ε are restricted to the interval [0, 1]. Hence, $0 \leqslant \hat{\varphi}, \hat{\varepsilon} \leqslant 1$ are distributed as shown in the left and middle panels above. However, λ is only restricted to be non-negative; hence, basically, there is no upper bound for $\hat{\lambda}$, and in our case $\hat{\lambda}$ ranges from a minimum of 0.01 to a maximum of 42,736. Out of the 120 $\hat{\lambda}$s, 37 are greater than 5, 6 are greater than 100 and 3 are greater than 10,000. With this wide distribution, it is hard to plot the entire histogram with the same scale. We decide to plot it only up to a maximum of 5, and show the rest together as "bigger than 5." Therefore, the rightmost spike in the right panel above actually refers to the distribution of "$\lambda \geqslant 5$."

Testing the zero-intelligence hypothesis

Reinforcement learning is a way to model learning from experience. Therefore, before we can further look at these statistics, a very basic question is: *had these subjects ever learned?* We can address this question by using the famous device of the *zero-intelligence agent* [Gode and Sunder 1993] as a benchmark. According to Gode and Sunder [1993], the zero-intelligence agent basically behaves in a random way with no ability to learn from the best. Hence, if our subjects did not learn the relevance of ILOs, then they may randomly choose any of the three actions for the interior case or any of the two for the corner case; in other words, we may expect that

$$p_{i,t} = \begin{cases} \frac{1}{3}, & 0 < ILO_{j,t-1} < 1 \text{ (the interior case)} \\ \frac{1}{2}, & ILO_{j,t-1} = 0 \text{ or } 1 \text{ (the corner case)} \end{cases} \tag{15}$$

If this is the case, the likelihood function is simply $(1/3)T_1(1/2)T_2$, where T_1 represents the frequencies of the interior case, T_2 represents the frequencies of the corner case, and $T_1 + T_2 = T$. Alternatively put, we have

$$\max_{\varphi,\varepsilon,\lambda} \mathbf{L}(\varphi,\varepsilon,\lambda) = L(\hat{\varphi},\hat{\varepsilon},\hat{\lambda}) = \left(\frac{1}{3}\right)^{T_1}\left(\frac{1}{2}\right)^{T_2}, \quad T_1 + T_2 = T \tag{16}$$

where $\hat{\varphi}_j$, $\hat{\varepsilon}_j$, and $\hat{\lambda}_j$ are the maximum likelihood estimates. This is equivalent to testing the following null:

$$\begin{aligned} &H_0 : L(\hat{\varphi},\hat{\varepsilon},\hat{\lambda}) - \left(\frac{1}{3}\right)^{T_1}\left(\frac{1}{2}\right)^{T_2} \leqslant 0 \\ &H_1 : L(\hat{\varphi},\hat{\varepsilon},\hat{\lambda}) - \left(\frac{1}{3}\right)^{T_1}\left(\frac{1}{2}\right)^{T_2} > 0 \end{aligned} \tag{17}$$

or

$$\begin{aligned} &H_0 : \frac{L(\hat{\varphi},\hat{\varepsilon},\hat{\lambda}) - \left(\frac{1}{3}\right)^{T_1}\left(\frac{1}{2}\right)^{T_2}}{\left(\frac{1}{3}\right)^{T_1}\left(\frac{1}{2}\right)^{T_2}} \leqslant 0 \\ &H_1 : \frac{L(\hat{\varphi},\hat{\varepsilon},\hat{\lambda}) - \left(\frac{1}{3}\right)^{T_1}\left(\frac{1}{2}\right)^{T_2}}{\left(\frac{1}{3}\right)^{T_1}\left(\frac{1}{2}\right)^{T_2}} > 0 \end{aligned} \tag{18}$$

Here, we compare the likelihood from the independent random choice model with the likelihood from the three-parameter RL model. We apply both the parametric *t* test and, due to the usual concern, also the non-parametric Wilcoxcon sign and signed-rank test. The test results, as shown in Table 3, reject the nulls of both (17)

Table 3 Test statistics of null (17) and (18)

Test	*P value of null (17)*	*P value of null (18)*
t test	0.0000	0.0000
Sign	0.0000	0.0000
Signed Rank	0.0000	0.0000

Note: The tests are conducted based on the 120 likelihood values derived from the 120 subjects. The null is that the likelihood of the zero-intelligence model is no less than the likelihood using the three-parameter Roth–Erev model. All the *P* values of the tests are less than 0.0001, and they are taken as 0.0000 after rounding.

and (18) significantly, which implies that the subjects' choice behaviors are better described by the RL model than by the zero-intelligence (purely random) model.

Heterogeneity in agents

Even though the zero-intelligence model is rejected as a whole, it does not mean that all agents had learned or learned in the same way. To see that, Figure 2 gives the distribution (histogram) of the $\hat{\varphi}$, $\hat{\varepsilon}$, and $\hat{\lambda}$ over 120 agents. A few noticeable features stand out. First, the estimates are widely distributed over the parameter space, which shows a great heterogeneity of agents. Second, the dispersed distribution also indicates the existence of extreme behavior among agents. For example, agents with a strong recency attribute ($\varphi \approx 0$) coexist with agents with a strong forgetting attribute ($\varphi \approx 1$); agents with strong attention control ($\varepsilon \approx 0$) coexist with agents with very weak attention control ($\varepsilon \geqslant 2/3$). Similarly, agents with a strong intensity of choice ($\lambda \gg 0$) are accompanied by agents with no intensity of choice ($\lambda \approx 0$). The empirical distributions of $\hat{\varphi}$ and $\hat{\varepsilon}$ are more or less U-shaped and hence bimodal with the two local maxima appearing in the corner. On the other hand, the empirical distribution of the $\hat{\lambda}$ is L-shaped.

IS OBSERVED HETEROGENEITY ENDOGENOUS?

The picture of 120 human subjects with such great heterogeneity in their reinforcement learning behavior motivates us to ask ourselves what is the cause of that. Is the heterogeneous behavior exogenous (innate) or endogenous? For the latter, we specifically ask whether different experimental settings may induce different observed learning behavior, and hence the observed agents' learning behavior may be affected by the experimental designs.

Behavioral parameters and market microstructure

We first look at the learning behavior under different market experiments. Table 4, the top panel, gives the medians of $\hat{\varphi}$, $\hat{\varepsilon}$, and $\hat{\lambda}$ taken over 20 agents in each of the six market experiments (from A to F). To determine whether these parameters are different under different market settings, we run the Wilcoxon rank-sum test for each pair of settings, and the null is that the corresponding two samples of the estimate are from the same distribution. The results, in terms of P values, are displayed in Table 5. The test results consistently indicate that the distributions of all three estimated behavioral parameters are the same under all pairs of market microstructures. Therefore, the observed heterogeneity in the learning parameters cannot be explained by the different market microstructures, and hence is not endogenously caused by them.

Behavioral parameters and traders' information status

Even though the behavioral parameters are not found to be statistically different under different markets, we notice that, within each market, there are generally three different types of traders, namely perfect insiders, imperfect insiders, and outsiders. It would then be interesting to know whether the initial information acquisition can impact the resultant learning behavior. At least, one would not exclude the

Table 4 Behavioral parameters, medians of each market setting, and trader setting

	φ	ε	λ
Market type			
A	0.174	0.420	0.229
B	0.013	0.497	1.810
C	0.198	0.401	0.433
D	0.316	0.667	0.333
E	0.102	0.614	0.295
F	0.328	0.476	0.217
Trader type			
Perfect insiders	0.417	0.602	0.311
Imperfect insiders	0.031	0.463	0.415
Outsiders	0.252	0.494	0.286

Note: Each number shown in the table refers to the median of the estimates of the corresponding parameter under the indicated market type or trader type. For the market type, each median is from a sample with 20 subjects. For the agent type, we have 34 perfect insiders, 34 imperfect insiders, and 52 outsiders.

Table 5 Difference in learning behavior with respect to market settings and trader settings, Wilcoxon rank sum test

Pair	φ	ε	λ	*Pair*	φ	ε	λ
A and B	0.7370	0.5456	0.2064	B and F	0.3268	0.9678	0.1279
A and C	0.3700	0.9249	0.6578	C and D	0.5632	0.2039	0.9037
A and D	0.9571	0.2312	0.3704	C and E	0.6756	0.3022	0.7170
A and E	0.6007	0.2942	0.7170	C and F	0.6940	0.5452	0.4778
A and F	0.2565	0.5620	0.9893	D and E	0.8190	0.7741	0.6773
B and C	0.6663	0.6194	0.2573	D and F	0.3685	0.6053	0.2064
B and D	1.0000	0.3140	0.3847	E and F	0.3685	0.6741	0.3994
B and E	0.8822	0.3277	0.3564				
I-P and I-NP	0.2106	0.1695	0.7278				
I-P and O	0.3757	0.1956	0.8845				
I-NP and O	0.4109	0.8010	0.8567				

Note: The numbers shown above are the *P* values of the Wilcoxon rank sum test. The null hypothesis is that the observed (estimated) behavioral trait (parameter) has an identical distribution in the respective pair of market settings (top panel) or trader settings (bottom panel).
I-P, I-NP, and O appearing in the bottom panel refer to perfect insiders, imperfect insiders, and outsiders, respectively.

possibility that the perfect insiders who know exactly what the expiration price is may have a different preference over the two trading implementations (the market order and limit order) than those who are quite uncertain about it. The bottom panel of Table 4 gives the median of the estimated parameters with respect to the three different types of traders.

As before, we run the Wilcoxon rank-sum test for each pair of the three types of traders, and the results are shown in Table 5 (the bottom panel). The results generally do not connect information heterogeneity to heterogeneity in behavioral parameters. Therefore, the observed heterogeneity in the learning behavior is not caused by the information asymmetry. By combining this result with the previous one, we conclude that neither of our two experimental settings can cause the observed heterogeneous learning behavior, which in turn rejects the endogeneity hypothesis. Therefore, subjects' differences in learning behavior, as captured by the three parameters in the fitted RL model, are likely to be exogenous (innate). If they are exogenous, which can be analogous to personal psychological traits, then we may further ask: *do they matter*? This leads us to the next section.

BEHAVIORAL TRAITS AND EARNING CAPACITY

If it is subjects' personal traits that lead to the observed heterogeneity, then it is important to know whether these personal traits may have effects on their earning performance.[12] We classify the agents into four groups based on their ranks of the accumulated profits: the top five as the first cluster, the sixth to the tenth as the second, and so on. Given that we have six experiments, each cluster has a total of 30 (5 × 6) subjects. The medians of $\hat{\varphi}$, $\hat{\varepsilon}$, and $\hat{\lambda}$ over these 30 subjects for each group are provided in Table 6.

From Table 6, we first notice that both attention control $(1-\varepsilon)$ and intensity of choice (λ) are positively correlated with the earning performance. They both monotonically decrease from Cluster I to Cluster IV. In particular, the top two clusters of subjects both have $(1-\varepsilon)$ strictly greater than one half and are biased toward one, which is evidence that the propensity of the chosen action is reinforced by its positive payoffs. On the other hand, the bottom two groups of subjects both have $(1-\varepsilon)$ less than one half. For Cluster III (rank 11th to 15th), this value is about one-third (0.336), a value showing that all three actions are equally treated (see equation (6)); in this case, there is essentially no learning. Cluster IV even has $(1-\varepsilon)$ equal to zero; credits are totally assigned to non-chosen actions. These two clusters of subjects may not have ILO as the target to learn, and may not even have a clue as to what to learn. Therefore, in terms of this specific model, their attention may be diluted over many things simultaneously happening in the experiment.

The parameter λ, a measure of speed of adjustment or responsiveness, is also interesting. The magnitude of this parameter has been shown to be able to account for some stylized facts observed in financial markets [Hommes 2006]. A large λ can cause significant swings among the fractions of different types of financial agents [Chen et al. 2010]. In this study, our empirical results seem to indicate that subjects with a greater degree of responsiveness tend to perform better than they do otherwise.

Of course, the above numerical properties need to be further examined with statistical significance. Here, we apply the Wilcoxon rank-sum test to examine

Table 6 Behavioral parameters, medians of each cluster of earnings

Cluster	*Ranks*	φ	$\varepsilon(1-\varepsilon)$	λ
I	Top 5	0.006	0.021 (0.979)	2.685
II	6th–10th	0.250	0.299 (0.701)	0.969
III	11th–15th	0.415	0.664 (0.336)	0.322
IV	Bottom 5	0.227	1.000 (0.000)	0.107

Note: Each cluster (first column) is composed of subjects with the respective ranks (2nd column) over all six experiments. Ranks are determined based on the accumulated profits. Since it is $(1-\varepsilon)$ that gives the degree of attention control, to make it easier to read, we present the value of $(1-\varepsilon)$ inside brackets immediately next to the value of ε.

Table 7 Differences in earning performance with respect to behavioral traits: Wilcoxon rank sum test

Pair of clusters	φ	ε	λ
I and II	0.2132	0.5418	0.9824
I and III	0.2003	0.0047	0.4152
I and IV	0.6909	0.0000	0.0005
II and III	0.8077	0.0118	0.1860
II and IV	0.6476	0.0000	0.0002
III and IV	0.4889	0.0059	0.0332

Note: The numbers shown above are the *P* values of the Wilcoxon rank sum test. The null hypothesis is that the observed (estimated) behavioral trait (parameter) has an identical distribution in the respective pair of clusters.

whether these behavioral traits (parameters) associated with different clusters of agents are in fact from the same distribution. As before, we do this pair by pair, and the results are shown in Table 7. Table 7 shows that there are no significant differences in φ among different clusters of agents. Hence, the recency parameter may not contribute to the observed income heterogeneity. On the other hand, the Wilcoxon rank-sum test does suggest that the attention parameter (ε) differs significantly in all pairs of clusters except the top two (I & II). With regard to the difference in the speed of adjustment (λ), it is not that significant compared to ε. That difference only exists between the bottom group (Cluster IV) and others. Hence, among the three behavioral traits, attention control seems to be the only variable that is able to distinguish subjects' earning capacities. The degree of responsiveness may also play such a role, but only a marginal one.[13] The recency parameter plays no such role.

CONCLUDING REMARKS

In the process of conducting an experiment, what the agent has tried to learn and how he has learned is sometimes not an easy issue to answer. It can become even harder in an open environment like the prediction market, where both *the target to learn* and *the method to learn* can change as time goes on. However, as long as we have reason to believe that learning did happen during the experiment, then understanding learning via *learning models* remains a worthy topic of research. This paper provides the first application of RL models to the choice of market order or limit order, the two basic trading operations.

In this paper, we assume the use of the limit order as a target to learn. By further focusing on the change in its intensity, a simple three-parameter RL model is applied and estimated to fit the data. Through this simple RL model, we can easily identify some evidence of learning. In particular, when compared to the benchmark of the zero-intelligence model, the performance of the RL model fits the data significantly better, which implies that agents did react to market feedbacks. Hence, while there are many alternative learning models to work with, a simple model like the RL model is good enough to make us "observe" learning and gain some insights into it.

The hypothesis that agents are heterogeneous (the heterogeneous agent paradigm) has recently played a dominant role in economic research. The simple RL model employed in this model can effectively communicate with this hypothesis. Our finding provides evidence of great heterogeneity among agents. Another nice virtue of the RL model is that it is a psychologically based model, and hence all of its parameters can be interpreted psychologically. In our case, the three parameters can be refereed to as recency, attention, and responsiveness. Given their psychological nature, the agents' heterogeneity can reflect their differences in personal traits, which are totally exogenous to our experimental settings. Our tests do lend support to the exogeneity hypothesis. Furthermore, our analysis also shows that two of these three personal traits actually matter for agents' earning performance, in particular, attention control.

There have been few studies on the learning behavior of individual subjects in the experimental asset market, in particular the order book-driven market. As we can see from this paper, the order book-driven market is more complex, and it is even not clear as to what subjects may learn in this environment. By assuming the decomposable structure of learning, we are able to reduce this otherwise multi-dimensional learning problem to a single-dimensional one. Needless to say, more work is needed to delineate the decomposable conditions. In addition, if the decomposable conditions are not satisfied, how the reduced model may deviate from a full-fledged model, and what a full-fledged model looks like need to be considered. These issues will be left for the direction of further studies.

Acknowledgements

An early version of this paper was presented at the Eastern Economic Association Meetings, New York City, February 27 – March 1, 2009, and the IAREP-SABE 2009 Conference, Halifax, July 7–11, 2009. The authors benefited significantly from the discussions with conference participants, in particular Monica Capra. Thanks also go to the organizers, Jason Barr and Leanne Ussher. This version has been substantially revised in light of three anonymous referees' very painstaking reviews, for which we are most grateful. The NSC grant 98-2410-H-004-045-MY3 is also gratefully acknowledged.

APPENDIX

Details of the experimental designs

In this section, we will detail the design of our experimental future market. We start with the institutional arrangements (the first subsection). We then describe the generation of the artificial spot prices (the second subsection) and finally the interaction platform between the institution and the subjects (the last subsection).

Institutional arrangements

Our experimental future market is *order book-driven*. Each subject is initially provided with NT$10,000 in cash and 20 future contracts. The resulting assets of each round will not be carried over to the next round. For each round, the assets are always renewed to 10,000 in cash and 20 shares. The contract has a value at the expiration date, but the expiration price is available only for informed traders. Normally, subjects can buy the contracts if they have reason to believe that the market price is below the expiration price, and sell them when they believe the opposite applies. Of course, another common reason to trade is to gain speculative profits, buying low and selling high. Subjects are, however, not allowed to sell short, neither to buy on margin. Furthermore, at each point in time, each subject can, at most, have one order in the order book. If he submits another one, then the previous one will automatically be considered to be obsolete. In this way, they can also cancel out their earlier submissions.

In parallel to the future market, there is a spot market. This market , however, is artificial. The generated spot prices, which are detailed in the next subsection, are available to all subjects one by one after every second. This information may or may not be helpful for predicting the future price. In the future market, the price is determined by *continuously matching* buy orders and sell orders. Therefore, the basic public information for all subjects consists of these two price series, the spot price and the future price. Order book information is also made public, but only up to the orders associated with the existing highest bid and the lowest ask. Others'/Other submissions are not available.

Subjects are then further distinguished into perfect insiders, imperfect insiders, and outsiders. Basically, the subjects are randomly allocated into roles of informed and uninformed (and/or partially informed) subjects. The roles were first randomly assigned to a sequence of numbers from 1 to 20, constrained by the given role composition of the respective market (numbers of perfect insiders, imperfect insiders, and outsiders). Then, before the experiment, each subject had to pick one of these numbers without replacement, which automatically matched a role to them. Once the role was assigned, the subject remained in that role for the entire experiment.

All subjects were allowed to participate in only one of the six experiments, and they were provided NT$300 for their 3-hour participation. The participation fee was 30 percent higher than the regular campus work available for undergraduate students. The additional rewards were performance-based. At the end of the experiment, we ranked the subjects by their accumulated profits, that is, the sum of profits earned in each round (see equation (11)). The top three classes of subjects, with sizes of 3, 3, and 4, were then given financial rewards of NT$200, NT$150, and NT$100, respectively. In other words, 10 out of the 20 subjects were eligible for these additional rewards.

Generation of spot prices

The spot prices in our experimental asset market are artificial. However, to make what one may experience from our experimental markets not totally independent of what one may experience from the real-world markets, the artificial data were generated by sampling the daily closing prices of 36 listed companies from Dow Jones, Nasdaq, and Taiex. Specifically, for each listed company we select a subseries consisting of 471 consecutive daily closing prices. We denote these original *real* series as $\{P_t^r\}_{t=1}^{471}$. These series will then be transformed into an *artificial* series $\{P_t^a\}_{t=1}^{471}$ based on equation (A.1) as given below.

$$\text{(A.1)} \qquad P_t^a = \begin{cases} 100 & \text{if } \frac{P_t^r}{P_{51}^r} \times 50 \geqslant 100 \\ \frac{P_t^r}{P_{51}^r} \times 50 & \text{if } 1 < \frac{P_t^r}{P_{51}^r} \times 50 < 100, \\ 1 & \text{if } \frac{P_t^r}{P_{51}^r} \times 50 \leqslant 1 \end{cases} \quad t = 1, \cdots, 471$$

Clearly, the intention is to rescale these series so that they are all bounded below by one and bounded above by one hundred. P_{51}^r is chosen as a reference point because the first 50 of this subseries will be revealed to the subjects at the beginning of each round as the historical information. The remaining 421 observations will then be considered as the "spot (minute) price" of the market, and released to the market on a per minute basis, from minute 0 to minute 420 (a total of 7 minutes for each round). According to equation (A.1), the spot price shall always then start with 50 ($P^a_{51} = 50$).

The 36 original series are chosen in such a way that half of them have the expiration price higher than the initial price $P^r_{471} > P^r_{51}$, called the *up series*, and half of them have the opposite inequality, $P^r_{471} < P^r_{51}$, called the *down series*. In the middle, the price may fluctuate with no further constraint. For each experiment, 14 series were randomly sampled from these 36 series. Half of them (seven) were sampled from the 18 up series, and half of them (seven) were sampled from the 18 down series. Each of these 14 series, after rescaling (equation (A.1)), became the spot price for one round of the experiment. Together there were seven up series and seven down series. To make sure subjects could not infer any pattern, for example a one up series followed by a one down series, these 14 series were presented to subjects round by round, in a shuffling order.

The spot price was updated on a second basis for all subjects. However, in addition to the perfect insiders exactly knowing the expiration price P^a_{471}, the imperfect insiders knew a range $[P^{min}_{471}, P^{max}_{471}]$ covering P^a_{471}. The determination of the two endpoints of these ranges is as follows:

$$P^{min}_{471} = max\left(min\,\{P^a_t\}^{471}_{t=1}, P^a_{471} - \frac{1}{2}\hat{\sigma}(\{P^a_t\}^{471}_{t=1})\right) \tag{A.2}$$

and

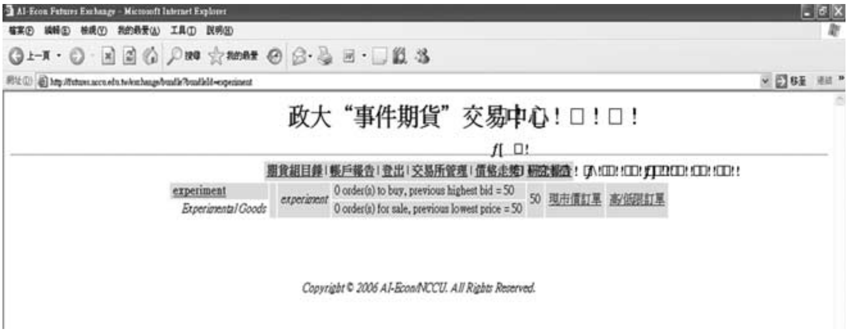
$$P^{max}_{471} = min\left(max\,\{P^a_t\}^{471}_{t=1}, P^a_{471} + \frac{1}{2}\hat{\sigma}(\{P^a_t\}^{471}_{t=1})\right) \tag{A.3}$$

The idea is to add and subtract one half of the standard deviation of the spot prices ($\{P^a_t\}^{471}_{t=1}$) to and from the expiration price to form the ceiling and the floor. However, to make this range informative, the ceiling and the floor are further constrained by the minimum price or the maximum price of the series. All of our subjects knew that the original date was from the real markets, but did not know from which company, which period, or how the transformation from the original data to the artificial data was made. The imperfect insiders also did not know how the range was determined.

Interaction platform

The information provided to the subjects and the actions available for them to take can all be seen from various interfaces in the transaction platform. The main page presented to the subject is given in Figure A1. The key information provided in this screen is only a surface of the order book information, that is, the highest bid and lowest ask available, as well as the associated desired size of transaction. Other bids and asks are not available. In addition, this page also has the latest transaction price. The subject can know more about the price history by clicking the bar "price movement" in the page, then he will be connected to a screen as shown in Figure A2.

What is shown in Figure A2 is the time series of the future price (the left panel), the spot price (the middle panel), and the cumulative trading volume (the right



AI-Econ Futures Exchange - Microsoft Internet Explorer
檔案(F) 編輯(E) 檢視(V) 我的最愛(A) 工具(T) 說明(H)
上一頁 搜尋 我的最愛
http://futures.nccu.edu.tw/exchange/bundle?bundleId=experiment
移至 連結

政大“事件期貨”交易中心！□！□！

ƒ[□!

期貨組目錄 | 帳戶報告 | 登出 | 交易所管理 | 價格走勢 | 研究報告 ! ①\!□□!!□□!ƒ[②□□!□□!□□!!

experiment *Experimental Goods*	*experiment*	0 order(s) to buy, previous highest bid = 50 0 order(s) for sale, previous lowest price = 50	50	現市價訂單	高/低限訂單

Copyright © 2006 AI-Econ/NCCU. All Rights Reserved.

Figure A1. Main page. English annotations as indicated by circled numbers are as follows: ① “Price Movement,” ② “Market Order,” and ③ “Limit Order.”

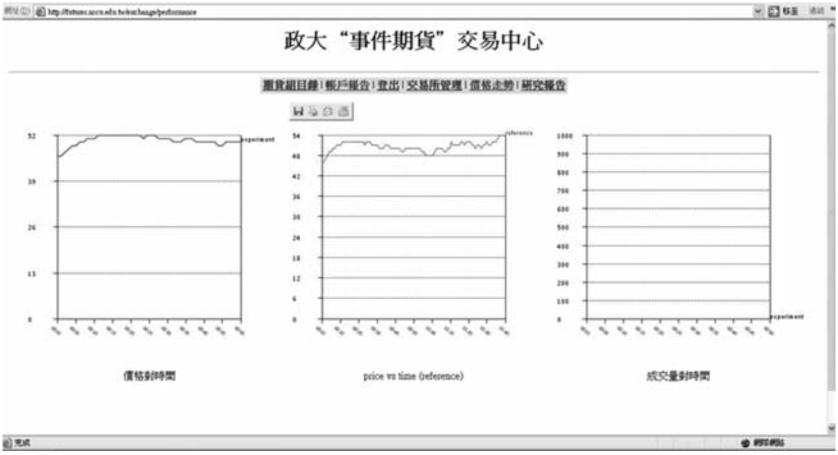
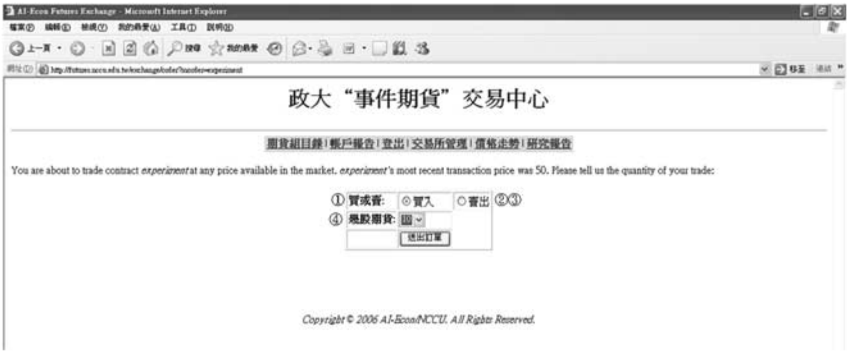


Figure A2. Future price, spot price, and accumulated trading volume. The three panels above give the time series of the future price (left), the spot price (middle), and the cumulative trading volume (right).

panel). The future price is the price generated by the subjects in our experimental asset market, and the spot price is the artificial price series as described in the previous subsection. If the subject wants to make a transaction, he can choose one of the two actions available in the main page, namely the market order and limit order. If he chooses the market order, he will be connected to the market order page (Figure A3); if he chooses the limit order, he will be connected to the limit order page (Figure A4).

In the market order page, the subject has to decide whether he would like to buy or sell, and how many units. These decisions have to be entered into the conversation box before he can submit the market order. In the limit order page, in addition to the units to buy or sell, the subject also has to specify the price limit, that is, the maximum price to buy and the minimum price to sell. Outside the conversation box, the background is the order book information, which is also available in the main page.[14]

After finishing the order form and submitting the order, the subject will be asked to confirm his decision. The subject can do this by clicking a confirmation key in the

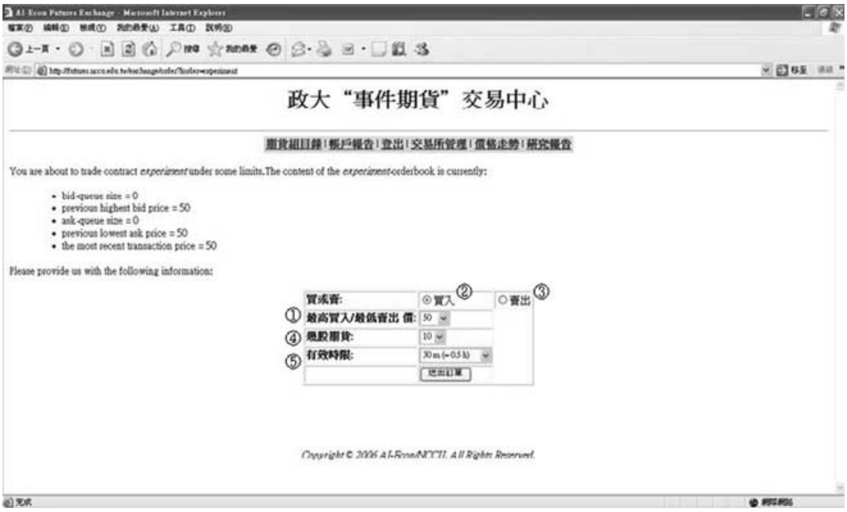

AI-Econ Futures Exchange - Microsoft Internet Explorer
政大“事件期貨”交易中心
期貨組目錄 | 帳戶報告 | 登出 | 交易所管理 | 價格走勢 | 研究報告

You are about to trade contract *experiment* at any price available in the market. *experiment*'s most recent transaction price was 50. Please tell us the quantity of your trade:

① 買或賣: ⊙買入 ○賣出 ②③
④ 幾股期貨: 10
送出訂單

Copyright © 2006 AI-Econ/NCCU. All Rights Reserved.

Figure A3. Market order page. English annotations as indicated by circled numbers are as follows: ① "Buy or Sell," ② "Buy Option," ③ "Sell Option," and ④ "Desired Units of Transaction."


AI-Econ Futures Exchange - Microsoft Internet Explorer
政大“事件期貨”交易中心
期貨組目錄 | 帳戶報告 | 登出 | 交易所管理 | 價格走勢 | 研究報告

You are about to trade contract *experiment* under some limits.The content of the *experiment*-orderbook is currently:

- bid-queue size = 0
- previous highest bid price = 50
- ask-queue size = 0
- previous lowest ask price = 50
- the most recent transaction price = 50

Please provide us with the following information:

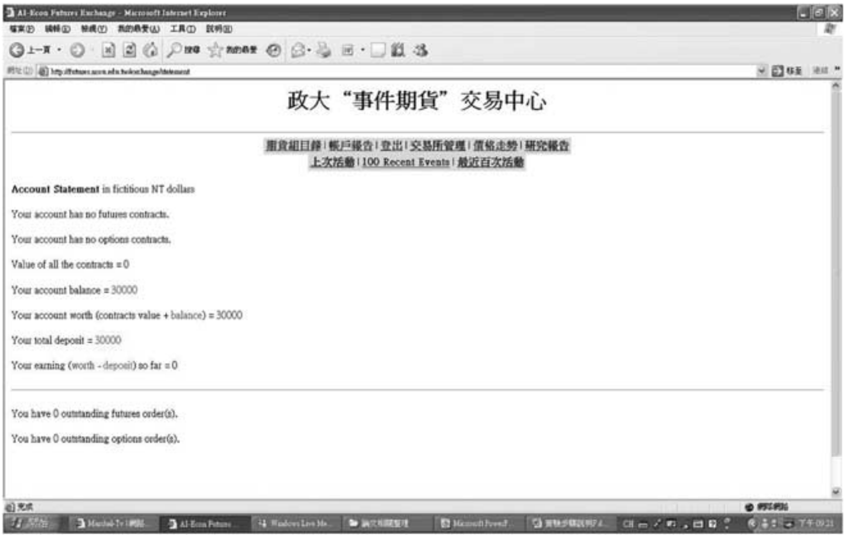
買或賣: ⊙買入 ② ○賣出 ③
① 最高買入/最低賣出 價: 50
④ 幾股期貨: 10
⑤ 有效時限: 30 m (= 0.5 h)
送出訂單

Copyright © 2006 AI-Econ/NCCU. All Rights Reserved.

Figure A4. Limit order page. English annotations as indicated by circled numbers are as follows: ① "Buy or Sell," ② "Buy Option," ③ "Sell Option," ④ "Limit Price," ⑤ "Desired Units of Transaction," and ⑥ "Effective Duration for this Offer."

screen or alter this decision by clicking the cancelation key. Once after clicking the confirmation, the transaction is finished and the subject will also receive a notification to indicate this.

In each round, a subject can also check his account at any point in time (Figure A5). After clicking the "Account Report," he is able to check his current wealth, that is, cash plus the value of his owned contracts. The latter is determined by the current market price, which of course may change over time. Profit or loss, which is simply the subtraction of the initial wealth (deposit) from the current wealth, is then also available and constantly updated. Finally, he can check his unfinished order in the market using this page as well. The "100 recent events" option shown in this page enables each subject to see his latest 100 transaction records in each round.



政大“事件期貨”交易中心

期貨組目錄 | 帳戶報告 | 登出 | 交易所管理 | 價格走勢 | 研究報告
上次活動 | 100 Recent Events | 最近百次活動

Account Statement in fictitious NT dollars

Your account has no futures contracts.

Your account has no options contracts.

Value of all the contracts = 0

Your account balance = 30000

Your account worth (contracts value + balance) = 30000

Your total deposit = 30000

Your earning (worth - deposit) so far = 0

You have 0 outstanding futures order(s).

You have 0 outstanding options order(s).

Figure A5. Account report.

Table A1 Profit distribution between insiders and outsiders: Wilcoxon rank sum test

Rounds	*B*	*C*	*D*	*E*	*F*
1	0.856	0.176	0.606	0.835	0.219
2	0.764	0.076	0.917	0.628	0.439
3	0.315	0.293	0.917	0.374	0.764
4	0.856	0.229	0.474	0.628	0.674
5	0.373	1.000	0.199	0.729	0.024*
6	0.120	1.000	0.474	1.000	0.180
7	0.439	1.000	0.606	0.491	0.219
8	0.373	0.176	0.917	0.450	0.764
9	0.062	0.460	0.606	0.677	0.019*
10	0.589	0.564	0.917	0.535	0.133
11	1.000	0.176	0.917	0.729	0.218
12	0.078	0.993	0.917	0.944	0.590
13	0.952		0.917	0.202	0.219
14			1.000	0.835	

The numbers shown above are the *P* values of the Wilcoxon rank sum test. The null hypothesis is that the profit earned by insiders and outsiders has an identical distribution in the respective round of the respective market. The symbol * refers to the statistical significance at a significance level of 5 percent.

ADDITIONAL RESULTS

Some additional results mentioned in the main text are given in this section. Table A1 is a summary of the results of the Wilcoxon rank sum test of the null that the profit distributions of insiders and outsiders are the same. The numbers shown in the table are the *P*-values of the null. Table A2 is the summary of the Wilcoxon rank sum test of the null that the distribution of ILOs is the same between insiders and outsiders. Table A3 provides the Spearman rank correlation between the rank of the personal traits (estimated parameter values) and the rank of the accumulated profits in each market.

Table A2 ILO distribution between insiders and outsiders: Wilcoxon rank sum test

Rounds	*B*	*C*	*D*	*E*	*F*
1	0.768	0.570	0.240	0.171	0.099
2	0.859	0.570	0.304	0.476	0.165
3	0.859	0.683	0.106	0.914	0.679
4	0.513	0.214	0.142	1.000	0.075
5	0.513	0.461	0.054	0.257	0.371
6	0.953	0.933	0.142	0.257	0.206
7	0.859	0.461	0.054	0.914	0.075
8	0.768	0.683	0.188	0.914	0.768
9	0.859	0.683	0.054	0.762	0.165
10	1.000	0.808	0.054	0.762	0.594
11	0.513	0.570	0.036*	0.762	0.859
12	0.768	0.933	0.106	0.914	0.679
13	0.953		0.106	0.610	0.679
14			0.240	0.610	

The numbers shown above are the P values of the Wilcoxon rank sum test. The null hypothesis is that the intensity of limit orders by which insiders and outsiders proceed to trade has an identical distribution in the respective round of the respective market. The symbol * refers to the statistical significance at a significance level of 5 percent.

Table A3 Spearman rank correlation between earnings and attention control (ε) and earning and responsiveness (λ)

Market	ε	λ
A	−0.490*	0.378*
B	0.364	−0.022
C	−0.356	0.278
D	−0.583*	0.395*
E	−0.828*	0.413*
F	−0.198	−0.102

The symbol * refers to the statistical significance at a significance level of 5 percent.

Notes

1. Cheung and Friedman [1997] is one of the few exceptions. They applied a three-parameter belief-based learning model to cover two extreme forms of behavior, namely Cournot and fictitious play, which signify different memory lengths of human subjects. The model is then applied to fit the behavior of 393 subjects *individually* in a number of normal-form games. They found that in terms of the estimated parameters subjects are quite heterogeneous.
2. Limit order and market order are the two most common types of order in the securities market. An order is called a limit order if there is a limit to accept the price to buy or sell a specific quantity of a security. The limit is also the limit price. Hence, any price above the limit price to buy will result in the non-execution of the action to buy, and any price below the limit price to sell will result in the non-execution of the action to sell. This ensures that a person will never pay more for or sell less of the security than the price that is set as his limit. In contrast to the limit order, an order is referred to as a market order if a buy order or sell order is executed based on the best currently available price.
3. We also investigated the correlation between $ILO_{j,t}$ and $\pi_{j,t}$ round by round. The result is very much the same as what is presented in Table 2. The correlation is significant in almost all rounds (75 out of a total of 80 rounds) at a significance level of 5 percent.
4. Markets B, D, and F have very similar features, and thus they will not be drawn here.
5. The first result is indeed interesting and a little surprising as well, which indicates that insiders might actively engage in some speculative trading and can suffer losses from that.

6. For example, this family of learning models plays little role in the macroeconomic literature [Evans and Honkapohja 2001].
7. The three actions considered in this paper are to increase the *ILO*, decrease it, or keep it the same (see the assumption that we make in the third section). Therefore, these three are by no means similar.
8. While we have a total of three actions, only two actions are effective when ILO_j goes to its two extremes. If $ILO_{j,t}=0$, then, in period $t+1$, the action "decrease" becomes unavailable, whereas at the other extreme ($ILO_{j,t}=1$), the action "increase" becomes unavailable. Therefore, in these corners, it is $\varepsilon=1/2$, rather than $\varepsilon=2/3$, that refers to a non-discriminatory updating.
9. There is a name for the parameter λ in the agent-based financial models, namely, *intensity of choice*, which measures how fast the subject will move from one action to another. Therefore, it may be alternatively recognized as the *speed of adjustment* or the *degree of responsiveness*. We shall use these terms interchangeably in this paper.
10. I am grateful to Monica Capra for drawing our attention to this question.
11. Differential evolution can be regarded as a variation of genetic algorithms, which is also a population-based global stochastic search algorithm. In econometrics, the application of genetic algorithms to optimizing nonlinear and rugged objective functions is first reviewed by Dorsey and Mayer [1995]. A list of contributions using genetic algorithms to parameter estimation can also be found in Chen and Kuo [2002]. Differential evolution is a hybrid approach incorporating simulated annealing (SA) in the selection process of GAs that can overcome the drawbacks of both approaches, that is, slowness in SA and perturbation disruption in GAs. For the details, see Price et al. [2005]. In this paper, differential evolution is implemented using *Mathematica*, Version 6.0.
12. This question is very much motivated by the recent studies on the economic significance of psychological traits. For a general survey, the interested reader is referred to Borghans et al. [2008].
13. We also carry out a similar analysis for each market to observe the effects of these two parameters over different markets. In this case, we rank the subjects of each market based on their personality traits (parameter values) and earning performance. The Spearman rank correlations between the two parameters and earning performance are obtained as in Table A3 (see Appendix). The significant effect of attention control and responsiveness are found in three out of the six markets, namely Markets A, D, and E. This may partially explain why the aggregate effect as shown in Table 7 is not that strong.
14. Notice that the order book information was only available up to this "surface." Information with regard to other orders in the order book was not made available to the subjects.

References

Ang, James S., and Thomas Schwarz. 1985. Risk Aversion and Information Structure: An Experimental Study of Price Variability in Securities Markets. *Journal of Finance*, 40(3): 825–844.

Arthur, William B. 1991. Designing Economic Agents That Act Like Human Agents: A Behavioral Approach to Bounded Rationality. *American Economic Review Papers and Proceedings*, 81(2): 353–359.

——— 1993. On Designing Economic Agents That Behave Like Human Agents. *Journal of Evolutionary Economics*, 3(1): 1–22.

Borghans, Lex, Angela L. Duckworth, James J. Heckman, and Bas ter Weel. 2008. The Economics and Psychology of Personality Traits. *Journal of Human Resources*, 43(4): 972–1059.

Bossaerts, Peter, Ulrik Beierholm, Cedric Anen, Helene Tzieropoulos, Steven Quartz, Rolando G. de Peralta, and Sara Gonzalez. 2008. Neurobiological Foundations for "Dual System" Theory in Decision Making under Uncertainty: fMRI and EEG evidence, Neuroeconomics 2008, September 25–28, Park City, Utah.

Bush, Robert R., and Frederick Mosteller. 1955. *Stochastic Models for Learning*. New York: Wiley.

Camerer, Colin, and Teck-Hua Ho. 1999. Experience-Weighted Attraction Learning in Normal-Form Games. *Econometrica*, 67: 827–874.

Chen, Shu-Heng, Chia-Ling Chang, and Yeh-Rong Du. 2010. Agent-based Economic Models and Econometrics. *Knowledge Engineering Review*, forthcoming.

Chen, Shu-Heng, and Tzu-Wen Kuo. 2002. Evolutionary Computation in Economics and Finance: A Bibliography, in *Evolutionary Computation in Economics and Finance*, edited by Shu-Heng Chen. Heidelberg: Springer, 419–455.

Chen, Shu-Heng, and Wei-Shao Wu. 2009. Price Errors from Thin Markets and Their Corrections: Studies Based on Taiwan's Political Futures Markets. *Advances in Econometrics*, 24: 1–25.

Cheung, Yin-Wong, and Daniel Friedman. 1997. Indiviudal Learning in Normal Form Games: Some Laboratory Results. *Games and Economic Behavior*, 19: 46–76.

Cross, John G. 1973. A Stochastic Learning Model of Economic Behavior. *The Quarterly Journal of Economics*, 87(2): 239–266.

Dorsey, Robert E., and Walter J. Mayer. 1995. Genetic Algorithms for Estimation Problems with Multiple Optima, Nondifferentiability, and Other Irregular Features. *Journal of Business and Economic Statistics*, 13(1): 53–66.

Duffy, John. 2006. Agent-based Models and Human Subject Experiments, in *Handbook of Computational Economics: Agent-based Computational Economics*, Vol. 2, edited by Leigh Tesfatsion and Kenneth Judd. Oxford: Elsevier, 949–1011.

Erev, Ido, and Alvin E. Roth. 1998. Predicting How People Play Games: Reinforcement Learning in Experimental Games with Unique, Mixed Strategy Equilibria. *American Economic Review*, 88(4): 848–881.

Evans, George W., and Seppo Honkapohja. 2001. *Learning and Expectations in Macroeconomics*. Princeton: Princeton University Press.

Gabaix, Xavier, David Laibson, Guillermo Moloche, and Stephen Weinberg. 2006. Costly Information Acquisition: Experimental Analysis of a Boundedly Rational Model. *American Economic Review*, 96(4): 1043–1068.

Gode, Dhananjay K., and Shyam Sunder. 1993. Allocative Efficiency of Markets with Zero Intelligence Traders: Market as a Partial Substitute for Individual Rationality. *Journal of Political Economy*, 101(1): 119–137.

Hommes, Cars. 2006. Heterogeneous Agent Models in Economics and Finance, in *Handbook of Computational Economics: Agent-based Computational Economics*, Vol. 2, edited by Leigh Tesfatsion and Kenneth Judd. Oxford: Elsevier, 1109–1186.

Price, Kenneth V., Rainer M. Storm, and Jouni A. Lampinen. 2005. *Differential Evolution: A Practical Approach to Global Optimization*. Heidelberg: Springer.

Roth, Alvin, and Ido Erev. 1995. Learning in Extensive-form Games: Experimental Data and Simple Dynamic Models in the Intermediate Term. *Games and Economic Behavior*, 8: 164–212.