1 Learning

Game theory has focused mainly on equilibrium concepts. Rational players have commonly known identical beliefs in equilibrium and by definition it is a self-enforcing state; once equilibrium is reached no player has incentives to leave the action or probability mixture over actions prescribed by the equilibrium strategy profile. However, equilibrium concepts do not explain how rational players get to have identical beliefs or, in other words, how this self-enforcing state arises.

A traditional explanation of equilibrium is that it results from the analysis, introspection and reasoning by the players in a situation where the rules of the game, the rationality of the players, and the players' payoff functions are all common knowledge. However, as mentioned in Fudenberg & Levine [22], these theories have many problems. First, a conceptual problem occurs when there are multiple equilibria since there is no explanation of how players come to expect the same equilibrium. Second, the hypothesis of exact common knowledge of payoffs and rationality might be arguable in many games. And third, equilibrium theory does a poor job explaining play in early rounds of most experiments, although it does much better job in later rounds.

Learning models on the contrary, explain how people learn, adapt or evolve toward equilibrium. Camerer[5] defines learning as an observed change in behavior owing to experience. Statistical learning models, the object of study in this literary review, predict how probabilities of future choices are affected by historical information.

However, it would be misleading if we describe learning models as models that explain how people learn or evolve toward equilibrium. That statement assumes both, that players always learn the equilibrium and that the learning models converge always to equilibrium, which neither of them is true. On one hand, there are many experiments in which people deviate from the equilibrium predictions, such as the ultimatum mini-game experiments or many of the games in which there is a unique mixed Nash equilibrium. On the other hand, the learning models do not always converge. Different learning models have different convergence and stability properties and often these properties depend on the properties of specific games. For instance, in games with a unique mixed equilibrium reinforcement learning does not converge. However, stochastic fictitious play converges in these games but the steady state to which it converges is not the Nash equilibrium.

It is important to note that the value of learning does not lie only in its ability to explain how people learn or evolve toward equilibrium. We will explain with specific experimental examples three situations in which learning models can have considerable value. The learning models mentioned will be described in the following section.

The first one is the one already mentioned. Suppose there are clear equilibrium predictions in an experiment and that subjects initially are far away from those predictions. However, as time passes by and as subjects play the game repeatedly they learn through experience to play the equilibrium. A nice example of this explanatory value of learning models is given by Cooper, Garvin

& Kagel[10,11]. The experiment is on a signaling game originally proposed by Milgrom & Roberts[28]. There are two types of monopolists, high cost type (MH) and low cost type (ML), who decide an output level and the entrant (E), after observing the output level and not knowing the type of the monopolist, decides to enter or stay out. In the original control experiment, where entrants have high cost, the intuitive equilibrium prediction is the pooling equilibrium where both high cost and low cost monopolists behave the same way. More specifically, MH plays strategically pretending to be ML. The results show that monopolists initially did not play the equilibrium prediction, more specifically both types chose the non strategic output. However, high rates of entrance showed to high cost type monopolists to play strategically avoiding entrance. The authors show that a fictitious play type model, adapted to play against one population can explain this behavior toward equilibrium. Therefore, learning models can explain how subjects learn the way to equilibrium.

A second useful application is when subjects' play converges but the steady state to which it converges is not Nash equilibrium. Learning models bring valuable contribution of how individuals might proceed in their way to the convergent state and therefore, why the equilibrium prediction fails. In games with unique equilibrium in mixed strategies, players usually do not converge to the equilibrium. Erev & Roth[20] collect all the experiments done on this type of games that were played for more than 100 periods. In half of these experiments, players do not show convergence to equilibrium. However, the authors show that reinforcement learning type models can explain fairly well the evolution of the play.

The third situation in which learning models are useful is when there are many equilibrium predictions. Looking at the equilibrium selected by individuals and learning models that predict this equilibrium selection might help to understand important factors that play a role in the selection criteria. One example is given by "continental divide game" carried out by Van Huyck, Battalio & Cook[33]. It is an order-statistic game in which each of seven players chooses integers from 1 to 14. A player's payoff depends on her number and the median number chosen by the players in her group. There are two pure-strategy Nash equilibria, at 3 and 12. Five groups started at a median at 7 or below; all of them flowed toward the low-payoff equilibrium at 3. The other five groups started at 8 or above and flowed to the high-payoff equilibrium. The experiment has three important findings. First, behavior bifurcates from initial choices in the range 4-8 toward the equilibria at 3 and 12, even though players who end up at low numbers earn half as much. Second, the currents of history are strong, creating "extreme sensitivity to initial conditions". And third, convergence is asymmetric; it is much sharper at the equilibrium of 12 than in the neighborhood of 3. As Camerer[5] points out, no concept in analytical game theory gracefully accounts for the fact that some groups flow to 3 and earn less, while others flow to 12 and earn more. Ho, Camerer & Chong[25] show that the Experienced-Weighted Attraction learning model captures the three characteristics fairly well, compared to other learning models.

Going back to the two alternative explanations of how equilibrium arises, it is important to note that learning models have different underlying assumptions from the introspection and careful reasoning explanation. There are two ways in which these alternative models are differentiated from each other. On one hand, the rationality attributed to the players is different. The introspection and careful reasoning assumes that players are rational and therefore they are capable of rational analysis of the game. However, learning models assume different levels of rationality. Learning models such as fictitious play assume rationality in the sense that they best respond with no mistakes to some beliefs. Others such as stochastic fictitious play assume that payoffs are perturbed and therefore, according to the perturbation degree players are more or less sensitive to expected payoffs associated with each strategy. Finally, some learning models are closer to animal learning behavior where no rationality on players is required, they behave according to reinforcement stimulus. On the other hand, the assumption on the information used is very different in the two alternative explanations. In the introspection of rational players' explanation everything is common knowledge from the payoff information to the rationality of the players. In learning models however, the information available and used also differs from model to model. Reinforcement learning assumes that no information of the opponents' play is used. Fictitious play however, explicitly models how beliefs about opponent's future play are built, based on his past actions so opponent's information is needed.

Learning models are behavioral in nature and therefore based on individuals and related to psychological features of individual behavior. Learning models analyzed here can be described by some initial values and two types of rules. A decision rule, which can be deterministic or probabilistic, that describes how actions are taken given the available information; and an updating rule, which can be in terms of beliefs, reinforcements or in general attractions attributed to each of the strategies. In general, we can say that the way attractions are updated is the differential element between learning models. This will be clear later on. Learning models have parameters that try to capture psychological features of individual reasoning such as how much weight to give to past experience, initial tendency to play each strategy, how much to experiment and so on. Apart from a theoretical description of learning models an important part of the literature focuses on applying these models to experimental data in order to explain individuals' behavior and therefore estimation issues arise. This raises the question of how these models should be estimated, at individual level or assuming that there is a representative individual in the population of players in the case of one population or a representative individual in each population of players in the case of two populations. Both analysis can be found in the literature, parameter estimation for each individual in Cheung & Friedman[9] or representative learning model estimation in most of them. Even if they are basically individual learning models, since the degrees of freedom are a concern in the estimation of the experimental data these models are usually estimated assuming representative agent model.

2 Learning models1:The Experienced-Weighted Attraction

2.1 The EWA by Camerer & Ho[7]

Experienced-Weighted Attraction, introduced by Camerer & Ho[7], is the central learning model we present in this literary review. It is a general model that tries to capture and isolate as many psychological features of statistical adaptive behavior. Other learning models that have been used in the literature for many decades can be derived as particular cases of the Experienced-Weighted Attraction model, EWA from now on. We proceed to present the general EWA model and particular cases are derived in the next section. This is done in Camerer & Ho[7] and therefore, we follow this paper very closely in the next two sections.

EWA assumes each strategy has a numerical attraction, which is going to determine, through a decision rule the probability of choosing a strategy. Think of attractions as positive numerical values associated with each of the strategies. As it will be clear later, this specification, that is, to associate each strategy with a numerical attraction is the most general specification, which permits particular interpretations of these numerical attractions such as expected payoffs given some beliefs or just propensities.

To describe EWA model, as any adaptive learning model, requires specifying three elements: initial values, an updating rule and a decision rule. Initial values are treated by the learning literature as black boxes that are unknown and therefore need to be estimated. Updating rules determine how initial values are changed from one period to another by experience and a decision rule defines how attractions in general determine the choice probabilities. Since learning models are individual learning models, EWA will be introduced for player i. Assume player i has N strategies and call s_i^j player i's strategy j, where $j = 1, 2, ...N_i$. Also, call x_i^j the probability that player i plays strategy j, that is, the mixed strategy probability of strategy j.

The main characteristic of this general learning model are two variables that are updated after each round, so two types of initial values are needed and also updating is governed by two rules. The first variable, $N(t)$, is interpreted as the number of observation-equivalent of past experience. The second variable is $A_i^j(t)$, player i's attraction of strategy j, or attraction associated with strategy s_i^j at time t.

Initial Values: the two variables start with some initial values, $N(0)$ and $A_i^j(0)$, that is, experience measure and attraction at $t = 0$. Later we will give a more specific interpretation of all the free parameters of the EWA model but we can say that these initial values reflect pre-game experience, either due to learning transferred from different games or due to introspection.

Updating Rules: both variables are updated each round.

$$(2.1) \qquad N(t) = \rho \cdot N(t-1) + 1 \qquad t \geq 1$$

ρ is the discount factor or depreciation rate of past experience. The second rule updates the level of attraction associated with the strategy of each player. The rule for updating attraction sets $A_i^j(t)$ to be the sum of depreciated experience-weighted previous attraction, plus the weighted payoff from period t normalized by the updated experience weight. A key feature of EWA is that attractions are not updated only when the action is taken but the model weights hypothetical payoffs that non-chosen strategies would have earned by a parameter δ.

$$(2.2) \qquad A_i^j(t) = \frac{\phi \cdot N(t-1) \cdot A_i^j(t-1) + [\delta + (1-\delta) \cdot I(s_i^j, s_i(t))] \cdot \pi_i(s_i^j, s_{-i}(t))}{N(t)} \qquad t \geq 1$$

$I(s_i^j, s_i(t))$ is an indicator function that takes value 1 when $s_i^j = s_i(t)$. Therefore, when a strategy is chosen, its attraction is updated summing the obtained payoff with weight 1. On the contrary, if a strategy is not chosen, its attraction is updated with the hypothetical payoff that this strategy would have yielded weighted by δ. ϕ is the discount factor or decay rate that depreciates past attractions.

Decision rules: they propose three different decision rules based on three choice probabilities: logit, power and probit choice probabilities. Notice that the three proposed decision rules are stochastic, in the sense that, what they predict is not a strategy but the probability of taking a strategy, that is, the mixed strategy probabilities. We will call x_i^j individual i's probability of playing strategy j. In any decision rule attractions determine the choice probability of each strategy or the mixed strategy probabilities.

The logit function describes the decision rule in the following way.

$$(2.3) \qquad x_i^j(t) = \frac{e^{\lambda \cdot A_i^k(t)}}{\sum_{k=1}^{N_i} e^{\lambda \cdot A_k^k(t)}} \qquad t \geq 1 \qquad j = 1, 2, ...N_i$$

The parameter λ measures sensitivity of players to attractions. The logit function has been used in many applications. Based on Harsanyi[24]'s purification theorem, one might consider that payoffs are perturbed, and therefore the best response is not a correspondence any more but a smooth function. Fudenberg & Kreps[23] develop a model of fictitious play along these lines. If we consider a specific form of the perturbation of the payoffs[1], then we get that

[1] Assume individuals maximize the following perturbed payoff function, in which perturbation to player i's payoff depends on the action he chose but not on the actions of the other players:

$u_i(x) + \beta \cdot \phi_i(x_i)$

Furthermore, assume that the perturbation take the form

$\phi_i(x_i) = -\sum_{x_i}(x_i) \cdot \log(x_i)$

Then it can be shown that the best response function has the same form as in (3.3) where $\lambda = 1/\beta$

the smooth best response has the form in (3.3). Quantal Response Equilibrium, introduced by McKelvey & Palfrey[27], has as special case the logistic quantal response where the errors follow a log Weibull distribution. The parameter λ is inversely related to the level of error: $\lambda = 0$ means that actions consist of all error, and $\lambda \to \infty$ means that there is no error. In learning it can have the same interpretation, especially when attractions are interpreted as expected payoffs, which will be explained later on. If $\lambda = 0$ then the actions are taken with equal probability and it does not depend on the attractions and if $\lambda \to \infty$ means that the actions are completely sensitive to attractions.

There are particular cases of this model that use an exact best response function. This decision rule is not stochastic and therefore it predicts which action is taken, that is, the mixed strategy predicted is always degenerate. This can be seen as a special case of the logit function where $\lambda \to \infty$, which will be cleared later on.

The power utility form is given by

$$(2.4) \qquad x_i^j(t) = \frac{(A_i^j(t))^\lambda}{\sum_{k=1}^{N_i}(A_k^j(t))^\lambda} \qquad t \geq 1 \qquad j = 1, 2, ...N_i$$

As they mention in the paper, the decision rule is not what distinguishes the EWA from other learning rules since any choice probability is compatible with EWA. Whether logit, probit or power forms fit better is an empirical question. In their 1998 paper[6], they conclude that for weakest-link type coordination data logit form fits better over the power form.

2.2 Interpretation of free parameters of EWA

First of all, notice that in order to fully specify EWA we need to specify initial values, $N(0)$ and $A_i^j(0)$ for each player and each strategy and the parameters ρ, ϕ, δ and λ if either the logit or the power choice probability is used. We can write EWA as a function of those parameters.

$$(2.5) \qquad EWA(N(0), \textstyle\sum_{j=1}^{N_i} A_i^j(0), \rho, \phi, \delta, \lambda)$$

One of the most important contributions of Camerer & Ho[7] paper is to provide with interpretation of the free parameters, to specify what general behavioral principles of learning these parameters capture.

The number of initial values, initial attractions ($A_i^j(0)$) and initial strength ($N(0)$), depend on the number of players and the number of strategies that each player has.

$N(0)$ is interpreted as the strength of initial attractions, relative to incremental changes in attractions due to actual experience and payoffs. $N(0)$ can be therefore thought of as a pre-game experience weight. The most intuitive interpretation is looking at different values of $N(0)$. For small values of $N(0)$, the effect of the initial attractions is quickly displaced by experience. If on the contrary $N(0)$ is large then the effect of the initial attractions persists. As the authors point out, setting $N(0) < N^* = \frac{1}{(1-\rho)}$ implies that the relative weight on decayed attractions, compared to recent reinforcement, is always increasing, the relative weight on observed payoffs is always declining. This implies a law of declining effect that is widely observed in research of learning.

Initial attractions also need to be specified. These might be derived from an analysis of the game, from surface similarity between strategies and from strategies that were successful in similar games.

As briefly mentioned above, initial values of learning models are treated as black boxes by the learning literature and therefore they are parameters to be estimated. How people play one shot games or how people discriminate among strategies without experience are very interesting questions that learning literature does not ask. Are people able to use simple rules such as never use a strictly dominated strategy or never think that the opponent might use a strictly dominated strategy? This question is answered by Costa-Gomes, Crawford & Broseta[16]. Subjects played 18 different games just one time using a computer interface that allowed subjects to search for hidden payoff information, while recording their searches. The authors could monitor subjects' information searches along with their decisions, which allowed them to better understand how their decisions are determined. They found a simple view of subjects' behavior. Two types, Naïve and L2 comprised 67-89% of the population and a third, D1, between 0-20% of the population. Naïve type assumes that the opponent will play their strategies with equal probability and best respond to that. L2 type assumes the opponent will play as Naïve and best respond accordingly. Finally, D1 is capable of one round of elimination of dominated actions by pure strategies and best responds to a uniform prior over its partner's remaining decisions.

Apart from the initial values, there are other free parameters. The discount factors ϕ and ρ depreciate past attractions and experience separately. This is also something novel. This allows learning models which are determined initially by pre-game experience, that is, very high $N(0)$, and depreciates past experience a lot, so initial persistence of $N(0)$ disappears quickly but which do not depreciate past attractions at all, $\phi = 1$.

The parameter δ measures the relative weight given to foregone payoffs, compared to actual payoffs, in updating attractions. This is the most important parameter of EWA and captures two basic principles of learning: the *law of actual effect* and the *law of simulated effect.* Behavioral psychologists call law of effect when successful chosen strategies are subsequently chosen more often, which has been observed mostly with animal subjects. They re-label this effect law of actual effect to distinguish it from simulated effect. The law of simulated effect, on the other side, is a similar effect but as the name tells, it refers to the effect that non-chosen strategies that would have yielded high payoffs are more likely to be chosen subsequently. Therefore, it is an imaginary or simulated success. EWA, when updating attractions, weights the payoff of the taken action by 1 and weights by δ the payoff when that action was not taken. In other words, EWA captures both law of actual effect and law of simulated effect. δ is a key parameter that differentiates reinforcement from belief-based learning which will be derived in the coming section.

3 Learning models2: derivation of other learning models

EWA is a relatively new learning model. It brought to our attention that two learning models, which had been in the literature for many decades and were thought to be competing models, were actually special cases of the general EWA learning model. Following the contribution of Camerer & Ho[7], we are going to derive as many learning models as possible from the EWA. The two competing learning models are belief-based learning and reinforcement learning, which are special cases of EWA.

3.1 Reinforcement learning models

Reinforcement learning models, also called stimulus-response or rote learning, have their roots in behaviorist psychology, which was popular from about 1920 to 1960. It assumes strategies are reinforced by their previous payoffs. Therefore, strategies that yielded high payoffs in the past will be more probable to be chosen in the future than strategies that yielded low payoffs. It is the simplest learning model we can imagine in the sense that it requires both limited information, only owns payoffs, and no rationality on individuals. In this sense, the main characteristic of reinforcement is that only chosen actions matter or influence future actions, non-chosen actions never get reinforced[2].

After behaviorist introduced reinforcement learning, Bush & Mosteller[4] formalized simple reinforcement rules and applied them to learning in decisions. Cross[17,18] applied reinforcement learning to economic decisions. Finally in the nineties, reinforcement learning was revived, first Arthur[1,2] applied to decision problems and later McAllister[26], Mookerjee & Sopher[47], Roth & Erev[30], Sarin & Vahid[32] applied reinforcement to games. In the presentation of reinforcement learning models we follow closely Roth & Erev[30] for the basic model and Erev & Roth[20] for the extensions and finally Sarin & Vahid[32] for the averaged version.

3.1.1 Cumulative cases: Roth & Erev[30] and Erev & Roth[20]

Roth & Erev[30]'s two-parameter reinforcement learning Roth & Erev[30]'s one-parameter reinforcement learning ,One-RL from now on, is easily seen as a special case of EWA because they have also numerical attractions, called propensities, associated with each of the strategies. The exact equivalence between the two models is stated in the following equation.

[2] With the exception of Erev & Roth's[29] three-parameter learning model, which includes experimentation and therefore strategies that are not chosen can be reinforced. This model is described later on.

$$Two - RL(s1, phi) = EWA(N(0) = 1, A_i^j(0) = s1 \cdot \frac{Xi}{Ni}, \rho = 0, \phi, \delta = 0, \lambda = 1) \tag{3.1}$$

There are two free parameters in this model is $s1$ and ϕ, initial strength parameter, which is assumed to be equal for every player and *recency* parameter.

The interpretation of $s1$ parameter is closely related to the experience-measure in EWA, it measures the sensitivity of propensities to reinforcements or payoffs. If $s1$ is high, then actual payoffs and reinforcements will not influence much the changes in subsequent choices. But if $s1$ is very low, then reinforcements will highly influence subsequent attractions and therefore the choice rule will be very sensitive to reinforcement. Notice that reinforcements are called the actual payoff a chosen strategy gets. Xi refers to average absolute payoff of player i and N_i is the number of strategies of player i. As it can be seen, Roth & Erev[30] assume that all the strategies have equal initial attractions that depend on the initial strength and averaged absolute payoff. If we have two strategies, both strategies will be chosen with the same probability the first period, 1/2 in this specific example, because they will have the same attractions or propensities.

ϕ allows past attractions or reinforcements to be discounted. They call this *recency* effect, recent experience may play a larger role than past experience in determining behavior. They also describe a One-Parameter reinforcement learning model in which ϕ takes value 1 and therefore there is no such effect.

Comparing to EWA, Two-RL sets $N(0) = 1$ and $\rho = 0$, then all the experience measures are equal to 1 every period. The reinforcement learning does not assume two variables in the updating rule but only one, attractions in general or propensities, in reinforcement learning jargon. δ is equal to zero, which means that there is no simulated law of effect, only strategies that have been chosen can be reinforced. This is the main characteristic of reinforcement learning, which assumes very low rationality on individuals. They respond only to chosen actions depending on their success. Finally, ϕ is equal to 1 in the One-RL, and therefore, past attractions are not discounted. Substituting everything in, we get the equivalence between the two models stated in equation (3.1).

There is a small difference that should be noted, the reinforcement is supposed to be positive, never negative. Roth & Erev[30] deal with this detail subtracting the minimum payoff in the game from all the payoffs, that is, payoffs are normalized in order to have minimum reinforcement equal to zero.

Roth & Erev[30], also specified a decision rule, they used power probability choice with $\lambda = 1$.

Erev & Roth[20]'s three-parameter reinforcement learning This specific model has no exact match and equivalence with EWA but the additional parameter introduced somehow relates to the EWA's δ parameter. Erev & Roth[20] introduce *experimentation* into the updating rule of the attractions. Even if most of the weight falls into the actually chosen strategy and its reinforcement, the other strategies are also reinforced by smaller and equal amount.

Therefore, the players experiment with other strategies too and therefore there is some kind of law of simulated effect. However, the law of simulated effect is captured by a different structure. On the contrary of EWA the chosen strategy is not reinforced by the full payoff but by a weighted payoff and the rest of the strategies are also reinforced.

They introduce ε, which is the weight given to non-chosen strategies, and with $(1-\varepsilon)$ the chosen strategy is reinforced. Using the EWA parameters the updating rule is given by

$$\begin{aligned} A_i^j(t) &= \phi \cdot A_i^j(t-1) + (1-\varepsilon) \cdot \pi_i(s_i^j, s_{-i}(t)) \quad if\ s_i^j\ chosen \\ A_i^k(t) &= \phi \cdot A_i^k(t-1) + \frac{\varepsilon}{(N_i-1)} \cdot \pi_i(s_i^k, s_{-i}(t))\ \ if\ s_i^k\ chosen\ k \neq j \end{aligned} \tag{3.3}$$

Therefore, there is not an exact equivalence between the EWA and three-parameter reinforcement learning, Three-RL. However, there are parameters that can be considered equivalent. We can write this equivalence considering δ equal to ε, abusing notation, even if this is not a true equivalence.

$$\begin{aligned} Three - RL(s1, phi, \epsilon) = EWA(N(0) = 1, A_i^j(0) = s1 \cdot \frac{Xi}{Ni}, \phi, \\ \rho = 0, \delta \sim \varepsilon, \lambda = 1) \end{aligned} \tag{3.4}$$

3.1.2 Averaged case: Sarin & Vahid[32]

Sarin & Vahid[32] propose a reinforcement model, in which previous payoffs are averaged rather than cumulated. They do not call attractions or propensities but *subjective assessments* associated with each strategy. There are two main differences from the Roth & Erev[30] reinforcement model. First, when a strategy is chosen its attraction is updated not just summing the payoffs to previous attraction but taking an average between past attraction and the obtained payoffs. The weight given when taking the average is the only parameter of this learning model, what they call it the learning parameter, which is between 0 and 1. Following their interpretation, the attractions are updated by adding a proportion of their surprise, that is, the difference between the observed payoff and their past subjective assessment. In their specification, the payoffs must be normalized to the range between zero and one and they argue that payoffs must have a different interpretation from the Neuman-Morgernsten utilities.

The second difference is the choice rule because they propose a deterministic choice rule. Individuals choose the strategy that has the highest attraction or subjective assessment. This can be seen as having a logit stochastic choice where $\lambda \to \infty$.

This is also a special case of EWA where we have the free parameter ϕ, their unique learning parameter, which is the weight given to the surprise effect.

$$SV - RL(phi, A_i^j(0)) = EWA(N(0) = \frac{1}{(1-\rho)}, \rho = \phi, \delta = 0, \tag{3.5}$$

$\lambda \to \infty)$

3.2 Belief-based learning models

Belief-based learning takes a different approach from reinforcement learning. Individuals are assumed to be rational, they are able to maximize and best respond, however, belief-based learning explicitly models how beliefs are formed and updated. According to these models, beliefs are formed by observing the history or past behavior of opponents. In deciding how to weight the past there are two extreme cases, Cournot and fictitious play. Belief learning dates back to Cournot[15] who suggested players choose a best response to observed behavior in the previous period. Instead of best responding to the most recent past action, fictitious play assumes that beliefs are formed as an average of all observed past behavior of the opponent. Theories of fictitious play were proposed by Brown[3] and Robinson[29]. These theories were initially proposed to compute Nash equilibrium algorithmically, providing a story about how mental simulation could lead to immediate equilibration in a kind of cognitive tatonnement. In the 1980s, Fudenberg & Kreps[23] reinterpreted it as a theory of how players might learn from periods of actual behavior.

There are three ways belief-based learning models can be defined: in terms of weights associated with each of the strategies, following Fudenberg & Kreps[23] and Fudenberg & Levine[22], in terms of beliefs, as Young[35] does, or in terms of expected payoffs, introduced by Shapley[31]. We will briefly mention the state variable and updating rule of the three different ways of presenting the special case of fictitious play belief-based learning before starting to show the equivalence between these models and EWA.

Fudenberg & Kreps[23] and Fudenberg & Levine[22] presented belief-based learning in terms of weights associated with opponents' actions and those weights are used to form beliefs, or in other words, the updated state variable are the weights. Call $w^k_{-i}(t)$ the weight associated with the player i's opponent's strategy profile k. A simpler interpretation is obtained when player i has one opponent with two strategies, then $k = 1, 2$. k takes a more complex form when player i's actions depend on more than one individual's actions, in which case k refers to all possible strategy combinations of opponent players. Just for an example, if player i has two opponents and each has 2 strategies, then k can take 4 different values $k = \{(1,1),(1,2),(2,1),(2,2)\}$. Therefore, the updating rule for weights given some initial weights and beliefs are formed as described in (3.6) and (3.7) It is easy to see that beliefs can be written in terms of past beliefs[3].

$$(3.6) \qquad w^k_{-i}(t) = w^k_{-i}(t-1) + I(s^k_{-i}, s_{-i}(t)) \qquad t \geq 1$$

[3] Write currents weights in terms of past weights and divide both sides by $\sum_{h=1}^{M-i} w^h_{-i}(t-1)$. The algebra is done in equation (4.12).

$$(3.7)\qquad B^k_{-i}(t) = \frac{w^k_{-i}(t)}{\sum_{h=1}^{M-i} w^h_{-i}(t)} = \frac{\sum_{h=1}^{M-i} w^h_{-i}(t-1)\cdot B^k_{-i}(t-1)+I(s^k_{-i},s_{-i}(t))}{\sum_{h=1}^{M-i} w^h_{-i}(t-1)+1}$$

$I(s^k_{-i}, s_{-i}(t))$ is the indicator function that takes value 1 when the strategy profile of opponents is equal to s^k_{-i}. So, only weights of those strategy combinations chosen by the opponents are changed from one period to another. Then beliefs are formed as a ratio between the weight associated with a profile k over the sum of weights of all possible opponents' strategy combinations. Notice that $B^k_{-i}(t)$ refers to individual i's belief about the probability of occurring strategy combination k.

A second way is mentioned by Young[35], where the beliefs are the state variable. This way requires initial beliefs to be defined, that is, in our notation $B^k_{-i}(1)$ must be defined for all $k = 1, 2, ...M_{-i}$.

$$(3.8)\qquad B^k_{-i}(t) = \frac{(t-1)\cdot B^k_{-i}(t-1) + I(s^k_{-i}, s_{-i}(t))}{t} \qquad t \geq 2$$

This way for estimating belief-based learning models has been used the most. Initial beliefs usually are taken as the first observed action of the opponent. This does not require to estimate initial weights associated with the strategies for every player, which saves a lot of degrees of freedom. The drawback is that this way does not allow as much flexibility as specifying initial weights.

A third way was introduced by Shapley[31], where the state variable are expected payoffs. These expected payoffs are updated with the payoff that would have obtained if they played this strategy. The advantage is that we can write all the updated expected payoffs for each of the strategies $j = 1, 2, ...N_i$ in a vector. Call the vector of initial expected payoffs for each strategy for player i, $E_i(0)$, then the updating rule is a function of opponents' action, $s_{-i}(t)$.

$$(3.9)\qquad E_i(t) = E_i(t-1) + (\pi_i(s^1_i, s_{-i}(t)), ..., \pi_i(s^{N_i}_i, s_{-i}(t)))' \qquad t \geq 1$$

The relation between the three ways will become transparent when we take the steps in order to show the equivalence between belief-based learning models and EWA.

3.2.1 Deterministic choice rule: fictitious play and Cournot model

Weighted-fictitious play is the general case that has as special cases the Cournot and the fictitious play mentioned above. The derivation of the weighted-fictitious play from EWA requires few more steps than the derivation of reinforcement learning.

We start by defining beliefs for every strategy combination that a player might encounter, given by s^k_{-i} , and define these beliefs as the ratio of hypothetical counts of observations of a strategy combination, denoted $N^k_{-i}(0)$. Then define the sum of all those $N^k_{-i}(0)$ for all the possible strategy combination that a player might see, $k = 1, 2, ..., M_{-i}$, where $M_{-i} : \Pi^P_{p=1} S_p$, P being the number of players and S_p the strategy space of player $p \neq i$. Notice that $N^k_{-i}(0)$ has exactly the same interpretation as the weights, $w^k_{-i}(0)$ associated with each possible strategy combination of opponents' play in Fudenberg & Levine[22]. In exactly the same way, we can define the initial beliefs:

$$B^k_{-i}(0) = \frac{N^k_{-i}(0)}{\sum_{k=1}^{M_{-i}} N^k_{-i}(0)} = \frac{N^k_{-i}(0)}{N_{-i}(0)} \quad \textit{with } N_{-i}(0), N^k_{-i}(0) \geq 0 \tag{3.10}$$

Then beliefs are updated by depreciating the previous counts by ρ and adding one for the strategy combination actually chosen by the other players. We can further simplify the denominator.

$$\begin{aligned} B^k_{-i}(t) &= \frac{\rho \cdot N^k_{-i}(t-1) + I(s^k_{-i}, s_{-i}(t))}{\sum_{h=1}^{M_{-i}} \left[\rho \cdot N^h_{-i}(t-1) + I(s^h_{-i}, s_{-i}(t))\right]} \\ &= \frac{\rho \cdot N^k_{-i}(t-1) + I(s^k_{-i}, s_{-i}(t))}{\rho \cdot N_{-i}(t-1) + 1} \qquad t \geq 1 \end{aligned} \tag{3.11}$$

Beliefs can be expressed in terms of previous-period beliefs. Divide both sides by $N_{-i}(t-1)$, simplifying the expression we get beliefs in terms of past beliefs.

$$\begin{aligned} B^k_{-i}(t) &= \frac{\dfrac{\rho \cdot N^k_{-i}(t-1) + I(s^k_{-i}, s_{-i}(t))}{N_{-i}(t-1)}}{\dfrac{\rho \cdot N_{-i}(t-1) + 1}{N_{-i}(t-1)}} \\ &= \frac{\rho \cdot N_{-i}(t-1) \cdot B^k_{-i}(t-1) + I(s^k_{-i}, s_{-i}(t))}{\rho \cdot N_{-i}(t-1) + 1} \qquad t \geq 1 \end{aligned} \tag{3.12}$$

This equation is exactly the same as the one presented in Fudenberg & Levine[22], where $N^k_{-i}(t-1)$s are the $w^k_{-i}(t)$s, and it is easy to see the two particular cases: Cournot when $\rho = 0$ and fictitious play when $\rho = 1$, for $0 < \rho < 1$ it is the weighted fictitious play. However, in order to show the equivalence between weighted-fictious play and EWA we need to go one step further and write the expected payoffs according the specified beliefs.

$$E^j_i(t) = \textstyle\sum_{k=1}^{M_{-i}} \pi_i(s^j_i, s_{-i}(t)) \cdot B^k_{-i}(t) \qquad t \geq 1 \tag{3.13}$$

Finally, we can express expected payoffs in terms of previous expected payoffs. Substitute the expression for beliefs and identify the previous period expected payoffs. Simplifying the expression we obtain the following expression.

$$E^j_i(t) = \frac{\rho \cdot N(t-1) \cdot E^j_i(t-1) + \pi_i(s^j_i, s_{-i}(t))}{\rho \cdot N(t-1) + 1} \qquad t \geq 1 \tag{3.14}$$

This equation makes clear the equivalence between EWA and weighted-fictitious play. Suppose initial attractions are equal to expected payoffs given initial beliefs that arise from the 'experience-equivalent' strategy counts $N^k_{-i}(0)$, then substitute $\delta = 1$ and $\rho = \phi$. This leads to attractions that are exactly the same as expected payoffs.

Weighted-fictitious play assumes that players choose the strategy that gives the highest expected payoff. Therefore, the actual choice of actions is not stochastic but deterministic. These deterministic models predict the strategy taken in each period, $X_i(t)$ and not the mixed strategy. To distinguish from the probability of taking a specific strategy we refer to this with a capital letter. This means that the decision rule is not given by a choice probability but by a discontinuous maximum function. However, this can be approximated by a probability choice model, assume we have the logit and that $\lambda \to \infty$. This would represent the max choice.

$$X_i(t) = max(A^1_i(t), A^2_i(t), \ldots, A^{N_i}_i(t)) = max(E^1_i(t), \ldots, E^{N_i}_i(t)) \tag{3.15}$$

.

The key assumption of belief-based learning models is that the initial attractions must be defined in terms of expected payoffs associated with prior beliefs, whereas these prior beliefs are formed as a ratio of initial experience-measures. Furthermore, looking at equation (3.14), notice that the simulated effect is fully in charge in the updating rule, that is, players completely ignore the payoff of the actually chosen strategy. In belief-based learning, when updating beliefs we do not need to look at what strategy we chose but updating is done looking at given what the opponent chose and to what we could have earned with all our possible strategies and choose the one which gives the highest expected payoff.

Notice that to show the equivalence between EWA and belief-based the easiest vehicle is the third way of describing fictitious play. The three ways are equivalent in a general sense, since all of them take past observed behavior of the opponent to update beliefs. Moreover, each period we can find equivalence expressions for each of the three ways. However, notice that the first way updates weights, the second way beliefs and the third way expected payoffs.

To finish, notice that belief-based learning models assumes a specific model of the opponent behavior. More explicitly, it assumes that the opponent actions are coming from a fixed distribution. The fictitious player tries to guess this distribution using past behavior and play optimally against it. As it will be pointed out in the following section, all the learning models presented in this literary review share this assumption. We can summarize the equivalence between belief-based learning models and EWA through these equations, using FP for fictitious play and WFP for weighted fictitious play.

$$FP(N^k_{-i}(0)) = EWA(N^k_{-i}(0), A^j_i(0) = E^j_i(0), \rho = \phi = 1, \delta = 1, \lambda \to \infty) \tag{3.16}$$

$$Cournot(N^k_{-i}(0)) = EWA(N^k_{-i}(0), A^j_i(0) = E^j_i(0), \rho = \phi = 0, \delta = 1, \lambda \to \infty) \tag{3.17}$$

$$WFP(N_{-i}^{k}(0), \rho) = EWA(N_{-i}^{k}(0), A_{i}^{j}(0) = E_{i}^{j}(0), \rho = \phi, \delta = 1, \lambda \to \infty) \tag{3.18}$$

3.2.2 Stochastic choice rule: stochastic weighted fictitious play

The derivation is exactly the same except for the choice probability. These models assume explicitly that the decision rule is stochastic and use normally the logit choice probability[4]. Therefore, we have one more parameter. But the rest is exactly the same. Again, we can summarize the equivalence relation between these models in the following equations, the S for stochastic.

$$SFP(N_{-i}^{k}(0), \lambda) = EWA(N_{-i}^{k}(0), A_{i}^{j}(0) = E_{i}^{j}(0), \rho = \phi = 1, \delta = 1, \lambda) \tag{3.19}$$

$$SCournot(N_{-i}^{k}(0), \lambda) = EWA(N_{-i}^{k}(0), A_{i}^{j}(0) = E_{i}^{j}(0), \rho = \phi = 0, \delta = 1, \lambda) \tag{3.20}$$

$$SWFP(N_{-i}^{k}(0), \rho, \lambda) = EWA(N_{-i}^{k}(0), A_{i}^{j}(0) = E_{i}^{j}(0), \rho = \phi, \delta = 1, \lambda) \tag{3.21}$$

4 Learning, "strategic teaching" and sophistication

In this section we point out that all the learning models described in the previous section hold a common assumption, which is a specific model of opponent's play. More specifically, those models assume that opponent's play is coming from a fixed and unknown distribution. Moreover, we claim that because of the specific assumption on opponent's play, all these models belong to the same class in the learning literature. This class is called the statistical learning models. In this sense, learning must be understood as a broader phenomenon that allows having different models of opponent's behavior. We further analyze when this assumption might be sensible both looking at experimental treatments and how these models fit the experimental data. This is not a new issue. Fudenberg & Kreps[23] mention explicit conditions under which this assumption might make sense. Reversing the point of view we also ask when it is not sensible to have such a simple model of opponent's play and we provide theoretical and experimental examples in which the statistical learning approach cannot successfully explain individual behavior because the assumption of opponent's model seems to be different from the one assumed by statistical learning models.

[4] Cheung & Friedman[12] uses a weighted fictitious play type learning model where the choice probability is the probit.

EWA learning model and its variants assume that opponent's play is coming from a fixed and unknown distribution. The objective of these learning models is to learn about this distribution and accordingly to learn to play optimally. More specifically, individuals learn about this distribution and optimal play either through their own past actions as reinforcement describes, through counting opponent's past actions and best responding to them or through a mixture of both ways according to a EWA model. This assumption about opponent's play has a clear implication: individuals do not try to influence the future play of their opponents. This might be shocking in a game-theoretical setting since statistical learning ignores the existence of strategic interaction, that is, not only my payoff depends on the action of other people but my action also enters on the payoff function of other players. This was pointed out by Fudenberg & Kreps[23] referring to fictitious play, which they called myopic behavior, meaning individuals maximize immediate expected payoffs given some beliefs. This can be generalized to all the models described here, since in the description of the model there is not any strategic analysis of the game except for the initial attractions or initial experience measures. Remember, however, that in the learning literature these free parameters are treated as black boxes which are unknown and therefore must be estimated.

After a quick assessment one might argue that statistical learning assumes individuals are passive learners, where the only learning source is coming from experience. In the same line, the argument says that in these learning models individuals never take an active role in playing strategically, making models of other's behavior and optimizing accordingly. The latter is called sometimes sophisticated behavior and it is differentiated from learning. We argue that this view is a misunderstood view of learning and therefore, the separation between learning behavior and sophisticated behavior is also mistaken, since both must be understood under the same phenomenon of learning. In order to reconcile the above statements we must understand learning as a description of behavior in a repeated setting that assumes a specific model of opponent's behavior. When a learning model describes individuals' behavior it assumes a specific model of opponent's play. The way to reconcile the statements above is to point out one more time the specific assumption that statistical learning models have: opponent's play is coming from a fixed and unknown distribution. Now, given this assumption two things can be underlined. First, given this assumption it is sensible to assume that individuals cannot influence other's actions. Given this fixed and unknown distribution, there is nothing to be influenced. Second, individuals are not passive learners but given a fixed and unknown distribution the only way of learning is through past experience, either own or opponent's past behavior or a mixture of the two.

We already mentioned that it can be troubling to have such a simple model of opponent's play in a game-theoretical setting. Taking one step further, a natural question arises: **when is it sensible to have this simple model of opponents' behavior?** In other words, **when can we expect that individuals do not try to influence other players' actions?** Fudenberg & Kreps[23] actually explain when this myopia might be justifiable. The first two reasons are based on individuals having large discount factors and individuals having beliefs that current action will have little effect in the future. As both are unsatisfactory they turn to reasons that have to do with the environment. They propose a situation in which there are a large number of players[5] who interact in small groups. They also propose three matching schemes that would justify this myopic behavior.

Story 1: at date t, one group of players is selected to play the game. They do so, and their actions are revealed to all the potential players. Those who play at date t are then returned to the pool of potential players and another group is chosen at random for date $t+1$.

Story 2: at each date t, there is a random matching of all the players, so that each player is assigned to a group with whom the game is played. At the end of the period, it is reported to all how the entire population played. The play of any particular player is never revealed.

Story 3: at each date t there is a random matching of the players, and each group plays the game. Each player recalls at date t what happened in the previous encounters in which he has involved, without knowing anything about the identity or experiences of his current rivals.

They argue that myopic behavior seems sensible under this environment. It is not a surprise that most experimental settings in which learning models described above are applied use mean matching or random matching schemes. Mean matching scheme, which has the same informational implications as *Story 2*, and random matching, which is exactly *Story 3*. Therefore, there are two key elements that make myopic behavior or simplistic model of opponents sensible: many players and anonymous random opponents. Notice that when an individual meets different and anonymous opponents along the play she can only keep track of observed opponent's actions, if given, and her own payoffs. She cannot have beliefs associated with each of the different opponent since their identity is never revealed. In this case, individuals model opponents' actions as coming from one individual. But the actions are certainly coming from different individuals which makes sensible to have the idea of a distribution. Also, notice that even if one individual might realize that her action actually influences other individuals' action, since she is only one player that others will encounter then she realizes that her influence is very limited depending on the number of players. Therefore, it can be argued that in such environments it might be sensible to assume a simple model of opponent's behavior and therefore, learning behavior in these environments can be consistent with statistical learning.

Notice however, that the lack of strategic sophistication characteristic of these models does not mean that individuals do not end up playing in a strategic way. A good example is Cooper, Garvin & Kagel[10,11]'s signaling game experiment mentioned in section 2. There are two players, monopolists and entrants, monopolist can be either high cost type (MH) or low cost type (ML) and choose after knowing their type a quantity. Entrants (E) decide to enter or stay

[5] In their example they talk about 5000 players 1 and 5000 players 2. Experiments suggest that a much smaller number of players might have the same effect.

out. In a specific setting where Es have high cost, there is one intuitive equilibrium prediction, which is a pooling equilibrium, where both MH and ML choose the same quantity and the entrants do not enter. This requires the MH to play strategically and imitate ML in order to pool and keep the entrant away. The monopolists start playing no strategically, ignoring the possibility of choosing the quantity in a way that does not reveal their type in order to avoid entrance. However, entrants learn that it is optimal to enter and as monopolists see high rates of entrance move to the strategic choice of actions where they keep the entrants out. They argue that an adaptive learning model based on fictitious play does a good job explaining this behavior. This example shows that through statistical learning individuals can learn strategic behavior even if this learning assumes individuals do not try to influence other's actions.

After seeing the conditions that make more sensible the simple model of opponent's play, the same question we asked but in negative is also an interesting question, even more interesting than the original one. That is, **when is it non-sensible to have such a simple model of opponent's play? When do subjects show a more sophisticated model of opponents?** As the questions point out two answers can be given. On one hand, there are theoretical analyses of such a sophisticated model of opponent's. We present Fudenberg & Levine[21]'s example that shows the advantage of having a more sophisticated model of opponent and Ellison[19]'s theoretical analysis of including a rational or sophisticated player in a population of statistical learners. On the other hand, there are experiments in which statistical learning models do not a good job because players show a more sophisticated behavior. Individuals seem to have more sophisticated models of other individuals and therefore, do not play as they are playing against an opponent who is playing according to a fixed distribution but as for example if these individuals were learning through a fictitious play. An interesting question is to see when a more sophisticated behavior becomes relevant and which environments or factors enhance a more sophisticated model of others. There are some answers already. We present briefly two experimental examples: Camerer, Ho & Chong[8]'s strategic teaching and Cooper & Kagel[12]'s cross-learning examples.

We start by Fudenberg & Levine[21]'s example. As they pointed out, in the case of small population a player may attempt to manipulate his opponent's learning process and try to "teach" him how to play the game. Suppose there are two players playing against each other and player2 is a myopic learner who is best responding to some beliefs about the other player's strategies. Then player1, rather than being a statistical learner, he should play as a Stackelberg leader and teach player1 how to play against him. The example uses a game represented by the following payoff matrix.

Figure4.1: Payoff matrix in Fudenberg & Levine[21]

		Column	
		1	2
Row	1	1,0	3,2
	2	2,1	4,0

Assume further that both players are learning through fictitious play. Player1 will play 2 since 2 is strictly dominating. Player2 observing this will end up playing 1, which is the unique equilibrium of this game. However, if Player1 guesses that player2 is learning through fictitious play, she can do better. She can manipulate player2's learning process choosing 1 and therefore player2 will end up doing action 2. Notice that both players end up in a non-equilibrium which gives them a higher payoff than the equilibrium play.

Notice that this environment of two players is very different from the one described by Fudenberg & Kreps[23], which justifies myopic behavior, that is, large population and anonymous random matching environment. However, Ellison[19] has pointed out that teaching the opponent can also happen in a large population with random matching due to "contagion effect". Ellison[19]'s paper is a theoretical paper that looks at the implications on the path of play when there is one rational or sophisticated learner in a population of statistical learners. Notice that a sophisticated learner means that this player assumes that other players are statistical learners, that is, he correctly guesses other players' behavior. This paper shows with a simple Pareto ranked coordination game and given specific initial weights that a sophisticated player might be able to manipulate just with one initial action and following fictitious play afterwards, the path of selection. Another interesting observation that Ellison[19] makes is that the incentive to "teach" opponents in a large population in the example is not robust to noisy play by the players. If players are using a smooth best response dynamics then contagion is likely to occur even without the intervention of a rational "teacher", and so the incentive to intervene is reduced.

Turning to the experimental evidence on sophisticated learning, one interesting case is the strategic teaching case mentioned by Camerer, Ho & Chong[8]. Their case is based on two specific treatments of Van Huyck, Battalio & Ranking[34]'s experimental data on two-person minimum-effort coordination games. In those games subjects repeatedly and simultaneously chose among seven efforts, with payoffs determined by their own effort and their pair's minimum effort. The games have seven symmetric Pareto-ranked pure-strategy equilibria, with all players preferring the one in which both choose the highest effort. Subjects were told their pair's minimum after each play, but nothing else about other subjects' efforts. The two treatments were identical, except that in treatment Cd, the subjects were randomly repaired for each period in a run of either three or five periods, whereas in treatment Cf their pairings, while random, were fixed for an entire run of seven periods. In each case, the structure of the environment was publicly announced, including whether or not the pairings were fixed and the fact that the subject population was fixed for the duration. This difference in pairing schemes led to very different outcomes. In treatment Cd, subjects' efforts were widely dispersed, with mean between 4 and 5, moderately inefficient outcomes, and no apparent convergence or time trend. In treatment Cf, by contrast, within seven periods, 12 of 14 pairs increased the minimum effort to its fully efficient level, often starting from much lower levels. The main difference is that, in treatment Cf subjects were evidently well aware that they could influence their partners' future efforts and many exploited this influence;

but treatment Cd's subjects apparently treated such influences as negligible. Therefore, sophisticated individuals deviate from the myopically optimal action to try to influence others' future actions when the setting makes this beneficial as in the case of the fixed partner. If one player is a belief-based type learner and the other has exactly this model of their opponent, then the sophisticated will teach the learner to play a specific equilibrium. If two sophisticated players are matched, the efficient outcome will appear from the very beginning. However, if two standard learners are matched then there is no guarantee that they will end up in the efficient outcome.

Another interesting case where there is sophistication is reported by Cooper & Kagel[12]. They focus on cross-game learning, the ability of subjects to take what has been learned in one game and transfer it to related games. The experiment is on signaling games and they look at how subjects learn to play strategically or limit pricing. A monopolist, whether low cost type (MH) or high cost type (ML) must choose a quantity and the entrant (E) after observing the quantity must decide whether to enter or stay out. The main treatment variable is the cost of the entrant which determines the existence of pooling and separating equilibria. The control group plays with high cost and low cost of entry. When high cost of entry, MHs start doing their non-strategic strategy, their myopic maximum, but after observing high rates of entry, they learn to limit pricing, which is to pool and imitate the MLs, and Es to stay out. When low cost of entry, no pooling equilibrium exists and ML should limit pricing separating themselves from the high type monopolists. However, this learning process is very slow, only experienced subjects learn to play the separating equilibrium. The authors explain that intuitively, limit pricing is more difficult to learn as a ML in the game with low cost entrants than as a MH in the game with high cost entrants because MHs can rely on imitating the choices of MLs, while low type monopolists have no such guide to follow in early rounds. In their experiment, after being playing in a low entry cost environment, they switch to a high entry cost environment, what they call the crossover. Given that all subjects have learned to play strategically as MH prior to the crossover, the critical question is whether this experience will help them to play more strategically following the crossover as MLs than inexperienced subjects, whether positive transfer exists or not. They find that after the crossover, MLs showed significantly more strategic behavior than in inexperienced control group and this strategic play is statistically indistinguishable from experienced MLs. The authors mention that the behavior of Ms in the control group is consistent with an adaptive learning based on fictitious play, Cooper, Garvin & Kagel[10,11]. However, after the crossover, this model cannot capture the positive transfer. Moreover, this learning model predicts negative transfer since Ms' beliefs do not change anticipating a different behavior on low cost Es. In order for the learning model to track the data, they added sophisticated learners who model their opponents as unsophisticated (fictitious players) and thereby anticipate the increase in entry rates following the crossover. Fitting this model to the data, they find a statistically significant fraction of sophisticated learners in the population, and that the fraction of sophisticated learners increases with experience. The authors in subsequent papers [13,14] further analyze the factors that help positive transfer and report two factors: meaningful context and team or group decision making. Meaningful context, contrary to generic or abstract context, help subjects to take advantage of the experience in similar games. Group decision making, two people making decision instead of individual decision making, also helped the positive transfer. This is explained by the "truth wins" norm from psychology literature. Once a strategic play is discovered and mentioned by somebody it becomes self-enforcing. Notice that in this experimental setting, there are many players and that different monopolists are matched with different entrants every round, so the usual key assumption that justified the regular myopic learning predicted by standard learning models are satisfied. However, cross-game learning occurs and individuals can predict what is going to happen and therefore best respond to that, showing sophisticated behavior.

5 References

[1]Arthur, B. 1991. Designing economic agents that act like human agents: A behavioral approach to bounded rationality. *American Economic Review Proceedings*, 81, 353-359.

[2]Arthur, B. 1994. On designing economic agents that behave like human agents. *Journal of Evolutionary Economics*, 3, 1-22.

[3]Brown, G. (1951). Iterative solution of games by fictitious play. In T. Koopmans (Ed.), Activity analysis
of production and allocation. New York: Wiley.

[4]Bush, R., and Mosteller, F. (1955). Stochastic models for learning. New York: Wiley.

[5]Camerer, C. F. (2003). Behavioral Game Theory: experiments in strategic interaction. Princeton, N.J.: Princeton University Press.

[6]Camerer, C.F., and Ho, T.H. (1998) Experienced-Weighted Attraction learning in coordination games: probability rules, heterogeneity and time variation. *Journal of Mathematical Psychology* 42, pp. 305-326.

[7]Camerer, C. F., and Ho, T.H. (1999). Experience-weighted attraction learning in normal-form games. *Econometrica*, 67, 827-874.

[8]Camerer, C.F., Ho, T-H. and Chong, J-K. 2002. Sophisticated Experience-Weighted Attraction learning and strategic teaching in repeated games. *Journal of Economic Theory* 104, 137-188.

[9]Cheung, Y.-W., and Friedman, D. (1977). Individual learning in normal form games: Some laboratory results. *Games and Economic Behavior*, 19, 46 76.

[10]Cooper, David J., Garvin, Susan and Kagel, John H. “Signaling and Adaptive Learning in an Entry Limit Pricing Game.” *RAND Journal of Economics*, Winter 1997, 28 (4), pp. 662-83 (a).

[11]Cooper, David J., Garvin, Susan and Kagel, John H. “Adaptive Learning vs. Equilibrium-Refinements in an Entry Limit Pricing Game.” *The Economic*

Journal, May1997, 107 (442), pp. 553-75 (b).

[12]Cooper, David J. and Kagel, John H. "Learning and Transfer in Signaling Games." Working paper, Case Western Reserve

University, 2001.

[13]Cooper, David J. and Kagel, John H. 2002. The Impact of Meaningful Context on Strategic Play in Signalling Games. *Journal of*

Economic Behavior and Organization.

[14]Cooper, David J. and Kagel, John H. 2003. Are Two Head Better than One? Team versus Individual Play in Signaling Games.

Working paper, Case Western Reserve University/Ohio State University.

[15]Cournot, A. (1960). Recherches sur les principes mathematiques de la theorie des richesses. (N. Bacon,Trans.). [Researches in the mathematical principles of the theory of wealth]. London: Haffner.

[16]Costa-Gomes, M.A., Crawford, V.P. and Broseta, B. 2001. Cognition and behavior in normal-form games. *Econometrica*, 69(5), 1193-1235.

[17]Cross, J. G. 1973. A stochastic leaning model of economic behavior. *Quarterly Journal of Economics*, 87, 239-66.

[18]Cross, J. G. 1983.A theory of adaptive economic behavior. London: Cambridge Univ. Press.

[19]Ellison, G.1997. Learning from personal experience: one rational guy and the justification of myopia. Games and Economic Behavior 19, 180-210.

[20]Erev, I., 6 Roth, A. (1998). Predicting how people play games: Reinforcement learning in experimental games with unique, mixed strategy equilibria. *American Economic Review*, 88(4), pp848-881.

[21]Fudenberg, D. and Levine, D. 1989. Reputation and Equilibrium Selection in Games with a patient player. *Econometrica,* 57, 759-778.

[22]Fudenberg, D., and Levine, D. 1998.Theory of learning in games. Cambridge, MA: MIT Press.

[23]Fudenberg, D. and Kreps, D. 1993. Learning mixed equilibria. *Games and Economic Behavior,* 5, 320-367.

[24]Harsanyi, J. 1973. Games with randomly disturbed payoffs. *International Journal of Game Theory,* 2, pp 1-23.

[25]Ho, T. Camerer, C., and Chong, K. 2002. Functional EWA: A one-parameter theory of learning in games. Caltech working paper.

[26]McAllister, P. H. (1991). Adaptive approaches to stochastic programming. Annals of Operations Research, 30, 45-62.

[27]McKelvey, R. D., and Palfrey, T. R. (1995). Quantal response equilibria for normal form games. *Games and Economic Behavior*, 10, 6 38.

[28]Milgrom, Paul. and Roberts, John. Limit Pricing and Entry Under Incomplete Information: An Equilibrium Analysis. *Econometrica*, March 1982, 50(2), pp. 443-59.

[29]Robinson, J. 1951. An iterative method of solving a game. *Annals of Mathematics*, 54, 296-301.

[30]Roth, A. E., and Erev, I. 1995. Learning in extensive-form games: experimental data and simple dynamic models in the intermediate term. *Games and Economic Behavior*, 8, 164-212.

[31]Shapley, L.S. 1964. Some topics in two-person games. In M. Dresher, L.S. Shapley, and A. W. Tucker (Eds.), Advances in Game Theory. Princeton, N.J.: Princeton University Press, 1-28.

[32]Sarin, R., and Vahid, F. 2001. Predicting how people pay games: a simple dynamic model of choice. *Games and Economic Behavior*, 34, 104-22.

[33]Van Huyck, J.B., Battalio, R.C. and Cook, J. 1997. Adaptive behavior and coordination failure. *Journal of Economic Behavior and Organization,* 32, 483-503.

[34]Van Huyck, J.B., Battalio, R.C. and Rankin F.W. 1990. Tacit cooperation games, strategic uncertainty and coordination failure. *American Economic Review,* 80, 234-248.

[35]Young, H. P. 2001. Individual Strategy and Socia Structure: an evolutionary theory of institutions. Princeton, N.J.: Princeton University Press.