DOES OBSERVATION INFLUENCE LEARNING?

By Olivier Amantier*

October 2001

Abstract

A common value auction experiment is run to compare the relative influence of observation and experience on learning. It is shown that the ex-post observation of opponents' actions and payoffs homogenizes behavior and accelerates learning toward the Nash equilibrium. Besides, experiential and observational learning are both relevant and of comparable magnitude. A general reinforcement model for continuous strategies, encompassing choice reinforcement learning, direction learning and payoff dependent imitation, performs well in explaining the experimental data and it dominates competing models such as the reinforcement of best response strategies.

Keywords: Learning, Observation, Continuous Reinforcement, Structural Estimation.
JEL Classifications : C13, D91, D83.

I would like to thank John Kagel for financing the experiments. I would like to thank Jean-Francois Richard, John Kagel, Alvin Roth, Thomas Muench, Debra Dwyer, Christopher Swann and Robert Slonim for helpfull discussions. I also would like to thank the editor and two anonymous referees for the quality of their comments. All remaining errors are mine.

**Department of Economics, State University of New York at Stony Brook, 11794-4384 Stony Brook, NY, USA. E-mail : olivier.armantier@sunysb.edu*

1. Introduction

Modern approaches of human behavior in psychology distinguish two principal sources of learning: personal experience and observation.[1] Experience shapes behavior as a response to a direct stimulus on an organism, while observational learning originates from observing the responses of other organisms. These basic principles of learning have been repeatedly tested and validated in psychology. Learning models developed in economics are largely inspired from the psychology literature. As we shall see in section 2 however, most learning models in economics rely exclusively either on experience or on observation. As a result, the following issues have been left essentially unexplored in the economic literature: firstly, the precise impact on learning of observation relative to experience has rarely been addressed; secondly, several competing hypotheses of observational learning have been proposed without being explicitly compared. The present paper attempts to address these issues by answering the following two sets of questions:

(1) How do observation and experience each affect learning in economics? Do subjects learn differently with observation? Does behavior converge faster toward an equilibrium when observation is possible? Is the magnitude of experiential and observational learning comparable?

(2) How is observation processed in the learning mechanism? Do people simply imitate others? Do they learn to predict their opponents' frequency of play? Do they lean toward strategies that would have been successful in the past?

To address these questions an experiment consisting of repeated common value auctions has been conducted under two extreme treatments. In the first treatment, players observe their opponents private signals, bids, and payoffs at the end of each auction period. In the second treatment, losing bidders receive no feedback, while the auction's winner observes only his own payoff. In the second treatment players can learn only from experience, while in the first treatment they can also learn from observation.

A model free approach is first applied to identify qualitatively the presence of learning and the influence of observation. Then, a learning model is developed to test competing hypotheses of observational learning. This model has five key characteristics: firstly, it explicitly combines experience and observation; secondly, it accounts for continuous strategies; thirdly, it is based upon a general reinforcement approach; fourthly, it captures the relevant features of a wide range of existing learning models in economics; finally, it is firmly grounded in

[1] See for instance the work of Whitehurst (1978) as well as Hikaru (1984).

the psychology literature.

In a nutshell, the experimental results show that observation homogenizes behavior and accelerates learning toward the Nash equilibrium solution. The structural estimation indicates that experimental subjects learn from experience but also copy successful behavior based on payoffs. Moreover, the experimental results suggest that observation influences behavior almost as much as experience.

The paper is structured as follow: section 2 briefly reviews the learning literature in economics. The model and the experiment are introduced in section 3. In section 4, regressions are run to identify the presence of learning and whether observation matters. The general reinforcement learning model is developed in section 5. The structural estimation strategy is presented in section 6. A benchmark is estimated in section 7 and it is compared to competing hypotheses in section 8. Finally, section 9 concludes.

2. Learning in Economics

Most economic models of learning rely either upon experience or observation. For instance, the Choice Reinforcement Learning approach (hereafter CRL) is the most popular in the class of experiential learning models (c.f. Bush and Mosteller (1955), as well as Erev and Roth (1998)). The model rests upon the assumption that individuals act randomly in ways that have produced high payoffs for themselves in the past. In other words, agents learn exclusively from their own experience.

In the observational learning literature, the Beliefs Based approach assumes that players form some beliefs about their opponents' future actions, based on past observations of plays. Then, players choose a best response strategy that maximizes their expected payoff given the beliefs they formed. In other words, the Beliefs Based approach assumes that players learn only from observing the frequency of play of their opponents. This class of models includes Cournot (1960) and Fictitious Play (Brown (1951)). Other models of observational learning include "imitation" (Schlag (1999)), where agents copy successful or popular actions of others; "learning direction theory" (Selten and Buchta (1999)), where players adjust their behavior in the direction of strategies that would have been successful; and finally, "Rule learning" (Stahl (1996), (2000)), assumes that players select among a set of predetermined behavioral rules based upon potential past performances. Note that these observational learning models assume that observation is processed differently within the learning mechanism.

The Experience Weighted Attraction model (hereafter EWA) of Camerer and Ho (1999) is a notable exception as it implicitly combines experiential and observational learning. The authors however limit observational learning to the reinforcement of unchosen strategies and competing alternatives are not tested.

In the experimental economics literature, few studies have investigated the effect of observation on learning. Erev and Rappoport (1998) run an entry game experiment and note that behavior is affected by information regarding other players' payoffs. Huck et al. (1999), as well as, Offerman et al. (2000) find that the experimental outcomes in Cournot games appear to be consistent with the prediction of an imitation model. However, experimental subjects are not randomly matched in these experiments, and one may wonder whether subjects are learning to play strategically, or whether they are learning to cooperate. Abbink et al. (1999) find that the responder rejection rate in ultimatum games increases when the actions of other participants are revealed. The authors suggest that responders try to create a group reputation for being "tough" by manipulating an "internal social norm". Mookherjee and Sopher (1994) find a significant effect of observation on behavior in a matching pennies experiment. The authors, however, were unable to derive an empirically relevant model to explain the nature of this effect. Finally, Duffy and Feltovich (1999) consider a similar design to analyze the effect of observation in ultimatum and best-shot games. The comparison of experimental outcomes with predictions simulated from a discrete CRL model yields mixed results.

This brief review of the literature indicates that the following two questions have not been clearly answered: firstly what is the relative impact of experience and observation on learning? Secondly, which observational learning approach prevails? The present paper propose to shed some light on these two issues.

3. Experimental Model and Treatments

3.1. The Model

An experimental session consists of T independent first price common value sealed bid auctions with N participants. At auction t, the true value of the item, x_0^t, is drawn from a uniform distribution on the interval $[\underline{x}, \overline{x}]$. Private signals ξ_i^t $(i = 1, ..., N)$ are drawn independently from a uniform distribution on $[x_0^t - \varepsilon, x_0^t + \varepsilon]$. The interval $[\underline{x}, \overline{x}]$, the distribution of true values, the value of ε, and the distribution of private signals are common knowledge from the beginning of the experiment. The true value x_0^t and the other bidders' private signals are not known to bidder i at the time she submits her bid b_i^t. The high bidder earns a profit equal to the true value of the item less the amount she bids. If bidders are symmetric, risk neutral, act non-cooperatively and if bids can be any real number and are subject to the boundary condition $B^{cv}(\underline{x}-\varepsilon)=\underline{x}-\varepsilon$, then the unique Bayesian Nash equilibrium bid function is:

$$B^{cv}(\xi_i^t)=b_i^{cv}=\xi_i^t-\varepsilon+Y \qquad\qquad if\ \underline{x}+\varepsilon\leq\xi_i^t\leq\overline{x}-\varepsilon \tag{3.1}$$

$$\text{where} \qquad Y=\frac{2\varepsilon}{N+1}\exp\left\{-\frac{N}{2\varepsilon}\left(\xi_i^t-\underline{x}+\varepsilon\right)\right\} \quad . \tag{3.2}$$

The bid function is more complex when ξ_i^t does not belong to the interval $[\underline{x}+\varepsilon,\overline{x}-\varepsilon]$.[2] To avoid drawing types in this interval, $[\underline{x},\overline{x}]$ is chosen large enough in practice so that x_0^t is unlikely to be generated outside the interval $[\underline{x}+2\varepsilon,\overline{x}-2\varepsilon]$. As a result, all ξ_i^t generated during the experiments belonged to the interval $[\underline{x}+\varepsilon,\overline{x}-\varepsilon]$.

The bid function (3.1) is linear in ξ_i^t up to a factor Y that decreases rapidly toward 0 as ξ_i^t increases. Garvin and Kagel (1994) ignore Y and approximate the strategy of player i by a bid factor $s_i^t\in\mathbb{R}$ such that

$$b_i^t=\xi_i^t-s_i^t\varepsilon \quad . \tag{3.3}$$

The strategy space is then fully characterized by a one dimensional continuous parameter s_i^t which can be directly inferred from the observation of an action. The corresponding Nash equilibrium bid factor is roughly $s^{NE}=1$. Player i can avoid the "winners curse" when $N=6$ (the number of bidders in the experiment) by using a bid factor $s_i^t\geq 0.714$, which leads to positive profits on average.

Kagel and Richard (1997) show that strategies of the form (3.3) are legitimate approximations of the bid functions since they are not only epsilon-equilibrium but also epsilon-best-responses to any strategy profile. This approximation can also be interpreted within the context of "Rule Learning" where one can use bounded rationality to argue that players limit themselves to linear rules of the form (3.3). Therefore, strategies are represented in the remainder by the bid factor, which is a tractable approximation of the bid function. Note that the objective is not to verify whether agents use the Nash equilibrium bid function, but rather to analyze how behavior changes over time. The bid factor is perfectly suited for this purpose.

[2] See Armantier (1999,a) for a complete derivation of the bid function over the entire support.

3.2. Experimental Treatments

The experiments are based upon the theoretical framework presented in section 3.1. Two extreme treatments are differentiated: in the "feed-back" (hereafter FB) treatment, every bids and the corresponding private signals, the winning bidder's profit as well as the true value x_0^t are announced to all players after each round. The information is revealed in such a way that it is not possible to identify individual bidders. In the "no-feed-back" (hereafter NFB) treatment the true value x_0^t and the winner's profit are announced only to the auction's winner. The remaining players are informed that they did not win the auction, and do not observe other bidders' information. Therefore, subjects are playing the same game, with the same unique Nash equilibrium in both treatments, and only the environment differs.

A common value auction experiment appears favorable to investigate the effect of observation on learning since i) common value auctions are complex games of incomplete information and substantial learning has been observed in previous experiments (see e.g. Garvin and Kagel (1994) or Selten and Buchta (1999)); ii) common value auctions may provide positive as well as negative payoffs which has been shown to intensify learning; iii) unlike Mookherjee and Sopher (1994) or Duffy and Feltovich (1999) the game has a unique symmetric equilibrium in pure strategy;[3] iv) auctions are "winner takes all" situations in which players should adjust their strategy only after winning an auction if they learn according to the CRL model.[4] On the other hand, if observation matters, the adjustment process should be smoother.

Subjects receive a starting balance and are no longer permitted to bid if their balance becomes negative during the course of the experiment. Subjects are randomly matched at the beginning of each auction period. In both treatments extra subjects were recruited in order to anticipate bankruptcies and keep the number of bidders constant. The occurrence of a bankruptcy and the introduction

[3] The matching pennies game in Mookherjee and Sopher (1994) has a mixed strategy equilibrium. Randomization is a non-trivial task for a subject and the optimal behavior is not easy to detect by other players (and also by the econometrician) even when observation is allowed. The games considered by Duffy and Feltovich (1999) possess several equilibria which, in addition, are not symmetric. Finally, it has been suggested that players may utilize strategically the information revealed in the ultimatum game used in Duffy and Feltovich (1999) not only to learn but also possibly to educate their opponents about their own "toughness".

[4] The experience of not winning an auction may lead players to modify their strategy. This possibility will be taken into account in the subsequent theoretic model.

of a replacement player was announced only in the FB treatment.[5] Note that replacement players enter the game without prior experience which may potentially create a discontinuity in behavior. Inspection of data, however, indicates that the impact of a replacement player is statistically insignificant.[6] Auction survivors were paid their end-of-session balance along with a participation fee in cash. Players who became bankrupt and extra subjects who did not participate received a participation fee. The experiment consisted of a FB and a NFB session. Each subject participated only in either one of the two sessions. Each session was divided in two sub-sessions lasting approximately an hour and a half each and taking place a week apart. The same subjects participated to the two FB (NFB) sub-sessions. Such design feature is commonly used to analyze the effect of experience (see e.g. Kagel and Richard (1997)). However, it may allow players to reflect or communicate between sub-sessions. Careful examination of data indicates that there is no significant difference between the actions taken at the end of sub-session 1 and the beginning of sub-session 2.

In practice the total number of periods played in a session was $T = 78$ in the NFB treatment (periods 1 to 38 in sub-session 1 and 39 to 78 in sub-session 2) and $T = 80$ in the FB treatment (periods 1 to 40 in sub-session 1 and 41 to 80 in sub-session 2). True values, and private signals were uniformly distributed in cents on the intervals $[\underline{x}, \overline{x}]$ and $[x_0^t - \varepsilon, x_0^t + \varepsilon]$ with $[\underline{x}, \overline{x}] = [15, 925]$ and $\varepsilon = 12$. In each session there were 24 active players in each period divided in 4 auctions of $N = 6$ bidders. A starting balance of \$25 was given to each player at the beginning of sub-session 1. In addition to their earnings subjects received \$20 to show up at both sub-sessions. Subjects were recruited primarily among undergraduates at the University of Pittsburgh. In both treatments, sub-session 1 started by reading the instructions aloud, followed with 2 dry runs in which the outcome did not count toward the players' final earnings.[7] The analysis of the data begin with the first auction period involving cash payoffs. In this experimental setting, the expected profit of the auction's winner was roughly \$3 per period. Descriptive statistics of the experimental outcomes are summarized in Table 1. The differences between the FB and NFB treatments are tested on the basis of the Mann-Whitney test.[8]

[5] In the subsequent analysis, a replacement player will be considered to play his first period at round j, if he enters the game after a bankruptcy in round $j - 1$.

[6] This result may be explained by the fact that bankruptcies are rare and essentially occur at the beginning of a session when subjects are actively learning and behavior is still very volatile.

[7] The instructions are similar to Garvin and Kagel (1994) and, therefore, are not reproduced here.

[8] The Mann-Whitney test is a non-parametric altenative to the two-samples-t-test that does

4. Adjustment Learning and Observation

The object of this section is to gather qualitative evidence of learning and to determine whether observation has a distinct effect on behavior. Table 1 indicates that subjects in the FB treatment earn significantly more money on average and they are less likely to suffer from the winner's curse. Players who receive feedback also participate for a significantly longer period of time as they are less likely to face bankruptcy. Finally, the item is bought more often by the high signal bidder in the FB treatment. In other words, Table 1 strongly suggests that there is a treatment effect, as observation appears to have a distinct influence on the experimental outcomes.

[Table 1 here]

Let us now examine the influence of observation on the learning process. The evolutions of the bid factor s_i^t in each treatment (averaged over 8 periods) are presented in figure 1 and 2.[9] The shape of these figures clearly indicates an increasing function. To formalize this result let us run the following regressions for each sample

$$s_i^t = \alpha_{FB} + \beta_{FB} t + u_{it} \qquad when\ s_i^t \in FB, \tag{4.1}$$

$$s_i^t = \alpha_{NFB} + \beta_{NFB} t + u_{it} \qquad when\ s_i^t \in NFB. \tag{4.2}$$

To account for the obvious heteroscedasticity across periods let us assume that $Var(u_{it}) = \sigma_d^2 \cdot t^{\gamma_d}$ where $d \in$ {FB,NFB}. Regression results related to this section are summarized in Table 2. Let us apply the standard one sided uniformly most powerful test to decide upon the null hypotheses $H_0 : \{\beta_d \leq 0\}$ for $d \in \{FB, NFB\}$.[10] The p-values in Table 2 indicate that one can reject the null hypotheses at the usual significance levels. This result however is not sufficient to conclude that the bid factor increases over time. Indeed, the evolution of strategies originates both from adjustment and/or market selection. Market selection eliminates, through bankruptcy, players who perform poorly. Such players use a bid factor below 0.714 which lowers the average strategy in the early periods. Consequently, the shape of figures 1 and 2 might represent "bad" players dropping

not require the assumptions of normality and equality of variances. For details see Gourieroux and Monfort (1995).

[9] The superscript t represents the number of periods actually played by a player and not the round of plays. The distinction is only relevant for replacement players.

[10] For details regarding the implementation of the test see e.g. Gourieroux and Monfort (1995).

out of the experiment. In order to differentiate the market selection effect from the adjustment process let us divide the original samples in two sub-samples of "good" and "bad" players. The latter regroups players who went bankrupt or lost more than 4/5 ($20) of their starting balance in less than 16 periods. Figures 3 to 6 compare the evolution of the bid factor for these sub-samples over the first 16 periods. Although these figures cannot be as clearly interpreted, they indicate that both types of players learn to increase their bid factor over time. This observation is confirmed at a 5% significance level by a pair of one sided tests applied to the good and bad players (see Table 2).

[Table 2 here]

To assess how observation affects the "speed" of learning, let us define the following two criteria: the number of periods before players learn to avoid the Winner's Curse and the number of periods before agents learn to consistently play the Nash equilibrium bid factor. Figures 1 and 2 indicate clearly that observation allows players to avoid the Winner's Curse sooner. The average bid factor is above 0.714 after (roughly) 20 periods in the FB treatment, versus 40 periods in the NFB treatment. Furthermore, the averages bid factors lay in a narrow band around the Nash equilibrium after 60 periods in the FB treatment, while in the NFB treatment strategies still oscillate after 80 periods in a large band below the Nash equilibrium, with no indication of a possible convergence. To confirm these results let us test first the one sided hypothesis $H_0 : \{\beta_{FB} \leq \beta_{NFB}\}$ in the joint estimation of the regression model (4.1) and (4.2) over the first 40 periods where most learning takes place. The test statistic is 4.583 corresponding to a p-value of 2.29E-6 and we can reject the null hypothesis at a 5% significance level. This test confirms that learning is initially slower in the NFB treatment. Symmetrically, let us test the one sided hypothesis $H_0 : \{\gamma_{FB} \geq \gamma_{NFB}\}$ where γ controls the multiplicative heteroscedasticity. The test statistic and p-value are respectively 3.181 and 7.31E-4. The null hypothesis is rejected at a 5% significance level which implies that strategies are significantly more volatile in the NFB sample. The speed of learning observed in figures 1 and 2 may be influenced by the market selection effect. However, a similar analysis restricted to good or bad players produces comparable results.

To summarize, subjects learn to increase their bid factor. This process is affected by experience as well as observation of others. Observation has a distinct influence on the experimental outcomes as it homogenizes behavior and accelerates learning toward the Nash Equilibrium. In the following sections we attempt to determine how observation is processed within the learning mechanism.

5. Reinforcement Learning Model and Observation

5.1. Existing Learning Models in Economics and Psychology

Popular learning models (CRL, EWA, Rule Learning, similarity...) are based upon four key principles: firstly, strategies (or learning rules) are discrete random variables drawn from a given distribution; secondly, strategy distributions are updated after each period based upon new information; thirdly, the updating rule reinforces the probability of a given played (CRL) or un-played (EWA) strategy (or learning rule) based upon a reinforcement factor representing either actual (CRL) or potential (EWA) payoff; finally, most parameters of the models represent some fundamental principle of learning that have been repeatedly tested in psychology.[11] Note that the learning models in economics have been almost exclusively developed for discrete strategy space. To accommodate continuous strategies, such as the bid factor in the theoretic model presented in section (3.1), one would need to discretize the strategy space. Such discretization, however, are often arbitrary and it may lead to inaccurate approximations.[12]

The subsequent section introduces a generalized reinforcement model explicitly accounting for observation and continuous strategies. The objective is not to formulate a new and/or competing model, but rather to capture most relevant features of existing learning approaches in one encompassing model. Moreover, the model proposed is consistent with the psychology literature where reinforcement is not limited to choices.[13]

5.2. General Reinforcement Learning Model

Before we formalize the continuous reinforcement rule, let us illustrate the learning process in the context of the traditional CRL model. At round t, player i draws

[11] For instance, Blackburn (1936) established "the Power Law of Practice", Watson (1930) confirmed the importance of "Recency" and reference points.

[12] Note that in the experiment, the bids and private signals are dicrete variables defined in cents over wide intervals. Therefore, the bid factor $s_i^t = \left(\xi_i^t - b_i^t\right)/\varepsilon$ takes a large but finite number of values on $[-1, 1.5]$ and it may be best described as a continuous variable.

[13] For instance, Social Learning emphasizes secondary reinforcements of unconditioned (neutral) stimulus (e.g. Houston (1991)) and perceived reinforcements of observed actions (e.g. Vicarious Reinforcement of Bandura (1977)).

her strategy s_i^t from a continuous distribution defined over the interval $[\underline{s}, \overline{s}]$, with mean m_i^t, variance v_i^t, and probability density function $g_i^t(.)$. For the ease of exposition, let us assume that the strategy distribution is symmetric (See figure 7). Then, m_i^t can be seen as the perceived "optimal" strategy at round t, while v_i^t represents the confidence in this "optimal" strategy. When v_i^t tends to zero the strategy distribution is concentrated around the mean m_i^t and player i is almost guaranteed to select m_i^t which is then a pure strategy equilibrium. On the other hand, when v_i^t gets larger the strategy distribution can be expected to be almost uniform and player i will experience different strategies with equal probability. v_i^t may be seen either as an experimentation factor or a bandwidth for computational errors.

After playing a strategy s_i^t subject i receives a payoff p_i^t and she updates her strategy distribution with a reinforcement rule $R(.)$. Heuristically, the effect of $R(.)$ in the CRL model is to add some mass to the strategy density $g_i^{t+1}(.)$ around or away from s_i^t, depending upon the reward p_i^t. Namely, when player i receives a positive payoff, the reinforcement rule has the shape of a normal density function with mean s_i^t and standard deviation $\sigma_i(t)$ (see figure 7). In this case, some mass is added to the density function $g_i^{t+1}(.)$ around the strategy played s_i^t. As the reward gets larger, the standard deviation $\sigma_i(t)$ gets closer to 0, and the reinforcement rule is more concentrated (see figure 8). As a consequence, more mass is added around s_i^t and player i is more likely to play the same strategy again. Symmetrically, when a player receives a negative payoff, the reinforcement rule has the shape of a reverse normal distribution (see figure 9). Intuitively, the negative reinforcement rule "subtracts" some mass from the strategy played s_i^t and redistribute it to strategies away from s_i^t. As a consequence, the probability of subsequently generating a strategy in the neighborhood of s_i^t decreases. The negative reinforcement rule is more concentrated around the strategy played as the reward p_i^t decreases and player i becomes less likely to draw the same strategy s_i^t in the future (see figure 10). When the payoff is zero, the strategy distribution stays unchanged as there is no experiential learning. Note that this model may converge toward a pure or a mixed strategies equilibrium (see Armantier (1999, b)).

Let us now formalize the model. Consider a general reinforcement learning model consisting of L different observational and experiential reinforcement rules. The reinforcement rule in model l $(l = 1, ..., L)$ is a continuous function denoted $R(s \mid m_{i,k}^l, r_{i,k}^l, t, \beta^l)$ where $s \in [\underline{s}, \overline{s}]$ and β^l is a vector of parameters. This rule is similar to the one previously introduced for the CRL model and it is such that at any period $k \leq t$, (where t is the current period) player i reinforces a strategy $m_{i,k}^{l}$ based upon a reinforcement factor $r_{i,k}^{l}$. The general reinforcement learning model combines all these reinforcement rules and the strategy distribution evolves according to the following law of motion,

$$g_i^{t+1}(s) = \frac{g_i^1(s) + \sum\limits_{k=1}^{t}\sum\limits_{l=1}^{L} R(s \mid m_{i,k}^l, r_{i,k}^l, t, \beta^l)}{1 + \sum\limits_{k=1}^{t}\sum\limits_{l=1}^{L}\int\limits_{\underline{s}}^{\overline{s}} R(u \mid m_{i,k}^l, r_{i,k}^l, t, \beta^l)du} \quad . \tag{5.1}$$

In the traditional CRL we have $L = 1$, $m_{i,k}^l = s_i^k$ and $r_{i,k}^l = p_i^k$ where s_i^k and p_i^k are the strategy played and the payoff received by player i at period $k \leq t$. The game is assumed to be such that payoffs are within a closed interval $\left(p_i^k \in \left[\underline{p}, \overline{p}\right]\right)$ and there always exists a strategy, s_i^k, for which the expected payoff is greater than zero. Note that the model verifies the "law of practice" stating that learning curves are steepest in early periods, thanks to the cumulative structure of the reinforcement rule. The function $g_i^1(.)$, commonly known as the initial propensity, may reflect players introspection or experience from previous games. How subjects select their initial strategy distribution is a question beyond the scope of the present paper. In the remainder $g_i^1(.)$ is a given distribution, possibly a function of parameters to be estimated.

The reinforcement rule adopted here is based upon a normal probability density function $f(. \mid \mu_i^k, \sigma_i^k(t))$ with mean μ_i^k and variance $\sigma_i^k(t)$,

$$R(s \mid m_{i,k}^l, r_{i,k}^l, t, \beta^l) = \begin{Bmatrix} f(s \mid \mu_i^k, \sigma_i^k(t)) & when\ r_{i,k}^l > 0 \\ f(\mu_i^k \mid \mu_i^k, \sigma_i^k(t)) - f(s \mid \mu_i^k, \sigma_i^k(t)) & when\ r_{i,k}^l < 0 \end{Bmatrix} \tag{5.2}$$

$$\text{where} \qquad \mu_i^k = m_{i,k}^l \qquad \text{and} \qquad \sigma_i^k(t) = \left(\alpha_0^l\right)^{(k-t-1)}\left(r_{i,k}^l\right)^{-2} \tag{5.3}$$

and $0 < \underline{\alpha_0} \leq \alpha_0^l \leq \overline{\alpha_0} < 1$. In this simple model the parameter β^l reduces to α_0^l. Note that the negative reinforcement rule in expression (5.2) is always positive. This formulation guarantees that $g_i^{t+1}(.)$ will be a density function.

The exponent of α_0^l acts as a discount or forgetting parameter that reduces the influence of past experiences. This effect, known as "Recency", has been shown to be robust in the psychology literature. As k gets distant from the current period t, $\sigma_i^k(t)$ gets larger which implies that the corresponding reinforcement rule is flatter. As a consequence, the outcome in a much earlier period k has less influence on the strategy distribution $g_i^{t+1}(.)$. Beyond a bounded memory interpretation, α_0^l

also takes into account the fact that other players are simultaneously learning to play the game. Player i's best response to other players' initial strategies will not be optimal anymore in later periods since player i's opponents will modify their strategies through learning. Therefore, subjects should deliberately forget distant past, and recent events should play a larger role in determining behavior.

5.3. Direction Learning

To illustrate the concept of direction learning consider the traditional CRL and the experimental auction model. When bidder i plays s_i^t and receives a negative payoff, he should realize that his bid was too high. He should also understand that a smaller bid factor would have generated a worse payoff, while a larger bid factor would have improved his situation. In that case, instead of negatively reinforcing strategies symmetrically around s_i^t, he should reinforce negatively strategies below s_i^t, and reinforce positively strategies above s_i^t. By doing so, player i is more likely to generate a strategy greater than s_i^t and get a better payoff in the future.

To incorporate direction learning within the reinforcement rule, let us replace μ_i^k in equation (5.3) by

$$\mu_i^k = m_{i,k}^l + \left(\alpha_j^l\right)^2 \left(\underline{s}.I_{\alpha_j^l<0} + \overline{s}.I_{\alpha_j^l>0} - m_{i,k}^l\right) e^{-\left(r_{i,k}^l\right)^2} \qquad , \tag{5.4}$$

where $-1 \leq \alpha_j^l \leq 1$, $j = 1$ $(j = 2)$ when the reinforcement factor $r_{i,k}^l$ is positive (negative), and $I_{\alpha_j^l<0}$ is the indicator function satisfying $I_{\alpha_j^l<0} = 1$ when $\alpha_j^l < 0$ (otherwise $I_{\alpha_j^l<0} = 0$). The parameter of the reinforcement rule now generalizes to $\beta^l = \left(\alpha_0^l, \alpha_1^l, \alpha_2^l\right)$. When α_j^l is positive (respectively negative), μ_i^k lays in the interval $\left[m_{i,k}^l, \overline{s}\right]$ (respectively $\left[\underline{s}, m_{i,k}^l\right]$) and strategies slightly greater (respectively smaller) than $m_{i,k}^l$ are primarily reinforced. In other words, the parameter α_1^l $\left(\alpha_2^l\right)$ accounts for any potential direction learning associated with a positive (negative) reinforcement factor. Note that μ_i^k gets closer to $m_{i,k}^l$ when $\left(r_{i,k}^l\right)^2$ increases. This implies that when the reinforcement factor is large and positive (negative) the center of the reinforcement rule is closer to the strategy reinforced, so that $m_{i,k}^l$ is more (less) likely to be played in the future. Finally, $\alpha_j^l = 0$ corresponds to no direction learning, while $\alpha_j^l = \pm 1$ can be considered full direction learning, since players essentially reinforce either $\underline{s}$ or $\overline{s}$ for a reinforcement factor $r_{i,k}^l$ close to zero. Unlike Selten and Buchta (1999), this formulation allows one to predict not only the direction of learning but also the amount of change.

Anderson et al. (1999) propose a different theoretic direction learning model where the direction and amount of change are based upon the gradient of the expected payoff function. The model proposed here is slightly more flexible since direction learning may vary with actual, expected or even another player's payoff, depending upon the reinforcement rule adopted. In addition, the effect of positive and negative rewards are differentiated which, as we shall see in section 7, is an important feature to describe learning in the experiment. Finally, the direction learning model adopted is simpler and it may be implemented easily to estimate structurally the general reinforcement model.

5.4. Reinforcement with Observational Learning

The reinforcement rule $R(s \mid m_{i,k}^{l}, r_{i,k}^{l}, t, \beta^{l})$ can account explicitly for different forms of observational learning. The structure of the rule essentially remains the same, but the variables $m_{i,k}^{l}$ and $r_{i,k}^{l}$ take different values depending upon the type of observational learning under consideration.

For instance, with payoff dependent imitation, player i partially reinforces the strategy played by player $j \neq i$ at period k $\left(m_{i,k}^{l} = s_{j}^{k}\right)$ based upon the reinforcement factor $r_{i,k}^{l} = \theta^{l} \cdot p_{j}^{k}$, where p_{j}^{k} is the payoff of player j, and θ^{l} is a discount factor that downweights reinforcement from other players' payoffs. In other words, when player i witnesses a large loss (profit) by player $j \neq i$, she revises her strategy distribution so that she will be less (more) likely to play s_{j}^{k} in the future. In "winner takes all" games such as an auction, players may choose to reinforce only the winner's strategy.

According to EWA players update the strategy played but they also use observation to look back and determine how un-chosen strategies and, most importantly, un-chosen best response strategies would have performed. In other words, at the end of round k, player i observes her opponents strategy profile s_{-i}^{k} and she calculates what would have been her best response strategy $BR(s_{-i}^{k})$. Then, player i updates her strategy distribution around $BR(s_{-i}^{k})$ based upon the profit (or expected profit) Π_{BR} this best response would have generated. Although analytically dissimilar, the generalized reinforcement model can capture the most relevant features of EWA. Indeed, in addition to a CRL rule, an observational reinforcement rule may be created by setting $m_{i,k}^{l} = BR(s_{-i}^{k})$ and $r_{i,k}^{l} = \theta^{l} \cdot \Pi_{BR}$. Note that θ^{l} represents the relative effect of foregone payoffs and it is closely related to the parameter δ in Camerer and Ho (1999).

One can define other forms of observational learning. For instance, pure imitation models reinforce the winner's strategy independently of payoff. In the context of auctions, Garvin and Kagel (1994) suggest that losing bidders apply the winner's strategy s_w^k to their own signal and learn from determining the profit this bid would have yielded had they won the auction. In other words, when player i does not win the auction he reinforces the strategy $m_{i,k}^l = s_w^k$ based upon the reinforcement factor $r_{i,k}^l = \theta^l \left(x_0^k - \xi_i^k + s_w^k \varepsilon\right)$. Finally, Erev and Rappoport (1998) suggest that observation of other players' payoffs influence the reference point. The latter hypothesis has been tested without success in an earlier version of the paper (see Armantier (1999, b)).

6. Estimation Procedure and Model Comparison

From an econometric perspective learning models have two interesting features. Firstly, the observed actions $S_N^T = \left(s_1^1, ., s_N^1, s_1^2, ., s_N^2, ., s_1^T, ., s_N^T\right)$ are neither identically nor independently distributed since strategy distributions are updated individually based on previous periods plays and outcomes. Secondly, learning models may converge toward a pure strategy equilibrium in which case the asymptotic strategy distribution $g_i^\infty(.)$ is degenerate. These two characteristics are such that the analysis of learning model estimators is non-trivial and rarely addressed in the literature with the notable exception of Cabrales and Garcia-Fontes (1999) and Armantier (2000). The Maximum Likelihood is a popular method to estimate learning models with a finite number of strategies (e.g. Camerer and Ho (1999), or McKelvey and Palfrey (1995)). However, as noted by Stahl (1996) and Armantier (2000) the application of Maximum Likelihood to continuous strategies may be hazardous. Indeed, as behavior converges toward an equilibrium the strategy distributions becomes more concentrated and later observations are much more influential on the likelihood function. This imbalance may result in a disproportionate contribution of later observations on the parameter estimate.

To circumvent this convergence problem Armantier (2000) proposes an M-estimator $\widehat{\beta}$ corresponding to an objective function of the form:

$$SOM_2(\beta) = \sum_{t=1}^{T} \sum_{i=1}^{N} \left\{ \left(\left[s_i^t\right] - \eta_{i,t}^1(\beta)\right)^2 + \left(\left[s_i^t\right]^2 - \eta_{i,t}^2(\beta)\right)^2 \right\} \tag{6.1}$$

where β is the parameter to be estimated and $\eta_{i,t}^p(\beta)$ is the moment of order p of player i's strategy distribution at period t. Heuristically, the objective is to reconcile observations with their theoretical moments conditionally on the history of play. This method offers the key advantage of allocating the same weight in the objective function to any observation. Armantier (2000) proves that the estimator of the general reinforcement model parameter $\widehat{\beta}$ is consistent and asymptotically normally distributed as T tends toward infinity. The expression (6.1) requires the derivation of the theoretical moments at each period and for all bidders. The strategy distributions do not have a tractable analytical form and the theoretical moments $\eta_{i,t}^{p}(\beta)$ are replaced by Monte Carlo simulation estimates, $\widehat{\eta}_{i,t}^{p}(\beta)$.

Nested learning models are typically compared on the basis of likelihood ratio tests (e.g. Stahl (2000) or Camerer and Ho (1999)). As previously mentioned however, this approach is not well suited here due to the inadequate properties of the likelihood function. Instead, we will consider traditional Wald tests based upon the unconstrained optimization of the objective function (6.1).[14] The Wald test possesses the appropriate asymptotic properties since the M-estimator is consistent and asymptotically normally distributed. The covariance matrix involved in the determination of the test statistic is evaluated with a Bootstrap technique based upon the estimated parameter $\widehat{\beta}$.[15]

7. Estimation of a Benchmark Model

Let us estimate first a benchmark model including the traditional CRL and payoff dependent imitation, both with direction learning. Alternative learning models will be compared and tested against this reference model in section 8.

The vector of parameters to estimate is $\beta = (\beta^{CRL}, \beta^{IM}, \mu, \sigma^2)$ where $\beta^{CRL} = \left(\alpha_0^{CRL}, \alpha_1^{CRL}, \alpha_2^{CRL}\right)$ and $\beta^{IM} = \left(\alpha_0^{IM}, \alpha_1^{IM}, \alpha_2^{IM}, \theta^{IM}\right)$ are the coefficients of the CRL and payoff imitation reinforcement rule and (μ, σ^2) are the parameters of the initial strategy distribution function. All strategies observed during the experiments are within the interval $[-1.0, 1.5]$.[16] Therefore, $[\underline{s}, \overline{s}]$ is fixed to $[-1.0, 1.5]$. Players are initially considered symmetric and strategies in period 1 are assumed to be generated from a normal distribution truncated on $[-1.0, 1.5]$ with parameters (μ, σ^2). The histograms of players' strategies in the first periods of play tend to support this assumption. The M-estimator presented in section 6 is applied to

[14] For details on the implementation of the Wald test see Gourieroux and Monfort (1995).

[15] The Bootstrap is a statistical technique consisting in repeatedly resampling the original data from the estimated distribution in order to make inferences from the resamples on parameters such as, for instance, the standard deviation of the estimated parameters. For details on the Bootstrap technique see Shao and Tu (1996).

[16] A bid factor of -95.8 was recorded but it was a mistake to the player's own admission.

the FB and the NFB samples. Results are presented in Table 3.

[Table 3 Here]

The estimation of the initial propensity parameters correspond to an expected strategy of 0.247 (0.249) and a standard deviation of 0.661 (0.692) in the FB (NFB) sample. Note that the first two moments of the initial strategy distribution are strikingly close in both treatments. Actually, a statistical test indicates that the parameters $\left(\widehat{\mu}, \widehat{\sigma}^2\right)$ cannot be distinguished in the estimation including both the FB and NFB samples. In other words, subjects appears to have a similar initial apprehension of the game independently of the treatment applied. Finally, the strategy selection in period 1 is very volatile and almost uniform over $[-1.0, 1.5]$. This may be explained by the complexity of the common value auction and/or by the subjects lack of previous experience with comparable games.

Let us examine first the parameters associated with the CRL model. The forgetting parameter $\widehat{\alpha}_0^{CRL}$ is close to 1 with both the FB and NFB samples. This implies that learning occurs at a slow pace, as past and recent experience have almost the same influence on the current strategy selection. The forgetting parameter is, however, slightly but significantly smaller in the NFB treatment. Therefore, behavior can potentially change somewhat more abruptly as the result of new experiences when information is low. The estimation of forgetting parameters is known to vary tremendously from game to game. Some estimates in Camerer and Ho (1999) and Stahl (2000) are actually greater than one suggesting “explosive dynamics” and/or a model’s misspecification. The forgetting parameters in the present paper are estimated without constraint, but they are all below one and seem consistent with previous studies.

The parameter α_1^{CRL} is insignificant, while α_2^{CRL} is close to 1, in both samples. This indicates that a negative reward leads subjects to increase the bid factor, while a positive payoff does not influence the direction of learning. Therefore, direction learning theory applies only in the experiment when payoffs are negative.

Let us now turn to the estimation of the payoff dependent imitation model parameters. First and foremost, one can observe that $(\alpha_0^{IM}, \alpha_1^{IM}, \alpha_2^{IM}, \theta^{IM})$ are insignificant (significant) when estimated with the NFB (FB) samples. Therefore, as expected, observational learning occurs only when subjects receive information about their opponents’ plays and outcomes. Note also that $\widehat{\alpha}_0^{CRL}$ is not significantly different than $\widehat{\alpha}_0^{IM}$ in the FB treatment. In other words, subjects appear to remember as well past personal experience and observation. On the other hand, α_2^{CRL} is slightly but significantly larger than α_2^{IM}. This implies that losing one’s own money has a more direct impact on the direction of learning. Finally, the parameter θ^{IM} is slightly but significantly smaller than 1 in the FB treatment. This indicates that losing bidders reinforce the strategy of the winner with almost the same intensity as if they had won the auction themselves. The magnitude of this effect is larger than in Camerer and Ho (1999) who find that (on average) subjects weigh observation half as much as experience. This difference might be explained by the fact that the common value auction is arguably more complex than the games considered by Camerer and Ho. In the common value experiment it appears that observation is a useful and necessary complement to experience.

To test for the presence of heterogeneity across players within the same treatment let us first estimate the model with a different parameter β_i $(i = 1, ..., N)$ for each player. Then, we can test the restriction $H_0 : \left\{ \left(\beta_i^{CRL}, \beta_i^{IM}\right) = \left(\beta_1^{CRL}, \beta_1^{IM}\right), \forall i = 2, ..., N \right\}$. The Wald statistics, the degrees of freedom and the p-values of all nested tests run in this section are presented in Table 4.[17] The p-values indicate that one cannot reject the null hypothesis at the usual significance levels. Unlike Stahl (1996) and (2000), we do not find conclusive evidence of heterogeneity across players. However, comparable tests clearly suggest heterogeneity across treatments.

Let us now test whether the parameters α_j $(j = 0, 1, 2)$ are common to both the experiential and the observational learning reinforcement rule in the FB sample. The null hypothesis is $H_0 : \left\{ \alpha_j^{CRL} = \alpha_j^{IM}, \forall j = 0, 1, 2, \right\}$. The p-value shows that we can reject the null hypothesis at a 5% confidence level. Finally, Table 4 also indicates that the pure CRL model $\left(H_0 : \left\{\beta^{IM} = 0\right\}\right)$ and the pure payoff dependent imitation model $\left(H_0 : \left\{\beta^{CRL} = 0\right\}\right)$ are strongly rejected by the data. In other words, observational and experiential learning are both present and of comparable magnitude but they are significantly distinct.

[Table 4 Here]

To evaluate the relevance of the estimates, let us run some Monte Carlo simulations. Figures 11 and 12 represent the average learning path (along with one standard deviation) of one thousand computer replications of the experimental sessions based upon the parameters estimated with the M-estimator. The overall simulated behavior mimic very accurately the behaviors of actual players, and is well within one standard deviation (Figures 1 and 2). An additional measure of fit is provided in Table 5 where Monte Carlo descriptive statistics are summarized. A comparison with Table 1 indicates that the general reinforcement learning model

[17] The Wald tests are based upon the estimations of the unconstraint models. With the exception of the heterogeneity test, the tests are performed only with the FB sample.

replicates fairly well the qualitative results observed during the experiment as well as the differences between the FB and NFB treatments.

[Table 5 here]

8. Models Comparison

The object of this section is to test competing learning models and in particular alternative hypotheses of observational learning. The different hypotheses can be nested within the general reinforcement model and comparisons will be conducted on the basis of Wald tests. Results are presented in Table 4.

8.1. Exogenous Adjustment and Nash Equilibrium with Errors

Learning is typically assumed to occur when the strategy selection of a player at a given period is influenced by his own or any other players past actions and/or outcomes. In contrast, a non-learning adjustment process is determined exogenously. Such adjustment includes for instance reducing mistakes while manipulating the computer, or making smaller mental calculation errors.

The general reinforcement rule can be modified to account for exogenous adjustment by imposing $m^l_{i,k} = \lambda_1 m^l_{i,k-1} + s_\infty(1-\lambda_1)$ and $r^l_{i,k} = \lambda_2 r^l_{i,k-1}$ where $\lambda_1 \in [0,1[$, $\lambda_2 > 1$, $s_\infty \in [\underline{s}, \overline{s}]$, $\alpha^l_0 = 1$, $r^l_{i,0}$ is normalized to 1 and $m^l_{i,k}$ $\left(r^l_{i,k}\right)$ is the strategy reinforced (the reinforcement factor) at period $k \leq t$. This exogenous reinforcement rule has three characteristics: firstly, the strategy reinforced and the reinforcement factor evolve exogenously; secondly, the reinforcement factor is an increasing unbounded function of t; finally, behavior converges toward a pure strategy equilibrium s_∞. The Nash equilibrium model with errors is a special case of exogenous adjustment where the asymptotic outcome s_∞ equals 1 (the Nash equilibrium in the auction model). This approach finds additional support from the fact that players in the FB sample appear to converge toward the Nash equilibrium.

The unrestricted model consists of the benchmark model combined with the exogenous adjustment process. Parameters of the unconstrained model are $(\beta, \lambda_1, \lambda_2, s_\infty)$ where $\beta = (\beta^{CRL}, \beta^{IM}, \mu, \sigma^2)$ is the parameter vector in the benchmark model. The model is tested twice under each of the following two null hypotheses H_0 : $\left\{\beta^{CRL} = 0, \beta^{IM} = 0\right\}$ and H'_0 : $\left\{\beta^{CRL} = 0, \beta^{IM} = 0, s_\infty = 1\right\}$. Table 4 indicates that both hypothesis are strongly rejected by the data. In other words, the adjustment process is not exogenous and it can be attributed to learning.

8.2. Imitation Learning

In a pure imitation model, losing bidders copy exactly the previous period winner's strategy. Such strict behavior is not observed in the data. However we can test weaker versions of the model such as pure imitation with errors and imitation of successful or popular behavior.

To incorporate the imitation of successful behavior, the general reinforcement learning model is modified by setting $m_{i,k}^l = s_w^k$ and $\sigma_i^k(t) = \left(\alpha_0^{PIM}\right)^{(k-t-1)}$ where $0 < \alpha_0^{PIM} < 1$ and s_w^k is the current period winner's strategy. Hence, every player reinforces the winner's strategy independently of the payoff it yielded. A pure imitation model with errors requires that $\sigma_i^k(t) = c - (c - \widetilde{\alpha}_0^{PIM})I_{k=t}$ where $I_{k=t}$ is the indicator function and c is a large real constant ($c = 100$ in practice). This formulation implies that the variance of the reinforcement rule equals $\widetilde{\alpha}_0^{PIM}$ when $k = t$. Otherwise, the reinforcement rule is flat for k strictly smaller than the current period t since the variance $\sigma_i^k(t)$ is large. As a result, the strategy distribution in period t is affected only by the winner's strategy in period $t-1$. In other words, subjects copy the strategy of the previous period winner but they are allowed to make mistakes and select a strategy different than s_w^t. Imitation of popular behavior is derived along the same line except that the reinforced strategy $m_{i,k}^l$ is now equal to the average $\left(s_a^k\right)$ or the mode $\left(s_m^k\right)$ of other players strategies.

The unconstrained parameters in the models including imitation of successful or popular behavior are $\left(\beta, \alpha_0^{PIM}\right)$, and $\left(\beta, \widetilde{\alpha}_0^{PIM}\right)$ in the models including pure imitation model with errors. The null hypothesis in each test is $H_0 : \left\{\beta^{CRL} = 0, \beta^{IM} = 0\right\}$. These hypotheses are strongly rejected by the data (see Table 4). Note that tests of the hypothesis $H_0 : \left\{\beta^{IM} = 0\right\}$ are also rejected indicating that payoff dependent imitation is not dominated by other forms of imitation learning. In other words, imitation appears to be a driving force in learning, as long as it incorporates the performance of the subject copied. Models including imitation of other players (not only the winner) based upon actual or potential payoff have also been tested without success. Finally, the Garvin and Kagel hypothesis presented in section 5.4 was also strongly rejected.

8.3. Reinforcement of Best Responses Strategies

Do best responses act as an attractor in the strategy selection? To test this hypothesis let us consider an observational learning model reinforcing the ex-post best response to the actual plays of the other players based upon the expected payoff this strategy would have generated.[18] The unrestricted model is defined by the parameters $\left(\beta, \beta^{RBR}\right)$ where $\beta^{RBR} = \left(\alpha_0^{RBR}, \alpha_1^{RBR}, \alpha_2^{RBR}, \theta^{RBR}\right)$ and the null hypothesis is $H_0 : \left\{\beta^{RBR} = 0\right\}$. Table 4 clearly shows that the reinforcement of best responses is dominated by a combination of CRL and payoff dependent imitation learning. This results may be partially explained by the fact that best responses and expected payoffs are complex to derive in common value auctions.

Camerer and Ho (1999) suggest that the observation of others serves only as a proxy to determine one's own best strategy. They also suggest that reinforcement of foregone payoffs should encompass imitation learning in symmetric games with a large number of players. The common value auction game is symmetric and the number of players may be considered reasonably large. The results obtained here, however, appear to partially dismiss the Camerer and Ho claim. The reinforcement of foregone strategies suggested by EWA has been found to have some explanatory power, but imitation should be seen as a distinct and sometime dominating force in observational learning that should not be ignored.

9. Conclusion and Discussion

The object of the paper was to shed light on the following two questions: What is the relative influence of observation on learning? Which among the observational learning models proposed in the literature prevails? A common value auction experiment demonstrates that i) players learn both from observation and experience; ii) observation may play a major role in the learning process; iii) observation may homogenize and accelerate learning toward the Nash Equilibrium; iv) observational and experiential learning can be of comparable magnitude; v) finally, a payoff depend imitation approach with direction learning dominates other observational models (such as pure imitation, or reinforcement of best strategies) to explain the experimental data.

These results however may not generalize to every game. In particular, the common value auction is a complex game of incomplete information, and observational learning may play a lesser role in simpler contexts. It has also been shown that other forms of observational learning perform well in certain games (e.g. Camerer and Ho (1999)). Finally, Offerman et al. (2000) suggest that the dissemination of information about others may facilitate (tacit) collusion in repeated games with the same players. Future research should therefore concen-

[18] In practice the expected payoff is evaluated by Monte Carlo simulations.

trate on identifying the determinants within the structure of the game and within the observed information that may influence learning. The contribution of the present paper was to demonstrate that observation, and more specifically payoff dependent imitation, may be a driving force in the learning process.

It should also be noted that subjects in the FB treatment receive total, non-noisy and immediate feedback regarding their opponents' private information, plays and outcomes. Such unique conditions are difficult to reproduce in real life. However, it is not uncommon in new or changing environments to observe that information about participants is revealed ex-post. Such exchanges of experience may be initiated by an official regulatory entity but it may also stem from the firms themselves through coalitions or trade associations. For instance, European and Australian governmental agencies and telecommunication companies, were made aware of the most minute details regarding both the strategies employed and the results obtained by their American counterparts before implementing and participating in their national spectrum auction. Providing information about others may be a particularly relevant policy in developing economies, new markets (e.g. the internet), or after a policy change. In such contexts, subjects may not have time to learn from their own experience and the social cost of failure may be quite high.

Finally, environmental factors, such as the ex-post observation of opponents actions and payoffs are typically ignored in theoretic modeling. The paper clearly shows that such factors may influence short term behavior. This research therefore fits within the contingent learning approach (Slembeck 1998) in the sense that we attempt to identify within the environment the determinants affecting learning. Some of these determinants have been widely studied in the psychology literature and these results should not be ignored by economists. The present papers shows that it is possible to develop empirically relevant models that are consistent with the two disciplines.

REFERENCES

Abbink K., Abdolkarim S. and S. Zamir, 1999, "Fairness and Public Good Aspects of Punishement Behavior", Mimeo, University of Bonn.

Anderson S., Goeree J. and C. Holt, 1999, "Stochastic Game Theory: Adjustment to Equilibrium Under Noisy Directional Learning", Mimeo, University of Virginia.

Armantier O., 1999, a, "Empirical Application of Game Theory : Three Essays", Ph.D. Dissertation, University of Pittsburgh.

Armantier O., 1999, b, "The Influence of Observation on Economic Behavior: Experimental Evidence", Mimeo SUNY Stony Brook.

Armantier O., 2000, "Estimation and Comparison of Learning Models", Mimeo SUNY Stony Brook.

Bandura, A., 1977, Social learning theory, Englewood Cliffs, NJ : Prentice-Hall.

Blackburn J., 1936, "Acquisition of Skill: An Analysis of Learning Curves", IHRB Report No. 73.

Brown J., 1951, "Iterative Solution of Games by Fictitious Play", In Activity Analysis of Production and Allocation, NY: John Wiley and Son.

Bush R. and F. Mosteller, 1955, "Stochastic Models of Learning", New York: Wiley.

Cabrales A. and W. Garcia-Fontes, 1999, "Estimating Learning Models from Experimental Data: Quadratic Deviation and Maximum Likelihood", Mimeo.

Camerer C. and T-H. Ho, 1999, "Experience-Weighted Attraction Learning in Normal Form Games", Econometrica 67, 827-874.

Cournot A., 1960, "Recherche sur les Principes Mathematiques de la Theories de la Richesse", Translated into English by N. Bacon, London: Haffner.

Duffy, J. and N. Feltovich. 1999, "Does Observation of Others Affect Learning in Strategic Environments? An Experimental Study", International Journal of Game Theory v28, n1 (February 1999): 131-52.

Erev I. and A. Rapoport, 1998, "Coordination, "Magic," and Reinforcement Learning in a Market Entry Game", Games and Economic Behavior v23, n2 146-75.

Erev I. and A. Roth, 1998, "Predicting How People Play Games: Reinforcement Learning in Experimental Games with Unique, Mixed Strategy Equilibria", American Economic Review, 8, 848-881.

Garvin S. and J. Kagel, 1994, "Learning in common value Auctions: Some Initial Observations", Journal of Economic Behavior and Organization, Vol 25, 351-372.

Gourieroux, C. and A. Monfort, 1995, "Statistics and Econometrics Model", Cambridge University Press.

Hikaru D., 1984, "Observational learning from a radical-behavioristic viewpoint", Behavior Analyst Vol 7(2), Fal 1984, 83-95.

Huck S., Normann H-T and J. Oechssler, 1999, "Learning in Cournot Oligopoly– An Experiment", Economic Journal v109, n454 C80-95.

Houston, J., 1991, Fundamentals of learning and memory (4th ed.), San Diego, CA, USA : Harcourt Brace Jovanovich, Inc.

Kagel J. and J-F. Richard, 1997, “Super-Experienced Bidders in First-Price Common Value Auctions: Rules of Thumb, Nash Equilibrium Bidding and the Winner’s Curse”, Mimeo, University of Pittsburgh.

McKelvey R. and T. Palfrey, 1995, “Quantal Response Equilibria for Normal Form Games”, Games and Economic Behavior, 10, July, 6-38.

Mookherjee D. and B. Sopher, 1994, “Learning Behavior in an Experimental Matching Pennies Game”, Games and Economic Behavior v7, n1 62-91

Offerman T., Potters J. and J. Sonnemans, 2000. ”Imitation and Belief Learning in an Oligopoly Experiment”, working paper, University of Amsterdam.

Schlag, K., 1999, “Which One Should I Imitate?”, Journal of Mathematical Economics v31, n4 493-522

Selten R. and J. Buchta, 1999, “Experimental Sealed Bid First Price Auctions with Directly Observed Bid Functions”, In Games and Human Behavior, Mahwah NJ, L.Erlbaum Associates Inc., 79-102.

Shao, J. and D. Tu, 1995, “The Jackknife and Bootstrap”, New York: Springer-Verlag

Slembeck T., 1998,“A Behavioral Approach to Learning in Economics. Toward an Economic Theory of Contingent Learning”, Mimeo, University of St. Gallen.

Stahl D., 1996, “Boundedly Rational Rules Learning in Guessing Games”, Games and Economic Behavior, 16, 303-330.

Stahl D., 2000, “Rule Learning in Symmetric Normal Form Games: Theory and Evidence”, Games and Economic Behavior, 32, 105-138.

Watson J., 1930, Behaviorism 2nd. ed. Chicago, University of Chicago Press.

Whitehurst G.J., 1978, “Observational Learning”, Hanbook of Applied Behavior Analysis: Social Learning Processes, NY Irvington Publisher, 142-178.

Figure 1
Evolution of the Bid Factor
Sample with Feed Back

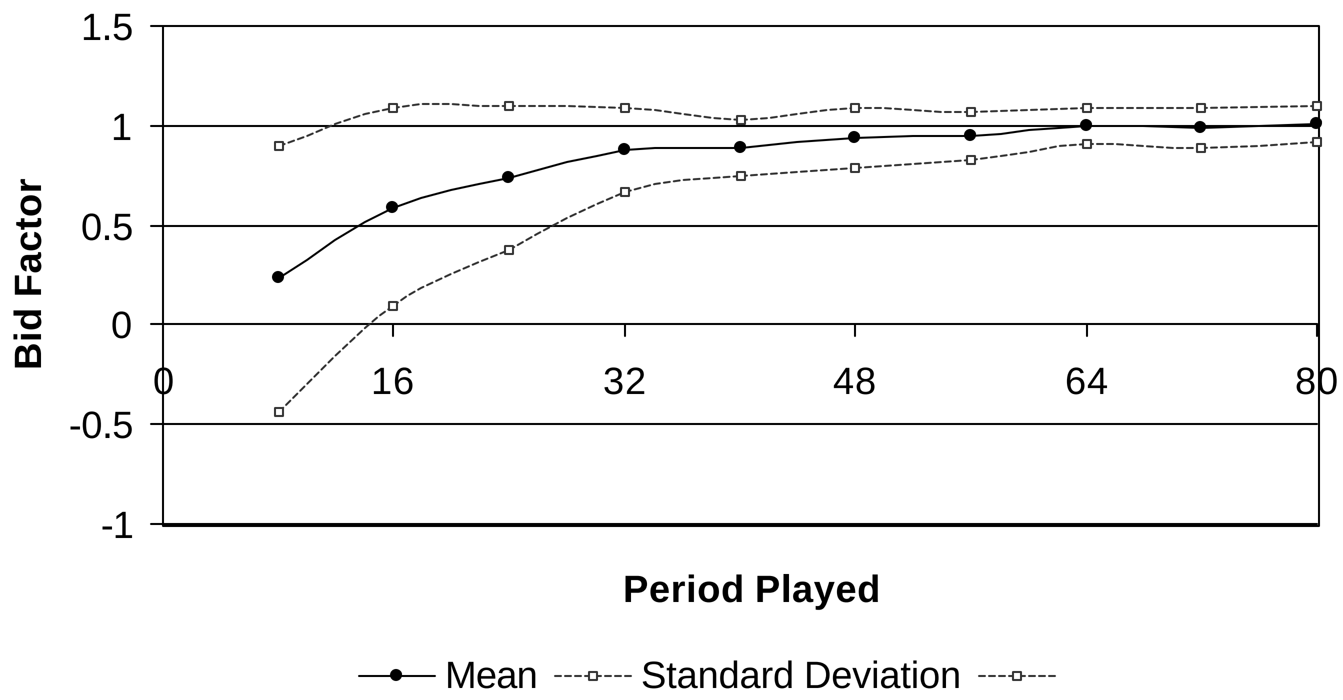
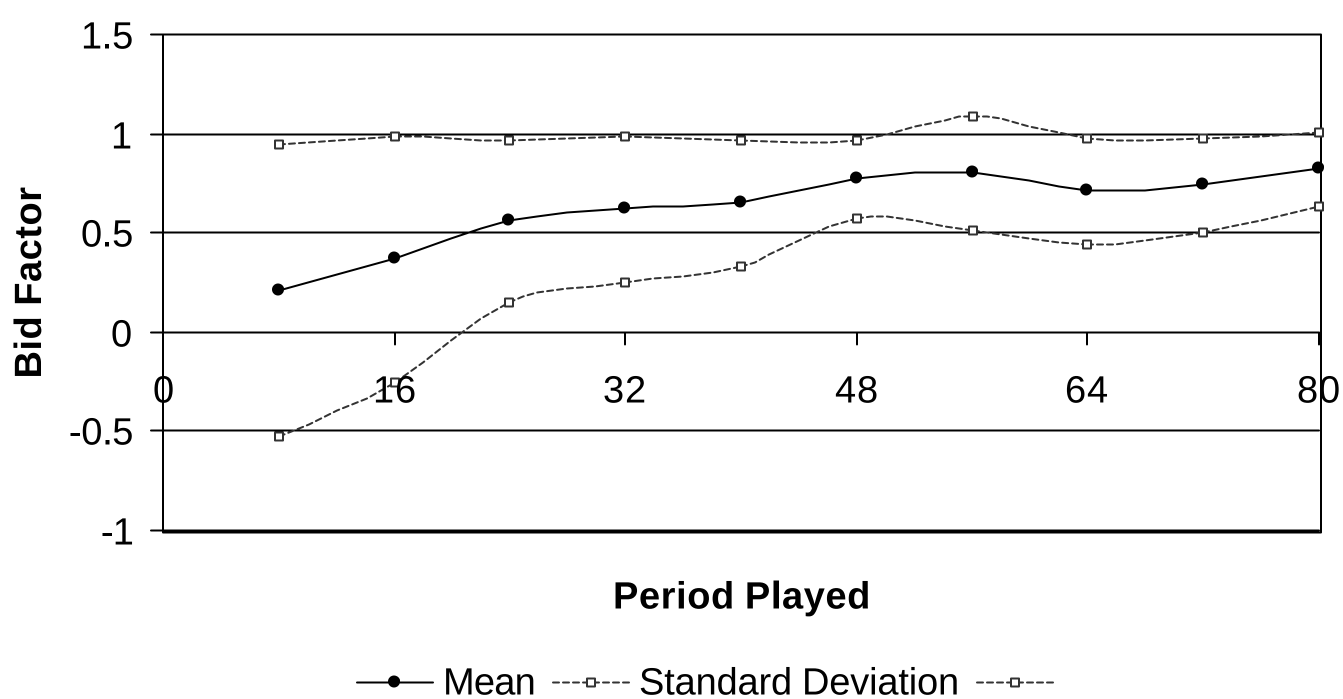


Figure 2
Evolution of the Bid Factor
Sample with No Feed Back



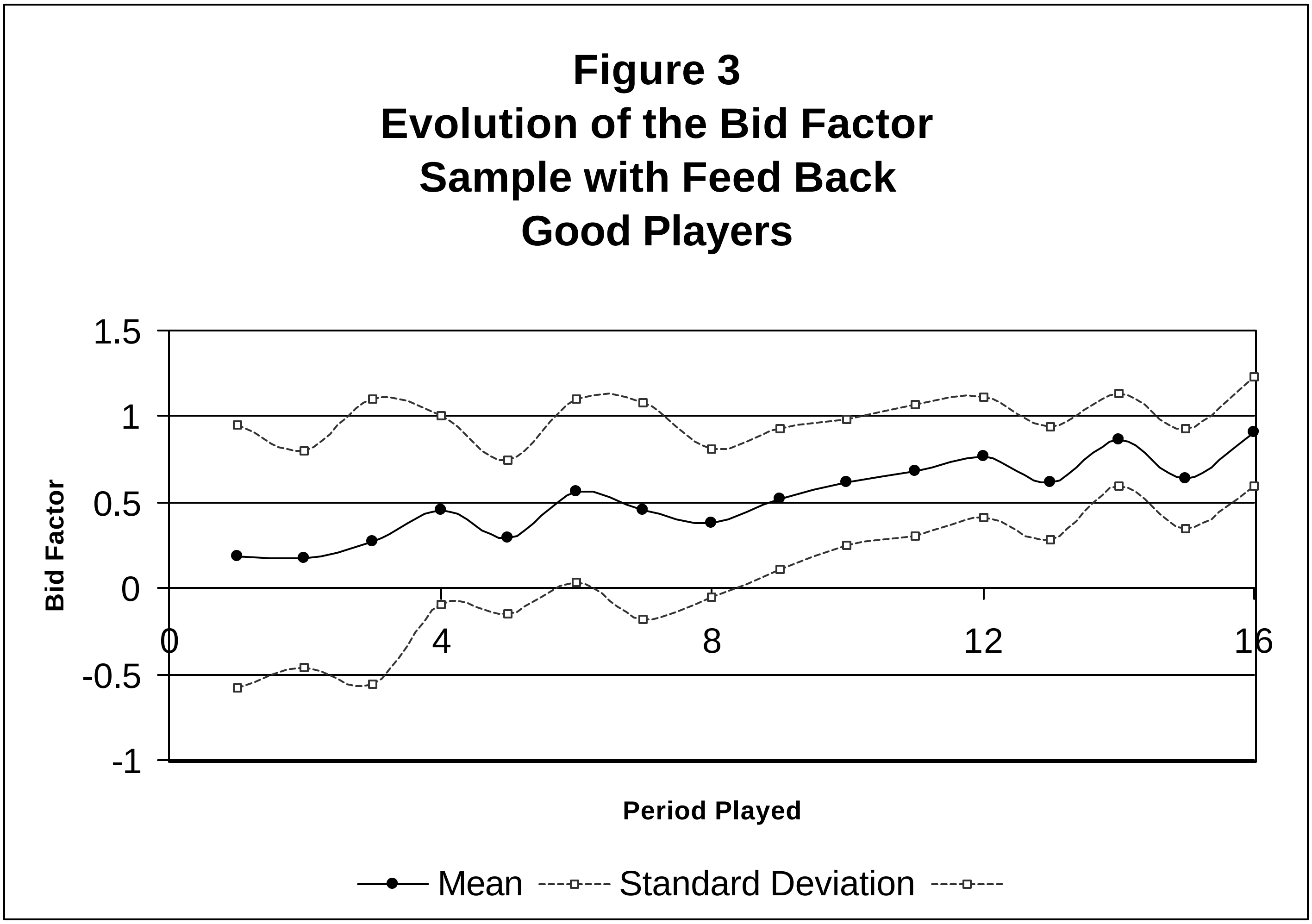


Figure 3
Evolution of the Bid Factor
Sample with Feed Back
Good Players
1.5
1
0.5
0
-0.5
-1
Bid Factor
0
4
8
12
16
Period Played
Mean
Standard Deviation

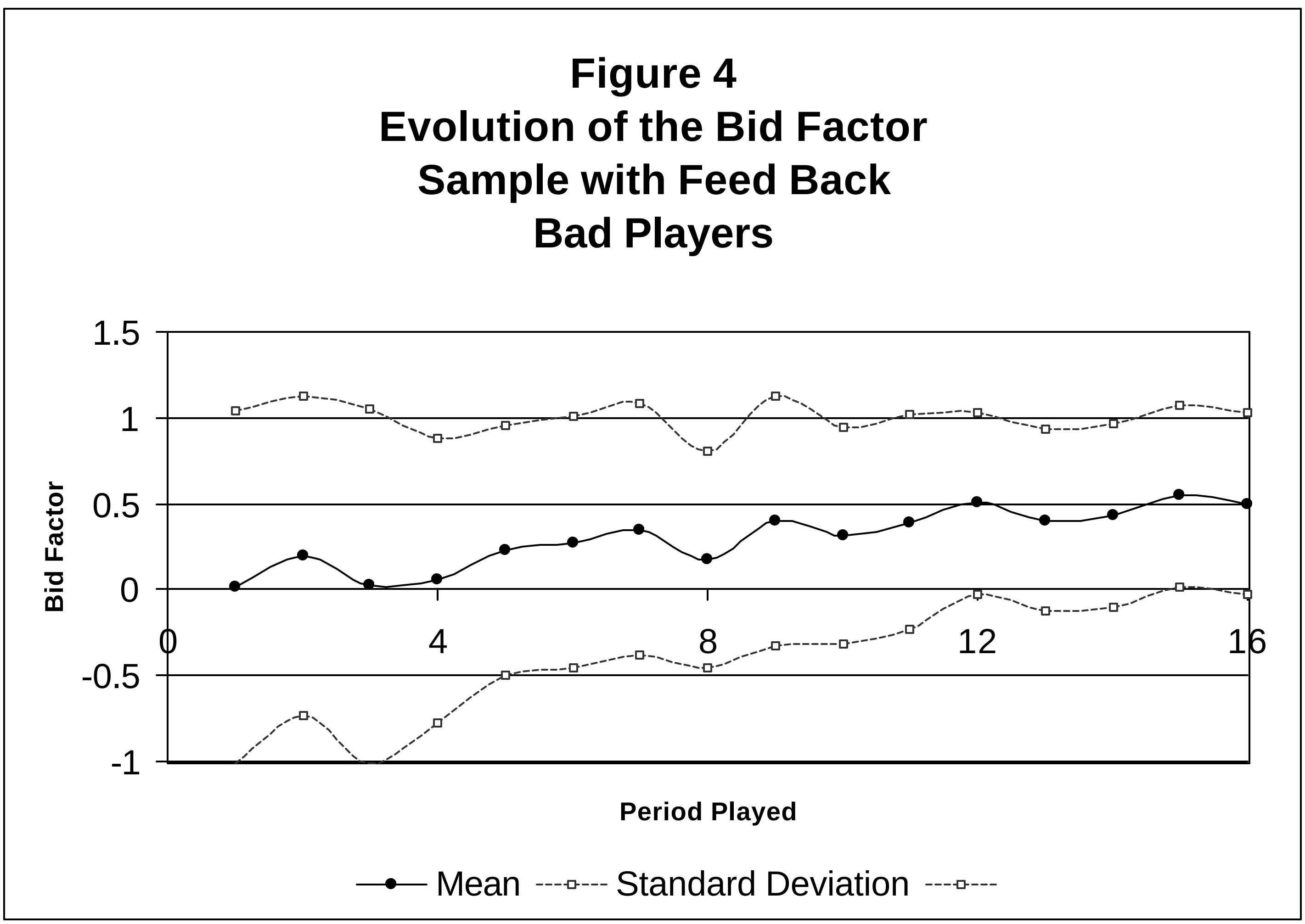


Figure 4
Evolution of the Bid Factor
Sample with Feed Back
Bad Players
1.5
1
0.5
0
-0.5
-1
Bid Factor
0
4
8
12
16
Period Played
Mean
Standard Deviation

Figure 5
Evolution of the Bid Factor
Sample with No Feed Back
Good Players

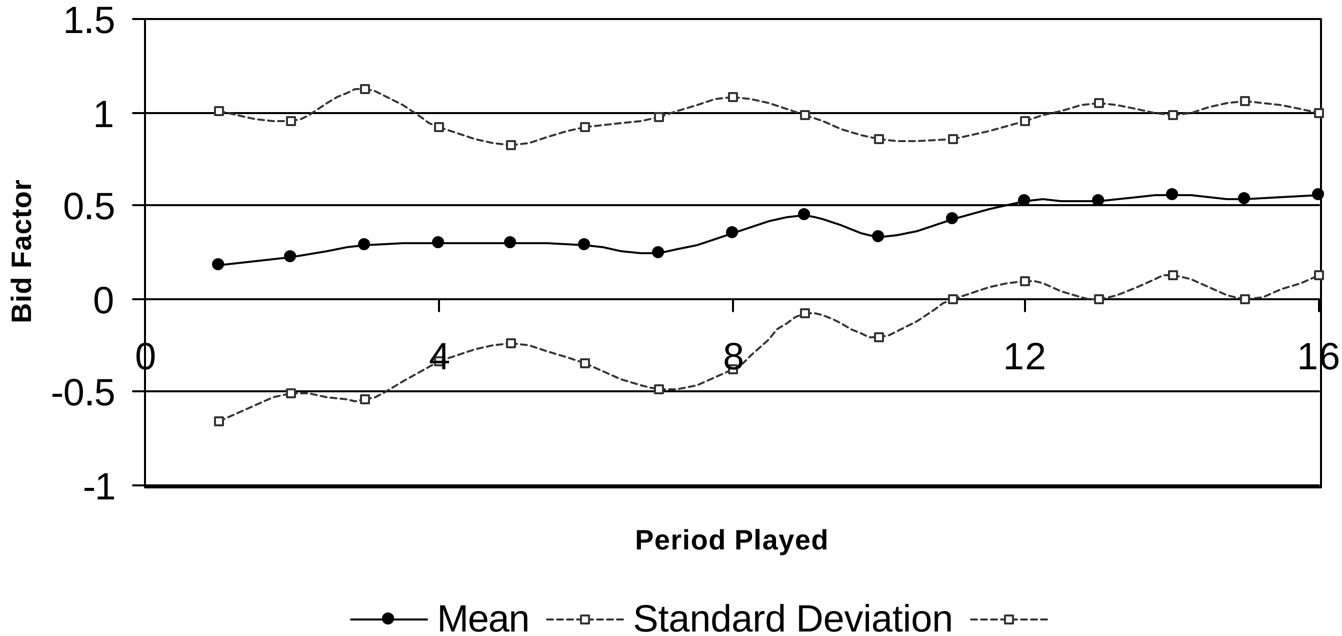


Figure 6
Evolution of the Bid Factor
Sample with No Feed Back
Bad Players

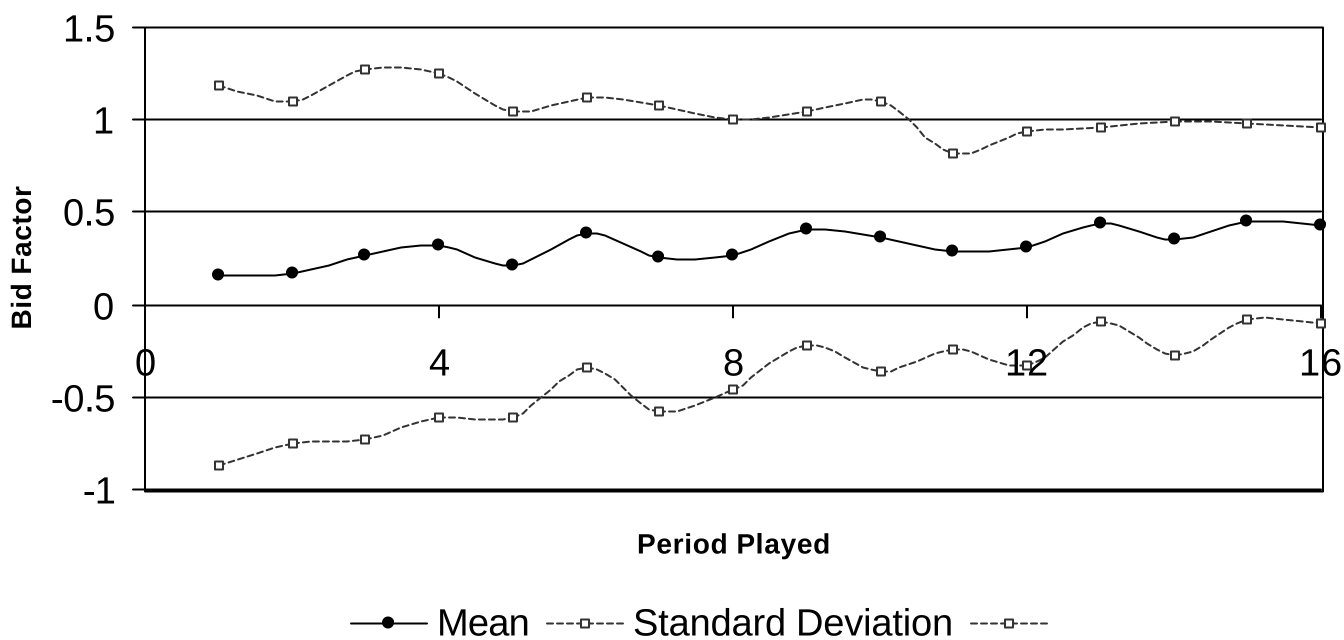


Figure 7
Continuous Reinforcement Model Positive

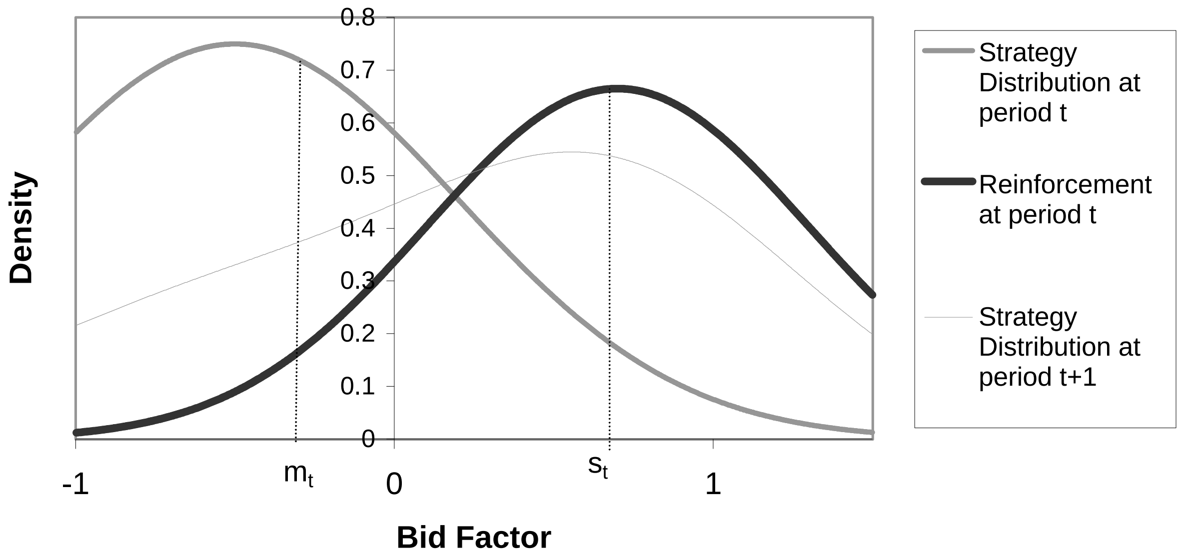


Figure 8
Continuous Reinforcement Model
Positive and Large Payoff

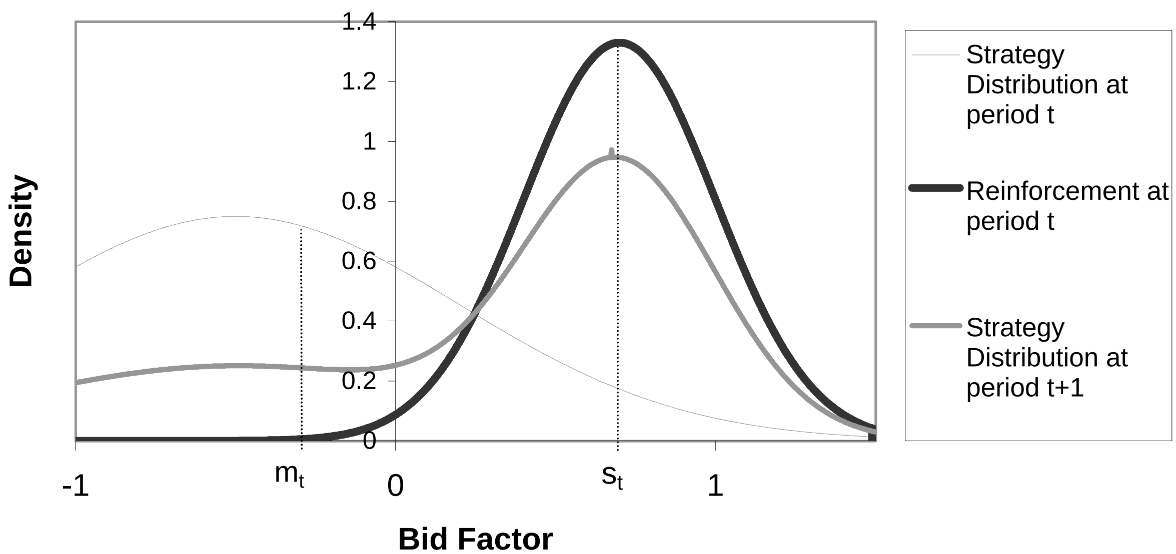


Figure 9
Continuous Reinforcement Model
Negative and Small Payoff

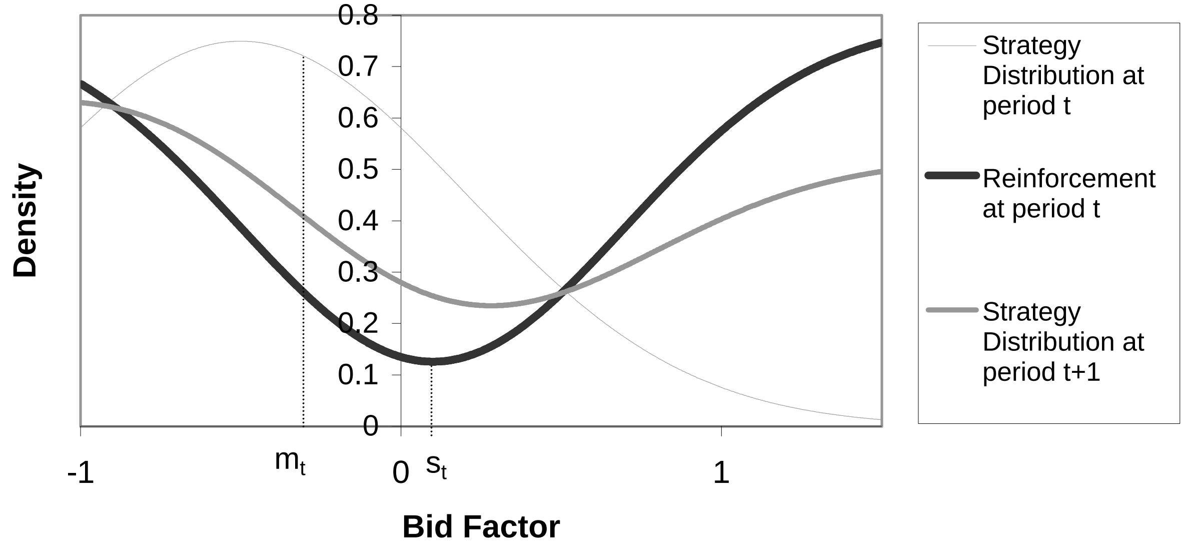
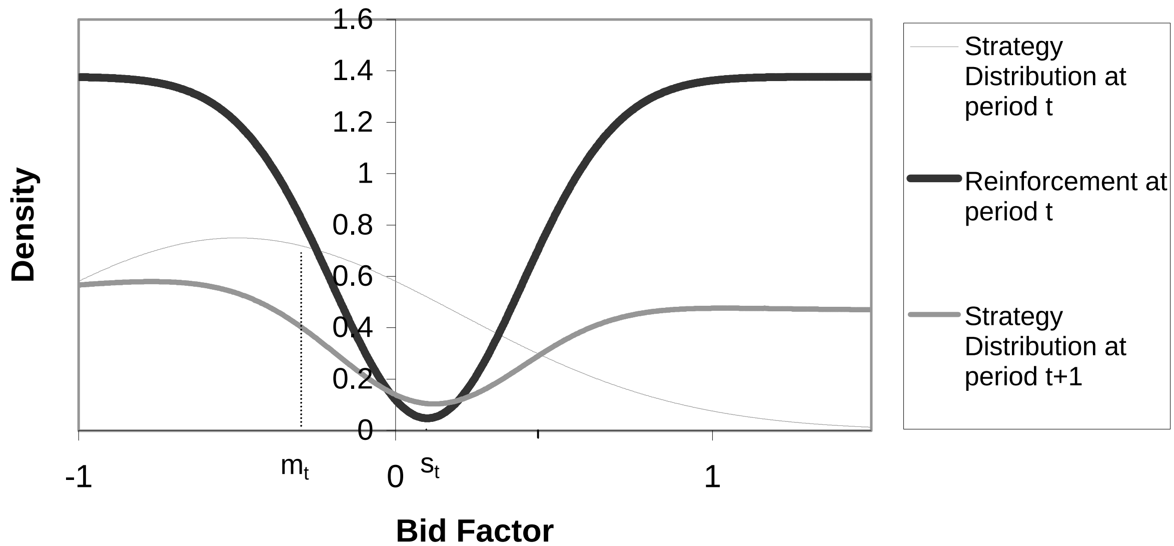
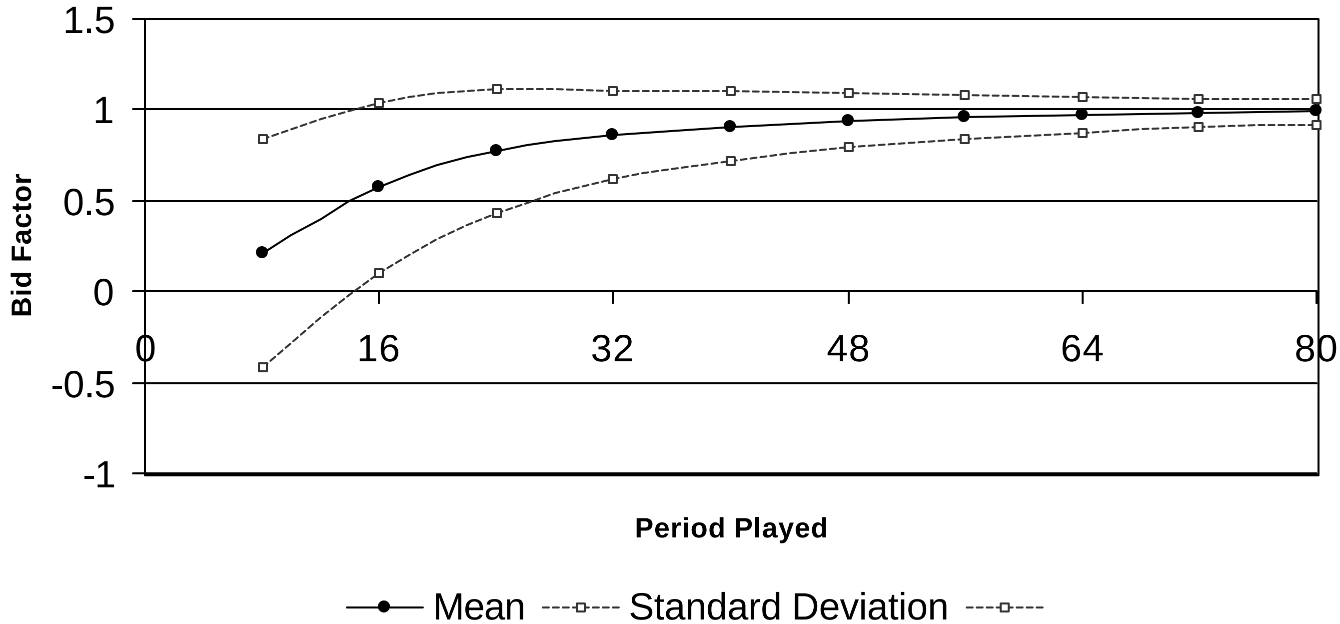


Figure 10
Continuous Reinforcement Model
Negative and Large Payoff



Graph 11
Evolution of the Bid Factor
Monte Carlo Simulations
FB Treatment



Graph 12
Evolution of the Bid Factor
Monte Carlo Simulations
NFB Treatment

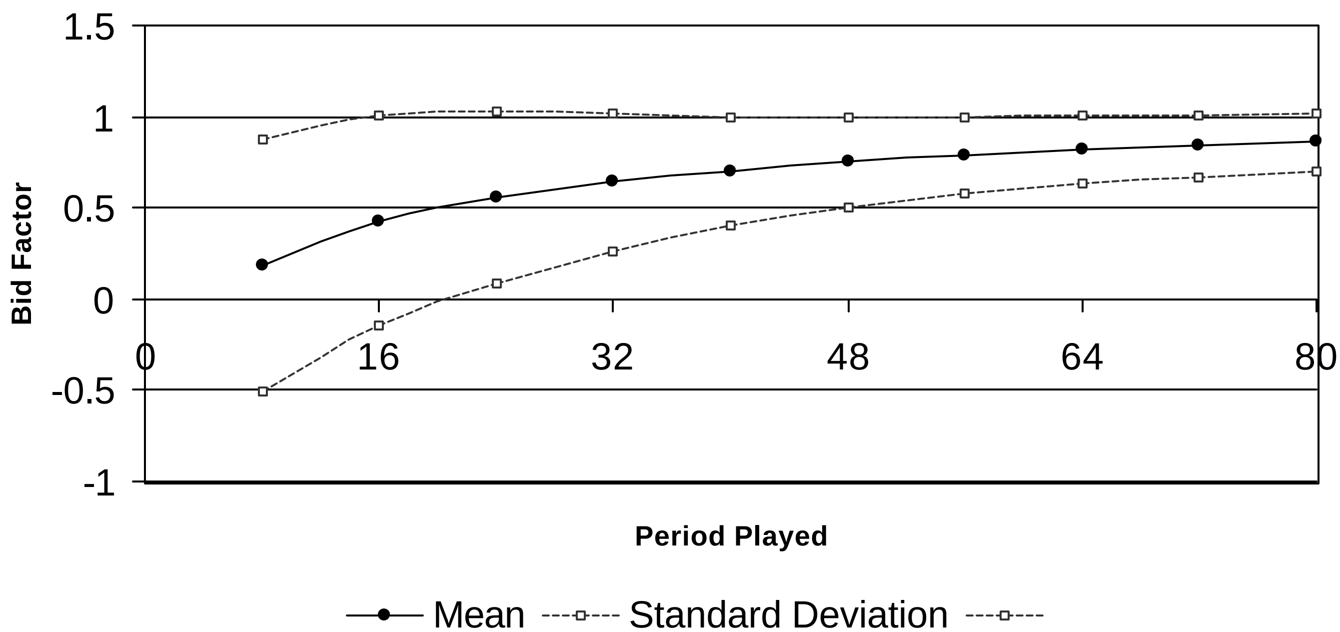


TABLE 1 EXPERIMENTAL OUTCOMES				
	FB Sample	NFB Sample	Mann-Whitney Test Statistic	*p*-Value
Total Number of Periods Played	80	78	.	.
Total Number of Participating Subjects	28	31	.	.
Average Number of Periods Played	68.57 (8.3)	60.38 (12.4)	-3.08	2.64E-3
Number of Bankruptcies	4	7	.	.
Number of Subjects who Participated for the Entire Session	20	18	.	.
Average Profit (Includes the Starting Balance)	54 (12.8)	43 (13.3)	-3.734	1.88E-4
Percentage of Auctions with Negative Profits	18	29	-3.234	1.23E-4
Average Number of Periods Before Bankruptcy*	13 (6.1)	22 (12.5)	-2.079	0.037
Average Profit of an Auction Winner	2.6 (0.7)	1.7 (0.7)	-3.544	3.84E-4
Percentage of Auctions Won by the High Value Bidder	36	28	-2.017	0.043

Numbers in parenthesis refer to standard deviations.
The FB and NFB samples have the same distribution under the null hypothesis in the Mann-Whitney test.
* The average is taken over the participants who experienced bankruptcy.

TABLE 2 THE EVOLUTION OF THE BID FACTOR						
	α	β	σ	γ	$H_0 : \{\beta \le 0\}$	
					Test Statistic	*p*-Value
Feed Back Sample	0.371 (0.241)	1.48E-2* (3.45E-4)	22.025 (17.123)	-1.796* (0.327)	5.262	7.12E-8
No Feed Back Sample	0.413 (0.471)	5.41E-3* (1.26E-3)	5.173 (8.735)	-1.069* (0.421)	4.277	9.48E-6
Good Players in Feed Back Sample	0.281* (0.101)	3.17E-2* (7.31E-4)	1.027* (0.368)	-0.988* (0.217)	3.832	6.36E-5
Bad Players in Feed Back Sample	0.126 (0.212)	1.94E-2* (2.11E-3)	1.785* (0.582)	-0.789* (0.225)	3.124	8.92E-4
Good Players in No Feed Back Sample	0.178* (0.049)	2.32E-2* (4.69E-3)	1.245* (0.514)	-0.800* (0.176)	3.668	1.22E-4
Bad Players in No Feed Back Sample	0.219* (0.096)	1.44E-2* (6.68E-3)	1.628* (0.627)	-0.559* (0.204)	2.127	1.67E-2

Numbers in parenthesis refer to standard deviations. * Indicates parameters significant at a 5% level.

TABLE 3
ESTIMATION OF THE BENCHMARK MODEL

	α_0^{CRL}	α_1^{CRL}	α_2^{CRL}	α_0^{IM}	α_1^{IM}	α_2^{IM}	θ^{IM}	μ	σ
FB Sample	0.928* (0.066)	0.084 (0.187)	0.734* (0.047)	0.902* (0.118)	0.154 (0.212)	0.638* (0.079)	0.831* (0.114)	0.226 (0.134)	3.281* (1.132)
NFB Sample	0.816* (0.097)	0.111 (0.179)	0.826* (0.104)	0.374 (0.478)	0.141 (0.206)	0.431 (0.294)	0.322 (0.216)	0.241 (0.171)	4.065* (1.469)

Numbers in parenthesis refer to standard deviations. * Indicates parameters significant at a 5% level.

TABLE 4
NESTED MODELS COMPARISON

H_0	Wald Test Statistic	Degrees of Freedom	Chi-square statistic ($\alpha = 0.05$)	*p*-value
No Heterogeneity FB Sample	197.776	189	222.074	0.316
No Heterogeneity NFB Sample	228.046	210	244.806	0.187
Experiential and Observational Parameters are Equal	16.245	3	7.814	1.01E-3
CRL	19.355	4	7.814	6.69E-4
Payoff Dependent Imitation	21.437	3	7.814	8.54E-5
Exogenous Learning	46.378	7	14.067	7.38E-8
Nash Equilibrium with Errors	59.719	8	15.507	5.29E-10
Imitation of Successful Behavior	52.957	7	14.067	3.78E-9
Imitation of Successful Behavior with Errors	59.719	7	14.067	5.12E-11
Imitation of Average Behavior	62.662	7	14.067	4.43E-11
Imitation of Average Behavior with Errors	73.084	7	14.067	1.06E-13
Imitation of Mode Behavior	41.084	7	14.067	7.80E-7
Imitation of Mode Behavior with Errors	53.120	7	14.067	3.51E-9
No Reinforcement of Best Strategies	6.083	4	9.488	0.193

TABLE 5
MONTE CARLO SIMULATIONS OUTCOMES

	FB Sample	NFB Sample
Total Number of Periods Played	80	78
Average Number of Participating Subjects	26.8 (3.8)	32.7 (5.7)
Average Number of Periods Played	71.6 (6.1)	57.3 (15.7)
Average Number of Bankruptcies	2.8 (1.6)	8.7 (3.3)
Average Number of Subjects who Participated for the Entire Session	21.2 (2.7)	16.1 (4.0)
Average Profit (Includes the Starting Balance)	59 (9.4)	40.4 (15.6)
Percentage of Auctions with Negative Profits	15.2	28.7
Average Number of Periods Before Bankruptcy	11.2 (4.8)	20.5 (10.5)
Average Profit of an Auction Winner	2.55 (0.53)	1.58 (0.79)
Percentage of Auctions Won by the High Value Bidder	38.4 (8.3)	25.7 (13.2)

Numbers in parenthesis refer to standard deviations