



Can they beat the Cournot equilibrium? Learning with memory and convergence to equilibria in a Cournot oligopoly

Thomas Vallée, Murat Yildizoglu

► To cite this version:

Thomas Vallée, Murat Yildizoglu. Can they beat the Cournot equilibrium? Learning with memory and convergence to equilibria in a Cournot oligopoly. 2010. hal-00526258

HAL Id: hal-00526258

<https://hal.science/hal-00526258>

Preprint submitted on 15 Oct 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Can they beat the Cournot equilibrium? Learning with memory and convergence to equilibria in a Cournot oligopoly

Thomas Vallée (*)
Murat Yildizoglu (**)

2010/22

(*) LEMNA – Université de Nantes
(**) GREQAM – Université d'Aix-Marseille

Can they beat the Cournot equilibrium? Learning with memory and convergence to equilibria in a Cournot oligopoly

Thomas Vallée^λ and Murat Yıldızoğlu^γ

September 18, 2010

^λLEMNA, IEMN - IAE
Université de Nantes
Chemin de la Censive du Tertre.
BP 52231. F-44322 Nantes, France

Thomas.Vallee@univ-nantes.fr

^γGREQAM (UMR CNRS 6579)
Aix Marseille University
Centre de la Vieille Charité.
2, rue de la Charité
F-13002 Marseille cedex 02
murat.yildizoglu@univ-cezanne.fr

Abstract

This article analyses the possibility of firms learning collusive solutions in a Cournot quantity game. Starting from the results of Vallée and Yıldızoğlu (2009) and of Alós-Ferrer (2004), we study the role of random experimenting, social learning (imitation), and (up-dated) memory in helping firms to discover more collusive market configurations than those of the Cournot equilibrium (CE). We show that long memory and its update is necessary to achieve such configurations.

JEL codes: L130; L200; D430; C630; C730.

Keywords: Cournot oligopoly; Learning; Evolution; Selection; Evolutionary stability; Nash equilibrium; Collusion.

Contents

1	Experimental setup and protocol	4
1.1	A simple oligopoly game	4
1.2	Agent learning	5
1.2.1	Random experiments	6
1.2.2	Imitation	6
1.2.3	Memory and updating	6
1.3	Simulation protocol	6
2	Computational results	7
2.1	Convergence with collective memory	7
2.2	Individual memory	9
2.3	Convergence with individual memory and imitation	10
2.4	Beating the CE: Convergence with individual memory and updating	11

3	Theoretical analysis of the learning mechanisms	12
3.1	Pure evolutionary social learning: experimentation and imitation	13
3.2	Experimentation and short individual memory ($K = 1$)	14
3.3	Experimentation and long memory ($K > 1$) with updating	16
3.4	Three main forces: a synthesis	18
3.5	Increasing the number of players and updating	18
4	Validation of analytical results through experiments	20
4.1	Convergence to collusive solutions	20
4.2	Role of the memory size	21
5	Conclusion	23
A	Appendix	26
A.1	Numerical specification	26
A.2	Pay-offs in the two-players game matrices	26

Introduction

Recent debate (Vriend (2000) and Arifovic and Maschek (2006), Vallée and Yıldızoğlu (2009)) has once again questioned the convergence to specific equilibria in the Cournot oligopoly game¹. This debate is a consequence of assumptions on the nature of the firms' learning in the game, contrasting individual and social learning. The underlying determinants of convergence are the necessary conditions for the emergence, in the game, of an evolutionary stable equilibrium (ESE). It is now accepted that, under the definition of ESE given by Maynard Smith (1982), the convergence to a specific equilibrium in a finite population game can be quite sensitive to the specifics of the selection process, (see, for example, Ficici and Polack (2000)).

An evolutionary process combines two basic elements: a mutation mechanism that brings diversity in the population and a selection mechanism that favors particular individuals in the population. The recent debate on individual *vs* social learning concerns the nature of the selection mechanism in the evolutionary Cournot game. The distinction between the individual and the social dimensions of learning implies a distinction between the different sources of fitness taken into account by the selection mechanism: selection solely based on individual performance, or selection that considers the performance of others, i.e. relative performance. Only in the latter case does Hamilton's *spite effect* (1970) play a role in the evolution of the strategies: will a firm adopt a strategy that increases its performance, if it improves competitors' performances even more ? The consequences of the spite effect can only be accurately taken into account in a one-to-one game based on relative performances. Morgan and Steiglitz (2003) develop the concept of 'spiteful replicator dynamics' using such a modeling strategy: replicator dynamics are no longer only based on a comparison of individual performance with average performance, but also on the comparison of the relative performance of the competing individuals. They show that such a selection mechanism can yield new dynamics that differ from those of the standard replicator equation. For example, it becomes possible to show that the spite effect can explain why an ESE is not necessarily a Nash equilibrium in a finite population.

Vallée and Yıldızoğlu (2009) use the framework of the Cournot oligopoly to study the importance of spiteful behavior and its determinants: mutation (innovation) and imitation (selection).

¹See Ania (2008) and Hehenkamp *et al.* (2010) for similar questions in Bertrand oligopoly games.

They show that the Walras equilibrium (WE) is quite robustly stable under general conditions, when learning is based on imitation and random experimenting (mutations) (see also Vega-Redondo (1997)). This result stems from the spite effect, which appears when learning possess a social dimension (imitation), and when dynamics are based on selection, hence on relative performance of firms. Using Genetic Algorithms (GA), they show that, when the GA represent social learning, convergence can only occur to WE. In the case of individual learning, convergence to Cournot equilibrium (CE) is only possible if the interactions of the firms allow them to discover the decreasing relationship between the market price and their quantities. This is, for example, impossible with learning based on hypothetical profits (as discussed by Arifovic and Maaschek (2006)).

These results clarify the conditions of convergence to equilibria under imitation and mutation based learning. One of the shortcomings of this learning mechanism is the absence of memory. In fact, we could not even consider the mutation based learning as a 'trial and error' process, because the agents cannot compare their performance obtained through the experiment with their previous results, and decide if the trial corresponds to an error or not. To be able to make this comparison, they must possess at least a one period memory of their previous strategy and performance.

Starting from this observation, other recent articles extend the initial model of Vega-Redondo (1997) by assuming that agents can remember their past strategies and corresponding performances. Alós-Ferrer (2004) considers a collective memory by assuming that agents can imitate past best strategies of the industry (composing the set that contains, for each period, the strategies that gave firms the highest profits). He shows that the learning process will converge, under this assumption, to any symmetrical quantity vector with $q_j \in [q^{CE}, q^{WE}]$. Bergin and Bernhardt (2004) show, on the other hand, that learning with individual memory implies convergence to CE. These results indicate that the only absorbing states of the learning dynamics with memory are $[q^{CE}, q^{WE}]$.

It is interesting to note that these conclusions are not always confirmed by experiments on Cournot oligopolies. The results of Huck, Normann, Oechssler (1999) and of Apesteguia, Huck, Oechssler (2007) conform with these conclusions (convergence to CE or WE as a function of the information given to the players). But, the initial oligopoly experiments of Fouraker and Siegel (1963) and later ones of Offerman, Potters, Sonnemans (2002) observe more collusive quantities than CE and, even a convergence towards the collusive solution (CS). In these latter experiments, the informational conditions and the possibility of memorizing past observations plays an important role in the possibility of *beating* the CE.

In consequence, we can ask if other solution can also be attained under alternative assumptions about the learning of firms with memory. One important point in the use of memory is the selection rule that is used by the players, when the same quantity appears in the memory with profits obtained in different periods. Alós-Ferrer adopts an imitate the best rule where the players adopt the quantity that gives the highest profit. In addition, when the memory is collective, there is a confusion between imitation and memory (since the player can adopt, from the memory, other firms' strategies). We know that imitation imposes a strong pressure towards the Walrasian equilibrium on the firms' strategies, even if the memory introduces a best reply orientation on the selection of strategies, through the trial and error process. The tension between these forces drives learning towards any symmetric solution between $[q^{CE}, q^{WE}]$. It is necessary to check if under alternative assumptions about memory, convergence to other solutions is possible. Huck, Normann and Oechssler (2004) show that in the absence of imitation, with individual learning based purely on a trial and error process, the only absorbing state is the collusive solution. For the trial and error process, and the corresponding results, memory is again necessary, but it is individual in this case. The absence of imitation (social learning yielding the

spite effect) eliminates the pressure towards WE and the system can converge below CE. Using again a discrete strategy space, but a different learning mechanism (Q-learning), Waltman and Kaymak (2008) show that the agents can converge towards more collusive solutions than CE. Q-learning is a mechanism that includes implicit individual memory (the strategies are selected on the base of the average of past performances)². These results point to the importance of individual memory, with some updating mechanism on the efficacy of the strategies in the memory, as an important mechanism for convergence to more collusive solutions. Our results below will aim to clarify these mechanisms in a more general framework with continuous strategies.

In this article, we use computational experiments with a continuous strategy space to show that:

- Under Alós-Ferrer’s original assumptions with collective memory, the convergence is towards WE.
- If we allow the updating in the collective memory of the profits associated with the quantities using the most recent observations, the Alós-Ferrer’s results are conserved. Hence, updating is not sufficient for deviating from WE under collective memory
- If the firms can only benefit from an individual memory, convergence to the CE is observed in accordance with Bergin and Bernhardt (2004).
- Convergence to more collusive solutions, in conformity with the experiments, can be observed only with individual memory, when the memory is updated with most recent observations. It is hence possible, for the firms, to *beat* the CE.

We also formalize the mechanisms behind these results and check the validity of our intuitions through a finer analysis of the simulation results.

The article is organized as following. The first section presents the framework of our analysis (the model, the equilibria and the simulation protocol). The main simulation results are discussed in the second section. The third section is dedicated to the formalization of the mechanisms driving these results. The results of the thirs section are compared with a finer analysis of the simulation results in the fourth section and the last section concludes the article.

1 Experimental setup and protocol

In this section, we first present a general Cournot oligopoly game and its equilibria. Learning of the agents will then be introduced. Finally, we summarize the simulation protocol.

1.1 A simple oligopoly game

We consider a standard symmetrical n –firms oligopoly model of quantity competition where all firms produce a homogeneous product. The inverse demand function for this good is given by $p = p(Q)$, where $Q = \sum_{i=1}^n q_i$ and $dp/dQ < 0$. The common cost function for the firms is $C(q_i)$, with $C' > 0$ and $C'' > 0$. The profit function of a firm is: $\pi_i(q_i, \dots, q_n) = p(Q)q_i - C(q_i)$. Since the interaction between the strategies (quantities) chosen by the firms only takes place through the common inverse demand function (and hence, through the sum of these quantities),

² In Waltman and Kaymak (2008), the updating is also different from the one used here. With Q-learning, the probability of mutation is updated by taking in account the fitness of the strategies that have been used in the past. In the present article, the probability of mutation remains constant over time, but the fitness of each adopted strategy is updated according to new market information.

a quantity profile (q_1, \dots, q_n) can be represented, from the point of view of a firm i , as (q_i, Q_{-i}) , where $Q_{-i} = \sum_{j \neq i} q_j$.

In this oligopoly, three different solutions can be defined: the Cournot-Nash equilibrium (**CE**), the Walrasian equilibrium (**WE**) and the Collective solution (**CS**).

Definition 1 A Cournot-Nash equilibrium (**CE**) is given by a quantity profile q^{CE} and a market price p^{CE} such as

1. Each firm maximizes its profit at this equilibrium

$$q^{CE} = \operatorname{argmax}_{q_i} \pi_i(q_i, Q^{CE}) \Rightarrow p(Q^{CE}) + q^{CE} p'(Q^{CE}) = C'(q^{CE}), \forall i = 1, \dots, n \quad (1)$$

2. The market clears: $p^{CE} = p(Q^{CE})$.

Definition 2 A Walrasian equilibrium (**WE**) is given by a quantity profile q^{WE} and a market price p^{WE} such as

1. Each firm uses marginal cost pricing (it is a price-taker)

$$q^{WE} = \operatorname{argmax}_{q_i} \pi_i(q_i; p^{WE}) \Rightarrow C'(q^{WE}) = p^{WE}, \forall i = 1, \dots, n \quad (2)$$

2. The market clears : $p^{WE} = p(Q^{WE})$.

Definition 3 A Collusive solution (**CS**) is given by a quantity profile q^{CS} and a market price p^{CS} such as

1. Quantities q^{CS} maximize total profit $\Pi = \sum \pi_i$

$$q_i^{CS} = \operatorname{argmax}_{q_i} \Pi \quad (3)$$

$$\Rightarrow p(Q^{CS}) + p'(Q^{CS})(q_i^{CS} + Q_{-i}^{CS}) = \sum C'(q_i^{CS}), \forall i = 1, \dots, n \quad (4)$$

2. The market clears: $p^{CS} = p(Q^{CS})$.

Since $dp/dQ < 0$ and $C'' > 0$, conditions (1), (2) and (3) imply the standard results on the comparison of these solutions:

$$\begin{aligned} Q^{WE} &\geq Q^{CE} \geq Q^{CS} \\ p^{WE} &\leq p^{CE} \leq p^{CS} \\ \Pi^{WE} &\leq \Pi^{CE} \leq \Pi^{CS} \end{aligned}$$

The collusive solution pareto-dominates other solutions for the firms.

1.2 Agent learning

The mechanisms that we consider for agent learning cover those discussed in the introduction: random experiments, imitation of others, collective or individual memory and the possibility of updating the memory. The firm can learn from others (imitation) or from the memory (collective/individual). It can also benefit from random experiments, which is the starting point of this article.

1.2.1 Random experiments

In each period, with a probability p^m , the quantity of each firm is changed using a truncated normal draw from $N(\bar{q}, \sigma)$ constrained to the interval $[q^{CS}(1 - \beta), q^{WE}(1 + \beta)]$. This interval contains all potential solutions of the Cournot game.

1.2.2 Imitation

In each period, with a probability p^I , each firm can imitate the best strategy in the present population:

$$q^I = \max_{q_j} \{\pi(q_j)\}_{j=1 \dots n}$$

1.2.3 Memory and updating

The memory contains observations (q, π) for the last K periods (size of the memory). When the memory is collective, it contains best strategies observed in each past period. When the memory is individual, the values only come from the observations of each individual firm (q_i, π_i) .

When update is possible, the data in the memory (q^0, π_τ^0) is replaced by (q^0, π_t^0) when the firm reuses the quantity q^0 in a period $t > \tau$.

In each period, each firm can use memory with a probability p^M . When the imitations is also present, the firm adopts the best strategy between imitation and memory.

$$\begin{aligned} q^M &= \max_{q_j} \{\pi(q_j)\}_{j=1 \dots K} \\ q_{t+1} &= \max \{q^I, q^M\}. \end{aligned}$$

Random experiments take place after imitation and memory.

1.3 Simulation protocol

We use the standard quadratic cost Cournot oligopoly with linear demand and quadratic costs:

$$\begin{aligned} P(q_i, Q_{-i}) &= a - b(q_i + Q_{-i}), \quad a, b > 0 \\ C_i(q_i) &= cq_i + dq_i^2 \\ c &\geq 0, d \geq 0. \end{aligned}$$

In the experiments, we adopt the following numerical values and the corresponding symmetric solutions: $n = 10$ firms, $a = 256$, $b = 1$, $d = 1$, $c = 56$

- **WE:** $q^{WE} = 16.7$, $p^{WE} = \mathbf{89.3}$
- **CE:** $q^{CE} = 15.4$, $p^{CE} = \mathbf{102}$
- **CS:** $q_i^{CS} = 9.09$, $p^{CS} = \mathbf{165}$

In order to analyze the convergence of the learning dynamics, and the role of different learning parameters in this convergence, we run Monte Carlo simulations with 500 runs of 10000 periods for each configuration. In each run the following parameters are randomly draw:

- the probability of imitation,
- the probability of mutation,

- the probability of memory being used,
- the size of the firms' memory;

Appendix A.1 gives the numerical values of other parameters of the model used in the simulations.

At the end of each run, we record the last period values for the relevant variables (market price, quantities, variance of quantities, etc.) and we compare these values with the theoretical values of the potential solutions given above. We use standard statistical techniques (plots, econometrics, regression trees and statistical tests) to establish our results and the influence of the parameters. All statistical tests are conducted using R-Project 2.8.1.

2 Computational results

Convergence under different assumptions about firms' learning is successively checked. For social imitation and mutation based learning alone, the results of Vallée and Yildizoglu (2009) directly apply and they are not reproduced here. This type of learning systematically converge to WE, in accordance with the theoretical literature, because of the *spite effect*: a firm deviating from the WE, in the direction of CE, increases its profit but it increases the profits of firms who remained on the WE even more, and imitation pulls the population back to WE. For the CE, we have the opposite result: a firm deviating from CE to WE decreases its profits but it decreases the profits of other firms even more, and these other firms will imitate it and migrate towards WE. This mechanism is very robust in theoretical terms, and imitation is a strong force that pushes the firms towards higher quantities. We consequently begin our analysis with the test of the results from Alós-Ferrer (2004) (**AF** henceforth). The main computational results are summarized here. The following two sections will explore the mechanisms that drive these results.

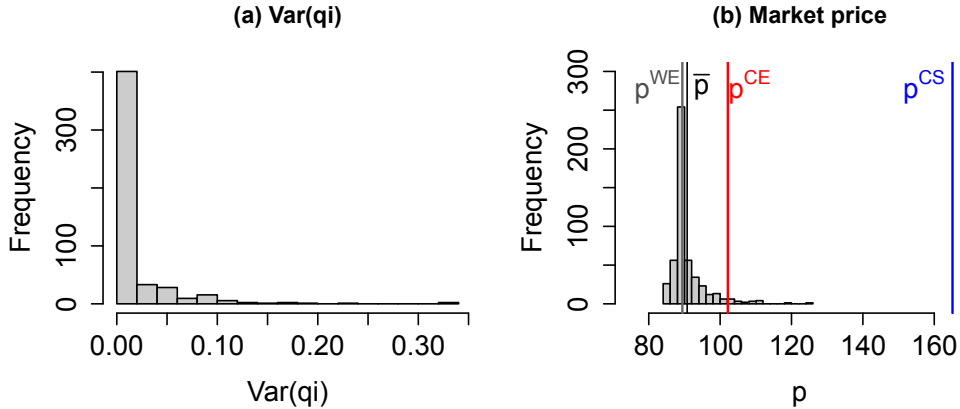


Figure 1: Convergence with collective memory ($t = 1000$)

2.1 Convergence with collective memory

AF proves, with discrete strategies and collective memory that the absorbing states correspond to the interval $[CE, WE]$. Our results with continuous strategies, given in Figure 1, mainly

confirm this conclusion (Graph (b)), but with a clearly higher rate of convergence to WE. In a very significant proportion of the runs, the strategies converged towards the WE, since the variance of the quantities is very small (Graph (a)) and the resulting market price is p^{WE} . But, we a convergence to above this interval is also observed in some marginal cases.

Result 1 *Convergence is observed even when the mutations are maintained during the whole history: the variance in the quantities is very small (Figure 1-(a)).*

Result 2 *Convergence is mainly to the WE price, and also to prices between p^{WE} and p^{CE} , but only in 46% of cases. Convergence to prices above p^{CE} can even be observed in 4.4% of the cases. As a consequence, our results do not strictly confirm the main result of **AF**. (Figure 1-(b)).*

What forces drive firms from the WE, towards higher prices? In order to check this, a probit estimation is carried out on the probability of observing $p > p^{WE}$ in the data. Following the standard convention for the significance of the explanatory variables (for α : * = 0.1, ** = 0.05, *** = 0.01, **** = 0.001), a weaker imitation is shown to be the main driving force. This result confirms the very robust role of the spite effect in the convergence towards WE.

Result 3 *The probability of attaining prices higher than the WE price is increasing with the size of the memory (**) and the probability of using memory (***), decreasing with the probability of mutation (***) and the size of the mutations (**). The probability of attaining prices higher than the CE price ($p > p^{CE}$) only depends on the size of the memory (***), and the influence is positive.*

Result 4 *The probability of having a final market price that does not verify the result of AF is increasing with the probability of mutation (****) and the size of the mutations (**), and decreasing with the probability of using the memory (***).*

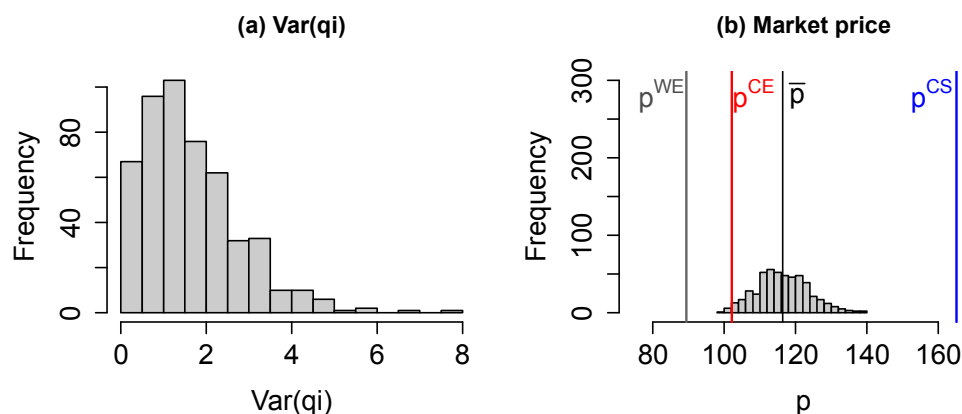


Figure 2: Results with individual memory ($t = 1000$)

2.2 Individual memory

When each agent is able to keep an individual memory of past strategies, Bergin and Bernhardt (2004, p.449) asserts that the industry should converge to the CE, assuming that the strategy space is continuous, the inverse demand function is linear and the cost function is strictly convex (the present experiment complies with all these conditions).

The results indeed change significantly in comparison with the preceding case, but a convergence towards the CE is not observed (Figure 2). A very large majority of cases, 99%, end with a market price higher than the p^{CE} : The firms are able to attain more collusive solutions.

Result 5 *Convergence is weaker with individual memory: the variances in the quantities of the firms are quite significant in the majority of the simulations (Figure 2-(a)).*

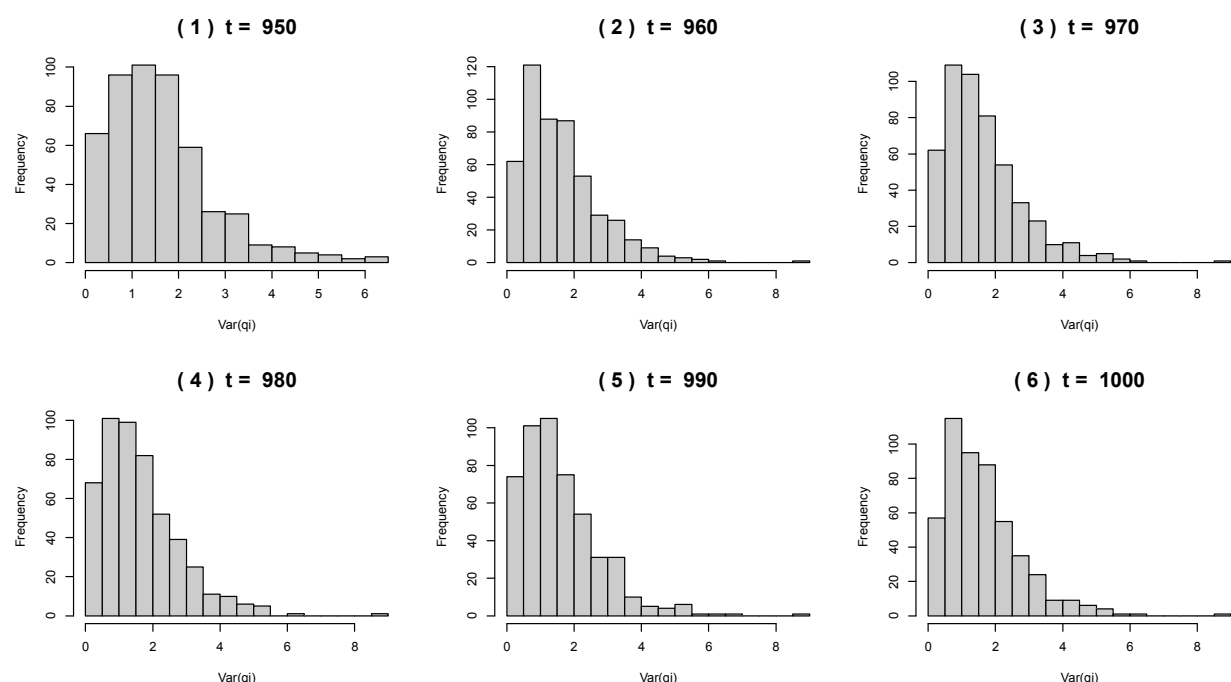


Figure 3: Absence of convergence with individual memory: Distribution of the variances in the individual quantities

An absence of convergence can be observed in Figure 3 where we represent the distribution of the variances in quantities over the last 50 periods. A significant decrease in the variances with the passage of time is not observed. Moreover, this non-convergence is quite robust and still exists even when the simulations run 10000 periods.

Result 6 *The industry finds the CE and even more collusive solutions (in 99% of cases (Figure 2-(b)). This is quite different from the original **AF** case. Only in 1% of cases is convergence to the region between the WE and CE prices.*

These results are clearly in contradiction with Bergin and Bernhardt (2004). The observed average price is clearly higher than p^{CE} , and Student's t-test rejects the null-hypotheses $H_0 : p = p^{CE}$, against $H_A : p > p^{CE}$. This negative result clearly questions the *practical* consequences of the stochastic stability concept.

Least square analysis of the determinants of the market price, in this case, gives the following results (for α we have $*$ = 0.1, $**$ = 0.05, $***$ = 0.01, $****$ = 0.001):

Result 7 *The final market price decreases with the probability of using memory (**), and with the probability of mutation (****).*

A standard source of convergence in evolutionary learning is imitation that favours the diffusion of successful strategies in the population (*cf.* Vallée and Yildizoglu (2009)). Since pure individual memory does not assure convergence, we can check if an imitation mechanism can allow firms to coordinate their production levels.

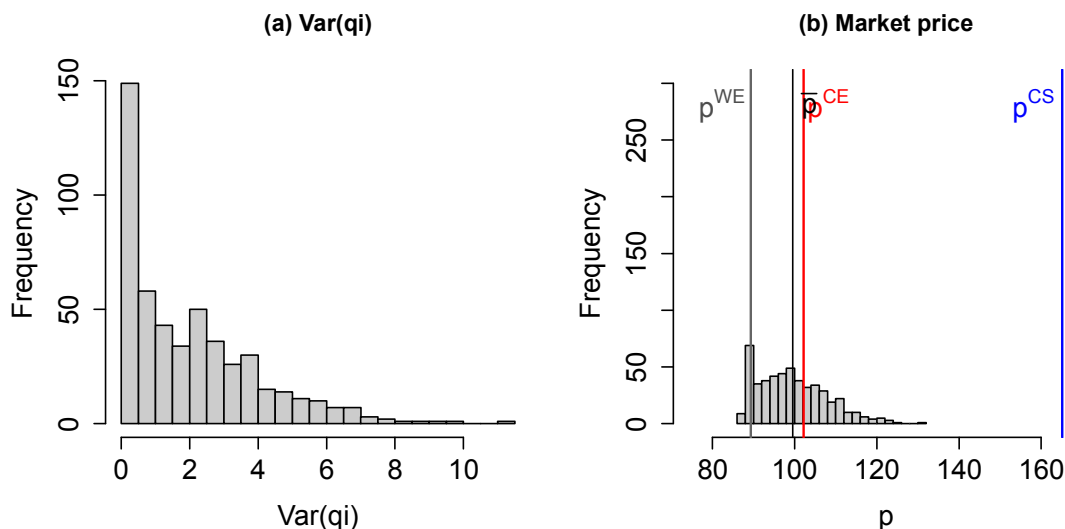


Figure 4: Convergence with individual memory and imitation of the best. Absorbing states emerge in $[WE, CE]$

2.3 Convergence with individual memory and imitation

We add to the preceding market framework by giving the firms the possibility of imitating the quantities of their competitors. We introduce this possibility in a very simple and standard (see above): In each period, each firm imitates the quantity of the competitor with the highest profit in the preceding period, with a probability p^I . This possibility introduces a restricted collective memory in the individual learning of the firms, and it should make the coordination between the firms easier.

Figure 4 shows that convergence can now be observed in many runs, where the variance in the quantities approaches zero. But, as shown in Vallée and Yildizoglu (2009), and discussed

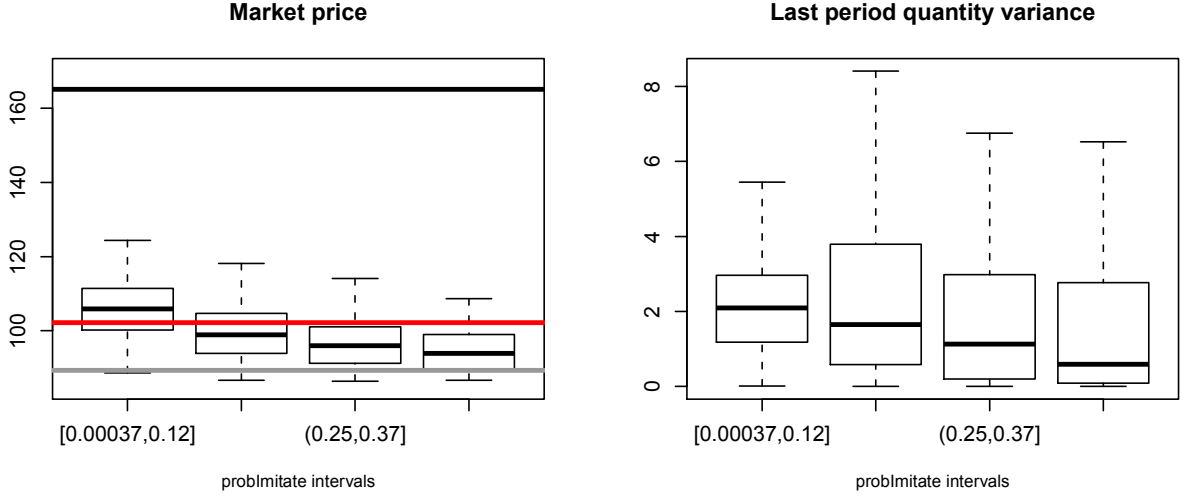


Figure 5: Convergence and efficiency as a function of p^I

below, imitation is a very strong source of spite effect and this pushes the firms toward the WE. Hence, convergence comes with a cost, firms have lower profits than with individual learning.

Figure 5 shows that the convergence is stronger with higher probabilities of imitation (the variances are clearly decreasing in the right hand diagram), but the market price is compressed toward the WE price at the same time: convergence takes place on lower prices and it yields lower profits for the firms. Hence imitation cannot allow the firms to consistently beat the CE. Another learning mechanism is necessary in order to obtain this outcome: the updating of individual memory.

2.4 Beating the CE: Convergence with individual memory and updating

A new result, a convergence towards more collusive solutions, emerges when firms can also update the performances in the memory, by substituting the most recent profit observed using one quantity for the older profit associated with this quantity (Figure 6).

Firms can only move towards the Pareto optimum (CS) by discovering that they must refrain from playing their best response against lower quantities of their competitors. This is only possible if the memory is long enough, since the temptation to play the best response is very strong. Even in this case, this effect is, in the end, neutralized by the best reply effect and firms stop before attaining CS.

Memory updating nevertheless displaces the zone of convergence to also cover a subset of $[p^{CE}, p^{CS}]$.

Result 8 *Convergence: the variances in the quantities are considerably smaller than in the preceding cases (Figure 6-(a)). In many cases, the firms do converge to a unique production level.*

Result 9 *Convergence is not toward a unique market solution. We also observe that the industry*

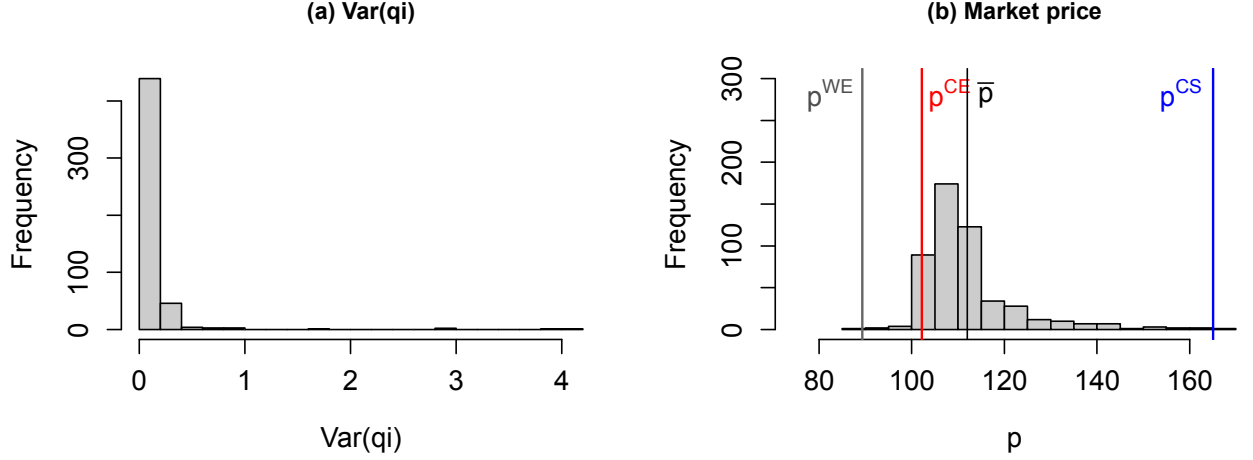


Figure 6: Convergence with individual memory and update. Absorbing states emerge in $[CE, CS]$

can robustly convergence towards prices higher than the CE price in this case (in 95% of the cases, Figure 6-(b)).

We can check if the cases with prices higher than p^{CE} do effectively correspond to more collusive situations. Figure 7 gives the profits of the firms in these cases, and we observe that the distribution of the quantities also perfectly corresponds to more collusive situations and to profits higher than the CE profits. Consequently firms can attain more collusive situations under individual learning with updating. How is this result possible? To answer this question, we now analyse explore the mechanisms that produce these results.

3 Theoretical analysis of the learning mechanisms

In order to understand better the general mechanisms behind the new results with individual memory and update, we first analyze these forces in a simple, two player, one mutation per period, version of the oligopoly game. We also consider the impact of higher numbers of players and mutations. If we assume that firms start from an initial symmetric point (q, q) and if they have the possibility of deviating, from this initial position, with a given perturbation ϵ , in both directions, their situation can be schematically represented by the corresponding 3×3 normal form game given in Table 1.

	$q - \epsilon$	q	$q + \epsilon$
$q - \epsilon$	(A, A)	(F, E)	(B, C)
q	(E, F)	(G, G)	(H, I)
$q + \epsilon$	(C, B)	(I, H)	(D, D)

Table 1: Normal Form of the 3×3 game

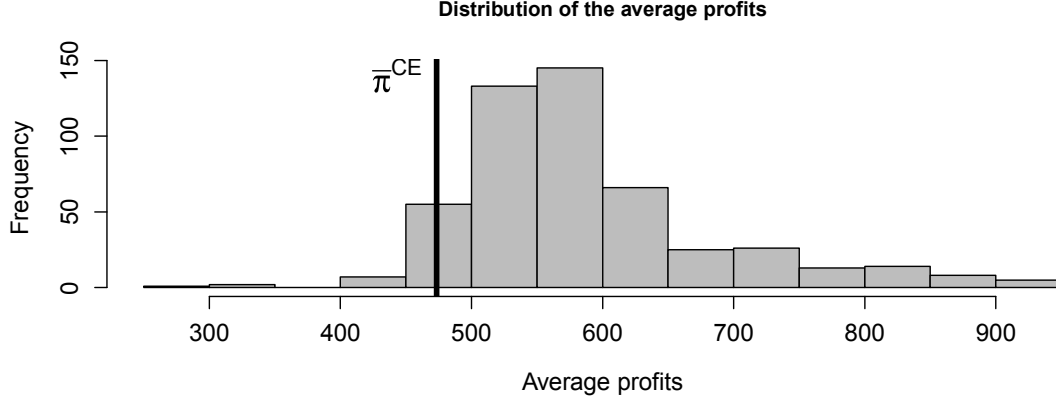


Figure 7: Emergence of more collusive absorptive states

The structure of the game depends on the relation between the pay-offs A, B, C, \dots . The values of the initial position (q, q) will determine this relation. The relevant pay-offs in the simple game analyzed in this article are given in Appendix A.2. For example, if initially $q = q^{CE}$, and assuming that $\epsilon > 0$ is sufficiently small to have $q - \epsilon \geq q^{CS}$ and $q + \epsilon \leq q^{WE}$, the following inequalities hold:

- $G > F$ and $G > I$: (q^{CE}, q^{CE}) is a Nash equilibrium.
- $E > A > G > F$ and $G > I > H > D$: first, CE is the unique Nash equilibrium and, second, the best reply dynamics will necessarily imply a movement from $(q - \epsilon, q - \epsilon)$ (*e.g.* CS) and from $(q + \epsilon, q + \epsilon)$ (*e.g.* WE) towards CE.

Using this matrix, we can analyze the impact of different learning mechanisms (memory, imitation, random experiment). We focus in the following sections on three mechanisms: mutation, imitation and individual memory with or without update. We will mainly discuss mainly the effects of these mechanisms in the zone $[q^{CS}, q^{CE}]$ since our new results concern convergence in this interval.

3.1 Pure evolutionary social learning: experimentation and imitation

Consider that, in every period, strategies can be modified (random experimentation), and that each firm can imitate any other, on the basis of the performance of this competitor in the previous period. Figure 8 represents the forces of convergence created by these two simple mechanisms. Notice that the imitation force can only play in an initially asymmetric state, and in this case, it necessarily pushes towards CE since, when $q \in [q^{CS}, q^{CE}]$, we necessarily have $\pi_i(q_i^{CS}, q_j^{CE}) < \pi_j(q_j^{CE}, q_i^{CS})$, $i \neq j = 1, 2$.

Figure 8 shows that, if the probability of imitation is not zero, the direction of learning is towards (q^{CE}, q^{CE}) , and that symmetrically, starting from (q^{CE}, q^{CE}) , these forces push the learning process towards (q^{WE}, q^{WE}) . These results can be generalized in the following proposition.

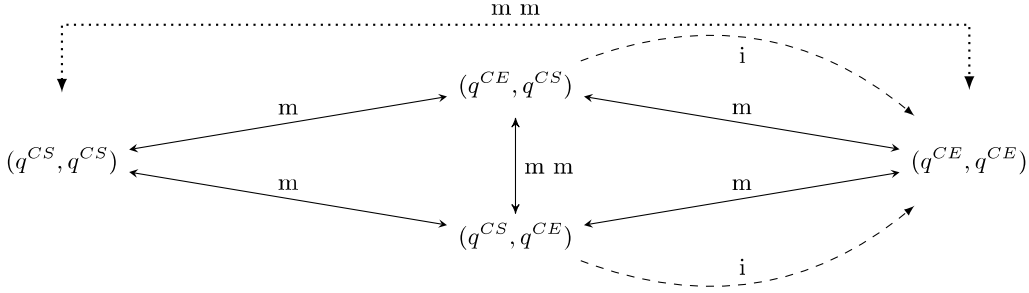


Figure 8: Forces: mutation (m) and imitation (i) and double simultaneous mutations(mm)

Proposition 1 *Imitation will favor an increase of the quantities, if the initial state before mutation corresponds to $q > q^{WE} + \frac{\epsilon}{2} > q^{CE}$.*

Proof. The strategy population will evolve towards lower quantities if the initial state favors the imitations of any individual mutation from q to $q - \epsilon$, and forbids the imitation of an individual mutation from q to $q + \epsilon$. These properties are observed if the initial allocation, q , respects the following two conditions:

$$F > E \implies \pi_1(q - \epsilon, q) > \pi_2(q - \epsilon, q) \quad (5)$$

$$H > I \implies \pi_1(q + \epsilon, q) < \pi_2(q + \epsilon, q) \quad (6)$$

It is easily verified that condition (5) requires that $q > q^{WE} + \frac{\epsilon}{2}$ while Condition ((6)) requires that $q > q^{WE} - \frac{\epsilon}{2}$. In the interval $[q^{CS}, q^{CE}]$, both conditions are false and the corresponding forces increase the quantities towards q^{CE} . ■

The effects of these mechanisms, representing the forces that result from imitation and random experiments, are summarized in Figure 9. These forces directly stem from Proposition 1. This figure shows how these forces play as a function of the initial quantity q and the mutation ϵ . Conditions (5) – (6) divide the quantity space into three zones.

In Zone *III*, we have $F > E$ and $H > I$. Any mutation from q to $q - \epsilon$ will thus be imitated, while any move from q to $q + \epsilon$ will not. In Zone *I*, we have $F < E$ and $H < I$, and only a mutation from q to $q + \epsilon$ will be imitated. Finally, in Zone *II* we have two opposing forces. However, if WE is reached, any mutation from q^{WE} to $q^{WE} - \epsilon$ will leave Zone *II*, since its width is ϵ . Once in Zone *I* imitation will push the quantities towards Zone *II* as a result of Proposition 1 (given that we have $A > G$, a lower collective quantity is associated with lower pay-offs, while, $G > D$ implies that a higher collective quantity is associated with lower pay-offs).

A direct consequence of Figure 9 is that imitation will enforce stabilization around the Walrasian equilibrium: the set of the absorbing states is in Zone *II* and convergence to WE will be observed when $\epsilon \rightarrow 0$.

3.2 Experimentation and short individual memory ($K = 1$)

We now consider that agents are able to recall their most recent strategy and its result. We have an individual memory of size $K = 1$. The agents are now able to fully benefit from a trial

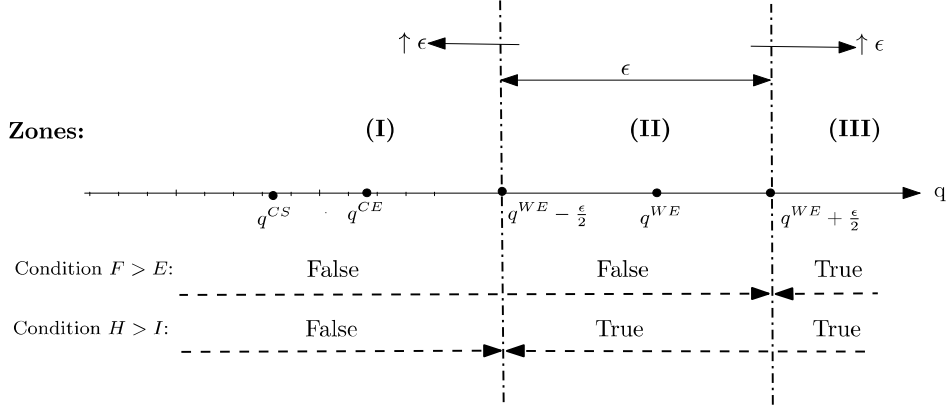


Figure 9: Imitation force

and error process in order to learn. Convergence forces that can result from such a memory are represented in Figure 10. The impact of memory depends both on the probability of using the memory, and on the previous state of the population. Consequently, we represent transitions that are possible with memory, only if the previous state allows their possibility.

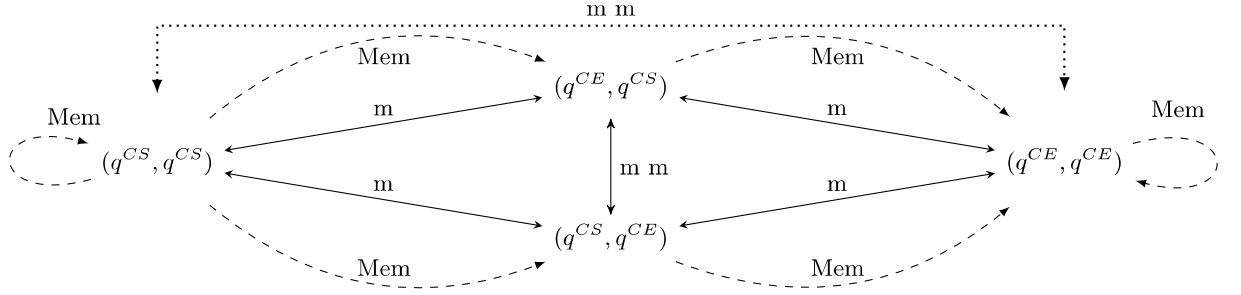


Figure 10: Forces: mutation (m) and memory (Mem) of size one

First, with a memory of size one, the information concerning strategy q^{CE} (resp. q^{CS}) will disappear if strategy q^{CS} (resp. q^{CE}) was used in the previous period. Second, a move from (q^{CS}, q^{CS}) to an asymmetric state cannot result from the use of memory, if the previous state was (q^{CE}, q^{CE}) , followed by double simultaneous mutations. Memory will incite the firms to switch back to (q^{CS}, q^{CS}) as a result of $A > G$. Moving from (q^{CS}, q^{CS}) thanks to memory is only possible, if the previous state was asymmetric. Figure 10 indicates that using a memory of size one will encourage convergence towards (q^{CE}, q^{CE}) , if the probability of using the memory is not zero. Similar reasoning can be used to show that, starting from (q^{WE}, q^{WE}) , the learning process converges towards (q^{CE}, q^{CE}) . The following proposition generalizes this result.

Proposition 2 *An individual memory of size one will favor movement towards the CE, if $q > q^{CE} + \frac{b+d}{3b+2d}\epsilon$ or if $q < q^{CE} - \frac{b+d}{3b+2d}\epsilon$: when $q \in [q^{CE} - \frac{b+d}{3b+d}\epsilon, q^{CE} + \frac{b+d}{3b+d}\epsilon]$, individual memory pushes any mutant to return to CE after mutation.*

Proof. With an individual memory of size one, the mutant firm will compare its actual pay-off with the pay-off before mutation (trial and error). Starting from any production level q , this mutant will move towards $q - \epsilon$, if its pay-off is increasing by unilaterally decreasing its production and, if its pay-off is decreasing by unilaterally increasing its production. Memory will encourage it to stay at q if and only if the two following conditions hold:

$$F > G \implies \pi_1(q - \epsilon, q) > \pi_1(q, q) \quad (7)$$

$$I < G \implies \pi_1(q, q) > \pi_1(q + \epsilon, q) \quad (8)$$

It is easy to verify that Condition (7) requires $q > q^{CE} + \frac{b+d}{3b+2d}\epsilon$, while Condition (8) requires $q < q^{CE} - \frac{b+d}{3b+2d}\epsilon$. ■

Figure 11 represents these results. In Zone I, $F < G$ and $G < I$, therefore there is no incentive to reduce production and any increase in production will be recorded as a good solution in the memory. In Zone III, we have $F > G$ and $G > I$. This is the inverse of Zone I: there is no incentive to increase production, and any decrease will be recorded as a good solution in the memory. In Zone II, both $F < G$ and $I < G$, which is equivalent to the standard Nash equilibrium conditions. Any increase of ϵ widens the zone and, any deviation of size ϵ will put the quantity outside of Zone II, since its width is $2\frac{b+d}{3b+2d}\epsilon < \epsilon$. However in this case, memory will encourage a return to this zone.

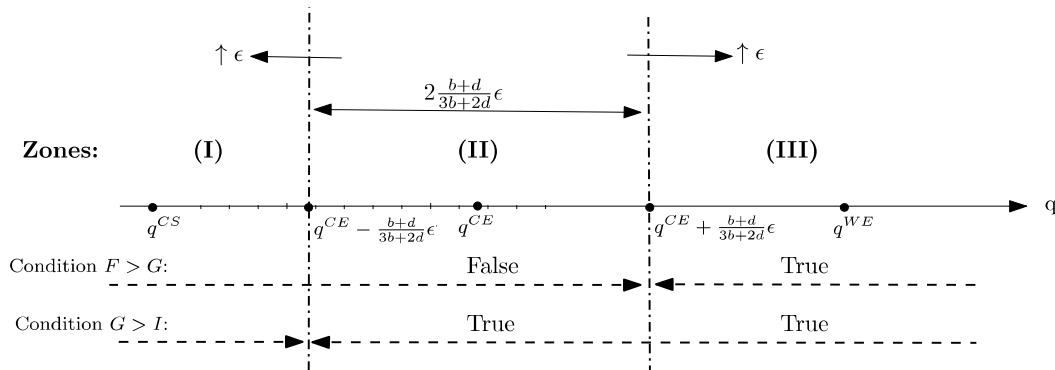


Figure 11: Forces with individual memory of size one

We now consider the case of longer memory.

3.3 Experimentation and long memory ($K > 1$) with updating

What are the consequences of a greater memory size? A direct consequence is that any strategy q_i may have a variety of values in the memory. In our example, q^{CS} can both appear with profit $\pi_i(q^{CS}, q^{CS}) = A$ and $\pi_i(q^{CS}, q^{CE}) = F$. Moreover, if the memory is very long, some of the performances could correspond to a very old population state and encourage decisions that are ill-adapted for the present state. These problems can be avoided if the agents can update the memory, in such a way that only the most recent information is associated to each strategy.

Our simulation results prove that only such an update can yield more collusive states that are statistically significant.

We can underline a condition that is necessary, but not sufficient, in order to move towards collusive solutions using a long memory. To move from (q, q) to $(q - \epsilon, q - \epsilon)$ will require at least two moves, either the transition $(q, q) \rightarrow (q - \epsilon, q) \rightarrow (q - \epsilon, q - \epsilon)$ or $(q, q) \rightarrow (q, q - \epsilon) \rightarrow (q - \epsilon, q - \epsilon)$. As a consequence, in order to compare results (payoffs) of the initial situation $\pi_i(q, q)$ and of the final situation $\pi_i(q - \epsilon, q - \epsilon)$, one must have a sufficiently long memory. This long memory facilitates a move towards the collusive situation, since, if it has been occurred in the past, the agents can remember that it was a favorable situation for a long time, including from an individual point of view. This would encourage the agent to return to more collusive quantities.

Proposition 3 *If the memory is long enough, a necessary condition for the memory to push the quantities towards the collusive solution is $q > q^{CS} + \frac{\epsilon}{2}$.*

Proof. We need to check that there is a diagonal convergence $D \leftarrow G \rightarrow A$ that lowers the quantities. We need to check under which conditions on initial quantities q we can observe $A > G$ and $G > D$ for both players, and, since the game is symmetric, the initial quantity must be compatible with following two conditions:

$$A > G \implies \pi_1(q - \epsilon, q - \epsilon) > \pi_1(q, q) \quad (9)$$

$$G > D \implies \pi_1(q, q) > \pi_1(q + \epsilon, q + \epsilon) \quad (10)$$

It is easy to verify that Condition (9) requires $q > q^{CS} + \frac{\epsilon}{2}$, while Condition (10) requires $q > q^{CS} - \frac{\epsilon}{2}$. ■

Figure 12 shows the results. In Zone I, $A < G$ and $G < D$ at all times. Consequently, since the lowest quantities are associated in the memory to the lowest pay-offs, there is an incentive to leave this area by increasing the quantities. In Zone III, since $A > G$ and $G > D$, any collective decrease of quantities is associated with higher payoffs. This creates an incentive to collectively move towards zone II. In Zone II the mechanisms are again more complex. Given that $A < G$ in this zone, a lower collective quantity is associated with lower pay-offs, while, because of $G > D$, a higher collective quantity is associated with lower pay-offs. Notice that the width of area II is ϵ , so again, any mutation will leave this zone.

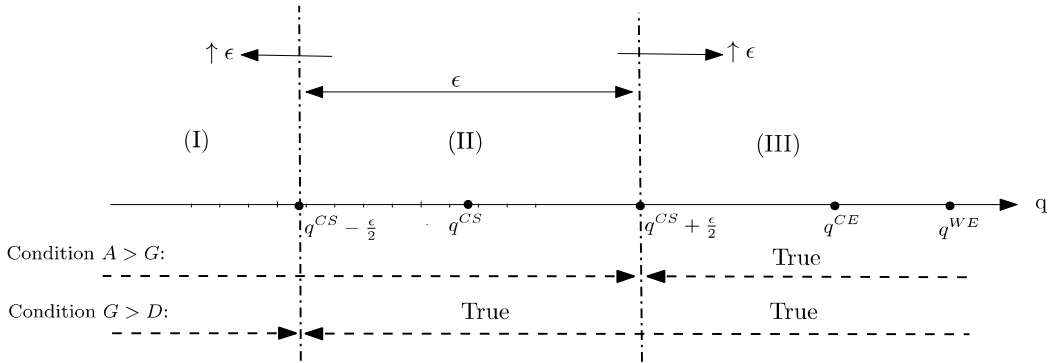


Figure 12: Long Memory Force with updating

3.4 Three main forces: a synthesis

The forces and mechanisms that drive the learning dynamics in accordance with propositions 2-3 are summarized in Figure 13. Whether or not the learning dynamics converge towards a particular equilibrium will depend on the relative strength of these forces. With short memory and a systematic use of imitation, convergence to WE is inevitable. With long memory, and very rare use of imitation, firms may overcome WE and CE, and converge to a more collusive solution. However, we observe that these forces will not allow firms to fully attain the CS solution. In fact, it is not possible for the firms to stabilize on quantities below $q^{CS} + \frac{\epsilon}{2} > q^{CS}$, since all the forces except long memory clearly favor higher quantities below this threshold.

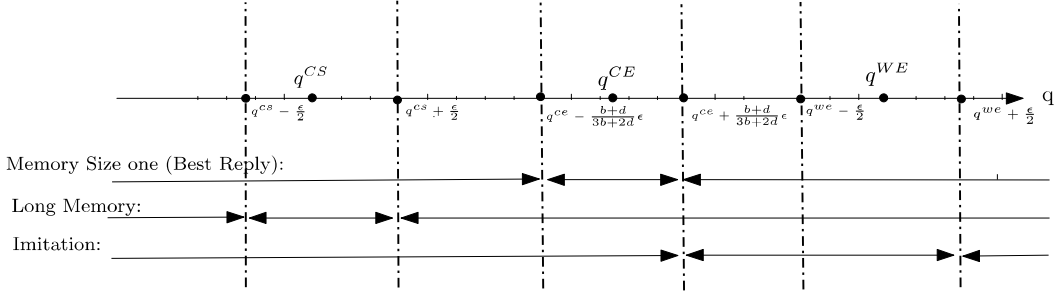


Figure 13: Forces: a synthesis

3.5 Increasing the number of players and updating

Recall that, if there is only one mutation per period, a move from (q, q) to $(q - \epsilon, q - \epsilon)$ will require at least two moves, that is either the transition $(q, q) \rightarrow (q - \epsilon, q) \rightarrow (q - \epsilon, q - \epsilon)$ or $(q, q) \rightarrow (q, q - \epsilon) \rightarrow (q - \epsilon, q - \epsilon)$. This gives rise to the next proposition.

Proposition 4 *In a two player game, a long memory by itself cannot allow convergence towards collusive equilibrium.*

Proof. Let first assume that only one mutation per period is possible. In order to converge to, and to stay at, the collusive equilibrium, the following conditions must be satisfied:

- $A > G$: diagonal convergence occurs (through imitation).
- $A > E$ and $F > G$: an individual incentive to move towards $q - \epsilon$ exists (Condition $F > G$) and no incentive to return to q (Condition $A > E$). If $A > E$ the updated memory will contain, for strategy q , a profit equal to E , which will be lower than the profit associated to strategy $q - \epsilon$, A . This will stabilize decreasing quantities towards q^{CS} .

These two conditions are satisfied if $q > q^{CE} + \frac{b+d}{3b+2d}\epsilon$. ■

Assume now that the oligopoly is composed of more than 2 firms, $(n > 2)$. Would this significantly change the results? If yes, can we qualify the impact of increasing n ? The impact would indeed be quite strong. The main impact would be that the game would lose its "symmetry" in this case.

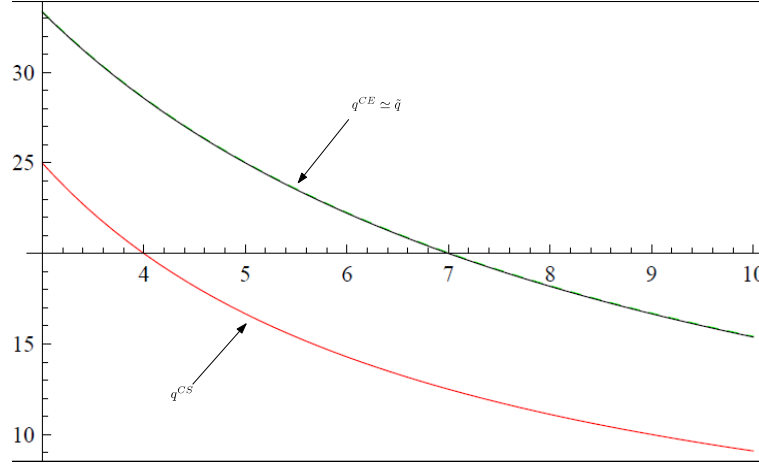


Figure 14: Conditions when $n_1 = 0$, that is 1 mutant per period in the same direction, for $n \in [3, 10]$

To illustrate this, consider a three player game, in which, at a given period, two firms mutate from q to $q - \epsilon$, while the third one sticks to the initial quantity. The total quantity becomes $3q - 2\epsilon$. The game, from the point of view of the non mutant firm, is equivalent to a 2-players game where the other firm is twice as big. We know, from proposition 4, that in a two player symmetric game the mutant firm m will not reduce its quantity if $\pi_m(q - \epsilon, q) < \pi_m(q, q)$, that is, when $q > q^{CE} + \frac{b+d}{3b+2d}\epsilon$. In a n -player asymmetric game, this condition must be modified, and it now depends of the number of mutants.

Define $\pi_i(q_i, n_m, q)$ as the profit of firm i when playing q_i , given that n_m firms are also playing q_i , and $n - n_m - 1$ firms are playing q . Hence, in a 3-player game, if two firms make the same move, from q to $q - \epsilon$, the new profit of each mutant firm i is $\pi_i(q - \epsilon, 1, q)$, while the profit of the non mutant firm j is $\pi_j(q, 0, q - \epsilon)$. Using these definitions, we can now state a proposition that considers the convergence to CS in this case.

Proposition 5 *When multiple mutations are allowed, a long memory by itself can efficiently provide for convergence towards a more collusive solution than the CE, but the industry cannot attain the CS.*

Proof. We need to verify the following conditions:

- $F > G : \implies \pi_i(q - \epsilon, n_1, q) > \pi_i(q, 0, q)$ (Condition 1);
- $A > E : \implies \pi_i(q - \epsilon, 0, q - \epsilon) > \pi_i(q, n_1, q - \epsilon)$ (Condition 2);
- $A > G : \implies \pi_1(q - \epsilon, 0, q - \epsilon) > \pi_1(q, 0, q)$ (Condition 3).

These conditions are individually satisfied if:

- $F > G : \implies q > \frac{a - c + (b + bn_1 + d)\epsilon}{b + bn + bn_1 + 2d}$.
- $A > E : \implies q > \frac{a - c + (bn + d)\epsilon}{b + bn + bn_1 + 2d}$.

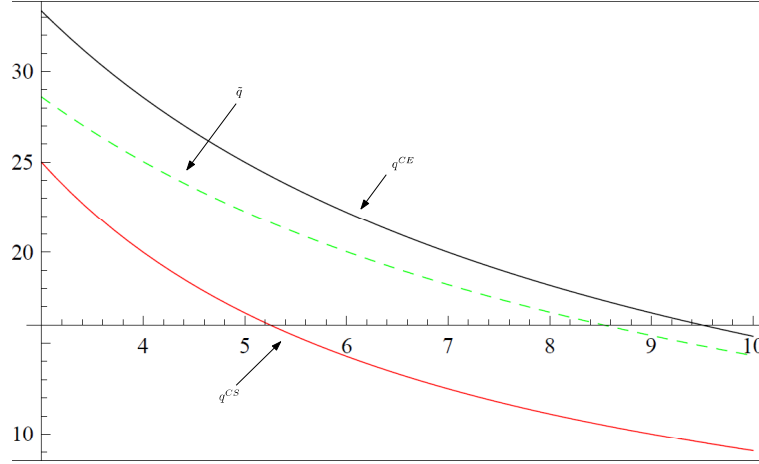


Figure 15: Conditions when $n_1 = 1$, that is 2 mutants per period, for $n \in [3, 10]$

- $A > G : \implies q > \frac{a - c + (bn + d)\epsilon}{2bn + 2d}$.

and they are collectively satisfied if $q > \tilde{q} \equiv \frac{a - c + (bn + d)\epsilon}{b + bn + bn_1 + 2d}$. Since $q^{CE} = \frac{a - c}{b + bn + 2d}$, and $q^{CS} = \frac{a - c}{2d + 2bn}$, one can check that if $n_1 > 0$ and $n \geq 3$ we have $q^{CS} < \tilde{q} < q^{CE}$. Also, for a given n_1 , $\lim_{n \rightarrow \infty} \tilde{q} = \epsilon \geq \lim_{n \rightarrow \infty} q^{CS} = 0$. As a consequence, Conditions (1-3) can necessarily be satisfied in the interval $]q^{CS}, q^{CE}[$, but not when the industry becomes very close to CS. ■

Figures (14 – 16) show the evolution of \tilde{q} for $n_1 = 0$ (individual mutation only), $n_1 = 1$ (double mutations) and $n_1 = n - 2$ (all the firms except one mutate) with $3 \leq n \leq 10$. We observe that \tilde{q} decreases with n , but it remains in the interval $]q^{CS}, q^{CE}[$, and never passes below q^{CS} .

4 Validation of analytical results through experiments

New results are presented in this section in order to confirm the role of the mechanisms discussed in the preceding section, in the global convergence results we have observed in Section 2. Supplementary experiments inspired by these theoretical results are also introduced.

4.1 Convergence to collusive solutions

Is the memory the main driving force of the results of Section 2? Figure 17 displays the distribution of prices as a function of different learning mechanisms. The horizontal lines indicate $p^{WE} < p^{CE} < p^{CS}$. The positive role played by the probability of using memory (a), and by its size (b) clearly appear in these diagrams. When the memory is used very frequently ($p^M \in]0.89, 1]$), around 75% of the observations correspond to prices higher than the CE price, which is a significant convergence towards prices more collusive than the CE. The negative impact of imitation is also clearly shown in graph (d). When p^I is very low, more than 75% of the

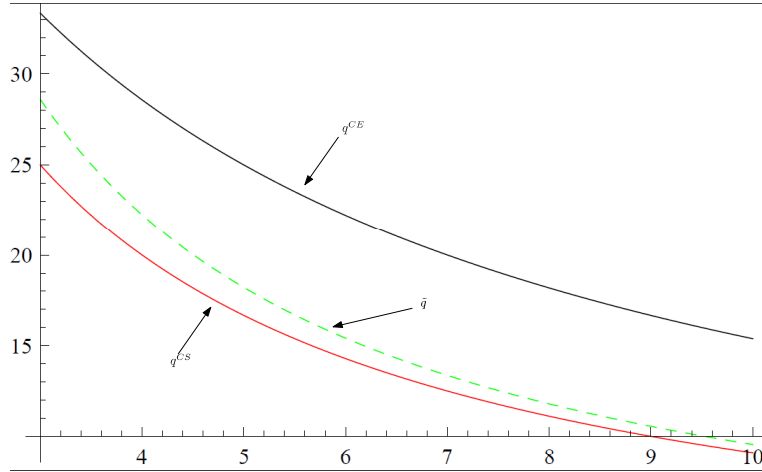


Figure 16: Conditions when $n_1 = n - 2$, that is 2 mutants per period if $n = 3$ and 9 if $n = 10$, for $n \in [3, 10]$

cases converge to more collusive prices. Graph (c) shows that experimentation is necessary, but not sufficient to secure convergence towards collusive prices.

These results can be confirmed using basic econometric analysis. The probit analysis of the probability of convergence to a price higher than p^{CE} confirms the role played by memory and its size (all factors are significative for $\alpha = 0.001$):

Result 10 *The probability of converging to a solution above the CE price increases with the probability of mutation, the probability of using the memory and the size of the memory. It decreases with the probability of imitation, and the size of the mutations. Memory and its size definitely play an important role in the capacity of firms to attain prices above the CE price.*

The regression tree³ in Figure 18 gives some quantitative insights on the conditions for convergence towards a price above p^{CE} , and on the interactions of different mechanisms.

The configurations in which the expected price is the highest are observed in the rightmost branches of the tree (we have $E[p] = 115$ in these cases). They correspond to the following simultaneous conditions: $p^I < 0.006408, K < 6$. On the left of the tree, when $p^I \geq 0.02153$, the highest expected price ($E[p] = 103.6 > p^{CE}$) is observed when $K \geq 6, p^M \geq 0.4982, p^m \geq 0.03607$, where random experimentation and memory compensate for the negative effects of higher imitation.

4.2 Role of the memory size

The role of the memory size with different numbers of firms is analyzed in this subsection. For this analysis, 3000 simulations of 1000 periods were run, where the number of firms was also randomized: $n \in [2, 5]$. Changing the number of firms, changes the equilibrium prices, as well as the distance between them. Figure 19 shows the distribution of the market prices for different numbers of firms, and memory sizes. The horizontal lines give equilibrium prices in each oligopoly type ($p^{WE} < p^{CE} < p^{CS}$). When the number of firms increases, a short memory becomes

³Computed using the rpart library of R-Project, with a complexity parameter of $cp = 1\%$.

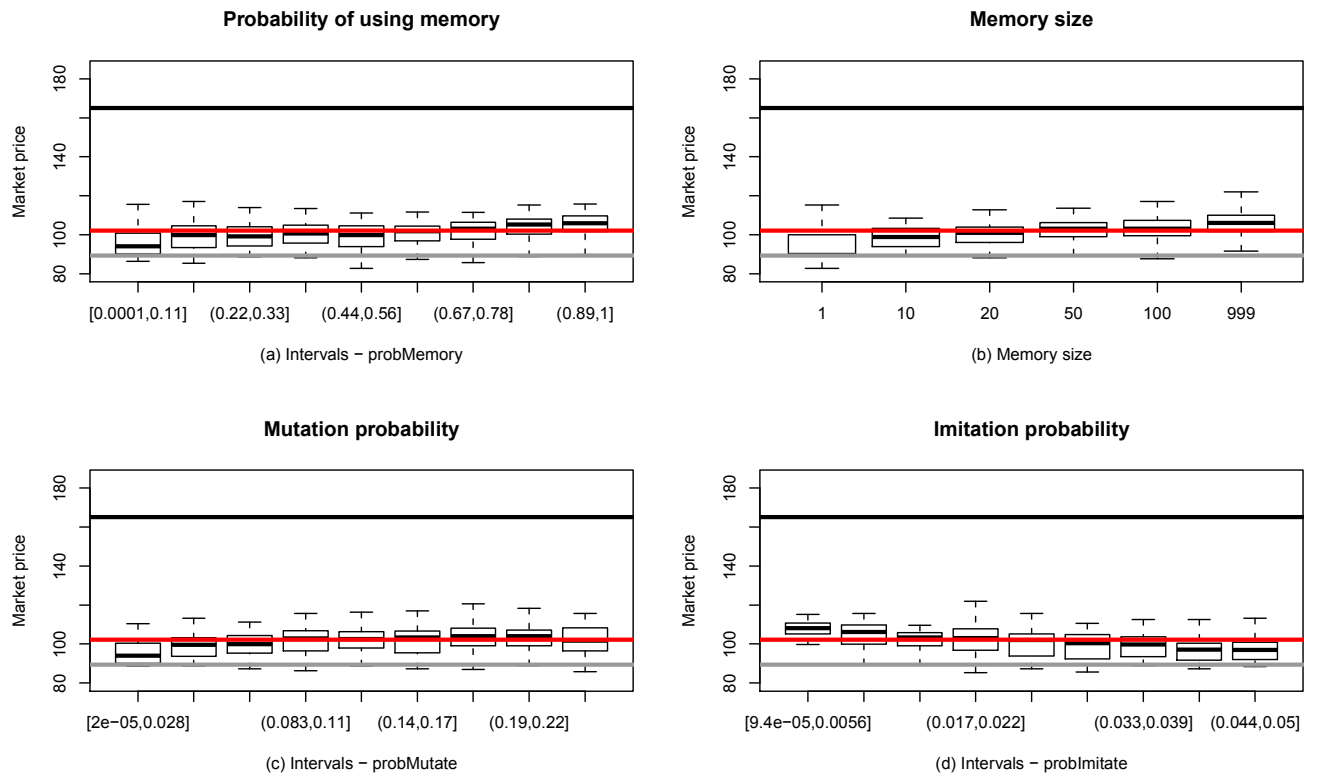


Figure 17: Role of learning parameters

Determinants of marketprice – Individual Memory, imitation and update

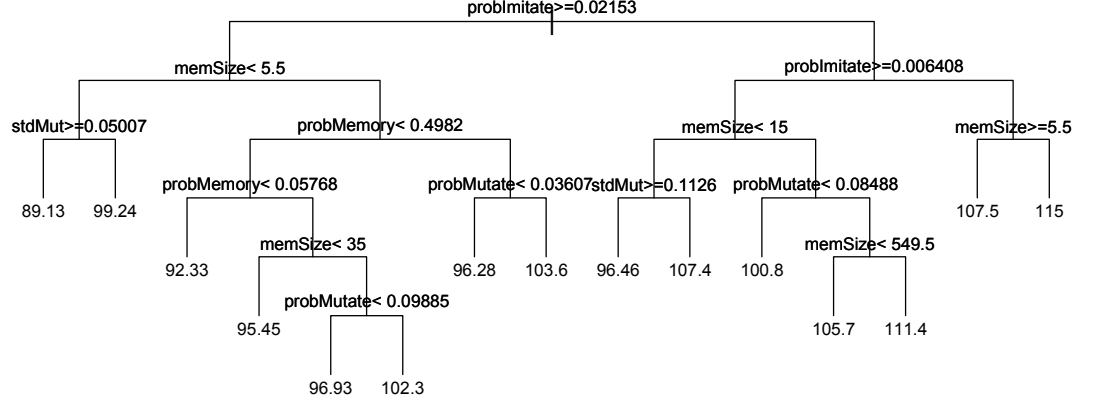


Figure 18: Determinants of the market price

insufficient to robustly attain prices higher than p^{CE} . $K = 1$ becomes particularly inefficient when $n \geq 4$. Only the two highest memory sizes warrant a convergence over CE in more than 75% of cases in all situations.

5 Conclusion

In this article, we have analyzed the possibility of firms learning collusive solutions in a Cournot quantity game. Starting from the results of Vallée and Yıldızoğlu (2009) and of Alós-Ferrer (2004), we study the role of random experimentation, social learning (imitation), and memory in helping firms discover more collusive market configurations than those of Cournot equilibrium (CE). We show that long memory and its update is necessary to attain such configurations.

First, we show these results using direct computational experiment. These experiments clearly emphasize the necessity of memory update to beat the CE. We clarify the mechanisms that are behind this global result, through an analytical exercise that establishes the conditions under which different components of the firms' learning can orient the convergence to the CE. This analysis indicates that it is possible for firms to convergence towards more collusive solutions, without being able to fully converge on the collusive solution (CS). Their capacity to beat the CE decreases as their number increases, which could seem unworthy of note, but, here is explained as resulting from the difficulty of discovering the CS in a system with a richer set of states. The supplementary simulations introduced in the last section of the article show that convergence to more collusive solutions becomes nearly impossible for firms as soon as their numbers is equal to five. The econometric and statistical analysis of the results indicate that the probability of converging to market prices higher then the CE price decreases with the probability of imitation, and increases with the probabilities of mutation and using memory, and with memory size. The

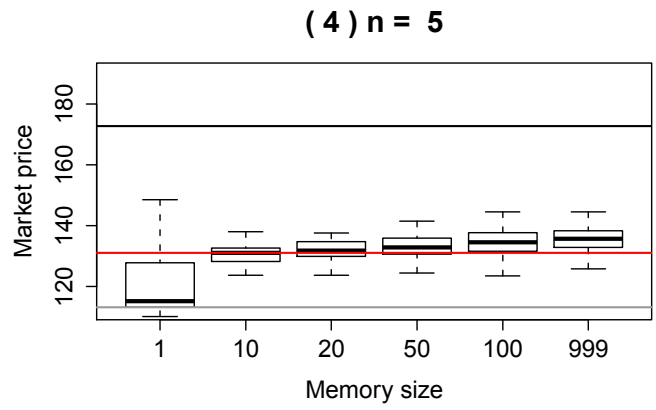
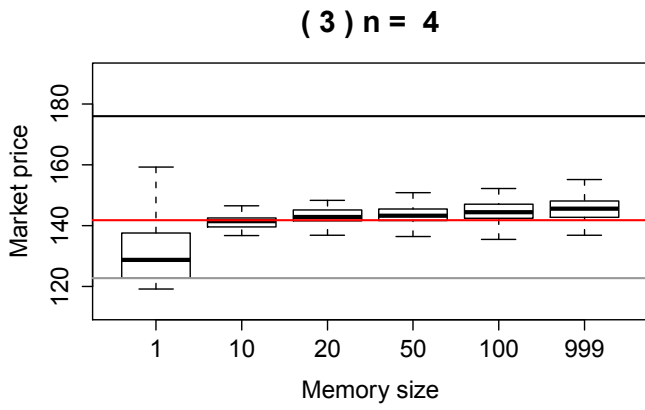
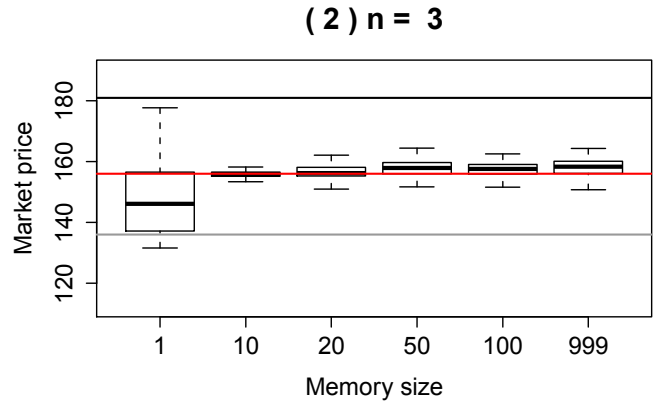
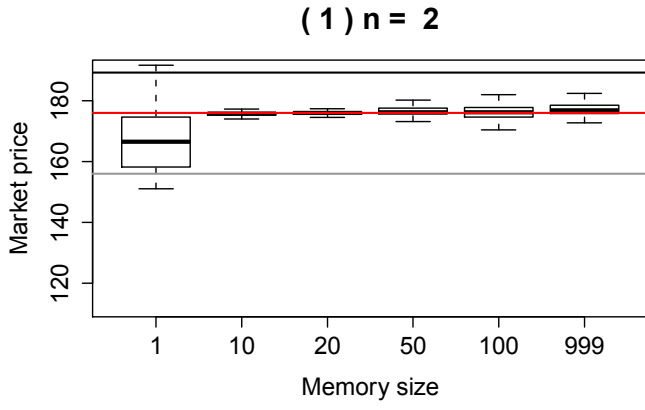


Figure 19: Memory size, number of firms and collusion. Light grey line: p^{WE} ; dark grey line: p^{CE} ; black line p^{CS} .

former confirms the tendency of social learning to push firms towards the Walrasian solution, while the latter establish the necessity of experimentation and updated memory for collusion.

References

- Ania A., (2008), Evolutionary Stability and Nash Equilibrium in Finite Populations, with an Application to Price Competition, *Journal of Economic Behavior and Organization*, vol. 65, 472–478.
- Arifovic, J., (1994), Genetic Algorithm Learning and the Cobweb Model, *Journal of Economic Dynamics and Control*, Special Issue on Computer Science and Economics, vol. 18:1, 3–28.
- Apesteguia J., Huck S. and J. Oechssler (2007) Imitation—theory and experimental evidence, *Journal of Economic Theory*, 136, 217 – 235.
- Arifovic, J. and M. Maschek, (2006), Revisiting Individual Evolutionary Learning in the Cobweb Model - An Illustration of the Virtual Spite Effect, *Computational Economics*, vol. 28, pp. 333–354.
- Alós-Ferrer, C. (2004), Cournot versus Walras in dynamic oligopolies with memory, *International Journal of Industrial Organization*, 22, 193– 217.
- Bergin, J. and D. Bernhardt, (2004), Comparative Learning Dynamics, *International Economic Review*, vol. 45:2, 431–465.
- Ficici S.G. and J. B. Pollack, (2000), Effects of Finite Populations on Evolutionary Stable Strategies, *Proceedings of the 2000 Genetic and Evolutionary Computation Conference, Las Vegas*, Morgan Kaufmann, 927–934.
- Fouraker, L. E. and S. Siegel, (1963) *Bargaining Behavior*, New York:McGraw Hill.
- Friedman, D., (1998), On Economic Applications of Evolutionary Game Theory, *Journal of Evolutionary Economics*, vol. 8, 15–43.
- Hamilton, W.D., (1970), Selfish and Spiteful Behavior in an Evolutionary Model, *Nature*, vol. 228, pp. 1218–1220.
- Hehenkamp B., Possajennikov A. and T. Guse, (2010), On the Equivalence of Nash and Evolutionary Equilibrium in Finite Populations, *Journal of Economic Behavior and Organization*, vol. 73, 254–258.
- Huck, S.; H. Th. Normann; J. Oechssler (1999) Learning in Cournot Oligopoly - An Experiment, *The Economic Journal*, vol. 109, No. 454, Conference Papers, pp. 80–95.
- Huck, S., T. Normann and J. Oechssler, (2004a) Through Trial and Error to Collusion, *International Economic Review*, 45, 205–224
- Huck, S., T. Normann, and J. Oechssler, (2004b), Two are few and four are many: number effects in experimental oligopolies, *Journal of Economic Behavior and Organization*, vol. 53:4, 435–446
- Kimbrough S. and F.H. Murphy, (2009), Learning to Collude Tacitly on Production Levels by Oligopolistic Agents, *Computation Economics*, vol. 33, 47–78.
- Kirman A. P., (1995), Learning in Oligopoly: Theory, Simulations and Experimental Evidence, in Kirman, A.P. and Salmon, M. (eds), *Learning and Rationality in Economics*, Blackwell Publishers, Oxford.
- Mailath G., (1998), Do People Play Nash Equilibrium? Lessons from Evolutionary Game Theory, *Journal of Economic Literature*, vol. 36:3, 1347–1374.
- Maynard-Smith J., (1982), *Evolution and the theory of games*, MIT Press, Cambridge MA.
- Morgan and Steiglitz (2003), Pairwise Competition and the Replicator Equation, *Bulletin of Mathematical Biology*, 65(6), 1163–1172.
- Offerman, T., J. Potters and J. Sonnemans (2002) Imitation and Belief Learning in an Oligopoly Experiment, *Review of Economic Studies*, 69, 973–997.

- Stegeman M. and P. Rhode, (2004), Stochastic Darwinian equilibria in small and large populations, *Games and Economic Behavior*, vol. 49, 171–214.
- Theocharis, R. (1960). On the stability of the Cournot solution on the oligopoly problem, *Review of Economic Studies*, vol. 73, 133–4.
- Vallée, T. and M. Yıldızoğlu, (2009), Convergence in the finite Cournot oligopoly with social and individual learning, *Journal of Economic Behavior and Organization*, 72(2), 670–690.
- Vega-Redondo, F. (1997), The evolution of Walrasian behavior, *Econometrica*, vol. 65, 375–84.
- Vriend, N., 2000, An illustration of the essential difference between individual and social learning, and its consequences for computational analyses, *Journal of economic dynamics and control*, vol. 24, 1–19.
- Waltman, L. and U. Kaymak (2008), Q-learning agents in a Cournot oligopoly model, *Journal of Economic Dynamics & Control*, 32, 3275–3293.

A Appendix

A.1 Numerical specification

$N = 500$; number of runs
 $T = 1000$; number of periods
 $p^m \in [0, 0.1]$; probability of mutation
 $p^I \in [0, 0.05]$; probability of imitation
 $p^M \in [0, 1]$; probability of using memory
 $\sigma \in [0.01, 0.5]$; standard deviation of perturbations before truncation
 $K \in \{1, 10, 20, 50, 100, T - 1\}$; size of the memory
 $n = 10$ and $n \in \{2, 3, 4, 5\}$; number of firms
 $\beta = 0.05$; parameter of mutation

A.2 Pay-offs in the two-players game matrices

Initial position corresponds to (q^{WE}, q^{WE}) :

$$\begin{aligned}\pi^i(q^{WE}, q^{WE}) &= \frac{(a-c)^2 d}{4(b+d)^2} \\ \pi^i(q^{WE} - \epsilon, q^{WE}) &= \frac{a^2 d + c^2 d - 2bc(b+d)\epsilon - 4(b+d)^3 \epsilon^2 + 2a(-cd + b(b+d)\epsilon)}{4(b+d)^2} \\ \pi^i(q^{WE} + \epsilon, q^{WE}) &= \frac{a^2 d + c^2 d + 2bc(b+d)\epsilon - 4(b+d)^3 \epsilon^2 - 2a(cd + b(b+d)\epsilon)}{4(b+d)^2} \\ \pi^i(q^{WE} - \epsilon, q^{WE} - \epsilon) &= \frac{a^2 d + c^2 d - 4bc(b+d)\epsilon - 4(b+d)^2(2b+d)\epsilon^2 + a(-2cd + 4b(b+d)\epsilon)}{4(b+d)^2} \\ \pi^i(q^{WE} + \epsilon, q^{WE} + \epsilon) &= \frac{a^2 d + c^2 d + 4bc(b+d)\epsilon - 4(b+d)^2(2b+d)\epsilon^2 - 2a(cd + 2b(b+d)\epsilon)}{4(b+d)^2}\end{aligned}$$

Initial position corresponds to (q^{CE}, q^{CE}) :

$$\begin{aligned}\pi^i(q^{CE}, q^{CE}) &= \frac{(a-c)^2(b+d)}{(3b+2d)^2} \\ \pi^i(q^{CE} - \epsilon, q^{CE}) &= \frac{(b+d)(a^2 - 2ac + c^2 - (3b+2d)^2 \epsilon^2)}{(3b+2d)^2} \\ \pi^i(q^{CE} + \epsilon, q^{CE}) &= \frac{(b+d)(a^2 - 2ac + c^2 - (3b+2d)^2 \epsilon^2)}{(3b+2d)^2} \\ \pi^i(q^{CE} - \epsilon, q^{CE} - \epsilon) &= \frac{a^2(b+d) - 18b^3 \epsilon^2 - 3b^2 \epsilon(c + 11d\epsilon) + a(-2bc - 2cd + 3b^2 \epsilon + 2bde)}{(3b+2d)^2}\end{aligned}$$

$$\begin{aligned}
& + \frac{b(c^2 - 2cd\epsilon - 20d^2\epsilon^2) + d(c^2 - 4d^2\epsilon^2)}{(3b + 2d)^2} \\
\pi^i(q^{CE} + \epsilon, q^{CE} + \epsilon) &= \frac{a^2(b + d) - 18b^3\epsilon^2 + 3b^2\epsilon(c - 11d\epsilon) + b(c^2 + 2cd\epsilon - 20d^2\epsilon^2)}{(3b + 2d)^2} \\
& + \frac{d(c^2 - 4d^2\epsilon^2) - a(2cd + 3b^2\epsilon + 2b(c + d\epsilon))}{(3b + 2d)^2}
\end{aligned}$$

Initial position corresponds to (q^{CS}, q^{CS}) :

$$\begin{aligned}
\pi^i(q^{CS}, q^{CS}) &= \frac{(a - c)^2}{4(2b + d)} \\
\pi^i(q^{CS} - \epsilon, q^{CS}) &= \frac{a^2 + c^2 + 2bc\epsilon - 4(2b^2 + 3bd + d^2)\epsilon^2 - 2a(c + b\epsilon)}{4(2b + d)} \\
\pi^i(q^{CS} + \epsilon, q^{CS}) &= \frac{a^2 + c^2 - 2bc\epsilon - 4(2b^2 + 3bd + d^2)\epsilon^2 - 2a(c - b\epsilon)}{4(2b + d)} \\
\pi^i(q^{CS} - \epsilon, q^{CS} - \epsilon) &= \frac{a^2 - 2ac + c^2 - 4(2b + d)^2\epsilon^2}{4(2b + d)} \\
\pi^i(q^{CS} + \epsilon, q^{CS} + \epsilon) &= \frac{a^2 - 2ac + c^2 - 4(2b + d)^2\epsilon^2}{4(2b + d)}
\end{aligned}$$