Learning in agent based models

Alan Kirman

GREQAM
Groupement de Recherche en Economie
Quantitative d'Aix-Marseille - UMR-CNRS 6579
Ecole des Hautes études en Sciences Sociales
Universités d'Aix-Marseille II et III

Document de Travail
n°2010-57

Learning in agent based models

Alan Kirman

December 2010

Learning in agent based models.

Alan Kirman
GREQAM, Marseille, Université Paul Cézanne, Ecole des Hautes Etudes en Sciences Sociales, Institut Universitaire de France.

Abstract

This paper examines the process by which agents learn to act in economic environments. Learning is particularly complicated in such situations since the environment is, at least in part, made up of other agents who are also learning. At best, one can hope to obtain analytical results for a rudimentary model. To make progress in understanding the dynamics of learning and coordination in general cases one can simulate agent based models to see whether the results obtained in skeletal models translate into the more general case. Using this approach can help us to understand which are the crucial assumptions in determining whether learning converges and, if so, to which sort of state. Three examples are presented, one in which agents learn to form trading relationships, one in which agents misspecify the model of their environment and a last one in which agents may learn to take actions which are systematically favourable, (or unfavourable) for them. In each case simulating models in which agents operate with simple rules in a complex environment, allows us to examine the role of the type of learning process used by the agents the extent to which they coordinate on a final outcome and the nature of that outcome.

Keywords
Learning, agent based models, simulations, equilibria, asymmetric outcomes.

J.E.L. classification codes
D83, C73, C63

Learning in agent based models

Introduction

In basic economic theory agents are supposed to optimise, given their constraints. This, in situations where they are anonymous, isolated and have no influence over the constraints with which they are faced. Any intelligent student, coming to economics for the first time, finds this unrealistic and several intensive courses, particularly in micro-economics are necessary to persuade him otherwise. What is it that is difficult to reconcile with economic reality ? The usual answer is that the calculations needed to take an optimal decision are too complicated. But this is not the real problem. An individual consumer, for example, has only to know his preferences over every possible choice with which he is faced and his budget which is determined by the prevailing prices. Then he simply choose his most preferred option. Thus, he has to make no calculations but simply to choose. Of course, as soon as he is faced with uncertainty the situation is a little more complicated but here again, he chooses in such a way that his choices are consistent with his subjective probabilities. Nevertheless, the assumptions of complete preferences, of complete information, and of coherent probabilities are very strong. In the case of the firm, specific calculations have to be made. All of this has led to alternative bases for explaining or justifying the choices that individuals and firms make. A typical example of how the assumption of optimisation may be justified, is that of Lucas who said the following,

"In general we view, or model, an individual as a collection of decision rules (rules that dictate the action to be taken in given situations) and a set of preferences used to evaluate the outcomes arising from particular situation-action combinations. These decision rules are continuously under review and revision: new decisions are tried and tested against experience, and rules that produce desirable outcomes supplant those that do not. I use the term "adaptive" to refer to this trial-and-error process through which our modes of behaviour are determined." Lucas (1988).

In other words, individuals simply learn to use rules which turn out to work well. However, Lucas goes on to say,

"Technically, I think of economics as studying decision rules that are steady states of some adaptive process, decision rules that are found to work over a range of situations and hence are no longer revised appreciably as more experience accumulates."

While this might sound like a reasonable justification for optimising behaviour, it is worth reflecting on the nature of the sort of learning process he envisages. Lucas' basic argument is that the evolution of the economic environment is very much slower than the speed at which agents adjust to that evolution. As a result, he suggests, we can safely ignore the impact of the agents' behaviour on the environment. Yet, a characteristic of economic situations is that they are made up of agents all of whom may be trying to learn about their environment and hence about what the other agents are doing. This rapidly becomes very complicated and it is not at all clear that the behaviour of individuals will « co-evolve » to something corresponding to the theoretical equilibrium solution.

One solution to this is to use a game-theoretic approach and to attribute game theoretic strategic reasoning to the participants in a market and to try to show that agents learn to coordinate on an equilibrium (see e.g.Fudenberg and Levine (1998) for a full account). In particular when there are many equilibria, of a game for example, one can find out if individuals, by learning, coordinate on a specific equilibrium.[1] Thus, in this view, learning is just an equilibrium selection device. But it is important to note that, in order to proceed in this way one has to specify precisely how individuals learn. This, of course brings us back to the standard dilemma in economics, should we find learning rules which are analytically tractable, or should we explore rules which are intuitively plausible. Whilst a consensus has developed in economic theory, rightly or wrongly, as to the definition of « rationality » and the associated axioms, no such consensus has developed about the appropriate definition of « learning ». Thus one is free to choose which is the appropriate learning model, but any theoretical results will be tied to that definition. Such results will be open to the criticism that they are « ad hoc ».

[1] In this approach, the standard axioms of rationality are not questioned but an alternative is to modify the notion of rationality while still assuming that people act strategically and this is the approach adopted by « Behavioural Game Theory » (see e.g. Camerer (2003)). In this approach it is not so much the extent of reasoning that is questioned but the existence of other motives that enter into peoples utility functions when they are participating in economic activities in general and markets in particular.

In fact, the standard axioms of rationality are also ad hoc but have become accepted. Given this, one can look in a new light at the idea of using a different approach to modelling learning, that of agent based modelling, (ABM). To quote Borril and Testsfastion (2010), « Roughly, ABM is the computational modeling of systems as collections of autonomous interacting entities ». The important word here is « comptutational ». The idea is to specify the nature of the agents involved, the rules by which they behave and by which they interact and then to simulate the model in order to observe the outcomes.

In fact, the agent based approach to learning, takes Lucas at his word and models individuals as using simple rules and choosing those which work best, and then sees what the result of the interaction between them will be. Rather than assuming that the individuals converge to « as if optimising » and thus to an equilibrium, the idea is to find out whether they, in fact, do so. Rather than starting out with an a priori equilibrium notion, one can examine what happens over time as agents learn to modify their behaviour together, whether they are conscious or not of what the other participants in the market or economy are doing.

Typically, it will only be in certain limited cases that one can simplify a model to the extent of being able to do a formal analysis of the phenomenon in question. However, one can then build a corresponding model in which agents follow the same rules as those in the analytical model, and then extend and expand it and, by simulating, see whether the results in the simplest cases hold up in the more general ones. ABM can thus provide a bridge between very simple stylised theoretical models and more general ones. Furthermore, it is often the case that apparently small modifications to the original theoretical model can lead to marked changes in outcomes and this can become apparent when simulating the associated agent based model. Thus, agent based models can help comfort theoretical results but can also reveal their limitations.

A somewhat more radical approach is to take ABM as a methodology in its own right and not to regard it as a way of checking the applicability to more general cases of limited theoretical results, (see Epstein (2006) and Borril and Testsfastion (2010)).

In this paper I will give three examples of situations in which agents learn about their environment, in each case, mentioning some simple theoretical results and then showing how agent based models help us to enrich and modify the latter. In the first case agents on a market

for a perishable good learn what prices to pay and to charge and with whom to trade. The two questions that are posed are will agents learn to form stable buyer-seller relationships and will there be price dispersion and discrimination ? In the second case agent learn about their environment but misperceive it. Their model of the environment is incomplete and the question is will they learn, nevertheless to find their way to the equilibrium or will they be comforted in their misperception by what they observe?

In the last example, identical individuals faced with a repeated situation learn, by reinforcement based on their previous experience, to adopt different strategies. Some will learn to adopt strategies which lead to a low pay-off and some will do the contrary. For an outside observer it might seem that, since the individuals are identical, some are just luckier than others. However, the agent based model presented shows that agents are not intrinsically fortunate or unfortunate, but learn to be « lucky » or « unlucky ».

Learning in Fish Markets.

As a first example of how agent based models can help us understand economic phenomena and particularly learning, let me start with examining buyer seller relationships on a wholesale fish market. The model was developed with Gerard Weisbuch and Dorothea Herreiner, (Weisbuch et al. (2000)). The market that we studied and for which we had detailed data was situated at Saumaty on the coast at the Northern edge of Marseille. It was open every day of the year from 2 a.m. to 6 a.m. Over 500 buyers[2] and 45 sellers came together, although they were not all present every day and they transacted more than 130 types of fish. Prices were not posted. All transactions were pairwise. There was little negotiation and prices can reasonably be regarded as take it or leave it prices given by the seller. The data set consisted of the details of every individual transaction made over a period of three years.

This market has many specific features and our idea was to try to explain some of these with simple models. The approach was first to develop an extremely basic model and to see if we could produce analytical results which would reproduce the stylised facts that we observed empirically. We then simulated an agent based models giving the agents similar rules to those

[2] In fact some 1400 buyers appear in the records but many of these were hardly present at all.

in the analytical model but in a more general context. We then, in Kirman and Vriend (2001), developed a model with very simple rules for the agents in the hope of reproducing some of the other stylised facts.

The first fact that interested us was that, on the Marseille market, buyers are split into two distinct groups. There are those who are essentially loyal to one seller and those who systematically shop around. Yet, there are almost no buyers who adopt an intermediate approach, that is who are basically loyal to one seller but also buy from others from time to time. To model the development of these relationships in a fully optimising game-theoretic context would have been a heroic task, and what is more this type of model, is a poor description of the observed reality. Therefore we are obliged to turn to a different type of model. So from the outset we assumed that agents use simple rules to make their choices and do not optimise..

As a first approach, in Weisbuch et al. (2000), we developed a simple theoretical model in which people simply learn from their previous experience and change their probability of visiting different sellers as a result of their experience. What I will argue is that models of this sort which attribute very little computational ability or general reasoning capacity to individuals may be capable of generating specific features of real markets. Although this sort of "bounded rationality" approach has received a lot of attention it is often dismissed for its lack of rigor. The reason for this is that choosing rules of thumb for agents is regarded as “ad hoc”. However, we have come to accept that the restrictions that we impose on the preferences of individuals, unlike other behavioural rules, are not ad hoc. Therefore, if we replace those assumptions, which by their very nature cannot be empirically tested, by other rules, we are subject to the criticism that we lose the rigor of "proper micro foundations". The answer to this, is, that we are interested in modelling the results of interactions between individuals following simple rules, not just as a way of justifying a theoretical equilibrium but rather as a vehicle for understanding empirical reality.

Trading relationships within the market.

In Weisbuch et al. (2000)), we considered a situation in which buyers do not anticipate the value of choosing sellers but rather develop relationships with sellers on the basis of their previous experience.

To be more precise, there are n buyers indexed by i and m sellers indexed by j. The buyers update their probability of visiting sellers on the basis of the profit that they obtained in the past from them. If we denote by $J_{ij}(t)$ the cumulated profit, up to period t, that buyer i has obtained from trading with seller j then the probability $p_{ij}(t)$ that i will visit j in that period is given by,

$$p_{ij}(t) = \frac{e^{\beta J_{ij}(t)}}{\sum_k e^{\beta J_{ik}(t)}} \qquad (1)$$

where β is a reinforcement parameter which describes how sensitive the individual is to past profits. This non-linear updating rule will be familiar from many different disciplines and is also widely used in statistical physics. It is known as the "logit" rule or, in game theory as the "quantal response" rule. The rule is based on two simple principles. Agents make probabilistic choices between actions. Actions that have generated better outcomes in the past are more likely to be used in the future. [3]This approach has the great advantage that it requires no specific attribution of rationality to the agents other than that they are more likely to do what has proved to be successful in the past. In the extremely simple case where there are just two sellers who yield identical profits for the buyers we were able to show that the latter became completely loyal if the weight they put on past experience, β was sufficiently high. However, the analysis rapidly becomes intractable when we wish to consider more general cases and then we have to resort to simulations of the model. This is a first step towards a full blown agent based model which allows us to introduce more variants on the behaviour of the individuals.

Consider, as a first example of a simple agent based model, the case in which there are three sellers in the market. Now we can fix the value of the β for the buyers and the profits π generated by the sellers and the discount rate γ and run the process[4]. At any point in time,

[3] Such a process has long been adopted and modelled by psychologists (see e.g., Bush and Mosteller, (1955)), and a more elaborate model has been constructed by Camerer and Ho (1999). . It is a special form of reinforcement learning. It has also been widely used in evolutionary and experimental game theory (see Roth and Erev (1995)). The particular form chosen here is used extensively. It is found in the model developed by Blume (1993), for example, to analyse the evolution of the use of strategies in games.

[4] Note that we could, in the simulated model, if we so wished, allow for heteroegeneity in the parameters of the individuals.

each buyer will have a probability of visiting each of the sellers. Thus he will be represented by a point in the three simplex or triangle as illustrated in figure 1 below. The nature of the relationships of all the buyers will be illustrated by a cloud of points. A buyer who shops around in a purely random way, that is who is equally likely to visit each of the three sellers will be represented as a point in the centre of the triangle. If, on the other hand, he visits one of the sellers with probability one then he can be shown as a point at one of the apexes of the triangle.

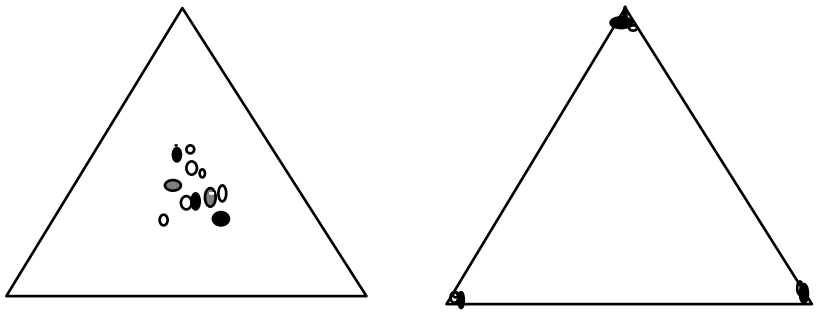

Figure 1 a Figure 1 b

Figure 1 (source Weisbuch et al. (2000))

The question now becomes, how will the cloud of points in the evolve over time? If buyers all become loyal to particular sellers then the result will be be that all the points, corresponding to the buyers will be at the apexes of the triangle as in figure 1b. This might be thought of as a situation in which the market is "ordered". The network linking buyers and sellers becomes deterministic. On the other hand, if buyers learn to search randomly amongst the sellers, then the result will be a cluster of points at the centre of the triangle, as in figure 1a. What we showed in Weisbuch et al. (2000), is that which of these situations will develop, depends crucially on three parameters β, the reinforcement factor which represents the importance attached to previous experience, γ the discount rate and π the profit per transaction. The stronger the reinforcement, the slower the individual forgets and the higher the profit obtained from sellers, the more likely is it that loyalty will emerge.

The question then is, whether or not the features of the actual market in Marseille do reflect the sort of behaviour predicted by this, admittedly primitive model? What the model suggests is that the transition from disorder to order, or more prosaically from shopping around to

loyalty as β changes, is very sharp. The change will depend on a critical value β_{ci} of β which will be different for each buyer *i* and will depend on the frequency of his visits to the market and his profit. It is easy to see why higher profits obtained will make a buyer become loyal to the seller that gave him those profits, but the importance of the frequency of visits needs a little explanation. If a buyer comes infrequently to the market then his information from previous visits is less pertinent than that the regular visitor got from his last visits. He will therefore discount previous experience more than his loyal counterpart. This will lead him to reinforce his shopping behaviour.

Before leaving the basic model, one observation is in order. The patterns illustrated in figure 1a were obtained with buyers all of whom had the same β_{ci} and the same is true for figure 1b where the value of β_{ci} is, of course, lower. But what happens in the simulations if the group is mixed as it is in reality? In the loyal buyers situation, the sellers learn to sell the correct amount because they have fixed and regular customers. In the random shopper situation, sellers cannot predict accurately the amount that they will sell. Therefore there is more waste, some buyers are unsatisfied and some sellers are left with fish on their hands. Now will it not be the case that the presence of random shoppers will interfere with the profits of the loyal and thus weaken their loyalty? Sellers will sometimes have insufficient stocks because they have been visited by random shoppers and this could have a negative effect on those who were normally loyal to those sellers. The advantage of the agent based model now becomes clear since having individuals with different β_{ci} is easy to handle whilst the formal analysis is very difficult. Consider the situation with equal proportions of loyal or low β_{ci} and high β_{ci} buyers. Interestingly, the presence of the random shoppers did not prevent the development of loyalty as is shown in figure 2. Those whose low β_{ci} led them to be loyal remain loyal and those whose high β_{ci} led them to shop around continue to do so.

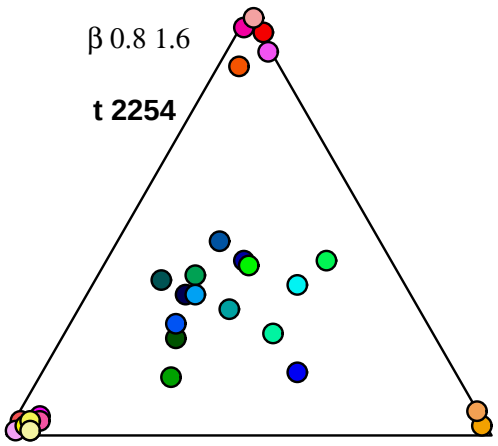

Figure 2 (source Weisbuch et al. (2000))

The existence of shoppers does not prevent those with a high β from becoming loyal.

What then should we have expected to observe on the real market if what our model suggested was right? As I have said, the important conclusion is that the division between loyal buyers and random shoppers should have been quite sharp and one should not have found individuals who shopped around to some extent but were somewhat more loyal to some sellers than to others. This is precisely what we observed on the Marseille fish market. The behavior of buyers was, in fact, highly bimodal. For each type of fish there was a concentration of buyers who only visit one seller and then a distribution of individuals who visited several sellers with a median of 4 per month as can be seen in Figure 3.

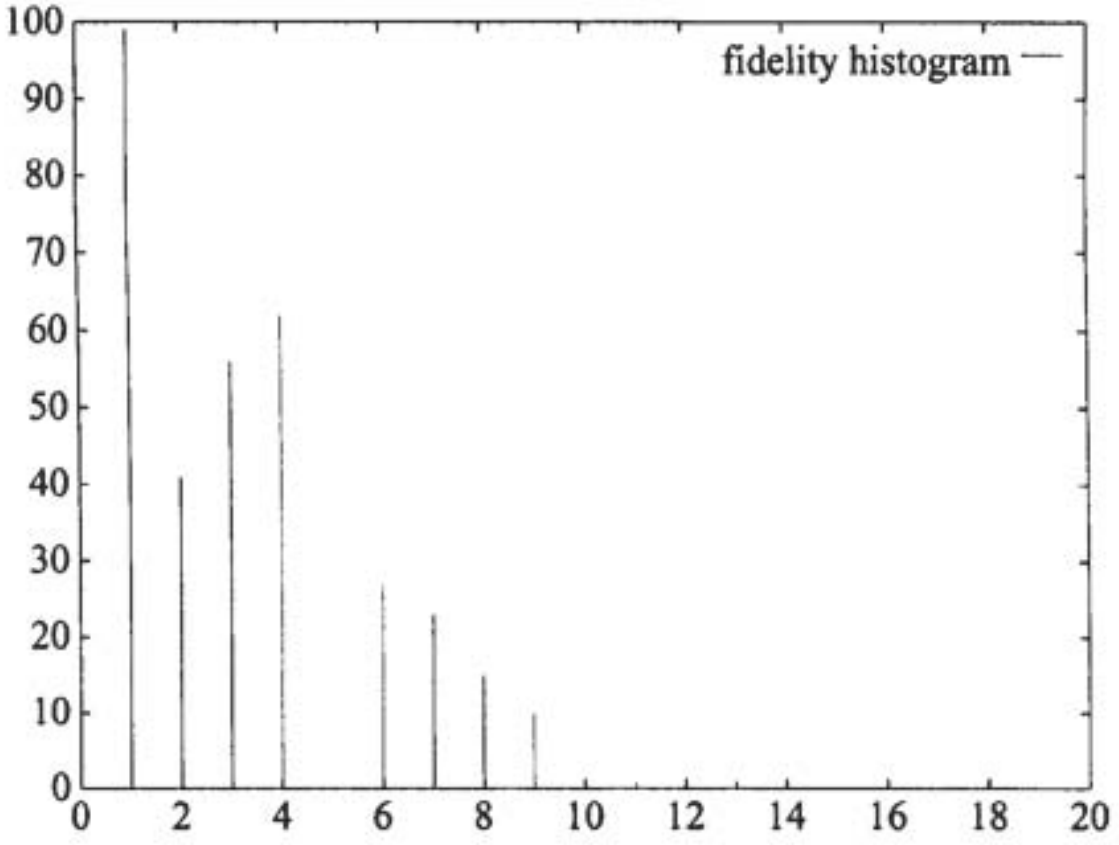

Figure 3 (source Weisbuch et al. (2000))

Loyalty Histogram

Number of sellers visited on the horizontal axis, and number of buyers on the vertical axis

The extent of the loyalty of customers for certain types of fish can be observed from the fact that, for example, 48% of all buyers bought more than 95% of their cod from one seller, the

seller of course, not being the same for all of these buyers. 33% of buyers bought more than 95% of their sole and 24% bought more than 95% of their whiting from one seller. In both the whiting and sole markets more than half of the buyers buy more than 80% from one seller. Furthermore, as the theory predicts, those sellers with the highest sales and those who come to the market most frequently are those who are most loyal.

Further analysis of an Agent Based Model

As should be clear by now, even in a relatively simple market structure, if one tries to incorporate some of the realistic features of the microscopic interaction between the actors, the model rapidly becomes analytically intractable. In Weisbuch et al. (2000) we started with a very basic theoretical model and then used that as a basis for simulating more general cases. If we go further and adopt the "agent based" modelling approach (see for example, Arthur et al. (1997), and Epstein (2007)) from the outset, we can analyse several features of the market at the same time. In such a model one hopes to find, as emergent features, a number of the salient aspects of the real empirical markets that interest us. In Kirman and Vriend (2000) we did this by developing a simple model which reproduced three of the features of the Marseille fish market. These are firstly, the division between loyalty and shopping behaviour on the part of buyers that we have already mentioned. Secondly, there is price dispersion even for the same species. Lastly, sellers learned in the model, to treat their clients in a way which corresponds to what actually happens in reality.

In the model we developed in Kirman and Vriend (1998), ten initially identical sellers and one hundred initially identical buyers met in the market hall for five thousand days for a morning and an afternoon session. They traded single individual units of a perishable commodity. Here we made two simplifications. The morning and afternoon sessions correspond to the idea that there are several opportunities to trade during the day. Using two periods allows us to take account of the idea that the possibility of trading later in the day has an influence on the prices that buyers will accept and sellers will propose early in the day. It would, of course, be more realistic to consider more trading opportunities in the day. The single unit assumption is frequently used but can be criticised on the grounds that when buyers require different amounts this may influence what they pay.

In the model, on each day the sequence of events was the following.

In the morning before the market opens the sellers purchase their supply outside the market for a given price that was identical for all sellers and constant through time. Thus, we assume that the participants on the Marseille market have no influence on what happens in the outside world. The market opens and the buyers enter the market. Each buyer requires one unit of fish per day. All buyers simultaneously choose the queue of a seller. The sellers then handle these queues during the morning session. Once the sellers have supplied all the buyers who are willing to purchase from them, the morning session ends. All those buyers who are still unsatisfied choose the queue of a seller in the afternoon. Sellers now sell to those buyers who are willing to purchase from them and the end of the afternoon session is then reached. All unsold stocks perish. Those buyers who did purchase fish, resell that fish outside the market at a given price that is identical for all buyers and constant through time. Each buyer can visit at most one seller in the morning and one seller in the afternoon.

What are the decisions with which the actors are faced ? Buyers have to choose a seller for the morning session. They then have to decide which prices to accept or reject during the morning session. If necessary, they also have to decide on a seller for the afternoon. Lastly, they must decide which prices to accept or reject during the afternoon session. Sellers have four decisions to make. They must decide what quantity to supply. They must decide how to handle the queues with which they are faced. They must decide which prices to set during the morning session and which prices to set during the afternoon session.

In the model described each individual agent uses a Classifier System for each decision and this means that each agent has four such systems "in his head". A simple stylised classifier system is presented in figure 4.

condition	action	strength
if	then	.....
.. ..	....	.....
.. ..	....	.....

Figure 4 (source Kirman and Vriend (2001))
A simple classifier system

Each classifier system consists of a set of rules. Each rule consists of a condition "if....." and an action "then....." and in addition each rule is assigned a certain strength. The classifier system decides which of the rules will be active at a given point in time. If the conditional part of the rule is satisfied it decides amongst all of those rules for which the condition is satisfied which to choose. This is done by a simple auction procedure. Each rule makes a "bid" to be the current rule and this bid = current strength + ε, where ε is white noise, a normal random variable with mean 0 and fixed variance. The rule with the highest "bid" in this auction becomes the active rule. The white noise is added to make sure that there is always some experimenting going on and hence a positive probability that a rule, however bad, will be chosen. At time *t* the classifier system updates the strength s_t of a rule that has been active and has generated a reward at time *t - 1* as follows:

$$s_t = s_{t-1} - c \cdot s_{t-1} + c \cdot reward_{t-1} \text{, where } 0 < c < 1 \text{ .} \qquad (2)$$

What the reward is, will depend on the rule in question. Supposing that in our market example the rule for the buyer is of the form, "*if the price proposed by the seller for one unit of fish in the morning is 11 euros then accept"*. The reward for using this rule would then be the profit

that is generated by using it. In this case the reward would be the price at which the unit of fish is sold on the retail market minus the price paid (11 euros). The model is initiated the with the strengths of all rules equal.

What the agents in this model are doing is learning by an even simpler version of reinforcement learning than that encountered previously. Details of the particular rules in the simulation model of the Marseille fish market can be found in Kirman and Vriend (2000).

The approach adopted here raises a fundamental question about agent-based models. At first sight it seems that almost no assumptions are imposed on the individuals, they are simply supposed to learn to use the best rule available. However, as modellers we already impose a great deal by determining the set of rules from which the agents choose. If left purely to their own devices perhaps our agents might have learned to behave otherwise. Yet, without having recourse to a full "artificial intelligence" approach it is difficult to give no structure at all to the set of choices of the agent. Thus in the sort of model we have developed our choice of rule set may have a significant effect on the outcome. Although such an approach seems to be innocent of theoretical pre-suppositions it should be noted that the very choice of the rules amongst which the agent chooses has an impact on the outcomes. Ideally, one would like to start with agents who are totally ignorant. However, this would imply that they would somehow generate a set of rules with which they would experiment. This means that they would have to learn at a very fundamental level. What is done here is in line with standard practice which is to provide the agents with a set of rules and simply note that this, to some extent, conditions the outcomes of the process. As an example, consider the fact that we would like agents to learn how to handle the queues with which they are faced. In an ideal world we would like the agents to realise that their handling of the queues is important and then for them to work out how to handle them. As it is, by giving different rules explaining how to handle queues the modeller is already biasing the behaviour of the seller by suggesting to him what it is that is important in generating his profit. However, what is not biased is the choice amongst the rules presented. Thus, the rule chosen will be the best available for handling queues amongst those presented, given the agent's experience, but he might well himself have focused on some other aspect of the market. Bearing this in mind, it is worth examining the results of the simulations and to see to what extent they reflect reality.

Once again it is important to bear in mind that in this market is that both sides are learning and what is crucial here as noted by Gale et al. (1995) and Roth and Erev (1995) is that the relative speed of learning on each side of the market will govern which outcomes occur. The importance of this will become clear as soon as we look at the results of the simulations.

Let us first look at the prices asked and accepted in the morning as shown in figure 5.

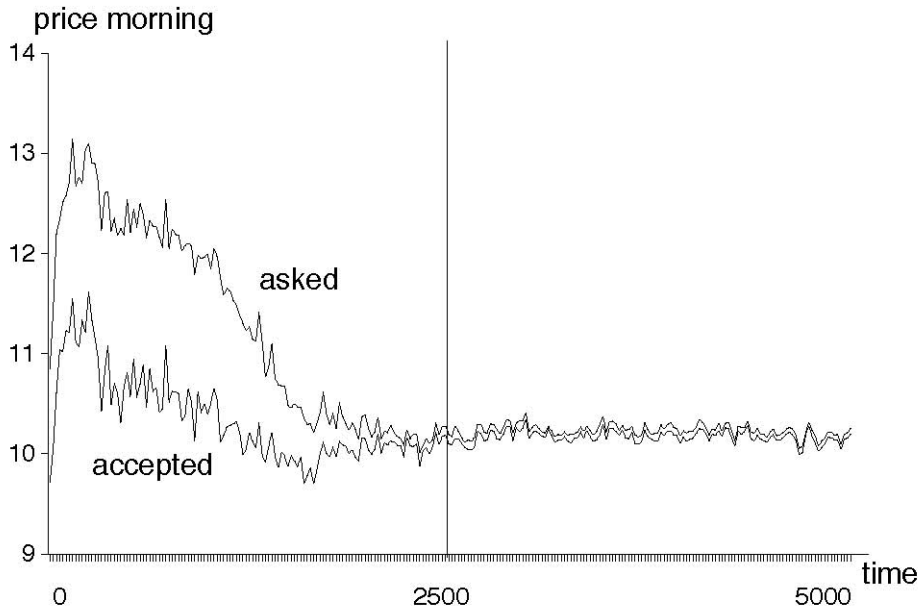

Figure 5 (source Kirman and Vriend (2001))

Prices asked and accepted at each period in time

There is, first of all, a considerable period during which learning takes place and then prices settle to 10 euros which is one euro greater than the price at which fish is bought by sellers outside the market and, hence, one greater than the perfectly competitive price. What is interesting is that during the learning period, which is what governs the final outcome, two things are going on. Sellers learn to ask prices close to the ultimatum price which is 14 euros, one less than the price at which fish can be sold on the outside market. However, buyers do not learn as quickly to accept such prices. Why is this? The answer is that initially some buyers will accept high prices having not learned to do otherwise. This will encourage

sellers to charge such prices. However buyers will start to find out that they can obtain higher profits by refusing high prices and accepting lower ones. There are always some such prices to be found. As sellers learn that buyers are not accepting their prices they start to decrease the prices asked and simultaneously as buyers observe that prices being asked are descending they start to decrease their acceptance levels. Sellers' learning "leads" that of buyers and as a result the prices converge. In the afternoon, there are also two separate learning processes going on and once again convergence occurs but to a higher price (11euros) than in the morning. The evolution of afternoon prices can be seen in figure 6.

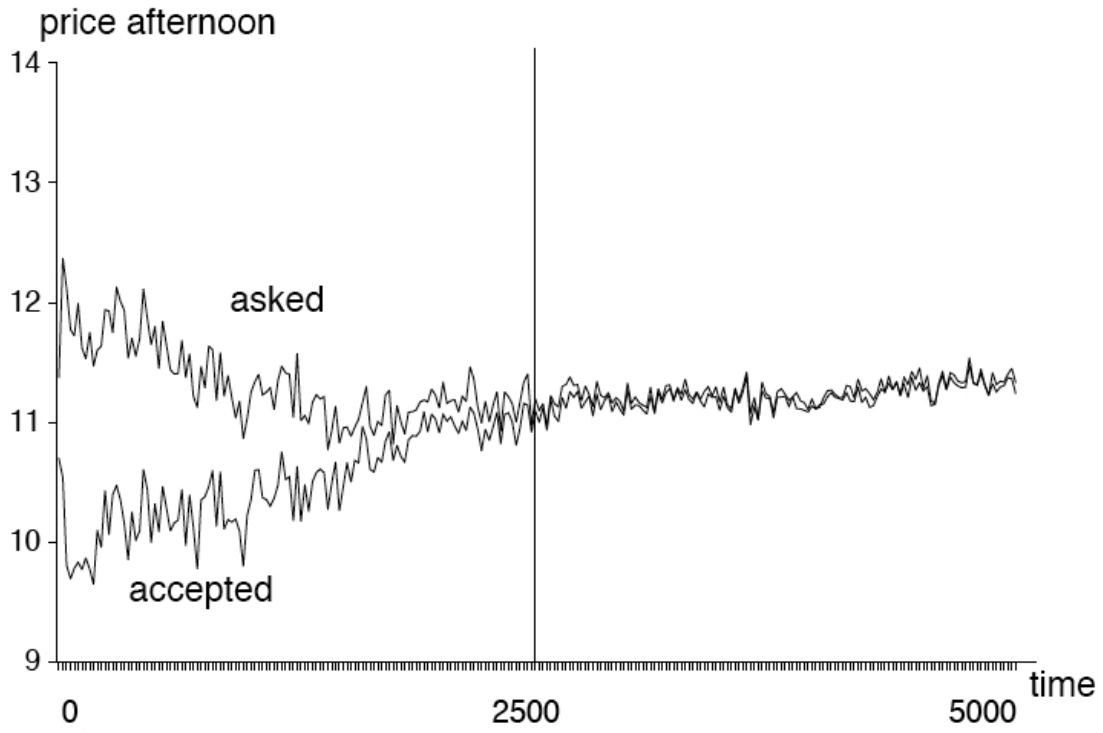

Figure 6 (source Kirman and Vriend (2001))
Prices asked and accepted in the afternoon session

This might seem extraordinary since, if buyers become aware that prices in the afternoon are higher than prices in the morning, they should presumably always buy in the morning. This is not correct. The reason is simple, those people who reappear in the afternoon have been selected. Typically, those buyers that encounter prices in the upper tail of the distribution reject in the morning. The result of this would be that the average price paid in the morning will be lower than the average price paid in the afternoon. The whole point here is that it is

not the average price that is rejected whereas what is shown in the figures is the average at any point in time. In figure 7 the price distribution over the last 2,500 days is shown and it can be seen that it does not become degenerate and concentrated on one price.

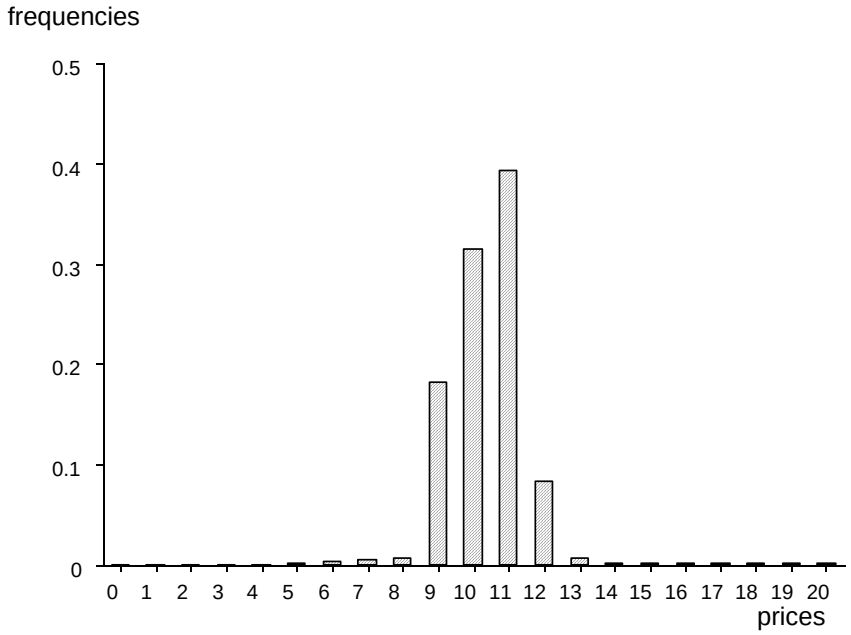

Figure 7 (source Kirman and Vriend (2001))
Price distribution over last 2500 days

Thus, a phenomenon which was observed on the real market in Marseille, as we saw earlier, emerged in our artificial fish market.

A second feature of the real market is that which has been discussed earlier, that of "loyalty". In the previous model we simply established the pattern of loyalty but did not suggest any macroeconomic consequences of that feature. To pursue this question we need to have some measure of loyalty and then to examine its impact. To do this we constructed an index of loyalty which has a value equal to one if the buyer is perfectly loyal to one seller and has a value equal to *1/n* where *n* is the number of sellers when the buyer systematically visits each seller in turn, that is, when he has the most extreme "shopping around" behaviour. More specifically, the loyalty index is given by:

$$L_{ij}(t) = \sum_{x=1}^{t} \frac{r_{ij}(t-x)}{(1+\alpha)^{t-x}} \qquad (3)$$

This is an indicator of how often buyer *i* visits seller *j*. It is a global statistic covering the whole period but there is a discount factor represented by *a*. The parameter $r_{ij}(t)$ is a counter which increases with each visit of *i* to *j*. Here we took *a =0.25* and $r_{ij}(t) = 0.25$ if buyer *i* visits seller *j* at time *t* and = *0* otherwise.

In the spirit of this approach, nothing was built into the rules of the sellers to make them privilege loyal buyers. We wanted to see whether this sort of behaviour would emerge.
The sort of rules they had were of the form:

"*If loyalty = a certain value then choose a certain probablity of serving that client",*
"*If loyalty =a certain value then charge p*"

What probability will be chosen depends on whether the seller learns to favour loyal customers or not. Which *p* is charged depends on how successful that choice turns out to be. In the simulated model it is easy to measure the profitabiity of being loyal for buyers and the profitabiity of loyal customers for sellers.

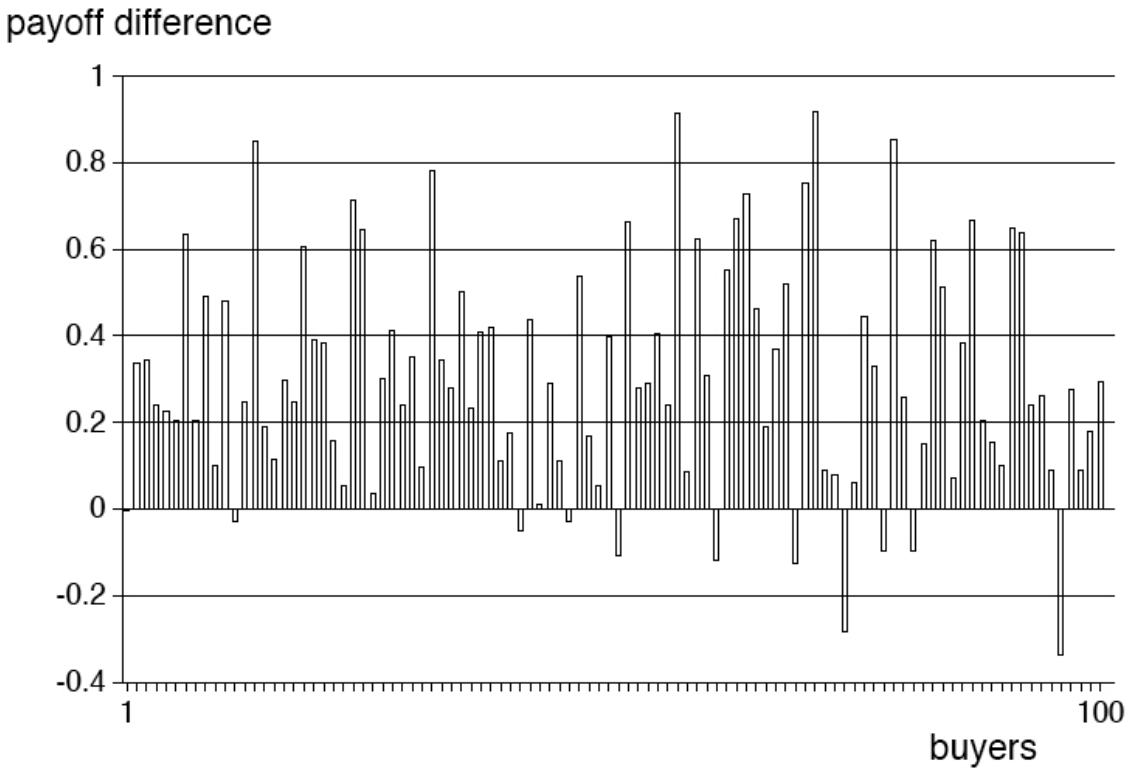

Figure 8 (source Kirman and Vriend (2001))

The relative advantage to buyers of being loyal.

What happens is that 90% of the buyers actually get a higher pay off by being loyal as can be seen in figure 8. What this means is that when basically loyal customers shop around, as they do stochastically from time to time, the profit realised on their own market is lower on average than when they buy from their regular supplier.

Furthermore, nine out of ten of the sellers get a higher profit when dealing with loyal buyers as shown in figure 9. In other words the profit, on average from dealing with a loyal customer is higher than from a random shopper. Here the difference in revenue represents the fraction of the average revenue from loyal customers above or below the average profit realised from transactions with casual buyers

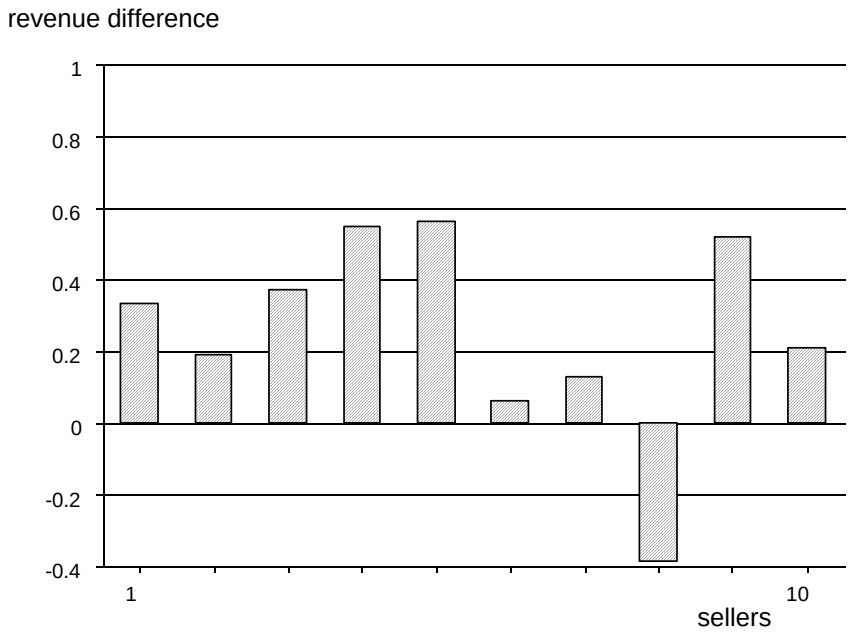

Figure 9 (source Kirman and Vriend (2001))

The relative profits obtained from dealing with loyal customers

This is a reflection of the fact that what is happening here is not a zero sum game. Only when a transaction takes place do buyers and sellers realise a profit. Thus, payoffs will be highly conditioned on acceptance and rejection, and on prices asked. The question then becomes how do loyal buyers tend to be handled by sellers? In all but one of the cases, sellers learn to give priority in the queue to loyal buyers but to charge them higher prices than random shoppers. Buyers learn that when they become loyal their profit is higher since they are more likely to be served even though they pay higher prices. Thus, loyalty is profitable both to buyers and sellers.

What about the one seller who did not find loyal customers more profitable than shoppers ? This seller learned to charge low prices to loyal customers but to give them low priority in the queue. One might ask why he did not learn to adopt the more profitable strategy learned by the other sellers. The answer here is that with the sort of local learning that is going on a move towards better service and higher prices for loyal customers can never develop. To make such a move would imply increasing prices and improving service. However, buyers will immediately observe the higher prices and will not necessarily immediately observe better service in terms of priority in the queue. This will lead them to reduce their probability of visiting this seller. As this seller observes that his customers are drifting away he will go back to his former behaviour and will therefore never learn his way to the more profitable strategy. However, it is interesting that, in the model this seller still makes profits, so he does not disappear. Thus, there is at least one explanation for the dispersion of profits that one observes on the market

This very simple rudimentary artificial fish market model manages then to reproduce some of the features of the real market. For example, it is interesting to note that, on average, in the Marseille fish market loyal buyers pay higher prices than shoppers. Those buyers who buy more than 50% of their fish per year from one seller pay, on average 5% more than the other buyers even though almost all the large buyers are loyal. Thus here we have a clear organisational feature which has emerged and which has had a very specific consequence for the market price distribution. The advantage of the type of model that I have described, is that it can always be extended to examine other aspects of the real market whereas to attempt to construct a theoretical model which incorporates all of these would be a more than ambitious task.

Getting the model wrong.

As a second example of how one can use agent based modelling consider the situation in which agents try to learn about their environment but have a mistaken vision of that environment. Two questions arise. First, will the agents learn to have a model of their environment which would yield data consistent with the empirical data ? Second, even if the answer is positive, will the model be correct ? The last question needs a little clarification. Another way to phrase it is to ask, if the agents, given the model that they have come to believe in, get to an equilibrium of that model, will it be an equilibrium corresponding to the « true » model had they known it ? As a very simple example, consider the case in which the agents start out by having an erroneous idea about the n of the n person game that they are playing. In Brousseau and Kirman (1993) we pursued an idea developed in Kirman (1985) and analysed a model to see whether, in that model, by learning, they might realize their mistake or alternatively might converge to a situation in which their beliefs are comforted by their observations (for early examples of this see Kirman (1983) and Nyarko [1991]). Such self-fulfilling expectations are of course related to those in the well known "sunspots" literature (see for example Woodford [1990]).

It might seem strange to think of players not being aware of the number of their opponents and thus not taking account of the actions of some of them. However in economic applications this is very plausible. What general equilibrium teaches us is that everything depends on everything else in the economy. In principle, at least in a finite economy where everybody has positive weight, an agent should therefore take account of the actions of all the other actors in the economy. This is clearly unrealistic, although it is possible to treat the full market problem as a game and examine its Nash equilibrium. However, even in much more circumscribed situations such as that of an oligopoly, who the players are, is far from clear. If products are differentiated, how different do they have to be from one's own before their producers can reasonably be ignored? How far does one supermarket have to be from another for it not to be considered as a rival? These are natural questions for the economist who applies game theory to his problems and make it worth considering what happens when a player fails to take account of the actions of some of his opponents.

The essential problem with learning in these situations, as in the first example, is that the action of the individual who is trying to learn, influences the outcomes that he is trying to

learn from. He is unable, when looking at his pay-offs, to distinguish the consequences of his own behaviour from those of the other players, or indeed from the consequences of "nature's" actions. This poses a very difficult theoretical question. As one learns, one is trying to improve one's pay-off and there is an implicit contradiction between this goal in the short run and learning more for the long run (see Green [1983] and Aghion et al. [1993] and Hanaki et al. (2005). This involves a problem of the optimal trade-off between « exploration and exploitation ». How much should I be prepared to sacrifice in profit today to have better information and thus to do better in the future, (see Weisbuch et al. (1998))? Even in the monopoly case with a linear demand curve and two unknown parameters (see Balvers and Cosimano [1990], Easley and Kiefer [1988] and Kiefer and Nyarko [1989]). The oligopoly problem is even more difficult, (see Huck et al. (2004) and Bischi et al. (2008). Thus we adopted a more limited approach and did no more than attribute rather simple learning behaviour to our individuals.

We considered a particularly simple example I developed in earlier papers (Kirman (1975 and 1983)) in which, in a duopoly game, each of the two players fails to take account of the other's existence and tries to maximise what he believes to be his short term monopoly profit. I shall just mention briefly a series of results which show that a whole class of self-sustaining equilibria can be attained from particular starting conditions if the agents use least squares learning. Whereas it was conjectured in an earlier paper that the learning process would converge, in general, we showed, developing results in Brousseau and Kirman (1992) that this apparent convergence is due to the slowing down of the process by the weight of the memory. This evokes the decelerating cyclical behaviour found by Shapley [1964]. As in his case, truncating the memory changes the behaviour radically. We examined, in some detail, the consequences of this for price dynamics. We also showed, in an agent based model, that reducing the weight on earlier observations and using a continuous time approximation makes the process, in general, unstable.

We found regions of stability and also complicated dynamics in the short memory game which evokes the idea of a chaotic process. Thus, even with a very simple example rather complicated phenomena may arise as individuals try to learn about an environment which they themselves influence. Learning about the wrong game may lead, by chance, to self-fulfilling expectations but may also lead to a complicated evolution of the pay-offs to the participants, particularly if they do not have very long memories.

It should be clear that the complexity of the dynamics of this simple example is due to the feedback from the players' own strategy and the unobserved strategy of his opponent into the variable that he observes. However, in the short memory game, where dependence on initial conditions become very indirect, price dynamics though complex are not completely unpredictable.

The prices are captured by a "pseudo-limit" but then move away from it. Thus there is, as in the case of full memory, apparent convergence. However, in the short term memory case, the convergence to a certain region cannot be attributed to a slowing down of the process. Indeed the "convergence" in the 5 period memory case, although slower than in the 2 period case, is stronger in the sense that deviations from the "pseudo limit" are smaller. Finally, this limit itself depends on the length of the memory.

Thus, as we will see, the convergence of the learning process in the simple game that we studied is heavily dependent on the behaviour of omitted players. This observation is of particular interest to those who wish to apply game theoretic reasoning in empirical economics.

A simple duopoly game

Consider a symmetric duopoly in which the demand functions for firms 1 and 2 are given by the "true model":

$$d_1\left(p_1(t), p_2(t)\right) = \alpha - \beta p_1(t) + \gamma p_2(t) \qquad (4)$$

$$d_2\left(p_1(t), p_2(t)\right) = \alpha - \beta p_2(t) + \gamma p_1(t) \qquad (5)$$

where $p_i(t)$ is the price set by firm i at time t. Assume, in the tradition of Cournot, that production is costless, and the pay-off to each firm is then given by

$$\Pi_i\left(p_1(t), p_2(t)\right) = p_i(t) d_i(p_1(t), p_2(t)) \qquad i = 1,2 \tag{6}$$

Now assume, as in Kirman (1983), that the two firms, through ignorance or inertia, are unaware that their demand depends on each other's actions. As I have suggested, in a duopoly situation, such an assumption is implausible, but it is more realistic in a several firm situation in which each firm feels unable to take explicit account of the behaviour of all the opponents and hence focuses on the "own-price" demand curve or on a demand curve involving only some of the prices of its opponents and adds a random term to take account of the, to him, unpredictable behaviour of the other firms. All the results can be generalised to an *n*-firm model, such as that developed by Gates et al. (1977).

The two firms will thus have the following "perceived model":

$$d_i\left(p_i(t)\right) = a_i - b_i p_i(t) + \varepsilon_i(t) \tag{7}$$

We did not make any specific assumptions about the distribution of the error terms although to be rigourous our agents should make specific assumptions if they adopt the learning procedure suggested.

Now look at how the players should learn about the parameters of their model with the two questions that I posed at the outset, in mind:

i) does the learning process converge?
ii) if so, does the limit correspond to a solution of the true game?

If ignorance is only partial, in the sense that players believe with certainty that $b_i = \beta$ the "slope of the true demand curve" and try to learn about a_i it is easily shown, using a fictitious play argument, (Kirman (1975)) that prices will converge to the Cournot-Nash solution

$$p_i^* = \frac{\alpha}{2\beta - \gamma} \tag{8}$$

Now consider the case where neither of the parameters a_i or b_i is assumed to be fixed. If players maximise one period payoffs [i] and if they believe that $E(\varepsilon_i(t))=0$ then the optimal price or strategy is given by

$$p_i(t)=\frac{\hat{a}_i(t)}{2\hat{b}_i(t)} \tag{9}$$

where $\hat{a}_i(t)$ *and* $\hat{b}_i(t)$ are the estimates at time *t* of the two parameters given the quantities and prices observed up to that point.

So, at each period, given its estimates, each firm will charge a price, and the demand realized as a result of these prices will, of course, be given by the true model specified by equations (1) and (2). This new observation of a price-quantity pair will lead to a revision of the estimates of the parameters and, in turn, to new prices and so forth.

It is now standard practice in the case where there is ignorance of both the parameters to assume that each firm tries to fit the observed data by means of least squares. This can be justified from several viewpoints such as the Bayesian (see Zellner [1971]) or as a special case of general updating processes (see Aoki (1976)). The model we considered is a special case of an early model developed by Gates, Rickard and Wilson (1977), in which they allowed for *n* firms and variable weights for preceding observations. They were obliged, however, to confine their attention to particular cases to obtain analytic results.

In our particular model, the ordinary least square estimates for a_i and b_i are given by:

$$\hat{b}(t)=-\frac{\sum_{k=1}^{t-1}\left(d_i(k)-\bar{d}_i(t)\right)\left(p_i(k)-\bar{p}_i(t)\right)}{\sum_{k=1}^{t-1}\left(p_i(k)-\bar{p}_i(t)\right)^2} \tag{10}$$

and

$$\hat{a}_i(t) = \bar{d}_i(t) + \hat{b}_i(t)\bar{p}_i(t) \quad (i = 1,2) \tag{11}$$

where

$$\bar{d}_i(t) = \frac{\sum_{k=1}^{t-1} d_i(k)}{t-1} \quad \text{and} \quad \bar{p}_i(t) = \frac{\sum_{k=1}^{t-1} p_i(k)}{t-1}$$

Observe that (7) can be rewritten for firm 1, for example, as

$$\hat{b}_i(t) = \beta - \gamma \frac{\sum_{k=1}^{t-1} [p_1(k) - \bar{p}_1(t)][p_2(k) - \bar{p}_2(t)]}{\sum_{k=1}^{t-1} [p_1(k) - \bar{p}_1(t)]^2} \tag{12}$$

and similarly for firm 2.

Given this, it is easy to see why estimates of the parameters of the perceived model are influenced by the behaviour of the opponent. The second term in (12) is the covariance of the prices or the bias due to the omission of a variable which is correlated with one of the included variables. Indeed, it is precisely the fact that the prices are inter-related in this way that generates problems in the evolution of the system.

The whole system is clearly recursive. Hence, from the equation for the true demand (4), we have for firm 1, for example,

$$p_1(t) = \frac{\left\lfloor (\alpha + \gamma\bar{p}_2(t)) \sum_{k=1}^{t-1} [p_1(k) - \bar{p}_1(t)]^2 \right\rfloor - \gamma\bar{p}_1(t) \sum_{k=1}^{t-1} [p_1(k) - \bar{p}_1(t)][p_2(k) - \bar{p}_2(t)]}{2\beta\left(\sum_{k=1}^{t-1} [p_1(k) - \bar{p}_1(t)]^2 - \gamma \sum_{k=1}^{t-1} [p_1(k) - \bar{p}_1(t)][p_2(k) - \bar{p}_2(t)]\right)} \tag{13}$$

and similarly for firm 2.

It is apparent that it is not a trivial matter to establish whether, for this recurrence relation, convergence does or does not occur. If we are interested in examining what happens, in

general, in this process we have to resort to simulations. But before doing that it is just worth making a few observations about this process.

Firstly mistaken beliefs can be sustained. In other words, there are equilibria in which the observations made by players confirm their mistaken beliefs about the model and hence their own behaviour leads such equilibria to persist. First, for any positive prices p_1 and p_2 there are estimates of the parameters a_1 and b_2 such that were the individuals to make these estimates they would never move from those prices. For given p_1^* and p_2^* these are given by

$$b_1^* = -\beta + \frac{\alpha + \gamma p_2^*}{p_1^*}$$

$$b_2^* = -\beta + \frac{\alpha + \gamma p_1^*}{p_2^*}$$

and

$$a_1^* = 2\left(\alpha - \beta p_1^* + \gamma p_2^*\right)$$

$$a_2^* = 2\left(\alpha - \beta p_2^* + \gamma p_1^*\right)$$

Clearly if at some $\bar{t}$ it is the case that

$$\hat{a}_i(\bar{t}) = a_i^* \text{ and } \hat{b}_i(\bar{t}) = b_i^*$$

then $p_i(t) = p_i^*$ for $i=1, 2$ and all $t \geq \bar{t}$.

Furthermore we can actually find initial conditions such that the least squares learning procedure will subsequently converge to values of the parameters, in the set defined above, (for details see Kirman (1983)).

One interesting feature of this type of equilibrium is that it contains the cooperative, joint monopoly solution but not the Cournot Nash equilibrium. The joint monopoly solution is achieved if in the first two periods, *t=1,2,*

$$p_1(t) = p_2(t)$$

then for all $t \geq 3$

$$p_1(t) = p_2(t) = \frac{\alpha}{2(\beta - \gamma)} \tag{14}$$

Thus, curiously, in this example, unconscious imitation leads to a cooperative solution. This emphasises the idea that the process can lead, by chance, to players coordinating on an equilibrium which is Pareto superior to the Nash equilibrium.

Simulations

The only theoretical results available on the learning process just described are that if the process does converge it must be to one of the equilibria that can be reached in two steps as just described. Suppose now that we simply simulate a model with agents who use linear least squares learning for the misperceived duopoly model. What we find is that, in general process has the curious property that it cycles but more and more slowly. Thus, apparently it converges but closer examination reveals that it does not do so. The origin of the phenomenon is the infinite memory attributed to the agents. However, it is not difficult to modify the model to examine this. But, again this has to be done by simulating the model. At the risk of being repetitive let me emphasise again that since any simulation must involve specifying the parameters or rules that characterise the agents, the choices can always be described as « ad hoc ». Nevertheless in our case there seems to be a reasonable solution to the difficulties involved with the unlimited memory least squares learning, and that is simply to truncate individuals' memories. Since this removes the direct influence of the initial conditions one might hope that this would increase the possibility of obtaining stability of equilibria. However, Shapley's (1964) example, in which he showed that « fictitious play » could lead strategies to cycle with finite memory, would seem to suggest the opposite. The idea of limiting the number of previous observations taken into account is not new and an early example was that of Gates, Rickard and Wilson [1977], who working in a similar context

considered the case in which agents used only two previous observations. However, curiously they assumed that at period t agents used the (t-1)st and first observation. When convergence occurs in this case, the limit clearly depends on the initial conditions. Smale (1980) emphasised the role of bounded memory and suggested that agents could only keep some summary statistics for the past.
In the context of the duopoly example there is a basic problem. Supposing the process were to converge after a finite number of periods, if the agents were to have only finite memories the process would no longer be defined. The reason for this is intuitively obvious. Once the process stops evolving prices are constant and agents only observe a point and their estimates of the parameters of their demand curve are no longer defined.

When we simulate the process with memories of differing but finite lengths we observe behaviour which suggests convergence in the sense that the process fluctuates close to certain values and then moves away, sometimes fairly far before again returning to the values in question. Thus when it does not cycle, it exhibits erratic behaviour, but around a well defined point.

To get an idea of what is going on, take the simplest case with individuals whose memory is limited to just two observations. Thus at each point in time they have a perfectly fitting demand curve. In this case the recursive relation defining the price process reduces to

$$p_i(t) = \frac{\alpha\left(p_i(t-2) - p_i(t-1)\right) + \gamma\left(p_i(t-2)p_j(t-1) - p_j(t-2)p_i(t-1)\right)}{\beta\left(p_i(t-2) - p_i(t-1)\right) - \gamma\left(p_j(t-2) - p_j(t-1)\right)} \quad (15)$$

i=1,2 $\quad i \neq j$

In this case the process oscillates around the values

$$p_1^* = p_2^* = \frac{\alpha}{2\beta}$$

Thus, it might seem that this would be a solution which would correspond to that in which the players knew the values of the parameters affecting their variable and simply attributed zero weight to the strategy of the opponent. However, unfortunately, this is not correct. In order

to sustain the prices p_i^* the values of the estimates $\hat{a}_i$ and $\hat{b}_i$ must be given by,

$$\hat{a}_i = \alpha + \frac{\alpha\gamma}{\beta} \qquad (16)$$

$$\hat{b}_i = \beta + \gamma \qquad (17)$$

Thus the rôle of the interference of the other player becomes very clear and indeed as γ diminishes we obviously return to the monopoly case.

In Figure 10 the time path of the process is shown.

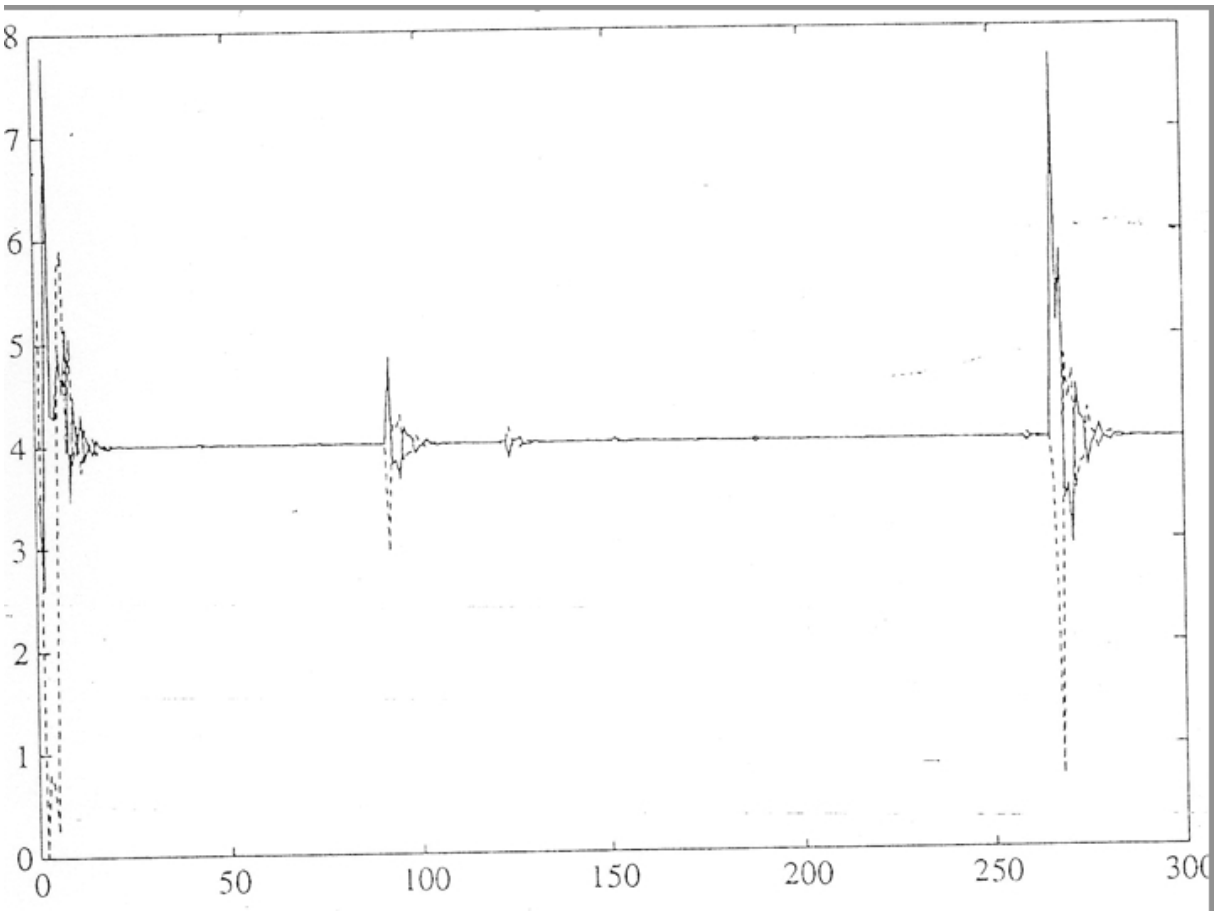

Figure 10 (source Brousseau and Kirman (1993))
The evolution of prices with two period memory

It should be observed that, although it seems to settle down after the first few periods, it does not in fact remain there and the movement which can be seen around period 80 recurs in

amplified form aroung period 260 and again later if the simulation is continued. It is not therefore, as has been remarked already, possible to argue that the process converges to $\left(\frac{\alpha}{2\beta}, \frac{\alpha}{2\beta}\right)$, but one could think of this as an accumulation point of the attractor of the process.

In some simulations the process briefly oscillated wildly, even attaining negative prices. However, in some special cases it is possible to show that there are "stable regimes" in the sense that the process will not leave that region once it has attained a point within it.

Next we tried increasing the length of memory to 5 periods. In Figure 11, the process is illustrated.

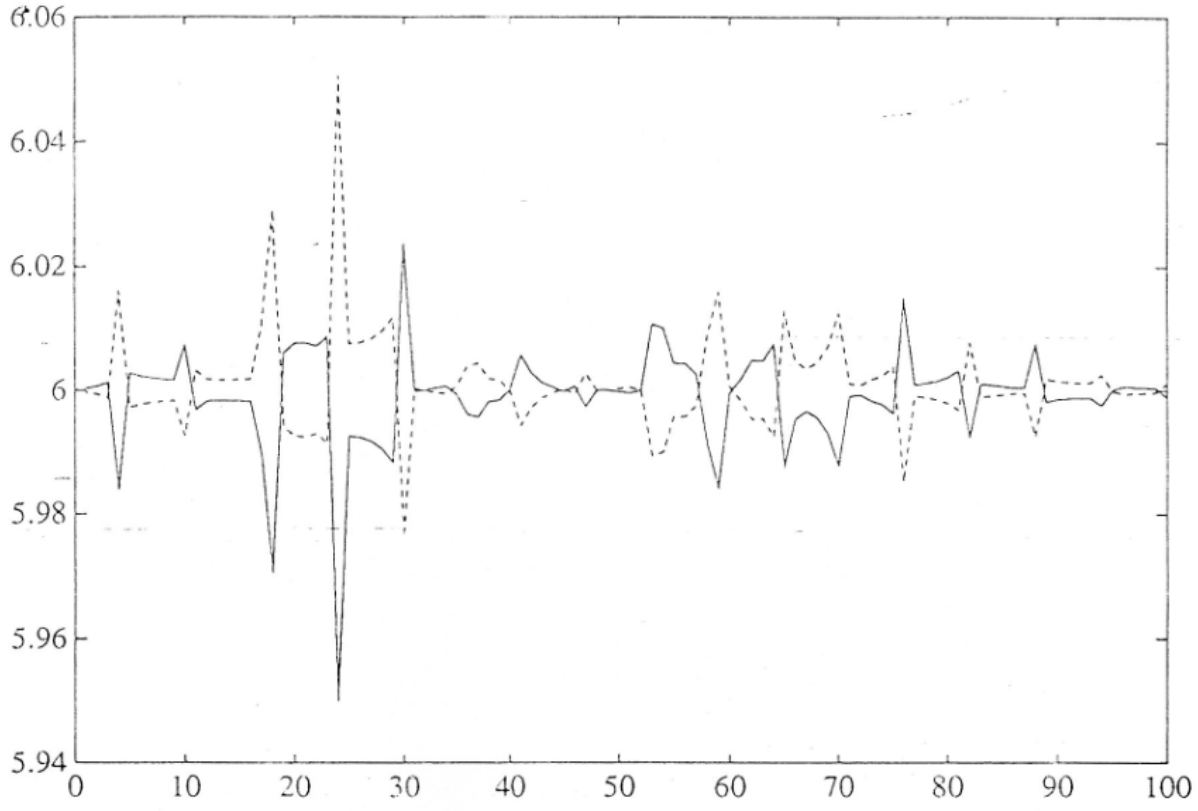

Figure 11 (Source Brousseau and Kirman (1993))
Evolution of prices over time with 5 period memory

Here, the prices came close to a point given by $\frac{\alpha}{2\beta - \gamma}$. Although once again the two prices given by the dotted and solid lines seem to converge after 100 periods, in fact sufficiently

long simulations cause movements away from these points again. This is due of course to the particular estimation procedure being used, and it is an open question as to whether this would happen with other types of learning. As a simple comparison, we showed an example with no limits on memory, and here we can see a clear apparent convergence of prices, but in this case however to an asymmetric solution (Figure 12).

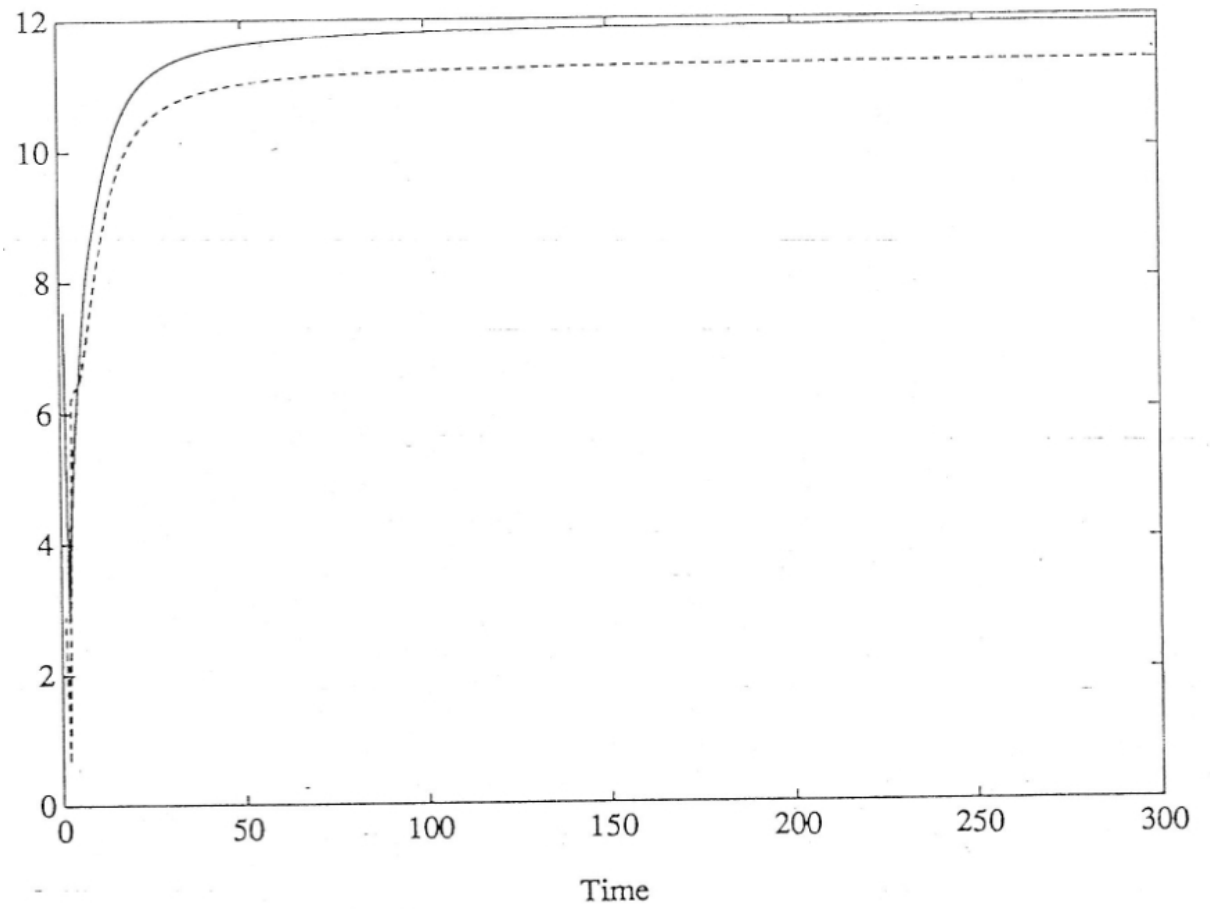

Figure 12 (Source Brousseau and Kirman (1993))

Evolution of prices with unlimited memory : convergence to an asymmetric solution

Here we found that the simulations often had a new "pseudo-limit" which was no longer $\frac{\alpha}{2\beta}, \frac{\alpha}{2\beta}$ but rather $\left(\frac{\alpha}{2\beta-\gamma}, \frac{\alpha}{2\beta-\gamma}\right)$ the Nash equilibrium. This is of particular interest, since it is, at first sight, at odds with the theoretical results, since the Nash equilibrium is not a self sustaining equilibrium. However, what is important to understand here is that the limit points of the process here are not « self sustaining » in the strict sense. But, what is more interesting is that in the model simulated in Hanaki et al. (2010)), discussed below, where

players in an oligopoly, were completely ignorant of each others' strategies pay-offs and the structure of the game, the prices converged rapidly to the Nash equilibrium.

One final observation concerning the simple example studied here is that the periodic divergence of the process seems to be due to the lack of movement of prices near the "pseudo limit" since, as I have said, when prices converge to a single value the demand curve becomes undefined. If this were the full explanation then it would seem intuitively that the periodic disturbance should disappear if small "trembles" were added to the game so that individuals would always have dispersed points from which to estimate their parameters. However simulations with white noise added to the true pay-off functions simply produced larger fluctuations around the "pseudo limit". Thus this noise did not affect the basic characteristics of the underlying process, and this is because the random movements generated by the error, come to dominate the process and the estimated demand curve will experience shifts which become unrelated to the underlying process and can lead to large changes in the corresponding prices. By simulating the agent based model this became apparent whilst the sporadic large changes in prices would not have shown up when trying to analyse the limit points of a theoretical model.

Complete ignorance

In the previous discussion individuals had the wrong model and this led them to converge to a solution which was not an equilibrium of the « true game ». However, the agents were far from being totally unsophisticated. Given their partial knowledge of the situation their best estimates and decisions were erroneous. Now, one might ask, what happens if they are much less sophisticated and have no idea what is going on ? In this case they might simply use reinforcement learning on the basis of their own experience.[5] In Hanaki et al. (2010) we performed simulations of models in which agents used simple reinforcement learning and examined what happened if they updated their probability of using rules on the basis of the standard exponential model.

To be more precise we simulated a model in which N agents were faced in an oligopoly model with the following demand curve,

[5] Vriend (2000) discusses the difference between situations in which agents use only their own experience and those in which they also benefit from the experience of others. In many situations it will not be clear to agents what actions the other players have chosen nor the payments obtained from so doing and that is why in Hanaki et al.(2010), we stick to using agents' own experience.

$$d_i = 2.0 - 1.5 p_i + \frac{1}{N-1} \sum_{j \neq i} p_j \qquad (18)$$

Once again the difficulty with using simulated agent based models is that one has to specify numerical values for the parameters of the model in question. However, one can do a systematic search of the parameter space in order to ascertain the robustness of the results. In our simple example the profit for firm *i* is taken to be $\pi_i = \max(p_i d_i, 0)$ and the Nash Equilibrium prices are $p_i=1$ for all *i.* Now, consider the possible choices for a buyer and let them be given by,

$$p^i \in \{0.0, 0.1, 0.2, ..., 2.9\}$$

So the agent has to choose which price to charge and, in the light of his experience attributes a weight $A_i^p(t)$ to the price *p* at time *t.* Given these weights the probabilities of choosing the prices are determined using the well-known exponential rule, that is,

$$\text{prob}(p^i(t) = p) = \frac{\exp(\lambda(t) A_i^p(t))}{\sum \exp(\lambda(t) A_i^k(t))} \qquad (19)$$

The weight that agents put on the prices that they have chosen is updated in such a way that the weight they put on experience in the far past does not take on too much importance, i.e.

$$\lambda(t) = \min(0.0025t, 200)$$

Given this the weights are updated as follows,

$$A_i^p(t+1) = w A_i^p(t) + (1-w) R_i^p(t) \qquad (20)$$

The only remaining thing to determine then is the profit to add at each step and here there are two possible choices, with and without discounting, So the general rule is given by,

$$A_i^p(t+1) = w A_i^p(t) + (1-w) R_i^p(t)$$

$$\text{where } R_i^p(t) = \begin{cases} \pi^i(t) \text{ if } p = p^i(t) \\ A_i^p(t) \text{ otherwise } \textbf{(}\textit{without discounting}\textbf{)} \\ 0 \text{ otherwise } \textbf{(}\textit{with discounting}\textbf{)} \end{cases} \qquad (21)$$

The role of discounting turns out to be very important, for without discounting the process converges to the Nash equilibrium but, as soon as discounting is introduced, a wide range of final prices becomes possible. This is illustrated in figures 13 and 14.

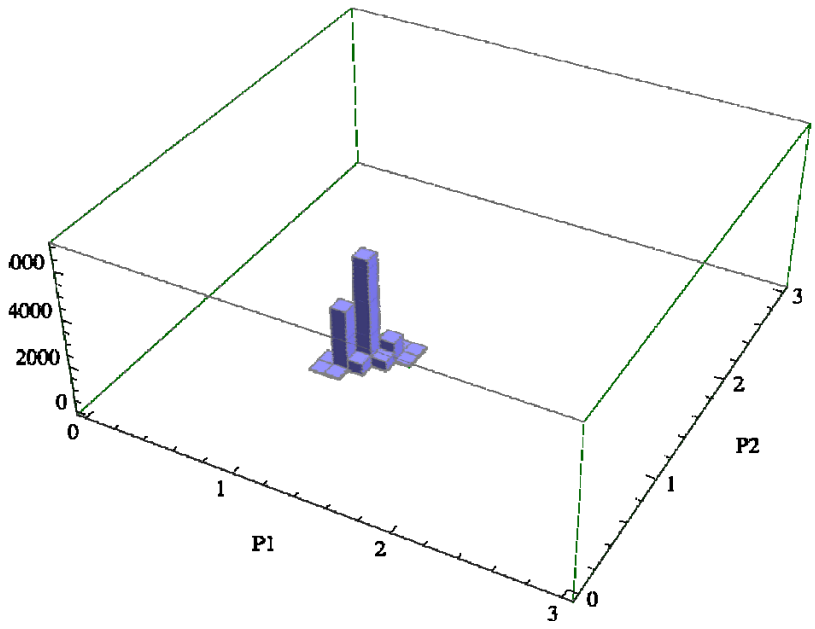

Figure 13 (Source Hanaki et al. (2010))
Price limits without discouning past pay-offs