UNIVERSITÀ DEGLI STUDI DI PADOVA

Dipartimento di Scienze Economiche "Marco Fanno"

INFORMATION AND LEARNING IN OLIGOPOLY: AN EXPERIMENT

MARIA BIGONI
Università di Padova

March 2008

"MARCO FANNO" WORKING PAPER N.72

Information and Learning in Oligopoly: an Experiment*

Maria Bigoni

March 14, 2008

Abstract

I report results of an experiment designed to study the relation between the process of information search and learning in a Cournot oligopoly, with limited *a priori* information. Different theories of learning have been applied to this setting, each yielding a specific market outcome in the long run, and postulating specific informational requirements. By allowing players to choose the information they wish to acquire, and controlling for these choices, I study the features of the learning model actually followed by the subjects, and the relation between the information they gather and the market behavior they adopt.

According to my results, learning appears to be a composite process, in which different components coexist. Belief learning seems to be the leading element, as subjects try to form expectations about their opponents' future actions and to best reply to them. When subjects also look at the strategies individually adopted by their competitors, though, they tend to imitate the most successful behavior, which makes markets more competitive. Finally, reinforcement learning also plays a non-negligible role, as subjects tend to favor strategies that have yielded higher profits in the past. I show that these different elements may be usefully incorporated into a more sophisticated learning model, shaped after self tuning EWA learning model.

Keywords: Information; Learning; Imitation; Collusion

JEL classification: L13; C92; C72

*I am indebted to Sven Olof Fridolfsson, Chloè Le Coq and Giacomo Calzolari for help in organizing and conducting the experiments. I am also grateful to Giancarlo Spagnolo, Ken Binmore, Colin Camerer, Andreas Ortmann, Magnus Johannesson and to participants to the 2007 AFSE-JEE meeting in Lyon for helpful comments and suggestions. The usual disclaimer applies.

1 Introduction

The experimental study I present concerns the relation between the process of information search and players' behavior in a repeated Cournot oligopoly. This research is aimed at finding out what happens when information acquiring and processing is too difficult or too costly for the agents to behave according to the perfect rationality paradigm. I investigate what pieces of information subjects look for and which heuristics they adopt when playing a Cournot game with these information constraints, in order to understand how different pieces of information may affect the subjects' learning processes and in turn also their market behavior.

The topic is not particularly new: the interest on it reached its apex after 1997, when in an article appeared in Econometrica, Vega-Redondo [26] proposed a theoretical model of behavior of Cournot oligopolists which leads to surprising conclusions. According to the author's theory, if firms tend to imitate the behavior that proved most successful in the previous period (that is: they produce the level of output that yielded the highest profit) but with positive probability experiment other strategies, Walrasian behavior can emerge in the long run within any Cournot oligopoly with homogeneous goods. In a number of following works [1] Vega Redondo's theory has been experimentally tested and compared with other learning models that make different assumptions about players' information and lead to different behaviors and market equilibria.

Despite the efforts made during those years, I do not believe that a final conclusion has been reached. Rather, it seems that the attention has been averted from this topic but a definitive answer to the questions it brings up has not yet been provided.

The main novel contribution of the experiment presented in this paper consists in combining the study of learning with an experimental analysis of the way subjects select the information they need before choosing their strategy. Instead of comparing subjects market behavior under different informational frameworks – which is the approach adopted in all the previous experiments about this topic – I provide the players with a broad range of information, but force them to choose only some pieces of it. The players' process of information gathering is strictly (but non obtrusively) controlled, by means of a special software, originally called MouseLab and developed by Eric J. Johnson et al. (1988) [17].

Paying attention not only to what players do but also to what they know, it is possible to better understand the mental mechanisms which guide their choices and consequently the impact that the informational framework has over their behavior.

I believe that the results of my experiments can contribute to the debate which has developed after the article by Vega-Redondo [26], because the technique I adopt enable me to investigate more deeply the relation between information and behavior.

Moreover, with my experiment I want to test the MouseLab technique as an experimental device that – despite some very insightful applications – is not yet widely used in experimental economics and that could be effectively adopted to investigate other interesting topics, such as the process of information acquisition in auctions and phenomena like informational cascades and herding behavior in financial markets.

The main results I get are that: (*i*) the attitude to best reply to the strategies adopted by the opponents in the previous period seems to be the most important driver of players' behavior, (*ii*) imitation plays a non negligible role in learning, and drives players' choices away from the best reply, (*iii*) information gathered by the subjects affects their behavior and (*iv*) it may be usefully incorporated into a rather sophisticated learning model, shaped after self tuning EWA learning model, proposed by Ho, Camerer and Chong (2007) [11].

The paper proceed as follows: in Section 2 I mention the main literature related with the topic I study and with the empirical methodology I adopt. Section 3 presents the experimental design. Three theoretical benchmarks meant to support the analysis of the results are illustrated in Section

[1] see for example Huck *et al.* 1999 [13], Rassenti *et al.* 2000 [22], Offerman *et al.* 2002 [20] and Bosh-Domènech and Vriend 2003 [3]. More recently, further experiments on Vega-Redondo's imitation model have been conducted by Apesteguia *et al.* [1, 2], but these experiments were not framed as oligopolies.

4. Section 5 contains the results of the experiment and Section 6 concludes, also mentioning possible future developments of this research.

2 Related Literature

2.1 Information and Learning in Oligopoly Experiments

Between 1999 and 2002, four articles were published, which presented experiments regarding information and learning in a Cournot oligopoly setting. In these works the same experiment is repeated under different treatments, varying the quality and quantity of information provided to the subjects. The authors then compare the actual behavior observed in the different treatments and make inference about the impact that the various informational frameworks have on players' choices. Nonetheless a number of details changes from one experiment to the other, and maybe this is the first reason why the results obtained by the authors are not at all unanimous, nor they are conclusive: for example, the experiments performed by Huck, Normann and Oechssler [13] and by Offerman, Potters and Sonnemans [20] provide a rather strong support to the theory proposed by Vega Redondo, mentioned before, while the works presented by Rassenti et *al* [22] and by Bosch-Domènech and Vriend evidence no trend towards the Walrasian equilibrium and do not find any clear indication that players tend to imitate the one who got the best performance in the previous period.

Huck et *al.*'s experiments (HNO from now on) study a 40-periods Cournot market with linear demand and cost, in which four symmetric firms produce a homogeneous good. Across their five treatments, they vary the information they provide to the subjects, both about market and about what other players in the same market do. In particular, information about market can be complete, partial or absent.

When information is complete, participants are informed about the symmetric demand and cost functions in plain words and they are provided with a 'profit calculator', which can compute market price and firm's profit when one enters the total output of other firms and his own output, and can also suggest to the subject the quantity which would yield him the highest payoff given the hypothetical total quantity produced by the competitors.

Information is said to be absent when participants do not know anything about the demand and cost conditions in the market nor do the instructions explicitly state that these would remain constant over time; in these treatments all subjects know is that they would act on a market with four sellers and that their decisions represent quantities. Finally, in treatments with partial information, participants are just told that market conditions remain constant for all periods and coarsely informed about demand and profit functions.

In three of the treatments, participants are also informed about competitors' individual quantities and profits in the previous period, while in the remaining two treatments they are told only the total quantity the others have actually supplied. HNO find significant differences in individual and aggregate behavior across the treatments, and collect data suggesting that increasing information about the market decreases total quantity, while providing additional information about individual quantities and profits increases total quantity. HNO also test other learning theories besides the one proposed by Vega Redondo, and they find that when subjects know the true market structure, their quantity adjustments depend significantly on the myopic best reply to the quantity produced by their competitors in the very last period. In general, though, none of the theoretical learning models they consider, *per se*, seems to fully explain the observed behavior.

Offerman et *al.* [20] conducted a similar computerized experiment, obtaining results which are consistent and complementary to those presented by HNO. In their setting, a triopoly with non-linear demand and cost functions is repeated 100 times, with complete information about market. The authors study how players' behavior changes across three treatments, which differ for the amount of information provided to the subjects about individual quantities and revenues of the other two competitors in their market.

In one treatment ($Qq\pi$) firms were provided with individualized information about the quantities and the corresponding profits of the other two firms; in a second treatment (Qq) they were just told the quantities produced by the opponents, but not their profits, and in the last treatment (Q) firm were only informed of the total quantity produced it their market. As HNO, they observed a substantial difference between the treatments, and the data they collected evidence that the feedback information provided to the subjects affects the behavioral rules they adopt. Moreover, in agreement with what reported by HNO, the Walrasian outcome is only reached quite often in treatment $Qq\pi$, where the players are informed about their opponents'profits. On the other hand, they observe that the collusive outcome seems to be a stable rest point only in treatment Qq and $Qq\pi$, but not in the treatment with no information about others' individual quantities and profits, in which the only rest-point is represented by the Cournot-Nash equilibrium.

The experiment performed by Bosch-Domènech and Vriend (BDV) differs from the previous two both for the setting and for the aims. While HNO and OPS compare the prognostic capability of different learning rules which lead to different theoretical outcomes, here the authors focus specifically on Vega-Redondo's behavioral rule and the main purpose is to investigate whether people are incline to imitate successful behavior and, in particular, whether this behavior is more prevalent in a more demanding environment. The authors study a series of 22-periods Cournot duopolies and triopolies with homogeneous commodity and linear demand and cost functions. They examine six treatments altogether: for both duopolies and triopolies they consider three different treatments that differ in the way information is provided and in the time pressure put on the players.
In the treatment denominated "easy", the players are given a profit table that conveniently summarizes all the information concerning the inverse demand curve and the cost function, and there is no time pressure on the players. After each period, each player gets information about the actions of each of the other players in the same market, but not about their profits.
In the 'hard' and 'hardest' treatments, players have just one minute to decide on their output level; after each period they receive feedback information both about the actions of all players and about the profits obtained by each of them, and the output decision which led to the highest profit is highlighted.
In the 'hard' version, the players get an inconveniently arranged enumeration of the market prices associated with all possible aggregate output levels and of all possible cost levels. The 'hardest' version differs from the 'hard' treatment in that the information about the demand side of the market is limited to the statement that 'the price level depends on aggregate output'.
The purpose of the 'hard' and 'hardest' treatments is to explore to what extent imitation is influenced by the bounds imposed on the subjects' choice capabilities and to check if it is actually more prevalent when the task of learning about the market becomes more difficult and at the same time the decision of the most successful firm is displayed more prominently, and the answer they give to this question is essentially negative. The data they collected show that as the learning-about-the-environment task becomes more complex, average output increases, but the Walrasian output does not seem to be a good description of the output levels observed in the experiment and if anything, imitation of successful behavior tends to decrease rather than to increase when moving to more complicated environments.

The fourth experiment has been conducted by Rassenti *et al.* (RRSZ); it represents an oligopoly with homogeneous product, in which five firms interact repeatedly for 75 periods, with fixed payoff conditions. The setting exhibits a substantial difference from the previous three since in this case the cost functions – linear, with constant marginal costs and no fixed costs – are private information and differ across the firms. The demand function is linear, and is public information among the players.
The authors perform two different treatments: one in which subjects were able to observe past output choices of each one of their rivals, the other in which they are informed only about the past total output of rivals.
They use their experimental results to test a number of learning models – such as best response dynamics, fictitious play and more general models of adaptive learning. None of these models receives strong support from the data they collected: the observation of actual movement of total output over time appears to be inconsistent with both best response dynamic and with fictitious play, for most

experiments. Moreover the authors show that their data do not provide any evidence neither in support for learning models based on imitation, nor for the more traditional hypothesis that information about competitors enhance the potential for collusion, because the treatment conditions involving provision of information about rivals' outputs and prior experience do not seem to have a significant effect on total output levels. The evidence relative to individual behavior is mixed, and no predominant models of learning emerge; the most prominent result is that in general observed behavior for individual subject sellers is not converging to the static Nash equilibrium predictions for individual output choices in these experiments.

In light of these results, I would conclude that it is worthwhile going on investigating on information and learning in oligopolistic markets, because the topic is interesting from a theoretical point of view and it also has interesting practical implications, but a theory consistent with experimental data is still far from being definitely developed. For this reason I decided to design an experiment which is similar to the four previously mentioned under many respects but introduces the use of an experimental technique that allowed me to monitor the information acquisition process through a computer interface. The main idea underlying this software – originally developed by Johnson *et al.* (1988) [17] – consists in hiding relevant pieces of information behind a number of boxes on the screen so that to access them the decision maker has to open the boxes and look at their content. He can open just one box at a time, and by recording the number and the duration of the look-ups the program provides precious information about the decision makers' learning process. To my knowledge, this technique has never been applied to the analysis of learning processes in repeated strategic games.
Next section summarizes four among the most famous experiments using Mouselab, with the aim of explaining how this program works in practice and of pointing out its strength points. A more detailed survey on the experimental study of cognition via information search can be found in Crawford (in press) [7].

2.2 Experiments Controlling the Information Acquisition Process

One of the most famous experiments using MouseLab has been performed by Johnson, Camerer, Sen and Rymon(JCSR) [16]: in this work the information acquisition process is observed with the aim of testing the game theoretic assumption of backwards induction. The subjects were asked to play eight three-round alternating-offer bargaining games, with a different anonymous opponent each time. In the first round one of the two players makes an offer to his opponent about how to share a given amount of money; if the other player accepts, than the game is concluded, otherwise he will have to make a counteroffer about how to share a new pie, smaller than the first one. Again, if the first player accepts, the game is over and each of them gets his part as established in the agreement; on the other hand, if the first player rejects the offer the pie shrinks again and he will have the opportunity to make one last offer to his opponent. If even this offer is rejected, nobody gets anything. The sizes of the three pies are represented on the computer screen in front of each player, but they are hidden under three boxes that can be opened only one at a time, simply by putting the mouse' cursor over the box itself. The box will stay open until the mouse is moved somewhere else.
The authors observe three measures of information search: the number of times each box is opened in a period, the total time each box stays opened in a period and the number of transitions from one specific box to another. They note that most of the looking time is spent looking at the first round pie size and contrary to the backward induction prediction there are always more forward predictions than backward ones. From the data collected through these experiments they conclude that people do not use backwards induction instinctively, even if an additional treatment in which players are previously trained to use backward induction shows that people are able to learn it when appropriately instructed.
They also found that there is a strong correlation between differences in information processing and differences in players' behavior. This and the other results presented in this paper testify that measuring attention directly can effectively contribute to the comprehension of both failure and successes of the game theoretic predictions and help to understand how information and learning can affect the

outcomes of different games.

Another seminal study on information acquisition processes has been done by Costa-Gomes, Crawford and Broseta [6](CGCB). They asked the subjects to play 18 two-players normal form games, with different anonymous partners. The payoff tables are hidden and MouseLab is used to present them: for every combination of strategies, subjects could look up their own or their partner's payoff as many times as they wanted, but they could only see one of these numbers at a time. Till the end of the series of games, no feedback was provided to the agents, in order to suppress learning and repeated game effects as much as possible.
In Johnson *et al.* the goal was to test a specific theory of behavior – namely backward induction. On the contrary, here the authors compare nine different decision rules (or types) and try to make inference about which one is more likely to inform players' behavior. As in JCSR, they assume that each decision rule determines both a player's information search and his decision once he gets the information he was looking for. Therefore, by observing both the information acquisition process performed by the agents and the choices they actually make when playing the games, it is possible to deduce what decision rule they adopt.
This study confirms the presence of a systematic relationship between subjects' deviations from search pattern associated with equilibrium analysis and their deviations from equilibrium decisions. Besides, according to Costa-Gomes *et al.*'s analysis most of the subjects are much less sophisticated than game theory assumes: between 67% and 89% of the population belong to two types, namely to the *Naïve* type, who best responds to beliefs that assign equal probabilities to each of their partner's possible strategies and to *L2* type, who best replies to *Naïve* subjects.

More recently, MouseLab has been used again in two experiments that provide further evidence about how the study of the information acquisition process can be useful to understand what behavioral rules and heuristics are adopted by subjects who display out of equilibrium choices.
One experiment has been conducted by Costa-Gomes and Crawford [5](CGC) and has the same theoretical and econometric framework of CGCB but it differs for the class of games submitted to the subjects. In this case, participants were requested to play 16 different two-person guessing games, with anonymous partners and no feedback till the end of the series. The games have been designed so that the space of possible behaviors is wide and there is a strong separation of the guesses and searches implied by the different decision rules considered in the paper. Results are consistent with those presented in CGCB [6], but they are significantly sharper: many subjects can be easily attributed to a particular type only by their guesses, and most of the others can be identified via an econometric and specification analysis keeping into account also their information search pattern.

Another interesting application of MouseLab has been recently presented by Gabaix, Laibson, Moloche and Weinberg [10], who experimentally evaluate the *directed cognition model*: a bounded rationality model that assumes that at each decision point, agents act as if their next search operations were their last opportunity for search. As in the other three experiments, the authors register the search pattern actually adopted by the subjects in two experiments and they compare it with with what is predicted by the *directed cognition model* and by the optimal search model (i.e. the Gittins-Weitzman algorithm), traditionally adopted in economics.
In the first experiment they asked the participants to choose among three projects whose outcome is uncertain, but could be discovered at a given cost. In the second experiment the subjects were requested to solve a highly complex choice problem in which the classical optimal choice model is analytically and computationally intractable: they had to choose one out of eight goods which each have nine attributes that could be discovered by opening different boxes on the computer screen. The players cannot collect all the information about the goods, because in this game time is a scarce resource. Individual information acquisition processes were recorded through the MouseLab interface, and the data collected this way reveal that the directed cognition model successfully predicts the empirical regularities observable in subjects' behavior.

The four experiments mentioned in this section evidence how the study of the information acquisition process is complementary to the observation of subjects' actual choices which traditionally constitutes the empirical basis for testing models of decision making or trying to develop new ones.

3 Experimental Design

The market environment I have chosen for my experiments is similar to the one proposed by HNO [13]; if possible it is even simpler. In all the sessions and treatments, the setting remains the same. Four identical firms compete à la Cournot in the same market for 40 consecutive periods. Their product is perfectly homogeneous. In every period t each firm i chooses its own output q_i^t from the discrete set $\Gamma = \{0, 1, ..., 30\}$, which is the same for every firm. The choice is simultaneous.
Price p^t in period t is determined by the inverse demand function:

$$p^t = \max(0, 81 - \sum_i q_i^t)$$

Let $C_i(q_i^t) = q_i^t$ be the cost function for every firm i; firm i's profit in period t will be denoted by

$$\pi_i^t = p^t q_i^t - C_i(q_i^t).$$

The shape of these functions has been chosen so that the three main theoretical outcomes – namely collusive, Cournot and Walrasian outcomes – are well separated one from the other and belong to the choice set Γ. More precisely, collusive equilibrium is denoted by $\omega^M = (10, 10, 10, 10)$, Cournot-Nash equilibrium is $\omega^N = (16, 16, 16, 16)$ and Walrasian equilibrium is $\omega^W = (20, 20, 20, 20)$.

A time limit of 30 seconds per round was introduced, so to force subjects to choose the information they are really interested in, and to reproduce an environment in which rationality s bounded because of external factors. If a subject failed to make his choice within the time limit, his quantity was automatically set equal to 0, granting him a profit of 0 for that period.

3.1 Information Provided to the Subjects

Participants knew how many competitors they had (anonymity was nonetheless guaranteed). Instructions explained in plain words that there is an inverse relation between the overall quantity produced by the four firms and market price and that a firm's production costs increase with the number of goods it decides to sell. Besides, players were told that per-period profit is given by market price times the number of goods sold by the firm, minus production costs (see the instructions in Appendix A).

Subjects were also endowed with a *profit calculator* similar to the one proposed by Huck *et al.* [13]. This device had two functions: (i) it could be used by a player to evaluate the quantity that would yield him the highest profit, given the quantity produced on the whole by his three competitors, and to compute the profit he would earn if he produced the suggested number of units; (ii) it could also be used to calculate the profit given both the quantity produced by the player himself and the overall quantity produced by his competitors.
The software I developed for this experiment recorded how many times each subject used the profit calculator and every trial he did.

The number of rounds was common knowledge among the subjects. According to game-theoretic predictions, cooperation should be sustainable only if our stage game were repeated in(de)finitely many times, but according to Selten *et al.* [23]

> Infinite supergames cannot be played in the laboratory. Attempts to approximate the strategic situation of an infinite game by the device of a supposedly fixed stopping probability are unsatisfactory since a play cannot be continued beyond the maximum time available. The stopping probability cannot remain fixed but must become one eventually.

In light of this consideration and of the results obtained by Normann and Wallace [19] – who show that the termination rule does not have a significant effect on players' behavior except for an end effect – I decided to adopt a commonly known finite horizon, for sake of transparency and practicality.

In every period after the first one, the profits earned in the previous period by the player himself and by each of his opponents were displayed. Three distinct buttons – each corresponding to one of the player's competitors – served to display the strategy they chose in the previous period, that is the quantity they decided to produce. Another button allowed the subject to open a window displaying, by means of a table and a couple of plots, the quantity chosen and the profits earned by the player himself in every previous period. It was also possible for the player to look at the aggregate quantity produced in each of the previous periods by his competitors. This information was conveyed through a table and a plot, if the subject pushed the corresponding button.

As mentioned before, it was not possible to access various pieces of information at the same time, since opening a new window automatically closed the previous one. The graphical interface granted me a deep control over the subjects' information search behavior, making sure that every look-up is intentional and allowing me to verify which piece of information is more interesting for the subjects.

On the computer screen there was a counter showing the running cumulative profits earned by the player since the game began, and a timer displaying how many seconds remained before the end of the current period. Figure 9 in the appendix shows how subjects' computer screen looked like.

4 Three Theoretical Benchmarks

I am interested in studying the market dynamics when the stage game so defined is repeated several times, and the firms do not have all the information (or the computational capabilities) to evaluate what the standard theory predicts is an optimal behavior for them.

	Required information	**Predicted equilibrium**
Best Reply Dinamics	Competitors' aggregate quantity and BR function	Nash ($q = 16$)
Imitate the Best	Last period individual profits and quantities	Walrasian ($q = 20$)
Trial and Error	Own past profits and quantities	Collusive ($q = 10$)

Table 1: *Theoretical benchmarks*

As a first benchmark to evaluate the experimental results, I individuated three theoretical learning models based on very different assumptions on the information available to the firms and yielding three well distinct market outcomes, namely the Cournot, Walrasian and joint profit maximizing outcomes, respectively. These models are summed up in table 1 and presented more in detail in the following paragraphs.

4.1 Best Response Dynamics

Following Huck, Normann and Oechssler [13] I consider here the simplest model of best reply dynamic. This model theorize that in every period each player myopically chooses his output as a best reply to the sum of the quantities produced by the other three in the previous period. More precisely, the best reply correspondence for player i maps $\sum_{j \neq i} q_j^{t-1}$ to the set

$$BR_i^t := \{q \in \Gamma : \pi_i^t(q, \sum_{j \neq i} q_j^{t-1}) \geq \pi_i^t(q', \sum_{j \neq i} q_j^{t-1}), \ \ \forall q' \in \Gamma\}.$$

Under the hypotheses I made on market structure, due to the discreteness and finiteness of the choice set, we have:

$$BR_i^t = \begin{cases} \{0\}, \text{ if } \sum_{j\neq i} q_j^{t-1} \geq 80 \\ \{30\}, \text{ if } \sum_{j\neq i} q_j^{t-1} < 20 \\ \left\{\frac{80-\sum_{j\neq i} q_j^{t-1}}{2}\right\} \text{ if } 20 \leq \sum_{j\neq i} q_j^{t-1} < 80 \text{ and } \sum_{j\neq i} q_j^{t-1} \text{ is even} \\ \left\{\frac{80-\sum_{j\neq i} q_j^{t-1}}{2} - 0.5,\ \frac{80-\sum_{j\neq i} q_j^{t-1}}{2} + 0.5\right\} \text{ otherwise.} \end{cases}$$

In this last case, it is assumed that the player chooses q_i^t from BR_i^t according to some probability distribution with full support.

The best reply dynamic defined this way yields a Markov chain over the state space $\Omega = \Gamma^4$ which does not necessarily converge to a stable equilibrium, consistently with what has been shown by Theocharis [25] for the case in which quantities are chosen in a continuous space.
To catch an intuition of this result, suppose for example that the system reaches one of the two states of the absorbing set $s = \{(30, 30, 30, 30), (0, 0, 0, 0)\}$: once this has happened, the system will keep on oscillating between this two states and will never be able to escape the set.
HNO state the following theorem:

> Theorem 1 The best reply dynamic with inertia converges globally in finite time to the static Nash equilibrium.

Within the framework considered here, this implies that the learning process brings the system to converge to the state $\omega^N = (16, 16, 16, 16)$.

For sake of completeness I replicate here HNO's demonstration, applying it to the specific context under exam.
To introduce inertia into the learning model, HNO simply hypothesize that in every period each player i chooses q_i^t from the set BR_i^t with some fixed probability $(1-\theta)$, while with probability θ he sticks to the quantity he chose in the previous period, so $q_i^t = q_i^{t-1}$. We will see that the best reply dynamic with inertia can be represented by an ergodic Markov chain having only one recurrent set, therefore the probability distribution over the state space approximate the unique invariant distribution of the process, regardless of the initial state, and that this invariant distribution puts probability one over the state ω^N, which is the only recurrent state in Ω.

Proof of Theorem 1: It is clear that ω^N is an absorbing state, since the process we have defined can never escape from it; in order to prove the result, it is necessary to demonstrate that no other state in Ω is recurrent, namely ω^N is accessible from any other state:

$$\omega' \to \omega^N \ \ \forall\, \omega' \in \Omega,\ \omega' \neq \omega^N$$

which means that there exist a $\tau \in \mathbb{N} \setminus \{0\}$ such that the probability $p^{(\tau)}_{\omega',\omega^N}$ of reaching state ω^N from ω' in τ periods is positive.

To prove this result we shall first show that the state ω^N is accessible from any $\omega^+ \in \Omega^+$, where $\Omega^+ = \{(q_1, q_2, q_3, q_4) \in \Omega : q_i > 0,\ i = 1, 2, 3, 4\}$, then we shall conclude by verifying that for any $\omega^0 \in \Omega \setminus \Omega^+$ there exists an $\omega^+ \in \Omega^+$ such that ω^+ is accessible from ω^0 ($\omega^0 \to \omega^+$), therefore, by the Chapman-Kolmogorov equation, it follows that $\omega^0 \to \omega^N$. $\forall \omega^0 \in \Omega$.

The first part of the proof requires the preliminary definition of the concepts of "ordinal potential", "ordinal potential game" and "improvement path", introduced by Monderer and Shapley [18].
Let $N = \{1, 2, ..., n\}$ be the set of players, Y_i denote the set of strategies of player i and $u_i : Y \to \mathbb{R}$ the payoff function of player i, where $Y = Y_1 \times Y_2 \times ... \times Y_n$ is the set of the strategy profiles.
A function $P : Y \to \mathbb{R}$ is an *ordinal potential* for the game $G = (N, Y, u)$ if, for every $i \in N$ and for every $y_{-i} \in Y_{-i}$

$$u_i(x, y_{-i}) - u_i(z, y_{-i}) > 0 \Leftrightarrow P(x, y_{-i}) - P(z, y_{-i}) > 0 \quad \forall x, z \in Y_i$$

An *ordinal potential game* is a game that admits an ordinal potential.
An *improvement path* in Y is a sequence $\gamma = (y^0, y^1, ...)$ of elements of Y such that, for every $k \geq 1$, there exists a unique player – say player i – such that the following conditions are simultaneously satisfied:

- $y^k = (x, y^{k-1}_{-i});\ x \in Y_i,\ x \neq y^{k-1}_i$
- $u_i(y^k) > u_i(y^{k-1})$

The proof of Theorem 1 relies on the following Lemma by Monderer and Shapley [18]:

Lemma 1.1 Every improvement path of a finite potential game is finite.

Proof of Lemma 1.1: For every improvement path $\gamma = (y^0, y^1, ...)$, by the definition of ordinal potential we have:

$$P(y^0) < P(y^1) < P(y^2) < \dots$$

As Y is a finite set, this sequence must be finite.

This result can be applied to our model, since – as shown by HNO – the function $P(\omega) = (p(\omega) - 1)\prod_{j=1}^{4} q_j$ is an ordinal potential for our game if the strategy set of each player is restricted to $\Gamma \setminus \{0\}$. Therefore, there is a finite improvement path departing from every state $\omega^0 \in \Omega^+$ where $\Omega^+ = \{(q_1, q_2, q_3, q_4) \in \Omega : q_j > 0,\ j = 1, 2, 3, 4\}$.
By definition this improvement path ends in a state ω^k such that no player can improve his own payoff by changing his strategy if the quantities chosen by the other players remain the same, i.e.

$$\nexists i \text{ s.t. } \pi_i(\omega^{k+1}) > \pi_i(\omega^k) \text{ where } \omega^{k+1} = (q', q^k_{-i}) \text{ for some } q' \neq q^k_i,\ q' \in \Gamma \setminus \{0\}$$

This condition is clearly satisfied only by $\omega^C = (16, 16, 16, 16)$, representing the unique Nash equilibrium of the stage game.

Finally, note that the Best Reply process with inertia can give rise to an improvement path over Ω^+, since with positive probability in every period only one player changes his strategy while the others stick to the quantity they previously chose. So, under the Best Reply process, ω^N is accessible from every state in Ω^+.

As mentioned before, to complete the proof of Theorem 1 it is enough to show that

$$\forall \omega^0 \in \Omega \setminus \Omega^+\ \exists \omega^+ \in \Omega^+ \text{ s.t. } \omega^0 \rightarrow \omega^+$$

Let $\omega^0 = (q^0_1, q^0_2, q^0_3, q^0_4) \in \Omega \setminus \Omega^+$ and ω^{BR} be a state in which every player i chooses $q_i \in BR^0_i$, giving a best reply to ω^0. By definition, ω^{BR} is accessible from ω^0 under the process defined by the Best Response Dynamics ($\omega^0 \rightarrow \omega^{BR}$).
If $\sum_i q^0_i < 79$ it is straightforward to see that $\omega^{BR} \in \Omega^+$ since the best reply to a quantity strictly smaller than 79 is always positive.
If $\sum_i q^0_i \geq 79$ it can be shown that the sum $\sum_i BR^0_i \leq 162 - \frac{3}{2}\sum_i q^0_i \leq 43.5$, since for every player i $BR^0_i \leq \frac{80 - \sum_{j \neq i} q^0_j}{2} + \frac{1}{2}$. Therefore $\exists \omega^+ \in \Omega^+$ s.t. $\omega^{BR} \rightarrow \omega^+$, thus the system moves from ω^0 to ω^+ with positive probability in at most 2 steps.
This concludes the proof.

The problem with the analysis I have just concluded is that it makes predictions for the long run outcomes, and for a positive, but undefined degree of inertia. Since it might be interesting also to check how the process behaves in the short run, for various degrees of inertia, I complete this and the following sections by presenting the results of simulations I have done for this and the other two models, under the setting presented in section 3. My simulation then reproduces a market with four firms, facing the demand and cost functions characterizing my experiments, and interacting for 40 consecutive periods. The quantities produced in the first period are randomly drawn from a uniform distribution over the set $\{0, 1, \dots, 29, 30\}$, while firms' behavior in the following periods – in this case –

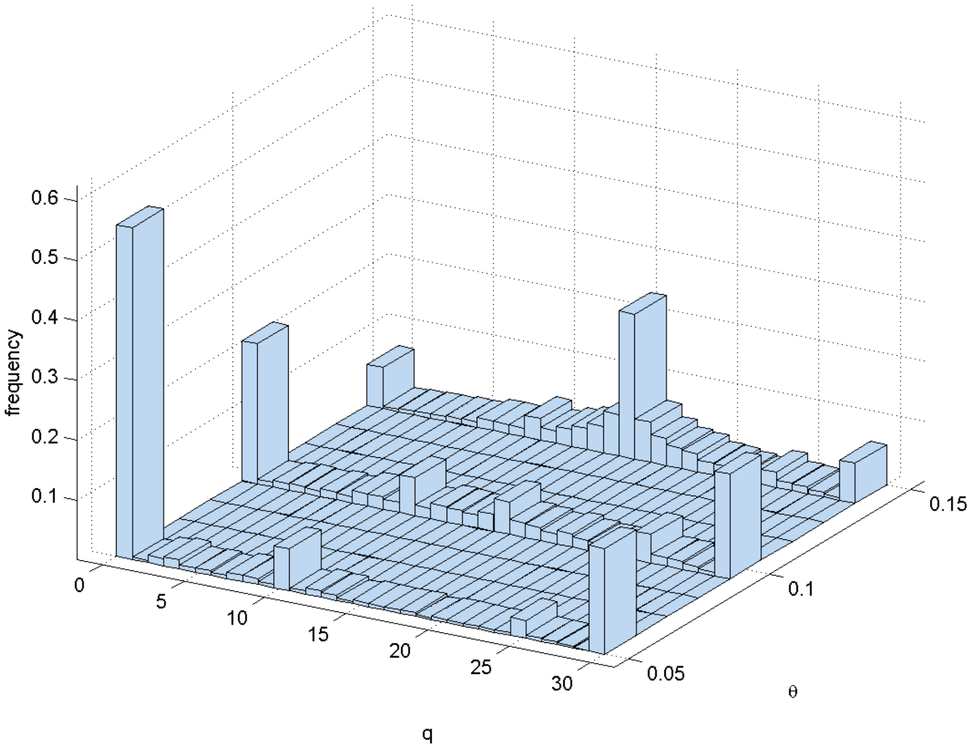

Figure 1: *Frequency distribution of individual choices in the 40th round; 10000 cycles of simulation.*

is determined by the best response dynamics with a degree of inertia equal to θ, taking values 0.05, 0.1, 0.15. I ran 10000 cycles of simulation per each value of θ. The frequency distributions of individually chosen quantities in period 40 are reported in figure 1.

According to my simulations, the higher is the degree of inertia, the faster the convergence to the static Cournot-Nash equilibrium. More specifically, we notice that a degree of inertia below 10% is not sufficient to obatin some convergence within 40 periods.

4.2 Imitate the Best

The learning model presented here has been originally proposed by Vega-Redondo [26]. The core of the model is represented by the *imitation dynamic*: a discrete time dynamic which assumes that at every time t each firm chooses its output q_i^t from the set:

$$B^{t-1} = \{q \in \Gamma : \exists j \in I \text{ s.t. } q_j^{t-1} = q \text{ and } \pi_j^{t-1} \geq \pi_i^{t-1} \; \forall i \in I, \, i \neq j\}$$

This learning process, when applied to the specific context of our fictitious market, defines a Markov chain over the state space $\Omega = \Gamma^n$ (where $n = 4$ in our case). Let ω_q stand for the *monomorphic state* $(q, q, ..., q)$ in which every firm chooses the same quantity $q \in \Gamma$. It is easy to verify that $\forall q \in \Gamma$ the *monomorphic state* ω_q is absorbing and that all the non-monomorphic states are transient. Therefore, the process has a number of recurrent sets equal to the cardinality of Γ, and there is a stationary distribution μ_q corresponding to each of them, which puts probability one over ω_q. Thus, the long run behavior of the evolutionary process consisting only in the imitation dynamic displays a large potential multiplicity, since it can rest forever in any monomorphic state.

To investigate the robustness of each of these multiple outcomes, Vega-Redondo introduced a perturbation into the process, assuming that in every period t each firm sets its quantity according to the imitation rule with probability $1 - \epsilon$, while with probability ϵ it departs from the rule and chooses its quantity according to a distribution with full support over Γ. The interpretation here can be that with small probability every firm makes an error or it experiments a different strategy. This *perturbed process* defines a Markov chain irreducible and ergodic – since each state is accessible from any other one and all the states are aperiodic. As a consequence, the chain has only one stationary distribution μ_ϵ, which clearly depends on ϵ; moreover, the τ-steps transition matrix $P^{(\tau)}$ converges to a rank-one matrix in which each row is the stationary distribution μ, that is:

$$\lim_{\tau \to \infty} P^{(\tau)} = \mathbf{u}\mu$$

where $\mathbf{u}$ is the unit vector: namely, the Markov chain converges to its stationary distribution, regardless where it began.

Recall that the perturbation has been introduced into the imitation process in order to test the robustness of the multiple outcomes of the unperturbed process. We are then interested in investigating the behavior of the perturbed process as $\epsilon \to 0$.

The crucial result for our application is a straight consequence of the theorem stated by Vega-Redondo:

> THEOREM 2 The limit distribution $\mu^0 = \lim_{\epsilon \to 0} \mu_\epsilon$ is a well defined element of the unit simplex $\Delta(\Omega)$. Moreover, μ^0 puts probability one over the state $\omega^W = (q^W, q^W, q^W, q^W)$ where q^W is such that $p(nq^W)q^W - C_i(q^W) \geq p(nq^W)q - C_i(q)\ \forall q \in \Gamma$.

. This implies that under the "imitate the best" dynamic, the only stochastically stable outcome in our setting is the Walrasian outcome $\omega^W = (20, 20, 20, 20)$, in which all the firms get zero profits.

PROOF OF THEOREM 2:

The proof of this theorem relies on the graph-theoretic techniques developed by Freidlin and Wentzell [9], therefore some basic concepts should be introduced in order to expose it.
A *directed graph* G is an ordered pair $G := (V, A)$ with

- V, a set of vertices or nodes,
- A, a set of ordered pairs of vertices, called directed edges, arcs, or arrows.

An edge $e = (x, y), x, y \in V$ is considered to be directed from x to y, so y is said to be a direct successor of x, and x is said to be a direct predecessor of y. More generally, if there exists a path leading from x to y, then y is said to be a successor of x, and x is said to be a predecessor of y.
To apply this idea to the situation under analysis, first let O be the directed graph having Ω as the vertex set and in which for every vertex $\omega \in \Omega$ there exists an edge $e(\omega, \omega')$, $\forall \omega' \in \Omega$, $\omega' \neq \omega$.
A *resistance* $r(\omega', \omega'')$ can be associated to every edge $e(\omega', \omega'')$, where

$$r \text{ s.t. } 0 < \lim_{\epsilon \to 0} \epsilon^{-r} P^{\epsilon}_{\omega', \omega''} < \infty$$

and $P^{\epsilon}_{\omega', \omega''}$ denotes the (one step) transition probability from ω' to ω'' according to the perturbed process defined by the Imitation Dynamics with a probability of error equal to ϵ. The resistance simply measures the total number of mistakes (or experiments) involved in the transition from state ω' to ω''.
An ω*-tree* H for any vertex ω of O is a tree spanning O so that for every $\omega' \neq \omega$, $\omega' \in \Omega$ there exists a unique directed path from ω' to ω. Let $r(H) = \sum_{(\omega', \omega'') \in H} r(\omega', \omega'')$ denote the resistance of the ω-tree H, and $\mathcal{H}_\omega$ be the set of all the ω-trees in O.
The *stochastic potential* of a state $\omega \in \Omega$ is:

$$\gamma(\omega) = \min_{H \in \mathcal{H}_\omega} r(H)$$

Let $\gamma^* = \min_{\omega \in \Omega} \gamma(\omega)$.

Now we can provide the proof for theorem 2, which follows directly from the following three lemmata:

Lemma 2.1 Let P^ϵ denote the Markov chain defined by the perturbed process, and μ^ϵ be its unique stationary distribution. Then $\lim_{\epsilon\to 0}\mu^\epsilon = \mu^0$ *exists* and μ^0 is a stationary distribution of P^0 – the Markov chain defined by the unperturbed process. Moreover, the probability μ^0_ω associated to the state ω by the limit stationary distribution μ^0 is strictly positive if and only if $\gamma(\omega) = \gamma^*$.

Lemma 2.2 The stochastic potential $\gamma(\omega^W)$ equals the cardinality of Γ minus 1.

Lemma 2.3 For all $q \neq q^W$, the monomorphic state ω_q has a stochastic potential

$$\gamma(\omega_q) \geq |\Gamma|$$

Proof of Lemma 2.1:.We shall follow the demonstration provided by Peyton Young [21] (Appendix), that we report here for sake of completeness.
First we can apply to P^ϵ a result established by Freidlin and Wentzel [9](Chapter 6, Lemma3.1) for every aperiodic, irreducible stationary Markov processes the unique stationary distribution is given by the formula:

$$\mu^\epsilon_\omega = p^\epsilon_\omega / \sum_{\omega'\in\Omega} p^\epsilon{}_{\omega'}$$

where

$$p^\epsilon_\omega = \sum_{H\in\mathcal{H}_\omega} \prod_{(\omega',\omega'')\in H} P^\epsilon_{\omega',\omega''}$$

Choose the ω-tree H with minimum resistance and consider the identity:

$$\epsilon^{-\gamma^*} \prod_{(\omega',\omega'')\in H} P^\epsilon_{\omega',\omega''} = \epsilon^{r(H)-\gamma^*} \prod_{(\omega',\omega'')\in H} \epsilon^{-r(\omega',\omega'')} P^\epsilon_{\omega',\omega''} \tag{1}$$

By the definition of r,

$$\lim_{\epsilon\to 0} \epsilon^{-r(\omega',\omega'')} P^\epsilon_{\omega',\omega''} > 0, \forall (\omega', \omega'') \in H \tag{2}$$

If $r(H) = \gamma(\omega) > \gamma^*$ it follows from (1) and (2) that

$$\lim_{\epsilon\to 0} \epsilon^{-\gamma^*} \prod_{(\omega',\omega'')\in H} P^\epsilon_{\omega',\omega''} = 0$$

therefore

$$\lim_{\epsilon\to 0} \epsilon^{-\gamma^*} p^\epsilon_\omega = 0.$$

Similarly, if $r(H) = \gamma(\omega) = \gamma^*$ we obtain

$$\lim_{\epsilon\to 0} \epsilon^{-\gamma^*} p^\epsilon_\omega > 0.$$

Since

$$\mu^\epsilon_\omega = \epsilon^{-\gamma^*} p^\epsilon_\omega / \sum_{\omega'\in\Omega} \epsilon^{-\gamma^*} p^\epsilon{}_{\omega'}$$

it follows that

$$\lim_{\epsilon\to 0} \mu^\epsilon_\omega = \begin{cases} = 0 \text{ if } \gamma(\omega) > \gamma^* \\ > 0 \text{ if } \gamma(\omega) = \gamma^* \end{cases}$$

Finally, since $\lim_{\epsilon\to 0} P^\epsilon_{\omega',\omega''} = P^0_{\omega',\omega''}\ \forall \omega', \omega'' \in \Omega$ and μ^ϵ satisfies the equation $\mu^\epsilon P^\epsilon = \mu^\epsilon\ \forall \epsilon > 0$, then $\mu^0 P^0 = \mu^0$. μ^0 is therefore a stationary distribution of P^0, hence it puts probability 1 over one of the monomorphic states. This concludes the proof for Lemma 1.
We shall now show that the monomorphic state having positive probability under the stationary distribution is precisely ω^W.

Proof of Lemma 2.2: here we shall apply the proof provided by Vega Redondo [26] to the specific context we are analyzing, to show that:

$$\gamma(\omega^W) = |\Gamma| - 1 = 30.$$

Note that

$$\forall \omega \in \Omega,\ \exists \omega_q \text{ s.t. } P^0_{\omega,\omega_q} > 0 \text{ and therefore } r(\omega, \omega_q) = 0.$$

Consider any monomorphic state ω_q, $q \in \Gamma$, and a state $\tilde{\omega}_q = (q_1, q_2, q_3, q_4)$ such that $\exists i \in I : q_i = q^W$ and $q_j = q \forall j \neq i, j \in I$. By the definition of r it is easy to check that $r(\omega_q, \tilde{\omega}_q) = 1$ since $P^{\epsilon}_{\omega_q,\tilde{\omega}_q} > 0$ if $\epsilon > 0$, according to the previously stated definition of the perturbed process.
To conclude the proof, it is enough to verify that, for any $q \neq q^W$, the resistance $r(\tilde{\omega}_q, \omega^W)$ is equal to zero. Indeed, it is straightforward to check that $P^0_{\tilde{\omega}_q,\omega^W} > 0$ under our hypotheses, because the firm producing q^W gets always the highest profit, regardless of the output produced by the others:

$$\pi(q^W, 3q) > \pi(q, 2q + q^W)\ \forall q \in \Gamma,\ q \neq q^W.$$

Proof of Lemma 2.3: to prove that $\gamma(\omega) > |\Gamma|\ \ \forall \omega_q \neq \omega^W$ it suffices to show that at least two mistakes are necessary to escape from the basin of attraction of ω^W, namely: there is no

$$\tilde{\omega}^W = (q_1, q_2, q_3, q_4) : \exists i : q_i \neq q^W,\ q_j = q^W\ \forall j \neq i,\ j \in I$$

such that $r(\tilde{\omega}^W, \omega_q) = 0$, since $P^0_{\tilde{\omega}^W,\omega_q} = 0$ because the profit of the firm producing $q \neq q^W$ is always lower than the profit earned by each of the other three firms, producing q^W:

$$\pi(q, 3q^W) < \pi(q^W, 2q^W + q)\ \forall q \in \Gamma,\ q \neq q^W.$$

As for the best response dynamics, I ran a simulation to depict the behavior of this model in the short run, with different levels of experimentation. Again, I ran 10000 cycles of simulation per each value of ϵ, taking values 0.01, 0.05, 0.1. The frequency distributions of individually chosen quantities in period 40 are reported in figure 2.
Simulation results, show that with any of the considered levels of experimentation, the process converges pretty quickly to the predicted outcome, but convergence is faster as experimentation becomes more probable.

4.3 Trial and Error

This model of learning has been firstly proposed by Huck, Normann and Oechssler in 2000 [14] then revised by the same authors in a subsequent article [15] where they present a continuous time version of it.
Both versions of the learning model are, in principle, very simple. Assuming that the strategy set of the player is ordered, the model predicts that every time a player changes the strategy he adopts, he will check whether his payoff has consequently increased or decreased. If he observes a raise, in the following period he will keep on changing his strategy in the same direction as before. On the contrary, if the payoff declines the player will change his strategy in the reverse direction. This is the model with the most lax hypotheses about information: it just requires that the firms know their own past actions and their own profits.
Huck *et al.* show that Trial and Error learning yields a collusive outcome. They proove this result analytically for the continuous time version, and for the discrete version with only two firms, and they extend it to the discrete case with more than two firms by means of simulations.

Since in the case we are analyzing both time and the strategy set are discrete, we will consider the discrete version of the model, applying it to the oligopoly setting described above.

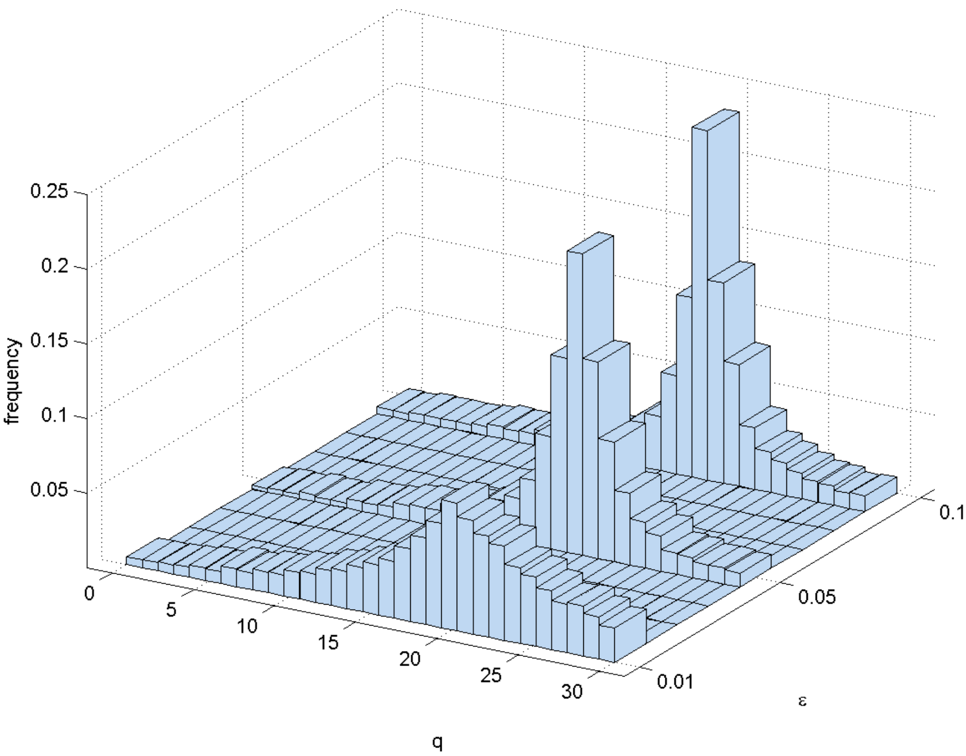

Figure 2: *Frequency distribution of individual choices in the 40th round; 10000 cycles of simulation.*

Given the quantity $q_1^i \in \Gamma$ chosen by any firm in period one, in every following period $t > 1$ each firm will set how much to produce according to the following rule:

$$q_t^i = \begin{cases} 0 & \text{if } q_{t-1}^i + s_{t-1}^i < 0 \\ 30 & \text{if } q_{t-1}^i + s_{t-1}^i > 30 \\ q_{t-1}^i + s_{t-1}^i & \text{otherwise.} \end{cases}$$

where the direction of change is given by

$$s_t^i = \text{sign}(q_t^i - q_{t-1}^i)\text{sign}(\pi_t^i - \pi_{t-1}^i)$$

if $(q_t^i - q_{t-1}^i)(\pi_t^i - \pi_{t-1}^i) \neq 0$; otherwise s is randomly chosen among the values $-1, 0, 1$, each having positive probability.
This defines a Markov chain over the state space $\Omega = \Gamma^4 \times \{-1, 0, 1\}^4$. As for the previous model, we assume the possibility of experimentation or mistakes, thus defining a perturbed process, in which with some small probability $\epsilon > 0$ each firm chooses an arbitrary direction of change s_i^t. This defines a Markov process which is irreducible and aperiodic, therefore has a unique stationary stable distribution. By contrast, in principle the unperturbed process may have many stationary distributions.

In what follows, I will show that *(i)* the unperturbed process has several absorbing sets, *(ii)* all the states belonging to these absorbing sets have the same stochastic potential, therefore *(iii)* they are all stochastically stable, meaning that all the absorbing sets of the unperturbed process belong to the support of the limit distribution $\mu^0 = \lim_{\epsilon \to 0} \mu^\epsilon$.

Recall that under the maintained assumptions, the cardinality of the state space is $31^4 \times 3^4$. If we disregard the order of the players, then the number of possible states reduces to $\binom{96}{4}$. Among these states, we want to individuate those, if any, which are recurrent. This was done by means of simulations and numerical analysis. First, the evolution of the unperturbed process was simulated over

first recurrent set	11 ↓ 11 ↓ 11 ↓ 11 ↓ 10 ↓ 10 ↓ 10 ↓ 10 ↓ 9 ↑ 9 ↑ 9 ↑ 9 ↑ 10 ↑ 10 ↑ 10 ↑ 10 ↑	average quantity: 10
second recurrent set	11 ↓ 11 ↓ 11 ↓ 10 ↑ 10 ↑ 10 ↑ 10 ↑ 11 ↑ 11 ↓ 11 ↓ 11 ↓ 12 ↓ 10 ↓ 10 ↓ 10 ↓ 11 ↓ 9 ↑ 9 ↑ 9 ↑ 10 ↓ 10 ↑ 10 ↑ 10 ↑ 9 ↑	average quantity: 10.25
third recurrent set	11 ↓ 11 ↓ 10 ↑ 10 ↑ 10 ↑ 10 ↑ 11 ↑ 11 ↑ 11 ↓ 11 ↓ 12 ↓ 12 ↓ 10 ↓ 10 ↓ 11 ↓ 11 ↓ 9 ↑ 9 ↑ 10 ↓ 10 ↓ 10 ↑ 10 ↑ 9 ↑ 9 ↑	average quantity: 10.77
fourth recurrent set	11 ↓ 10 ↑ 10 ↑ 10 ↑ 10 ↑ 11 ↑ 11 ↑ 11 ↑ 11 ↓ 12 ↓ 12 ↓ 12 ↓ 10 ↓ 11 ↓ 11 ↓ 11 ↓ 9 ↑ 10 ↓ 10 ↓ 10 ↓ 10 ↑ 9 ↑ 9 ↑ 9 ↑	average quantity: 10.47

Note: the numbers indicate the quantity produced by each of the four firms, the arrows the direction of change.

Table 2: *Recurrent sets of the unperturbed process*

200 iterations of the stage game, replicating this cycle for 10000 times. Regardless of the initial state of each cycle, which was randomly chosen from a uniform distribution over the whole state space, it emerged that during the last ten iterations the process rests over only 22 of the possible states, and that these 22 states belong to four recurrent sets, as displayed in table 2. Through a following numerical analysis I checked that all the other states are transient under the unperturbed process, meaning that the t-steps transition probability from each of these states to at least one of the states belonging to the recurrent sets is positive, for a finite t [2].

It is easy to verify that one and only one mistake is sufficient to transit from a recurrent set to the following one. It follows that all the states in the four recurrent sets have the same stochastic potential, which equals the number of recurrent sets minus one. As a consequence following directly from the aforementioned Lemma 2.1, all the states in the recurrent sets are stochastically stable.

A quick look at the quantities produced by the four firms in all of the recurrent states confirms the result previously stated by Huck *et al.*, namely that the Trial and Error process converges to a neighborhood of the joint profit maximizing outcome, which in this case is obtained when all the firms produce a quantity equal to 10.

I conclude this chapter by presenting the results of the simulation I did to study the short run behavior of the trial and error model, with different probabilities of mistakes. As in the previous two models, I ran 10000 cycles of simulation per each level of the error probability ϵ, taking values 0.01, 0.05, 0.1. Alike the imitate the best process, also trial and error turns out to converge to states which are close to the stochastically stable one within the number of repetitions that will take place in my experiments, regardless of the probability of error. In particular, we observe that as the probability of error decreases, the variance of the distribution decreases too and convergence is more precise.

[2] The analysis was performed with Matlab ©. The code is available from the author upon request.

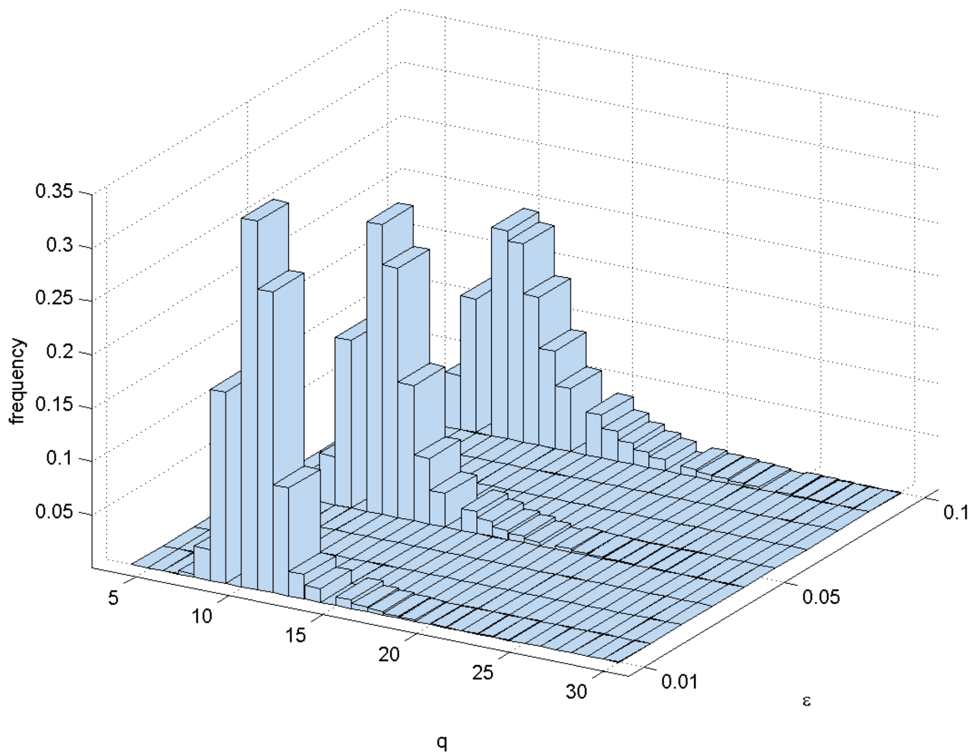

Figure 3: *Frequency distribution of individual choices in the 40th round; 10000 cycles of simulation.*

5 Results

The experiment was run on November 29 and 30, 2007, in the computer lab of the faculty of Economics, at the University of Bologna, in Italy. It involved 48 undergraduate students in Business Administration, Law and Economics, Commercial Studies, Economics, Marketing and Finance. Three identical sessions were organized, with 16 participants each. The length of the sessions ranged from one hour to one hour and fifteen minutes, including instructions and payment. The average payment was 13 €with a maximum of 17 and a minimum of 9, including a show-up fee of 4 €.

At the beginning of each session, subjects were welcomed into the computer room and sat in front of personal computers, and they were instructed not to communicate in any way with other players during the whole experiment. They received a printed copy of the instructions [3], which were read aloud so to make them common knowledge. Thereafter, they had the opportunity to ask questions, which were answered privately. Before starting the real game, subjects were also asked to complete a test on their computer, aimed at checking their understanding of the graphical interface they would have had to use during the game.

Only when all the players managed to answer correctly to all the questions in the test, the real game began. Each subject was randomly and anonymously matched with other three participants, who were to be his "opponents" throughout the whole game. At the and of the game, subjects were paid in cash, privately, in proportion to the profits they scored during the game.

The experiment was programmed and conducted with the software z-Tree (Fischbacher 2007 [8]).

In what follows I will first present some qualitative results about the output choices made by the subjects, and about their information search pattern. I will then try to establish a relation between the information acquired and the choices made by the subjects by means of two econometric models. Finally, I will briefly comment on the effects of education on the learning model adopted by the

[3] A translation of the instructions can be found in Appendix A

subjects.

5.1 Quantities

Figure 4 displays the frequency distribution of the individual output choices in all the periods and sessions of the experiment. First, we observe that the average (17.76)[4] is higher than the Cournot

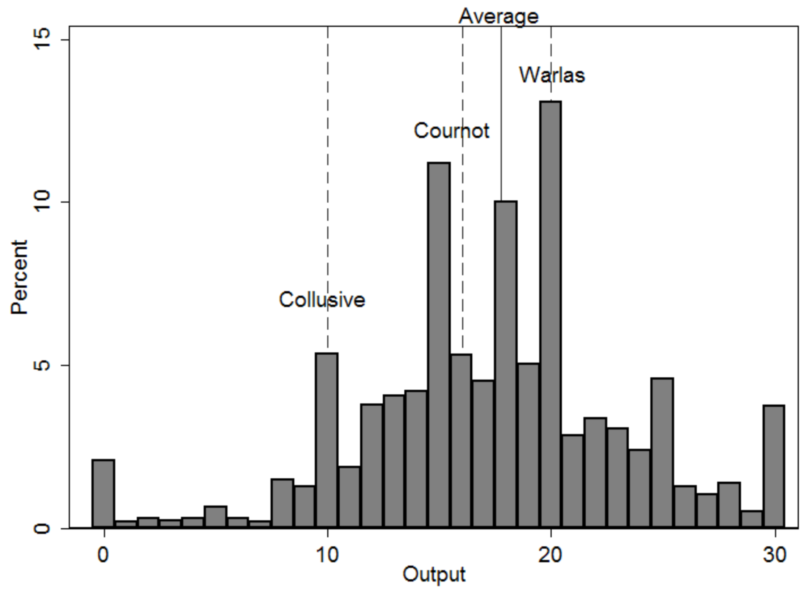

Figure 4: *Frequency Distributions of Individual Output Levels.*

output (16), but lower than the Walrasian one (20), which is instead the modal output. We also observe peaks corresponding to multiples of five, revealing a tendency to simplify the game focusing only on some of the available strategies, which can probably explain why 15 is chosen more often than 16, representing the Nash equilibrium in the stage game. The Pareto-dominant collusive outcome of 10 is chosen only in 5.36% of the cases.

Looking at table 3, we observe an increase in the average output as the game proceeds. For all the

Table 3: *Average individual output choice*

	Session 1	Session 2	Session 3
Periods 1-10	17.238	17.013	16.968
Periods 11-20	17.101	17.385	18.038
Periods 21-30	17.585	17.623	19.815
Periods 31-40	17.478	17.912	18.911
Total	17.353	17.487	18.436

three sessions, though, a non parametric Wilcoxon rank-sum test fails to reject (at the 1% significance level) the hypothesis that observations for the first and the last ten periods are drown from the same distribution. We also notice that the average quantity produced in the third session is significantly higher than in the other two[5].

[4] In this figure and in the following ones, the average is evaluated dropping the 40 observations in which the outcome was zero because a subject did not answered in time.

[5] According to a Wilcoxon rank-sum test, at 1% significance level.

Figure 5 presents the aggregate quantity produced in each group across all the periods of a game. We notice that the variability in total outcome remains high even towards the end of the game, with fluctuations between the Cournot and Walrasian equilibrium outcomes. In general, therefore, we cannot speak of convergence.

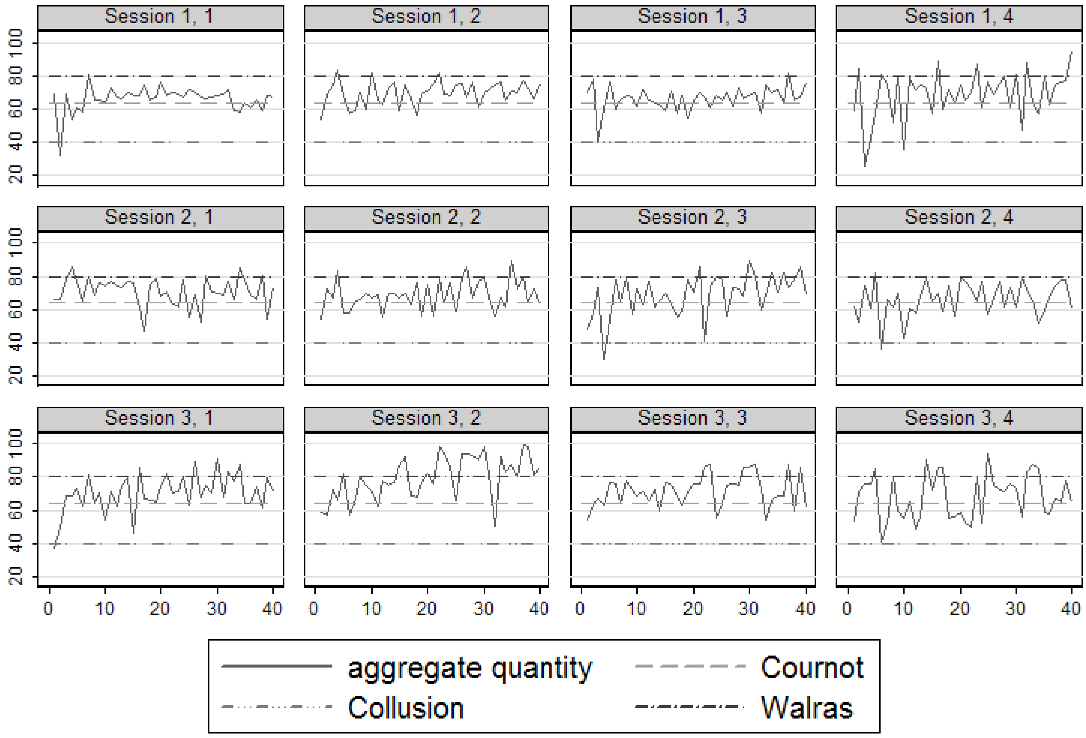

Figure 5: *Aggregate quantity produced, by session and group.*

5.2 Attention

Figure 6 shows that most of players' attention is devoted to the profit calculator. Indeed, the share of look-up time spent using the profit calculator is on average significantly higher than the time spent looking at any other piece of information, according to a Wilcoxon matched-pairs signed-ranks test (1% significance level). In line with what observed for the first experiment, in all the sessions players paid limited attention to their own past, so the trial and error learning model – which only requires this sort of information to be applied – finds weak support in these data.
On the other hand, if we observe the results for the three sessions, we notice that some differences emerge. In particular, in the third session the time spent using the profit calculator is significantly less that in the other two sessions (according to a Wilcoxon rank-sum test, at 1% significance level), while more attention is paid to the outcome individually chosen by each of the player's opponents in the previous period. We have already observed above that the average outcome in the third session was significantly higher than in the other two sessions. So, these data provide some support to Vega-Redondo's idea that information about the strategies chosen by the opponents yields a more aggressive competition between players. In the following, we will see that this impression is supperted also by a deeper econometric analysis.

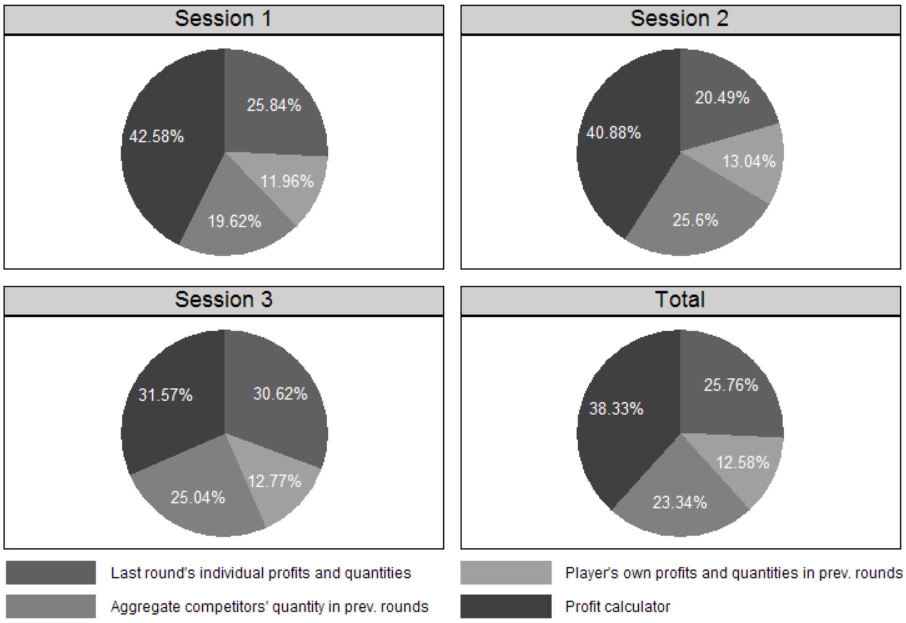

Figure 6: *Distribution of players' attention in the three sessions.*

Another noticeable fact emerges from figure 7: the fraction of look-up time dedicated to the profit calculator decreases along the game, and on average is significantly higher in the first than in the last 20 periods (see table 4), while the opposite is true for the time spent looking at the output

Table 4: *Fraction of look-up time dedicated to the profit calculator and to competitors' output choices in the first and in the second half of the game.*

		Session 1	Session 2	Session 3	Total
profit calculator	Periods 1-20	0.384	0.414	0.319	0.371
		$\sim$	$>^{**}$	$>^{***}$	$>^{***}$
	Periods 21-40	0.370	0.309	0.261	0.315
competitors' output choice	Periods 1-20	0.276	0.172	0.271	0.241
		$\sim$	$<^{***}$	$<^{***}$	$<^{***}$
	Periods 21-40	0.293	0.295	0.338	0.308

Note: the symbols ***, ** and * indicate significance at the 1%, 5% and 10% level, respectively.

individually chosen by the player's competitors in the previous period. Again, this shift in players attention together with the previously observed increase in the average output level seems to be in line with Vega-Redondo's model, even if the evidence is weak due to the lack of significance of the increase in quantities.

5.2.1 Use of the Profit Calculator

We have already noted that players made wide use of the profit calculator in this game. In particular, this device has been used in almost half of the observations collected through the three sessions, mostly to evaluate the myopic best reply to some aggregate quantity hypothetically produced by the player's opponents (see table 5).

Suppose a subject followed the aforementioned Best Response Dynamics rule and best replied to the aggregate output chosen by his opponents in the previous period: before using the profit calculator he should have gathered information about his competitors aggregate output in previous periods, by opening the appropriate box. When this happened, I claim that the look-up sequence is consistent

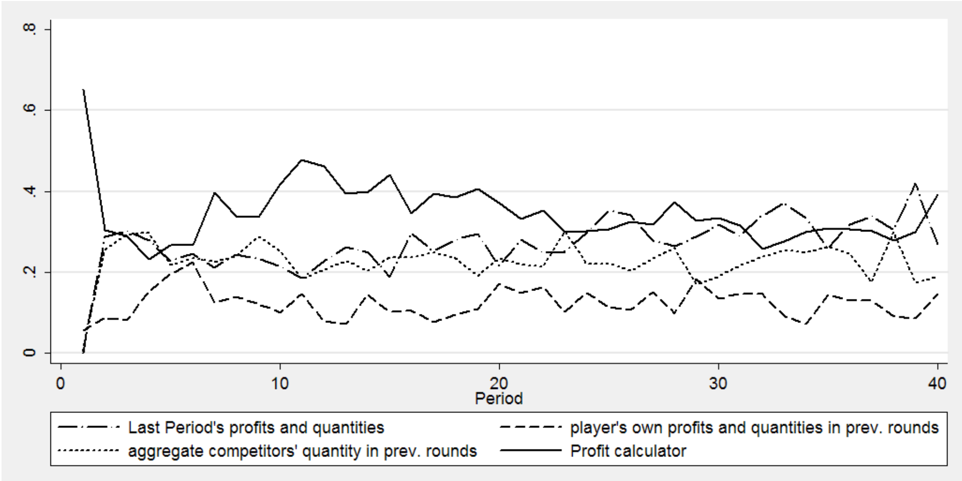

Figure 7: *Allocation of players' attention along the game.*

Table 5: *Use of the two functions of the profit calculator, and percentage of observations in which the look-up sequence is consistent with a myopic best reply.*

Use of the p.c.	N. obs (%)	% of L.U. sequences consistent with BRD
both functions	130 (13.56%)	91.54%
1st function (best reply) only	584 (60.90%)	82.86%
2nd function only	245 (25.55%)	79.62%
total	959	82.06%
Profit calculator not used	961	–

with best response dynamics (BRD), and it turns out that this is the case in more than 80% of the times the profit calculator was used.

When using the profit calculator, subjects had to enter a number corresponding to the hypothetical aggregate quantity produced by their opponents. This quantity can be seen as a proxy for their expectations about their competitors' future strategies. Figure 8 presents the frequency distribution of the difference between this quantity and the aggregate quantity actually chosen by competitors in the previous period.

According to these data, more than half of the times the profit calculator was used the quantity inputed belonged to the interval $[Q_{-1}(t-1)-3, Q_{-1}(t-1)+3]$, where $Q_{-1}(t-1)$ represents the sum of the quantities produced by the player's opponents in the previous period. This provides further support to the best response dynamics as a model of learning in this setting.

5.2.2 Interest for the strategies adopted by opponents.

As we have seen above, a considerable amount of attention is dedicated to the boxes showing the output individually chosen by each of the player's opponents in the previous period. Table 6 shows

Table 6: *Average look-up time*

	Best		Not Best
Opponent 1	2.07	$>^{***}$	1.53
Opponent 2	1.44	$>^{***}$	0.78
Opponent 3	2.45	$>^{***}$	1.69

Note: The statistical test is a two sample Wilcoxon rank-sum test. The symbol *** indicates significance at the 1% level

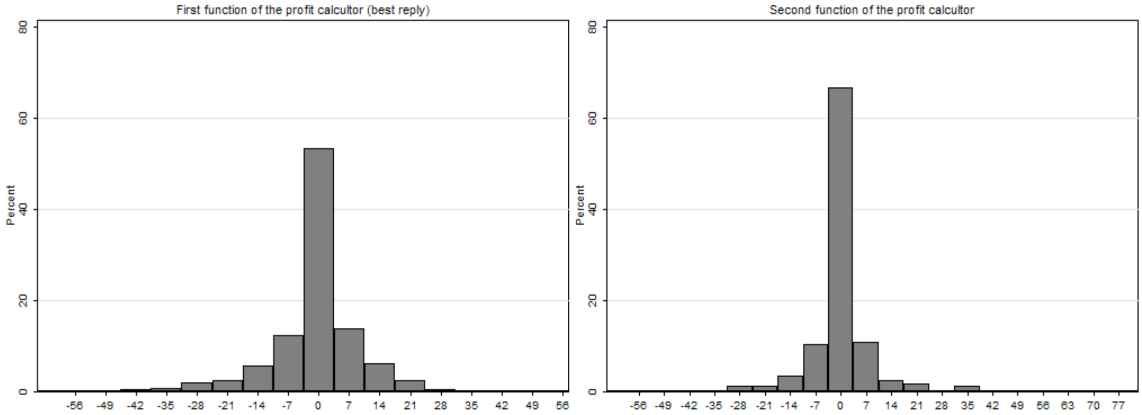

Figure 8: *Frequency distribution of the distance between the opponents' quantity entered in the profit calculator and the one observed in the previous period.*

that on average the look-up time dedicated to each of the opponents was greater when he had gotten the highest profit in the previous period. Still, if we compare the time spent looking at the strategy adopted by "the best" with the total time dedicated to the strategies adopted by the other opponents, we notice that the latter is significantly higher (see table 7).

So, in partial contradiction with what suggested by Vega-Redondo's theory of imitation, players in this experiment seem to be concerned not only with the choice made by the competitor who best performed in the previous period, but also with the output chosen by each of the others.

Table 7: *Average look-up time*

Session	**Best opponent**		**Other opponents**
Session 1	1.700	$<^{***}$	3.088
Session 2	1.317	$<^{***}$	2.146
Session 3	2.309	$<^{***}$	3.211
Total	1.775	$<^{***}$	2.815

Note: The statistical test is a Wilcoxon signed-rank test. The symbol *** indicates significance at the 1% level

5.3 Learning

I start the analysis of the learning mechanisms adopted by the players by comparing the explanatory power of the three aforementioned simple learning models: trial and error, imitation of the best and best response dynamics.

In a first attempt to have a picture of the learning model adopted by the players in this game, I adopted a measure proposed by HNO to assess to which extent the three simple learning rules presented in section 4 are able to predict each single choice the subjects made.

Let

$$z_i^t = \frac{q_i^t - q_i^{t-1}}{a_i^t - q_i^{t-1}}$$

where a_i^t is the quantity predicted, in turn, by Imitate the Best (IB), Trial and Error (TE) and Best Response Dynamics (BR).

According to the figures presented in table 8, best response dynamics is the rule that provides the most precise forecast ($0.5 <= z_{BR}^t < 1.5$) and predicts the right direction of change ($z_{BR} > 0$) with the highest frequency. It is also worth noting that “imitate the best” in general overshots, when it predicts the right direction of change, meaning that it forecasts a variation in quantity that is more than twice as big as the actual one ($0 <= z_{IB} < 0.5$); the reverse is true for “trial and error” model. Table 9 shows how many subjects report positive z_i^t values in at least 70% of the rounds and how many

Table 8: *Hit ratios (%)*

	$z < 0$	$0 <= z < 0.5$	$0.5 <= z < 1.5$	$z >= 1.5$
IB	45.45	26.70	21.38	6.47
BR	29.61	27.74	28.29	14.36
TE	47.09	13.21	6.14	33.55

present hits close enough to 1 ($0.5 \leq z < 1.5$) at least 30% of their decisions. None of the subjects seems to adopt the “trial and error” rule, while nine players behave according to what predicted by “imitate the best” in at least 12 periods. Yet, “best response dynamics” is the model that seems to inform the behavior of most of the players. Being interested also in how information affects the

Table 9: *Hit ratios at the individual level*

	$0.5 <= z < 1.5$	$z > 0$	**total**
IB	9	3	48
BR	18	25	48
TE	0	0	48

learning model players adopt, I grouped observations by the type of information subjects spent most of their look up time on and measured the percentage of observations in which ($0.5 <= z_i^t < 1.5$), for each of the three basic learning rules (Table 10). We notice that in general “best response dynamics” prevails over the other two rules, but when players dedicated the greatest part of their attention to the output individually chosen by each of their competitors in the previous period, then is “imitate the best” rule that gets the highest score, and this effect is even more pronounced when the time spent looking at the strategy adopted by the best among the opponents is longer than the time dedicated to the others. This is one more element in favor of Vega-Redondo’s theory claiming that when subjects have information about their opponents strategies and payoffs, they tend to become more competitive since they are tempted to imitate the one who got the best result.

5.3.1 Ordered probit estimation

Estimation procedure I now consider a model in which the sign of a player’s output change, Δq, is a function of the direction, x, indicated by the target output levels according to each of the three learning rules. This way, I should be able to determine to what extent each behavioral rule affects the way players adjust their output in every round of the game.
Let:

$$\Delta q_i^t = sign(q_i^t - q_i^{t-1})$$

and, for every learning rule r, let

$$x_{r,i}^t = sign(a_{r,i}^t - q_i^{t-1})$$

where $a_{r,i}^t$ denotes the quantity predicted for player i at round t by rule r, as above. In this experiment, I also control for the information gathered by players in every period. As mentioned above, each of the three basic learning models considered here requires a different type of information: more precisely, to

Table 10: $0.5 <= z < 1.5$ *(%)*

	Longest L.U. Time	IB	BR	TE	N. obs.
opponents' individual output	mostly to the best	36.73	19.05	4.08	147
	mostly to the others	23.79	23.45	5.52	290
profit calculator	L.U. seq. consistent with BR	17.77	37.50	8.43	664
	L.U. seq. not consistent with BR	20.37	31.48	1.85	108
	opponents' aggregate past output	22.66	23.56	4.83	331
	player's past output and profits	20.32	23.53	6.95	187
	no information acquired	14.43	15.46	3.09	97
	Total	21.38	28.29	6.14	1824

imitate the best one must have looked at the output produced by the best competitor in the previous period, to play according to trial and error a player needs to remember the strategy he has chosen and the profits he has obtained in the last two periods, meaning that he probably would have to open the box containing information about his own history of play, while to best reply he must know the aggregate output of his competitors in the previous period and he must use the profit calculator. For this reason, I created three dummy variables $d^t_{b,i}$, which indicate respectively if player i in period t had the information *(b)* necessary to best reply (InfoBR), to imitate the best (InfoIB) or to follow the "trial and error" model (InfoTE).
These dummy variables enter the model per se, but are also interacted with the variables x denoting the sign of the output variation predicted by the three learning models considered here.

Since subjects repeatedly interact with the same three opponents throughout the whole game, a critical point in this analysis is how to control for repeated observations of the same individuals or the same group. Moreover, I also wanted to check for possible correlation between data collected within each of the three sessions. For this purpose, I adopted a multilevel model with a random effect at the subject level nested within a random effect at the group level, which in turn is nested within a random effect at the session level.

More specifically, I assume that the latent response variable, z, be a linear function of the independent variables plus a subject specific error term $\zeta_{i,g,s}$, a group specific error term $\eta_{g,s}$, a session specific error term θ_s and finally and i.i.d. error term $u_{t,i,g,s}$. Random intercepts are assumed to be independently normally distributed, with a variance that is estimated through the regression.

The full model is then:

$$z = \sum_r \beta_r x_r + \sum_b \gamma_b d_b + \sum_r \sum_b \delta_{b,r} d_b x_r + \zeta + \eta + \theta + u$$

where subscripts r and b both take values in $\{IB, TE, BR\}$. For simplicity I omitted the subscripts for individual, group, session and period.

The dependent variable is derived in the standard way for an ordered probit given the latent variable and cutoffs between categories. Maximum likelihood estimation is used to fit values for the cutoffs, β, γ and δ, and for the variances of the subject, group and sessions specific error terms. The model was estimated using GLLAMM [6], a software specifically designed to provide a maximum likelihood framework for models with unobserved components, such as multilevel models, certain latent variable models, panel data models, or models with common factors.

I first estimated the full model, then I progressively obtained a more compact model using Likelihood-Ratio tests with a significance level of 5%. First, I checked for significance of the vari-

[6] see Rabe-Hesketh and Skrondal, 2004 [24] and http://www.gllamm.org

ances at the session and group levels; the null hypothesis that these were not significant was not rejected, so I adopted a more compact model with a random effect only at the subject level. Then, in steps, I eliminated the dependent variables that turned out not to be significant. For sake of simplicity, I present only the last estimate of this reduced model in table 11.

Table 11: *Ordered Probit Model: Estimations*

	Coefficient	**Standard Error**
βs		
TE	-0.022	0.041
BR	0.134**	0.066
IB	0.484***	0.053
γs		
InfoIB	0.327***	0.080
δs		
InfoTExTE	0.148**	0.068
InfoTExBR	0.151**	0.067
InfoBRxBR	0.520***	0.067
InfoIBxBR	0.258**	0.075
cut1	-0.041	0.065
cut2	0.369***	0.066
N	1824	
logL	-1576.005	

Note: In this table and in the following ones, the symbols ***, ** and * indicate significance at the 1%, 5% and 10% level, respectively.

Results We notice that the rule based on trial and error does not find strong support in these data even if it seems to guide, at least in part, players' behavior when they acquire the information necessary to apply it. We also observe that when players looked at the strategies adopted in the previous period by their opponents, and at the relative profits, they tended to increase their quantity (the estimated coefficient for InfoIB is positive and highly significant).
Remarkably, if the impact of information is not taken into consideration the "imitate the best" model seems to account for the greatest part of output variations; the coefficient for "best response dynamics" is also positive and significant, but is smaller in magnitude. This relative weakness of the best response model disappears if we consider the effect of information: indeed, according to these statistics if a subject acquire any of the three pieces of information considered here he will be more incline to move in the direction predicted to best reply, and particularly so if he uses the profit calculator after having looked at the aggregate output produced by his competitors in previous periods (namely, when InfoBR= 1).
One possible reason why the coefficient for x_{IB} is higher than the one for x_{BR} is that subjects were always informed about the profits individually obtained in the previous period by each of the players in their group. It is possible, then, that any time they realized that their profit was not the highest they tended to increase their output, then moving in the direction predicted by "imitate the best", even if they did not know the exact output chosen by the player who had got the best profit.

The first impression is that the hypothesis that subjects follow some very simple heuristic to choose their strategy in our game should be rejected. Learning through trial and error does not seem to be a plausible explanation of subjects behavior, both because the players pay too little attention the their own past profits and quantities, which is the only information required to apply this learning rule, and because their choices are not in line with what is theoretically predicted according to this model. On the other hand, imitating the best performer – per se – is not always able to forecast the observed choices correctly, even if subjects' look up patterns are consistent with this learning model. Myopic best reply seem to drive players' choices, at least partially, and is supported by the information they acquire, on average. In fact, to apply this learning rule the subjects need to know the sum of the quantities produced by their competitors in the last period – an information they almost always look at – and they must be able to compute a best reply, which means that either they use the profit calculator or they have used it extensively in the past and already know what the best reply is. Still, this model does not fully explain the observed variations in players' behavior.

According to the ordered probit regression, players' behavior is rather driven by the interplay of best reply and imitation, which in a sense confirms Vega-Redondo's idea: even if subject are incline to adopt the best reply when they know the market structure sufficiently well, if they are provided with information about their rivals' strategies and profits, they will be tempted to imitate those who are more successful, which yields more competitive outcomes.

Still, the model I estimated is essentially based on three extremely simple learning rules, and could therefore be too rigid to encompass all the facets of players learning behavior. For this reason, I decided to estimate a second, more complicated learning model, based on the self-tuning experience weighted attraction learning model proposed by Ho, Camerer and Chong (2007) [11].

5.3.2 EWA Learning Model

The parametric version of the Experience-Weighted Attraction (EWA) model was first proposed by Camerer and Ho [4] and Ho *et al.* [12]. It is a model that hybridizes features of other well known learning rules, such reinforcement learning and belief learning, and that thanks to its flexibility has proven to fit data better than other models. This model is based on the idea that every player assigns to each strategy a given level of attraction, which can be represented by a number. Attractions are updated after every period, according to the players' experiences, and determine every player's probability distribution over his or her choice set.

In the original model, attractions are updated using the payoff that a strategy either yielded, or would have yielded, in a period: the rule for updating attraction $A_i^j(t)$ attached by player i to strategy j in period t is

$$A_i^j(t) = \frac{\phi N(t-1)A_i^j(t-1)}{N(t)} + \frac{[\delta + (1-\delta)I(s_i^j, s_i(t))]\pi_i(s_i^j, s_{-i}(t))}{N(t)} \tag{3}$$

where s_i^j denotes strategy j of player i, $s_{-i}(t)$ the strategy vector played by player i's opponents in period t and $N(t)$ is a measure of the weight players put on past attractions relative to present ones; it can be interpreted as the number of "observation-equivalents" of past experience relative to one period of current experience.

$I(x, y)$ is an indicator function which takes value 1 if $x = y$ and value 0 otherwise. So, according to this model, it is assumed that players are able to evaluate the foregone payoffs they would have earned in period t had they chosen a different strategy s_i^j.

The parameter δ measures the relative weight given to hypothetical payoffs, compared to actual payoff $\pi_i(s_i(t), s_{-i}(t))$. The second parameter to be estimated is ϕ: a discount factor that depreciates previous attractions.

The variable $N(t)$ is also updated after every period according to the rule:

$$N(t) = \phi(1-\kappa)N(t-1) + 1 \quad t \geq 1 \tag{4}$$

where parameter κ determines the growth rate of attractions, which reflects how quickly players lock into a strategy: a third parameter that has to be estimated. When $\kappa = 0$ attractions are weighted averages of lagged attractions and past payoffs, so that attractions cannot grow outside the bounds of the payoffs in the game. When $\kappa = 1$ attractions cumulate, so they can be much larger than stage-game payoffs.

Attractions determine probabilities. More specifically: the probability $P_i^j(t+1)$ that player i chooses strategy j in period $t+1$ is monotonically increasing in $A_i^j(t)$ and decreasing in $A_i^k(t),\ k \neq j$. The relation between attractions and choice probabilities is represented by a logistic stochastic response function:

$$P_i^j(t+1) = \frac{e^{\lambda A_i^j(t)}}{\sum_k e^{\lambda A_i^k(t)}} \tag{5}$$

where the parameter λ measures sensitivity of players to attractions.

One of the main criticisms on parametric EWA concerns the number of parameters to be estimated. To solve this issue, Ho, Camerer and Chong (2007) [11] developed a simpler version on the model in which some parameters are fixed at plausible values, while others are replaced with functions of experience, that no longer need to be estimated. More specifically: parameter k is set equal to 0, because this capture almost all familiar learning models, and also because according the previous work by the same authors this parameter does not affect fit much. The initial experience $N(0)$ is restricted to be equal to 1, which is not a crucial assumption since in general subjects come to the experiment with weak priors, whose influence gradually disappears as the experiment proceeds. Finally, parameters δ and ϕ are replaced with functions of players' experience.
Parameter δ – representing weight of foregone payoffs – is substituted with the function:

$$\delta_i^j(t) = \begin{cases} 1 & \text{if } \pi_i(s_i^j, s_{-i}(t)) \geq \pi_i(s_i(t), s_{-i}(t)) \\ 0 & \text{otherwise.} \end{cases}$$

meaning that subjects reenforce, by a weight of one, only chosen strategies and all the other strategies that would have yielded a weakly higher payoff.
The discount factor ϕ is instead replaced by the "change detector" function $\phi_i(t)$ varying across time within the same game. The hypothesis made by the authors, here, is that the weight put on previous experiences should be lower when the player senses that the environment is unstable or that the strategies adopted by her opponents are changing. They then build a "surprise index" $S_i(t)$ measuring the difference between opponents' most recently chosen strategies and the strategies they adopted in all previous periods, and let $\phi_i(t) = 1 - \frac{1}{2}S_i(t)$. The surprise index is made up by two main elements: a cumulative history vector $h_i(t)$ and a recent history vector $r_i(t)$. The vector element $h_i^j(t) = \frac{\sum_{\tau=1}^t I(s_{-i}^j, s_{-i}(\tau))}{t}$ measures the frequency with which strategy s_{-i}^j was adopted by player i's opponents in period t and in all the previous ones. Vector $r_i(t)$ instead has all the elements equal to 0 but the k-th, where $s_{-i}^k = s_{-i}(t)$. The surprise index $S_i(t)$ simply sums up the squared deviations between the cumulative history vector $h_i(t)$ and the immediate history vector $r_i(t)$:

$$S_i(t) = \sum_j (h_i^j(t) - r_i^j(t))^2.$$

Self-tuning EWA model has two important advantages: first, it is particularly flexible since the functions $\delta_i^j(t)$ and $\phi_i(t)$ naturally vary across time, people, games and strategies; second, it can shift from a learning model to a different one as the game proceeds.

A modified version of self-tuning EWA model According to EWA learning model, attractions are updated keeping into account also foregone profits, but in the experiment I present here foregone payoffs from unused strategies are not known by the players. Subjects, though, can use the profit calculator to discover the profit a particular strategy would yield, given the strategies chosen by the other players. As explained in section 3.1, the profit calculator can be used in two different ways:

1. the profit calculator can be used by the players to evaluate the quantity that would yield them the highest profit given the aggregated quantity produced by their competitors, and inform them about the profit they would earn if they produced the suggested amount of good.

2. it can be also used to know the profit given both the quantity produced by the player and the sum of the quantities produced by his opponents.

By checking how a player used the profit calculator in each period, I know precisely which information he used to evaluate each strategy in every period.

If they wish, players can also access information about the profits earned in the previous period by their competitors. If they wanted to imitate the strategy chosen by the player who got the highest profit in the previous period – as suggested by Vega Redondo – they would attach a higher attraction to that strategy.

Keeping this peculiar characteristics of the game in mind, I decided to change the attraction updating rule, so that attractions in every period t are modified considering three elements:

- the profit $\pi_i(s_i^j, s_{-i}(t-1))$ actually obtained by the player in period $t-1$;
- the profits $\pi_{i,imit}^j(t-1)$ obtained by each of the player's opponents playing strategy s^j in the previous period;
- the profits $\pi_{i,PC1}^j(t)$ and $\pi_{i,PC2}^j(t)$ evaluated by the player using the first and the second function of the profit calculator respectively, given his or her expectations about the competitors' choices [7].

While the player always knew the strategy he played in the previous round and the profit he obtained, π_{imit}, π_{PC1} and π_{PC2} may be known or unknown to the player, depending on the pieces of information he or she decided to look up.
To check for the information the subject is aware of, I define four dummy variables:

$$d_{i,PC1}^j(t) = \begin{cases} 1 & \text{if in period } t \text{ player } i \text{ used the first function of the profit calculator, and this device indicates strategy } s^j \text{ as the best reply to the strategies played by the three opponents, and associates it to some profit } \pi_{i,PC1}^j(t) \\ 0 & \text{otherwise.} \end{cases}$$

$$d_{i,PC2}^j(t) = \begin{cases} 1 & \text{if in period } t \text{ player } i \text{ used the second function of the profit calculator, and this device associates strategy } s^j \text{ to some profit } \pi_{i,PC1}^j(t) \text{ given the opponents' strategies.} \\ 0 & \text{otherwise.} \end{cases}$$

$$d_{i,h}^j(t) = \begin{cases} 1 & \text{if player } i \text{ in period } t \text{ knew that his opponent } h \text{ had played strategy } s^j \text{ in the previous period} \\ 0 & \text{otherwise.} \end{cases}$$

$$b_{i,h}(t-1) = \begin{cases} 1 & \text{if player } h \text{ had the highest profit in period } t-1 \text{ among the opponents of player } i \\ 0 & \text{otherwise.} \end{cases}$$

These dummy variables, in a sense, replace the function $\delta_i^j(t)$ representing the weight of foregone profits in the original version of self-tuning EWA learning.
Now, it is possible to state our modified updating rule for attractions:

[7] If the second function of the profit calculator is used more than once by player i in period t, the profit $\pi_{i,PC2}^j(t)$ is calculated as an average of the various profits associated to strategy s_i^j by the device (different profits correspond to different hypotheses about the other players' behavior).

$$
\begin{aligned}
A_i^j(t) = &\frac{\phi_i(t)N(t-1)A_i^j(t-1) + \alpha\pi_i(s_i^j, s_{-i}(t-1))}{N(t)} + \\
&+ \frac{\beta d^j_{i,PC2}(t)\pi^j_{i,PC2}(t) + \gamma d^j_{i,PC1}(t)\pi^j_{i,PC1}(t)}{N(t)} + \\
&+ \frac{\epsilon\sum_{h\neq i} d^j_{i,h}(t)b_{i,h}(t)\pi^j_{i,imit}(t-1) + \zeta\sum_{h\neq i} d^j_{i,h}(t)(1-b_{i,h}(t))\pi^j_{i,imit}(t-1)}{N(t)}
\end{aligned}
\tag{6}
$$

In this formula, attractions in period t are updated according to a weighted average between (some of) the information the player has about the profit each strategy yielded in the previous period and the profit it may yield in the future.
Parameters α, β, γ ϵ and ζ then measure, respectively, the relative weight given to player's own experience, to the results potentially provided by the two functions of the profit calculator and to the profits earned by player's opponents, if observed.

Note that, in our model, $A_i^j(t)$ depends on the profits actually earned and (possibly) on the profits the opponents achieved *in the previous period*, and may be updated in period t with the profits evaluated by the profit calculator. So, the probabilities $P_i^j(t)$ depend on $A_i^j(t)$, and not on $A_i^j(t-1)$, as in the original model. The updating rule is then:

$$
P_i^j(t) = \frac{e^{\lambda A_i^j(t)}}{\sum_k e^{\lambda A_i^k(t)}} \tag{7}
$$

In order to test whether the difference between the estimated weight attributed to the profits achieved by the best competitor and the profits earned by other competitor is significant, I also estimated a restricted version of EWA model, in which equation (6) is replaced by equation (8)

$$
\begin{aligned}
A_i^j(t) = &\frac{\phi_i(t)N(t-1)A_i^j(t-1) + \alpha\pi_i(s_i^j, s_{-i}(t-1))}{N(t)} + \\
&+ \frac{\beta d^j_{i,PC2}(t)\pi^j_{i,PC2}(t) + \gamma d^j_{i,PC1}(t)\pi^j_{i,PC1}(t) + \delta\sum_{h\neq i} d^j_{i,h}(t)\pi^j_{i,imit}(t-1)}{N(t)}
\end{aligned}
\tag{8}
$$

.

Results Table 12 displays estimation results for the unrestricted and restricted version of the EWA model. Learning in this setting appears to be a blended process in which different components play

Table 12: *Results of the EWA learning models*

	unrestricted		restricted	
	b	se	b	se
α	0.476***	0.097	0.477***	0.148
β	0.845***	0.171	0.848***	0.263
γ	1.187***	0.241	1.189***	0.373
δ	–	–	0.356***	0.114
ϵ	0.338***	0.077	–	–
ζ	0.383***	0.089	–	–
ln(λ)	0.499**	0.197	0.497	0.308
Log-Likelihood	-5680.350		-5680.603	
Sample size	1920		1920	

an important role. The component related to belief learning seems to predominate: subjects attach

the highest weight to strategies hypothetically tested by means of the profit calculator. A Wald test fails to reject the hypothesis that β and γ are equal, so according to my results subjects tended to attribute less importance to the results from the first function of the profit calculator – evaluating best reply – than from the second one – which computes the profit given the output choice of the player and the aggregate output hypothetically produced by his opponents.
Both a Wald test on the results of the unrestricted model, and a likelihood ratio test between the restricted and the unrestricted model fail to reject the hypothesis that ϵ and ζ are equal. This means that subjects, when evaluating a strategy, do take into account the profits realized by other players choosing that strategy (the coefficient is always positive and highly significant), but do not attach more weight to the profit realized by the best among their competitors.
Finally, we notice that the estimates for α are positive, significant, and higher than those for parameters δ and ϵ and ζ, respectively for the restricted and unrestricted models. This suggest that reinforcement learning, based on player's own past experiences – plays a role, which seems to be even more important than the one played by "imitation".

5.4 Individual Characteristics

Finally, I would like to point out some interesting differences emerging among students having different educational backgrounds. Table 13 shows some facts about the composition of my subjects pool.

Table 13: *Subjects' education*

	master	bachelor
Business Administration	8	3
Law and Economics	1	
Commercial Studies		6
Economics	6	3
Finance	2	7
Marketing		12

First, it is worthwhile noting that master students pay relatively less attention to the strategies individually adopted by their opponents, and more to the profit calculator (see table 14). According to our previous results, this should make them less incline to imitate and more to best reply, which in theory should be viewed as a more "rational" behavior. In fact, master students turn out to be less aggressively competitive: the average quantity they choose is significantly lower, and so are their profits. They seem to adopt a follower behavior, best replying to opponents who tend to keep their own output high.

	bachelor		master
Average share of L.U time			
competitors' individual output	0.314	$>^{***}$	0.204
player's own history of play	0.119	$<^{***}$	0.133
competitors' past aggregate output	0.213	$<^{***}$	0.259
profit calculator	0.329	$<^{**}$	0.368
Average output	17.76	$>^{***}$	16.72
Average profit	176.82	$>^{***}$	158.14

Table 14: *Comparison between bachelor and master students*

Second, sizable differences emerge between students with different curricula of studies (table 15).

In particular, students in Finance, Marketing and Law and Economics show much less interest in the profit calculator, and much more in their competitors' output. This is somehow surprising, since all curricula except Commercial Studies and Law and Economics envisage an introductory course in microeconomics – including elements of the theory of oligopoly – for first year bachelor students, while only master students in Economics have a specific training in Industrial Organization and Game Theory. So, it is not clear whether this different attitude towards information and learning – which as we have seen has an impact on the level of market competition – derives from a different approach to economics characterizing different curricula or from individual attitudes that in turn have affected also subjects' choice for a certain course of studies.

Table 15: *Allocation of attention: a comparison between different curricula*

	Business Admin.	Comm. Studies	Economics	Finance	Marketing	Law and Econom.
Competitors' individual output	0.234	0.219	0.123	0.428	0.340	0.476
Player's own history of play	0.118	0.130	0.113	0.102	0.145	0.176
Competitors' past aggregate output	0.227	0.207	0.262	0.188	0.251	0.152
Profit calculator	0.383	0.409	0.476	0.258	0.244	0.136

6 Conclusion

In this paper I presented an experiment in which subjects were asked to play a repeated Cournot game with incomplete information. The first aim of the experiment was to check what feedback information subjects are really interested in, and to test how information is linked to the learning model adopted and in turn to the market outcome.

According to the data I collected, learning appears to be a composite process, in which different components coexist. The leading element seems to be a sort of belief learning, in which subjects form expectations about their opponents' future actions and try to best reply to them. It is also noticeable that in most of the cases the opponents' output inputed in the profit calculator – a proxy for players'expectations – is pretty close to the aggregate opponents' output observed in the previous period, meaning that either subjects expected their opponents not to change their strategy much or that they decided to use the profit calculator only when the opponents' strategy was stable enough to let them make predictions about the future.

Second, a considerable amount of look-up time is dedicated to the strategies individually adopted by competitors. As predicted by Vega-Redondo's theory, this piece of information generally boosts competition. Yet, my results suggest that players are not only interested in output produced by the most successful competitor, but by all of their opponents. These results are confirmed by the estimates obtained via my modified version of EWA learning model, suggesting that there is no difference between the weight attached to the profits collected by the most successful opponent and by the other competitors by subjects assessing the "strength" of a particular strategy. Anyhow, all tests I have done agree on that imitation is not the main driving force in the observed learning process.

Third, the "trial and error" learning model which was found to perform quite well in HNO does not find strong support in my data. Subjects are not interested in their own past history of play, which is the only piece of information required by this learning rule, and the model often fails to predict even the direction of change in players' output.
Fourth, the model I derived from Camerer and Ho's EWA learning stresses the importance played by reinforcement learning in this setting: when assessing the strength of a strategy subjects seem to take into greater consideration their own experience than what they know about other players' results.

Finally, from an analysis of players' individual characteristics it emerges that subjects's specific training in economics might affect their behavior both in terms of information search pattern and in terms of actual choices. This aspect deserves further investigation and suggests that it could be interesting to repeat the experiment with market professionals, in order to see whether their experience in the field affects their approach to the game.

With my experiment I meant to contribute to the understanding of learning mechanisms in game-like situations. I also wanted to test experimental devices based on the "Mouselab" technique as scientific instruments that might be usefully adopted in other experiments on learning and to investigate other interesting situations in which imperfect information of some of the agents plays a crucial role, or in which reputation is an asset. Examples might be auctions and financial markets, but also markets where hiding some attributes of the good being sold or the price of its add-ons may enable the sellers to get profits well above the competitive level.
In situations like those, a better comprehension of the relation between the data and stimuli provided to economic agents and their choices might help the regulator to set rules of information disclosure that bring the market outcome toward a more efficient equilibrium.

A Instructions

Welcome to this experiment about decision making in a market. The experiment is expected to last for about 1 hour and 15 minutes. You will be paid a minimum of 4€for your participation. On top of that you can earn up to 20 €if you make good decisions.

We will first read the instructions aloud. Then you will have time to read them on your own. If you have questions, raise your hand and you will be helped privately. From now on, you are requested not to communicate with other participants in any way.

Your task. During this experiment, you will be asked to act as the manager of a firm which produces and sells a given product: your task consists in deciding how many product units to put on the market in every period.

Your firm has **three competitors** that sell on the same market a product which is exactly identical to yours. Your competitors are three among the participants to the experiment taking place today in this room, but you will not have the opportunity to discover who they are, not even at the end of the game. Your identity will be kept secret as well.

The experiment consists in **40** consecutive **periods**. In every period, you will be asked to choose how many units to produce (between 0 and 30), and the same will be done by your competitors. Your choices affect both your firm's profits and the ones of your three competitors.

Every period lasts **30 seconds**: if in a period you fail to make your choice within the time limit, the computer will automatically set the number of units produced by your firm in that period equal to 0, and your profit in that period will be equal to 0 too.

Price, costs and profits. The market **price** at which you will be able to sell your product will be the higher, the smaller the total number of product units your firm and your competitors put on the market; if the total number of product units sold on the market is sufficiently high, the price will be equal to zero.

No product unit remains unsold: all the product units you put on the market will be purchased by consumers at the market price.

To produce, you will have to bear a **production cost** which will be the higher, the more product units you put on the market.

Your **profit** will be equal to the market price times the number of units you sell, minus production costs.

Earnings and Payment. You will receive an initial endowment of 2000 points. At the end of each period, your per-period profits or your possible losses will be added to your total profit, which will be always displayed in the top right corner of the screen. Notice that your total profit cannot become negative.

At the end of the game, your total profit will be converted in Euros, according to the rate:
1000 points = 1 Euro
The corresponding amount of money will be payed to you in cash, privately, at the end of the session. Remember that, in addition, you will be payed 4 €for your participation.

Information at your disposal. At the top of your computer screen you will read:

1. the number of periods elapsed since the game began (top left corner)

2. your total profit (top right corner)

3. the number of seconds (top, center) you still have at your disposal to take a decision. Remember that every period lasts 30 seconds, and if you do not take a decision in time it will be as if you decided to produce 0 units and in that period your profit will be equal to 0.

Before choosing how many units to produce, you will have the opportunity to look at some information on market characteristics and on what happened in the previous periods.

In particular, in every period following the first one, you will be informed about the profits obtained in the previous period by your firm and by your competitors. Moreover, you will be able to get more information about:

1. the quantity produced in the previous period by each of your competitors;
2. the quantities produced and the profits obtained by your firm in each of the previous periods: this information will be displayed both by means of a plot and in a table;
3. the quantity produced on the whole by each of your three competitors in the previous periods: this information will also be presented both by means of a plot and in a table.

In addition, you will have the opportunity to use a **profit calculator**, a device you can use to better understand how the market works. the profit calculator has two functions:

1. evaluate your profit, given the number of units produced by your firm and the number of units produced on the whole by your competitors.
2. evaluate the maximum profit you could earn – and the number of units your firm should produce in order to get such profit – given the number of units produced on the whole by your competitors.

Progress of the experiment. When the reading of these instructions is over, you will have the opportunity to ask for clarifications about the aspects of the experiments which are unclear.

When we have answered all the possible questions you will be asked to complete a test on your computer, which will allow us to check that you have fully understood the instructions, and you to get to grips with the software used in this experiment. The answers you give in this test will not affect your earnings in any way, nor they will influence any other aspect of the experiment. During the test, you will still have the possibility of asking questions, always raising your hand.

When all the participants have completed their test, the real experiment will begin. The computer will randomly generate groups of four persons; every participant to the experiment will belong to one and only one group during the whole experiment. The other three members of the group you belong to are your competitors, who then remain the same over all the 40 periods of the game.

Every period lasts at most 30 seconds. The maximum length of the game therefore is approximately 20 minutes.

At the end of the fortieth period the game will end, and the points scored by each of the participants will be converted into Euros.

Before being paid privately, you will be asked to answer a short questionnaire about the experiment, and you will have to hand back the instructions.

THANK YOU VERY MUCH FOR PARTICIPATING IN THIS EXPERIMENT AND GOOD LUCK!

B Graphical Interface

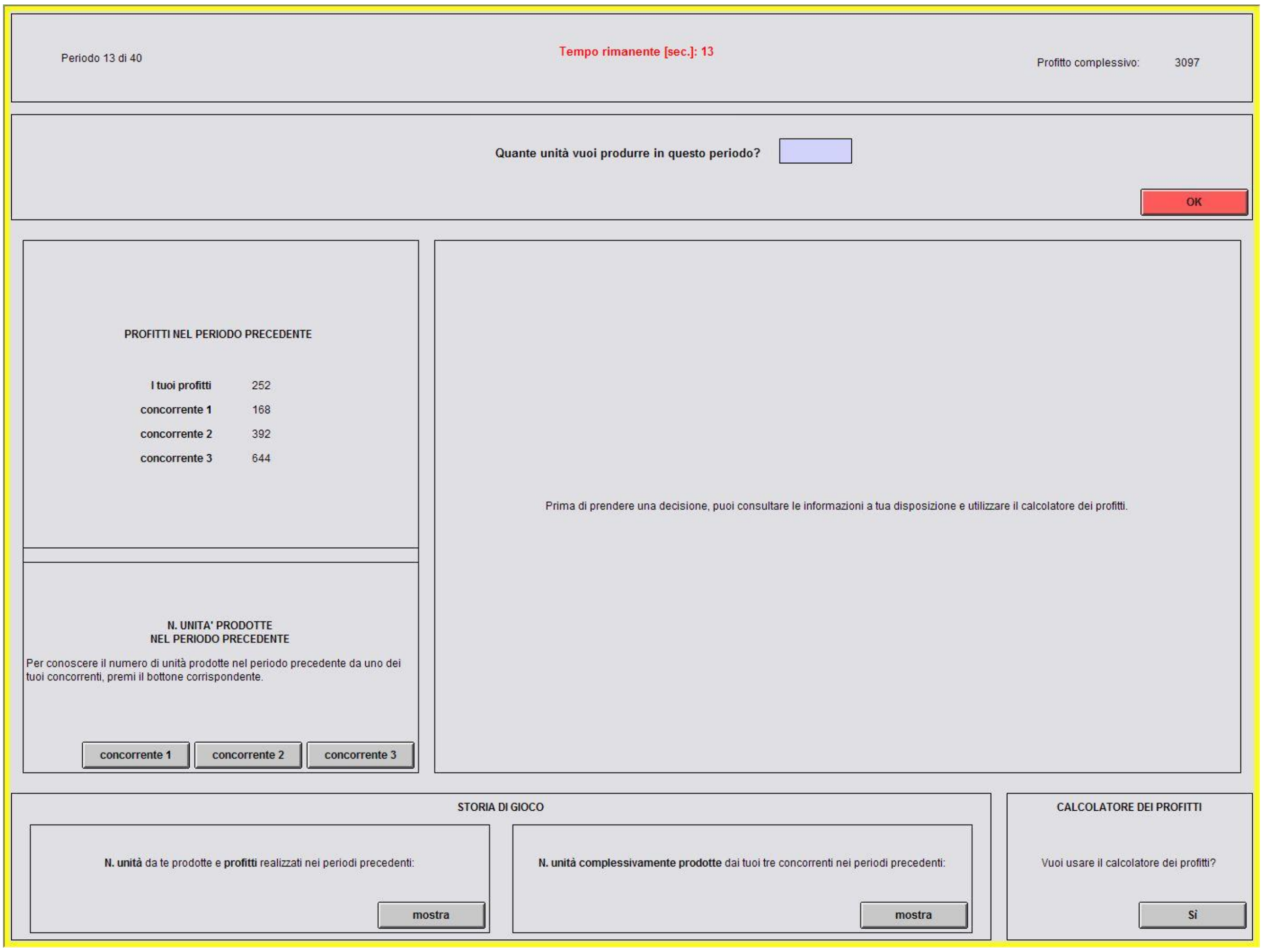

Figure 9: *Graphical interface*

Translation From top to button, left to right. [] indicate a button.

bar at the top: period 13 out of 40, remaining time [sec.]: 13, total profit: 3097

box at the top: how many units do you want to produce in this period? [OK]

first box on the left: Profits in the previous period
your profit: 252
competitor 1: 168
competitor 2: 392
competitor 3: 644

second box on the left: # of units produced in the previous period.
To know the number of units produced in the previous period by one of your competitors, push the corresponding button. [competitor 1] [competitor 2] [competitor 3]

center-right box: before taking a decision, you can look at the information at your disposal and use the profit calculator.

bottom-left box: history of play

# of units you produced and profits you obtained in the previous periods [show]

# of units produced on the whole by your three competitors in the previous periods [show]

bottom-right box: profit calculator
do you want to use the profit calculator? [yes]

References

[1] Apesteguia, J., Huck, S., and Oechssler, J. Imitation—theory and experimental evidence. *Journal of Economic Theory 136*, 1 (2007), 217–235.

[2] Apesteguia, J., Huck, S., Oechssler, J., and Weidenholzer, S. Imitation and the Evolution of Walrasian Behavior: Theoretically Fragile but Behaviorally Robust. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1095135, February 2008. CESifo Working Paper Series No. 2224.

[3] Bosch-Domènech, A., and Vriend, N. J. Imitation of successful behaviour in cournot markets. *Economic Journal 113*, 487 (2003), 495–524.

[4] Camerer, C., and Ho, T. H. Experience-Weighted Attraction Learning in Coordination Games: Probability Rules, Heterogeneity, and Time-Variation. *Journal of Mathematical Psychology 42*, 2-3 (June 1998), 305–326.

[5] Costa-Gomes, M., and Crawford, V. Cognition and Behavior in Two-Person Guessing Games: An Experimental Study. *The American Economic Review 96*, 5 (2006), 1737–1768.

[6] Costa-Gomes, M., Crawford, V. P., and Broseta, B. Cognition and Behavior in Normal-Form Games: An Experimental Study. *Econometrica 69*, 5 (September 2001), 1193–1235.

[7] Crawford, V. Look-ups as the Windows of the Strategic Soul: Studying Cognition via Information Search in Game Experiments. In *Perspectives on the Future of Economics: Positive and Normative Foundations*, A. Caplin and A. Schotter, Eds. Oxford University Press, in press.

[8] Fischbacher, U. z-Tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics 10*, 2 (2007), 171–178.

[9] Freidlin, M., and Wentzell, A. *Random Perturbations of Dynamical Systems.* Springer Verlag, New York, 1984.

[10] Gabaix, X., Laibson, D., Moloche, G., and Weinberg, S. Costly Information Acquisition: Experimental Analysis of a Boundedly Rational Model. *American Economic Review 96*, 4 (Sept. 2006), 1043–1068.

[11] Ho, T., Camerer, C., and Chong, J. Self-tuning experience weighted attraction learning in games. *Journal of Economic Theory 133*, 1 (2007), 177–198.

[12] Ho, T. H., Wang, X., and Camerer, C. Individual Differences in EWA Learning with Partial Payoff Information, October 2006.

[13] Huck, S., Normann, H.-T., and Oechssler, J. Learning in Cournot Oligopoly–An Experiment. *Economic Journal 109*, 454 (1999), C80–95.

[14] Huck, S., Normann, H.-T., and Oechssler, J. Trial & Error to Collusion - The Discrete Case, May 2000. Bonn Econ. Discussion Papers.

[15] Huck, S., Normann, H.-T., and Oechssler, J. Through Trial and Error to Collusion. *International Economic Review 45*, 1 (02 2004), 205–224.

[16] Johnson, E. J., Camerer, C., Sen, S., and Rymon, T. Detecting Failures of Backward Induction: Monitoring Information Search in Sequential Bargaining. *Journal of Economic Theory 104*, 1 (May 2002), 16–47.

[17] Johnson, E. J., Schkade, D. A., and Bettman., J. R. Monitoring Information Processing and Decision: The Mouselab System, 1988. working paper, Graduate School of Industrial Administration, Carnegie-Mellon University.

[18] Monderer, D., and Shapley, L. Potential Games. *Games and Economic Behavior 14* (May 1996), 124–143.

[19] Normann, H.-T., and Wallace, B. The Impact of the Termination Rule on Cooperation in a Prisoner's Dilemma Experiment, may 2005.
http://personal.rhul.ac.uk/ulte/003/termination.pdf.

[20] Offerman, T., Potters, J., and Sonnemans, J. Imitation and Belief Learning in an Oligopoly Experiment. *Review of Economic Studies 69*, 4 (October 2002), 973–97.

[21] Peyton Young, H. The Evolution of Conventions. *Econometrica 61*, 1 (January 1993), 57–84.

[22] Rassenti, S., Reynolds, S. S., Smith, V. L., and Szidarovszky, F. Adaptation and convergence of behavior in repeated experimental Cournot games. *Journal of Economic Behavior & Organization 41*, 2 (February 2000), 117–146.

[23] Selten, R., Mitzkewitz, M., and Uhlich, G. R. Duopoly Strategies Programmed by Experienced Players. *Econometrica 65*, 3 (1997), 517–556.

[24] Skrondal, A., and Rabe-Hesketh, S. *Generalized Latent Variable Modeling: multilevel, longitudinal, and structural equation models*. CRC Press, 2004.

[25] Theocharis, R. On the Stability of the Cournot Solution on the Oligopoly Problem. *Review of Economic Studies 73*, 2 (February 1960), 133–134.

[26] Vega-Redondo, F. The Evolution of Walrasian Behavior. *Econometrica 65*, 2 (1997), 375–384.