

## Equilibrium Dynamics Part 2

### Introductory Comment

Most of the class we have spent looking at how we can predict individual behavior using alternative equilibrium concepts. These equilibrium concepts, which have been the primary focus of game theory in the recent past, are steady state concepts. Once players find equilibrium there should be no tendency to change. Unfortunately, these concepts have little to say about how players ultimately find themselves in some equilibrium. This is not to say that economists have completely ignored the question of how equilibrium is obtained. For example, Cournot proposed a model for learning in games as early as 1838, a model that was embellished by Brown in 1951. Still, until more recently, models of learning that shift the focus from equilibrium to dynamics were primarily the purview of psychologists. The recent interest in equilibrium dynamics in game theory has been fostered by the increasing acceptance of experimental methods. This increased interest seems a natural response to the often poor performance of equilibrium predictions in the lab.

### **Objective: Understand the similarities and differences of different equilibrium dynamic models.**

*Evolutionary:* Assumes players are born with a strategy that they always play. Players are then randomly matched against each other. The number of players that use successful strategies increases over time, while the number of players that use unsuccessful strategies decreases over time. With this approach, there is essentially no thinking.

*Reinforcement:* Players choose strategies that have been successful in previous play more often than those that have not. With this approach there is thinking, but reinforcements are limited to a player's own experience with alternative strategies.

*Belief:* Players update beliefs about what others will do based on history. These beliefs are then used to evaluate the attractiveness of alternative strategies. More attractive strategies given ones beliefs are played more often. Belief learning is like reinforcement learning in that people think. However, reinforcement learning is purely reactive, dependent on the success of previous actions. Alternatively, belief learning is in a sense proactive, as well as reactive. The evolution of beliefs is typically modeled as reactive to past play, while the evolution of the attractiveness of a strategy is proactive in the sense that it is forward looking given backward looking beliefs.

*Experience-Weighted Attraction (EWA):* EWA was developed as a hybrid of reinforcement and belief learning that reinforces based on forgone as well as actual payoffs. Reinforcement and belief learning represent separate special cases of the model. So it can be interpreted as a thinking model that is both reactive and proactive.

*Anticipatory/Sophisticated:* These models are similar in nature to belief learning, but they use more than just past experience to update beliefs. For example, a sophisticated learning model might assume that players identify strictly dominated strategies and rule them out.

*Imitation:* Players mimic the strategies of other more successful players.

*Learning Direction:* Players adjust their strategy toward the ex post best response.

*Rule Learning:* Players have decision rules conditioned on the history of play, not strategies (e.g. the tit-for-tat rule in the prisoner's dilemma). They learn to use rules that are the most successful, possibly in an evolutionary or reinforcement fashion.

**Objective: Understand a general framework for modeling equilibrium dynamics in normal form games.**

There have been a wide variety of models proposed to explain out of equilibrium dynamics in games. Many of these models can be cast in a fairly general mold for normal form games. So let us take a moment to formally describe that mold.

Consider the normal form game  $G$ . There are  $m = 1, \dots, M$  players. Each player has a set of pure strategies  $S_m$ , which we will number from 1 to  $J_m$  for convenience. Let  $s_m \in S_m$  be some pure strategy for the  $m$ th player and  $s = \{s_1, \dots, s_M\} \in S = \times_{m=1, \dots, M} S_m$  be a strategy profile. Players' payoffs are defined over strategy profiles:  $u(s) = \{u_1(s), \dots, u_M(s)\}$  where  $u_m(s)$  is the  $m$ th player's payoff given strategy profile  $s$ .

Assume  $G$  represents a single stage of a game that is repeated  $T$  times.

Define  $s_t$  as the strategy profile played in period  $t$  and  $s_{mt}$  as player  $m$ 's part of strategy profile  $s_t$ . Define  $h_t = \{s_1, \dots, s_t\}$  as the history of play up to period  $t$ .

Call  $A_{mj}(t)$  the attraction or attractiveness of strategy  $j$  to player  $m$  in period  $t$ , where  $A_{mj}(0)$  is some initial attraction. One way to interpret  $A_{mj}(t)$  is as expected utility. The term attraction is used however to emphasize that this expected utility can be much richer than  $E(u_m(s))$ . The collection of all attractions for a player is  $A_m(t) = \{A_{m1}, \dots, A_{mJ_m}\}$ .

Now to describe the dynamics of play in this game we need to describe two more things (i) how attractions are transformed into strategy choices and (ii) how attractions evolve over time.

Typically, attractions are used to define a probability distribution over a player's strategy set:

$$s_m(t+1) = \{s_{m1}(t+1), \dots, s_{mJ_m}(t+1)\} = F(A_m(t))$$

where  $s_{mj}(t+1)$  is the probability player  $m$  chooses strategy  $j$  such that  $s_{mj}(t+1) \geq 0$  and

$$\sum_{j=1}^{J_m} s_{mj}(t+1) = 1.$$

Generally, we can describe the updating or reinforcement of attractions as

$$A_m(t) = R(A_m(t - 1), h_t, G),$$

so future attractions depend on past attractions, the history of play, and the structure of the game.

**Objective: Review the alternative specifications used for transforming attractions into strategy choices.**

To some extent, how attractions are transformed into strategy choices depends on your purpose.

In the context of simulating or predicting behavior, we want a transformation that leads to a unique strategy prediction. We can treat attraction just like utility and say that the strategy with the highest attraction is the one chosen. However, this deterministic approach can produce rather boring dynamics for many of the reinforcement rules explored in the literature. An alternative stochastic approach is to choose a player's strategy randomly based on  $F(A_m(t))$  producing some probability distribution over a player's strategies. With this approach, more attractive strategies are chosen more often, but not always, which can lead to richer dynamics. With such a stochastic approach, repeated simulations are necessary to understand the performance of the system.

In the context of fitting observed data, we want probability distributions, not unique strategies. As mentioned previously, if we treat attractions like utility, we can draw on the rich econometric literature for discrete choice random utility models. For example, if a game only has two strategies, we can use a probit or logit model based on the difference in attractions between the two strategies. For games with more than two strategies, a multinomial logit model is often

employed,  $s_{mj}(t+1) = \frac{e^{a_{mj} + I_{mj}A_{mj}(t)}}{\sum_{k=1}^{J_m} e^{a_{mk} + I_{mk}A_{mk}(t)}}$  where the  $a$ s and  $I$ s are estimable parameters. The  $a$

parameters in this specification serve to capture a player's predisposition to certain strategies given equal attractions, while the  $I$  parameters capture the sensitivity of choice probabilities to alternative attractions. *Note: In this multinomial logit model not all of the  $a$  and  $I$  parameters can be identified. Only relative values can. But this is not idiosyncratic to models of equilibrium dynamics.*

While there are a variety of ways to specify the transformation of attractions into the probability of alternative strategy choices, the more novel and interesting work has focused on the specification of attraction updating or reinforcement.

**Objective: Review a variety of methods that have been used to update attractions.**

First, let us consider a simple reinforcement model:

$$A_{mj}(t) = \begin{cases} A_{mj}(t-1) + u_m(s_t) - \bar{m}_m(t) & \text{for } j = s_{-m} \\ A_{mj}(t-1) & \text{otherwise} \end{cases}$$

where  $\bar{m}_m$  is some reference payoff (e.g.  $\bar{m}_m = \min_{s \in S} u_m(s)$  in Erev and Roth (1998)). A key feature of this updating function is that the attraction for a strategy can change only if the strategy is played. The direction and magnitude of change depends on some reference payoff that may or may not change over time. But we must be careful here, even though the attraction of a strategy does not change, the probability it is played can change because  $F(A_m(t))$  aggregates all attractions into choice probabilities. Using the multinomial logit model, if  $A_{mj}(t)$  increases, the probability strategy  $j$  is played will increase while the probability all other strategies are played will decrease. It is also important to realize that the updating function depends only on a player's observed payoffs and some reference payoff, so the information requirements are pretty minimal. When it comes to using this model, it should also be noted that there is at least one free parameter,  $A_m(0) = A_{mj}(0)$  for all  $j$ , but could be as many as  $J_m$ ,  $A_{mj}(0)$  for all  $j$ . This parameter can be estimated with observed data using a variety of methods.

Now let us consider a simple beliefs based model. Define the probability of strategy profile  $s$  in

$$\text{period } t \text{ given history } h_t \text{ as } p(s, h_t) = \frac{s_t^d + \sum_{r=1}^{t-1} g^r s_{t-r}^d}{1 + \sum_{r=1}^{t-1} g^r} = \frac{s_t^d + p(s, h_{t-1}) \sum_{r=1}^{t-1} g^r}{1 + \sum_{r=1}^{t-1} g^r} \text{ where } s_t^d \text{ is equal to 1}$$

if  $s = s_t$  and 0 otherwise and  $g$  is some weighting parameter. The attraction updating equation can now be written as:

$$A_{mj}(t) = \sum_{s_{-m} \in S_{-m}} p(j, s_{-m}, h_t) u_m(j, s_{-m}) = \frac{u_m(j, s_{-m}) + A_{mj}(t-1) \sum_{r=1}^{t-1} g^r}{1 + \sum_{r=1}^{t-1} g^r}$$

where  $s_{-m}$  and  $S_{-m}$  are a strategy profile and the set of all possible strategy profiles exclusive of player  $m$ . Note that  $p(s, h_t)$  is simply a weighted average of the number of times strategy profile  $s$  was played in the past. So attraction  $A_{mj}(t)$  is just the expected payoff from playing strategy  $j$  given the weighted historical probability that  $s_{-m}$  was played in the past. If we assume  $g = 0$ , we have the model originally proposed by Cournot. This model says that players only care about what happened in the last period. Assuming  $g = 1$ , yields the model proposed by Brown (fictitious play), which states players weight all previous experience equally. For  $1 > g > 0$ , players place more weight on their most recent experience. For  $g > 1$ , players place more weight on older experiences. The interpretation of a negative  $g$  is problematic. It should be noted that this updating rule requires players to have more information than the simple learning model. Players must know their entire payoff matrix and must know the strategy choice of other players. Also, to use this model, there is some freedom in choosing initial attractions, but there is also some freedom in choosing the weighting parameter  $g$ .

Comparing this model to the simple learning model highlights two important differences. A strategy's attraction changes regardless of whether or not it was used in the past, due to how it would have performed in the previous period. That is, strategies are reinforced based on forgone payoffs as well as realized payoffs. Second, reinforcement and previous attractions are not given equal weight. Furthermore, with positive  $g$ , each new period reduces the importance of reinforcement, while increasing the importance of previous attractions.

One of the nice features about this belief model is that it can discount past experience. This is particularly important in games because a player's opponents will also be learning or updating their beliefs, so more recent experiences are likely to be more representative than more distant experiences. But a similar feature can be built into a learning model:

$$A_{mj}(t) = \begin{cases} fA_{mj}(t-1) + u_m(s_t) - \bar{m}_m(t) & \text{for } j = s_{mt} \\ fA_{mj}(t-1) & \text{otherwise} \end{cases}$$

where  $f$  represents the rate of decay (or growth) of previous attractions. With this specification comes an additional degree of freedom, the specification of  $f$ . With this specification, it is also possible for the importance of previous attractions to exceed the importance of reinforcement, something that was not possible in the beliefs model.

Erev and Roth also throw in what they call generalization/experimentation to the reinforcement dynamic. Suppose strategies are ordered by some degree of similarity and define

$$A_{mj}(t) = \begin{cases} fA_{mj}(t-1) + (u_m(s_t) - \bar{m}_m(t))(1-e) & \text{for } j = s_{mt} \\ fA_{mj}(t-1) + (u_m(s_t) - \bar{m}_m(t))\frac{e}{2} & \text{for } j = s_{mt} \pm 1 \text{ or} \\ fA_{mj}(t-1) & \text{otherwise} \end{cases}$$

$$A_{mj}(t) = \begin{cases} fA_{mj}(t-1) + (u_m(s_t) - \bar{m}_m(t))(1-e) & \text{for } j = s_{mt} \\ fA_{mj}(t-1) + (u_m(s_t) - \bar{m}_m(t))\frac{e}{J_m - 1} & \text{otherwise} \end{cases}$$

where  $e$  represents the proportion of reinforcement devoted to similar/other strategies. With this definition, some or all strategies are reinforced even if they are not played, which adds another degree of freedom to the model, the specification of  $e$ .

On a personal note, the case for generalization is compelling, but Erev and Roth's treatment is rather ad hoc. Calling Erev and Roth's modification to traditional reinforcement learning experimentation seems to be a stretch. When I think of experimentation, I think of trying something new on a whim to see what happens. If the outcome is good, I am more likely to do it again. If the outcome is bad, I am less likely to do it again. Therefore, I think experimentation is best modeled through the choice function and not through attraction updating. For example,

$$s_{mj}(t+1) = \frac{e^{a_{mj} + I_{mj} A_{mj}(t)}}{\sum_{k=1}^{J_m} e^{a_{mk} + I_{mk} A_{mk}(t)}} (1 - e(t+1)) + \left( 1 - \frac{e^{a_{mj} + I_{mj} A_{mj}(t)}}{\sum_{k=1}^{J_m} e^{a_{mk} + I_{mk} A_{mk}(t)}} \right) \frac{e(t+1)}{J_m - 1},$$

where  $e(t+1)$  is the probability player  $m$  experiments randomly with some other strategy. Of course, this type of experimentation in the choice function will still indirectly influence attraction updating through the history of strategy choices.

Camerer and Ho (1999) develops experience weighted attraction (EWA) as a hybrid of reinforcement and belief learning. Define  $N(t) = 1 + rN(t-1)$ , which are referred to as experience equivalents. For  $r = 1$ , all past experience is weighted equally. For  $r < 1$ , recent experience is more important. For  $r > 1$ , early impressions are most important. They then define attraction updating as

$$A_{mj}(t) = \begin{cases} \frac{N(t-1)fA_{mj}(t-1) + u_m(s_t)}{1 + rN(t-1)} & \text{for } j = s_{mt} \\ \frac{N(t-1)fA_{mj}(t-1) + du_m(j, s_{-mt})}{1 + rN(t-1)} & \text{otherwise} \end{cases}.$$

This model melds important features of the beliefs model with important features of the simple learning model. First, it depreciates past attractions for  $1 > f > 0$ . Second, it weights past attractions and reinforcements. Finally, it considers forgone payoffs through  $d$ . Contrary to the beliefs model, forgone payoffs can be either more or less important than realized payoffs.

If we let  $r = 0$  and  $d = 0$ , we get back our simple reinforcement dynamic. Alternatively, if we set  $r = f$  and  $d = 1$ , we can get back a simple beliefs model. Therefore, EWA is a generalization of the two models. With this model, we have freedom to choose  $N(0)$ ,  $A_{mj}(0)$  for all  $j$ ,  $f$ ,  $d$ , and  $r$ .

**Objective: Summarize empirical performance of reinforcement, beliefs, and EWA models.**

Renewed interest in models of equilibrium dynamics was sparked by experimental data that showed steady state equilibrium predictions more often than not fail to predict behavior. One explanation for the failure of steady state equilibrium concepts was that these concepts missed the mark, which suggests a better steady state model could make better predictions. However, the experimental evidence also clearly shows that the steady state assumption is a mistake, especially in early rounds of experiments where subjects have little experience. Furthermore, experimental observations did not seem to arrange themselves purely randomly. There are clear systematic and robust time trends in repeated game experiments, suggesting that well crafted dynamic models may perform better than the predictions of static steady state models. As researchers delved into the issue, this expectation has been confirmed. Most dynamic models

whether they are simple or complex do a better job explaining behavior than the steady state models.

But which models do best? It depends. In some instance, simple beliefs models do better than simple reinforcement. However, in other instances simple reinforcement does better than simple beliefs. EWA will always do at least as good as simple belief or reinforcement models because it contains each as a special case. However, EWA may not perform as well as more complicated belief or reinforcement models (for example, Erev and Roth's reinforcement model with generalization).

In a few cases (fictitious play for example), the general implications of the dynamic have been explored theoretically. However, for many others, the general theoretical implications remain a mystery.

As a final note here, another interesting question that has popped up with these models is the consistency of parameter estimates across different games and different subjects. Erev and Roth argue the parameter estimates for their learning model are consistent across games, though they do not test the hypothesis formally. Alternatively, Cheung and Friedman find significant differences. Cheung and Friedman also report significant differences in learning and choice parameters across subjects, which echo the findings of the Stahl and Wilson.

### **Objective: Understand what is missing from reinforcement and belief models.**

In the simple learning model, players only care about their payoff from strategies they have played in the past. These payoffs will be influenced indirectly by a players own payoff matrix and the payoff matrix of an opponent because of the strategy choices made by each player. However, none of this information is necessary because it is not used explicitly in the model. Belief models require more information on a player's own payoffs and a player's opponents' strategies, but still no information on opponents' payoffs. What does this mean? It means that in these models there is no wondering about an opponent's explicit reasoning. In the monopoly experiment we did, I saw most of you doing a lot of pencil pushing between your price and quantity choices, suggesting you were thinking carefully about what your previous choices and the outcomes of these choices were really telling you about the demand you faced. Presumably, your goal was to predict demand, so you could make better strategy choices. It was also clear to me that some of your strategy choices were made strategically, with the thought that a good choice of P and Q would reveal more information even though it might not reveal a higher payoff. This is the notion of sophistication. People are smart. Instead of always responding reactively to past payoffs or best responding to available information, people often make choices that they know might be costly in order to figure out how to make better choices in the future.

Most dynamic models are built around certain information assumptions. Since information can change from one environment to the next, a general model should be able to adjust to the information environment.

Strategy specifications in the typical dynamic model have been limited to stage games. But for models like the repeated prisoner's dilemma, the stage game specification may be too limiting. For example, a subject might want to play a strategy that specifies cooperation if an opponent has cooperated in the past, but three periods of defection for punishment if an opponent has defected.

**Objective: Provide an overview of models of belief elicitation.**

The simple belief models we have focused on up to this point, construct beliefs based on previous play. There is an alternative: belief elicitation. A variety of methods have been proposed to elicit peoples' beliefs. For example, we can simply ask them what they think the probability of each strategy profile is or we can ask them to predict which strategy profile will be chosen. Either way, the result is direct information regarding a subject's beliefs, which we can presumably use to predict behavior. A recent study (Nyarko and Schotter, 2002) used belief elicitation and then compared fictitious play beliefs to these stated beliefs. They concluded the two were not equal. Furthermore, they show there is a significant difference between model parameters estimated with stated and fictitious play beliefs. Finally, their model using stated beliefs outperforms a reinforcement model and EWA model.

This paper raises an interesting question. Can the variability in the parameters of the choice model across different games and subjects be explained by a systematic bias in using past strategy profiles to model beliefs?

**Objective: Explore reinforcement, belief, and EWA dynamics in stylized 2 × 2 games.**

The excel file "EWA Simulation" models EWA dynamics for general two player games by simulating the play of 25 pairs of players over 500 periods. The "Parameters" worksheet allows you to change:

- (i) The initial attraction ( $AR(0)$  and  $AC(0)$ ), which can differ across players.
- (ii) The initial experience weight ( $N(0)$ ), which is assumed to be identical across players.
- (iii) The weight on previous attractions,  $f(\phi)$ , which is assumed identical across players.
- (iv) The depreciation of experience weights,  $r$  ( $\rho$ ), which is assumed identical across players.
- (v) The weight of forgone payoff,  $\delta$  ( $\Delta$ ), which is assumed identical across players.
- (vi) Choice probability parameters  $a_m$  ( $\alpha$ ) and  $I_m$  ( $\lambda$ ) for each player.
- (vii) The 2×2 game payoffs.

Consider the Hawk-Dove game:

|            |                | Column Player |                 |
|------------|----------------|---------------|-----------------|
|            |                | Left ( $y$ )  | Right ( $1-y$ ) |
| Row Player | Up ( $x$ )     | -2            | 0               |
|            | Down ( $1-x$ ) | 8             | 4               |

where  $x$  and  $y$  are the probability the row player chooses Up and the column player chooses Left. There are two pure strategy equilibrium for this game  $x = 1$  and  $y = 0$ , and  $x = 0$  and  $y = 1$ . There is also a mixed strategy equilibrium for the game,  $x = 2/3$  and  $y = 2/3$ .

Let us start with simple reinforcement learning ( $AR(0) = 0$ ,  $AC(0) = 0$ ,  $N(0) = 1$ ,  $f = 1$ ,  $r = 0$ , and  $\delta = 1$ ). Unless stated otherwise, we will also assume  $a_m = 0$  and  $I_m = 1$ . Figure 1 illustrates the dynamics for the first 100 periods. Notice that these dynamics settle down pretty fast where  $x \approx 0.5$  and  $y \approx 0.25$ , but these probabilities do not match up well with either of the equilibrium predictions. That is, the prediction from this simple learning model is much different than the Nash prediction. So, what is going on here? Looking at the simulations for individual pairing's shows that the dynamic seems to converge to  $x = 1$  and  $y = 0$ ;  $x = 0$  and  $y = 1$ ; and  $x = 0$  and  $y = 0$ . The first two of these are not surprising because they are pure strategy Nash. But even with individual pairings, there is little support for the mixed strategy Nash,  $x = 2/3$  and  $y = 2/3$ , within the context of this version of the learning dynamic. The strategy combination frequently observed in individual pairings is not Nash. Why does this last strategy combination persist? The learning dynamic ignores forgone payoffs. Since the payoff to  $x = 0$  and  $y = 0$  is relatively good for both players, it will tend to be strongly reinforced for both players. For  $x = 1$  and  $y = 0$ , and  $x = 0$  and  $y = 1$ , one player is strongly reinforced, but the other is not.

Now let us look at a version of Cournot's belief model (Figure 2 assumes  $AR(0) = 1$ ,  $AC(0) = 1$ ,  $N(0) = 1$ ,  $f = r = 0$ , and  $\delta = 1$ ). Unlike the reinforcement model, the Cournot belief model does not appear to settle down much even after 500 periods. Furthermore, there appears to be a general cyclical trend. Looking at the simulation for individual pairs suggests why. For individual pairs, things start by bouncing around, but then settle down to long stretches of pure strategy Nash equilibrium play. However, since the Cournot model does not allow this equilibrium play to reinforce itself, eventually one player deviates and everything starts bouncing around again. In some cases there are long stretches of players cycling between  $x = 0$  and  $y = 0$ , and  $x = 1$  and  $y = 1$ . Changing to the Fictitious play model by setting  $f = r = 1$ , reduces the general cyclical trend, but the system still doesn't settle down over 500 periods.

To get the system to settle down again, we can reduce  $r$  relative to  $f$  (Figure 3 assumes  $AR(0) = 1$ ,  $AC(0) = 1$ ,  $N(0) = 1$ ,  $f = 1$ ,  $r = 0.75$ , and  $\delta = 1$ ), which is a possibility in the EWA framework. Within 20 periods the model has settled down, but not to any of the Nash predictions. A look at the individual pairings reveals that play converges to one of the two pure strategy Nash equilibria. So you might be asking yourself, how come there are different proportions of aggregate play between the Row and Column players? After all, the game is symmetric. The

problem is we are only simulating 25 pairings. With 1,000 simulated pairings, I would expect equal probabilities.

Figure 1: Reinforcement learning in the Hawk-Dove game ( $AR(0) = 0$ ,  $AC(0) = 0$ ,  $N(0) = 1$ ,  $f = 1$ ,  $r = 0$ , and  $\delta = 1$ ).

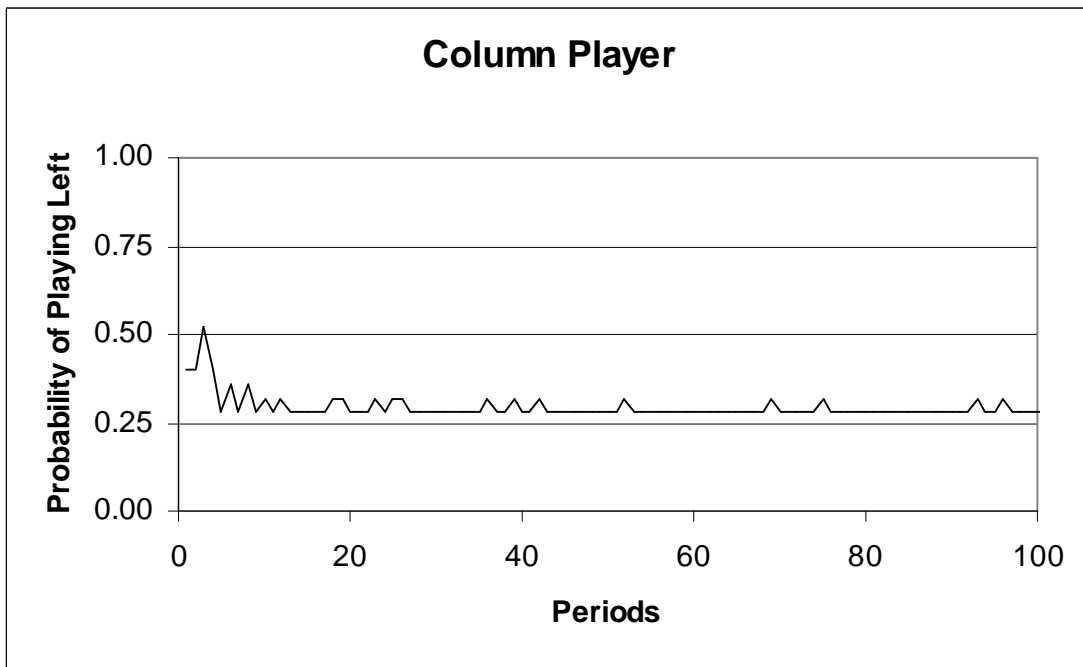
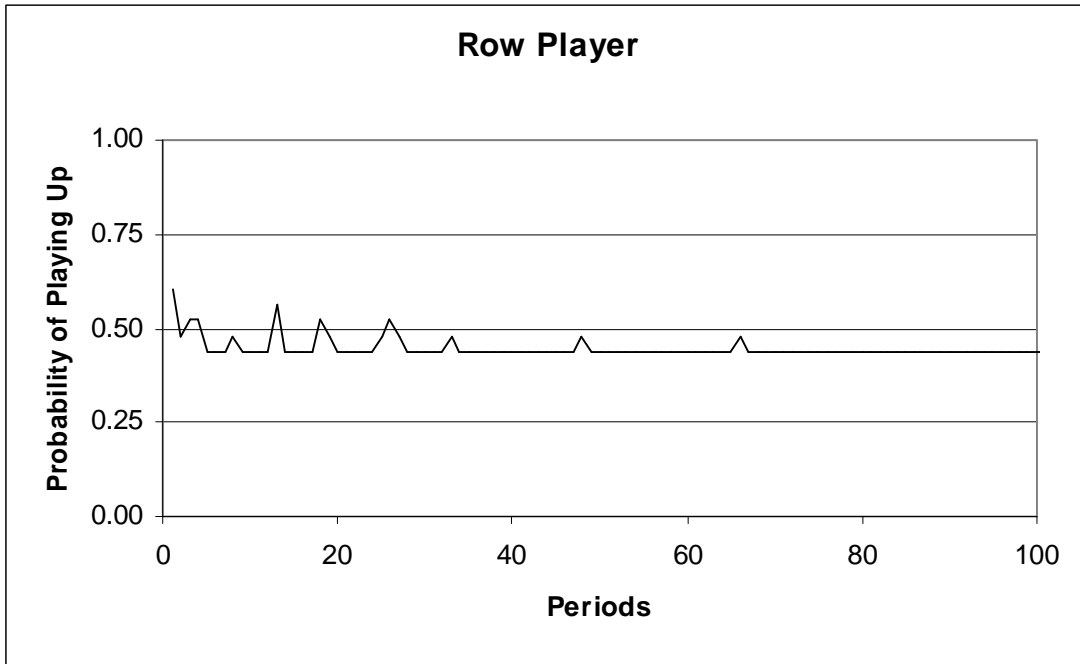


Figure 2: Cournot belief model in the Hawk-Dove game ( $AR(0) = 1$ ,  $AC(0) = 1$ ,  $N(0) = 1$ ,  $f = r = 0$ , and  $\delta = 1$ ).

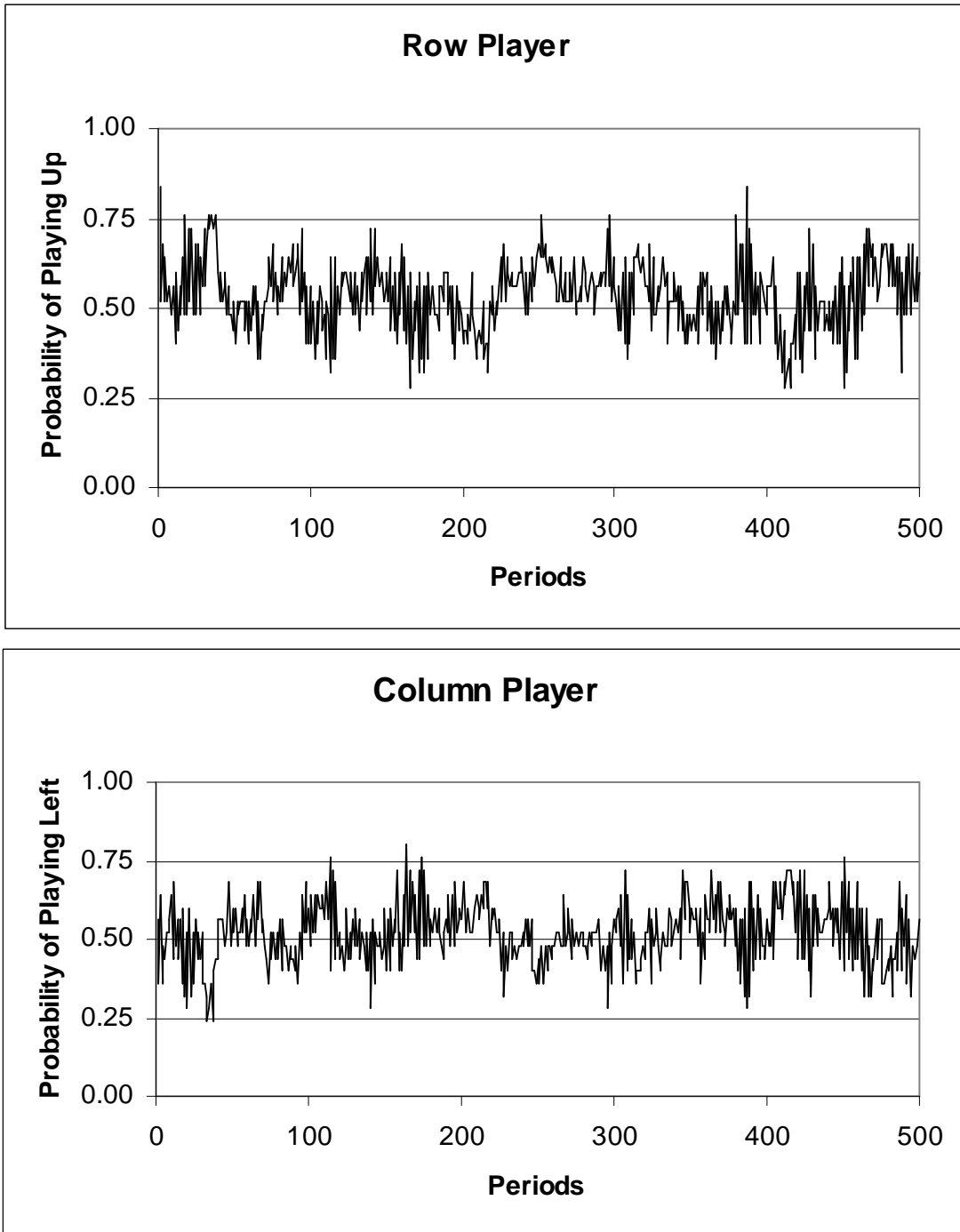


Figure 3: EWA model in the Hawk-Dove game ( $AR(0) = 1$ ,  $AC(0) = 1$ ,  $N(0) = 1$ ,  $f = 1$ ,  $r = 0.75$ , and  $\delta = 1$ )

