Available online at www.sciencedirect.com
ScienceDirect
Research in Economics 60 (2006) 155–167
RESEARCH IN ECONOMICS
www.elsevier.com/locate/rie

Learning to collude: An experiment in convergence and equilibrium selection in oligopoly☆

Huw D. Dixon[a], Patrizia Sbriglia[b,*], Ernesto Somma[c]

[a] *Economics Department, University of York, Y010 5DD, United Kingdom*
[b] *Dipartimento di Diritto e Economia, Facoltà di Economia, Università di Napoli II, SUN - Via Gran Priorato di Malta, 83043, Capua, Caserta, Italy*
[c] *University of Bari, United Kingdom*

Abstract

The paper considers a simple oligopoly model where firms know their own and the average pay-off in the industry. Firms choose decision rules for trading. The theory predicts that there are three types of Nash equilibria in this game (collusive, Cournot and Stackelberg). Our experiments test the selection process. We find that there is clear evidence of convergence to an equilibrium, and whilst both Cournot and collusive outcomes were selected, the collusive equilibrium is more common. The experimental results also give insights into the process of individual learning, confirming that subjects follow *aspiration rules* rather than *reinforcement rules*.

Keywords: Oligopoly; Learning; Experiment

0. Introduction

The issue of learning in oligopoly has recently attracted a great deal of attention. In most cases, the aim of the experimental research has been to assess the role of alternative learning paradigms in the open question of equilibrium selection in markets with few sellers. In fact, it is generally argued that the rules that individuals use to approach and play oligopoly games might be the key to explaining deviations (e.g., more competitive or more collusive outcomes) from the standard Nash–Cournot equilibrium concept.

The experimental tests have mainly been focused on the comparison between alternative learning rules: *imitation* (Bosh-Domènech and Vriend, 2003; Huck et al., 2000), qualitative response models (Nagel and Vriend, 1999), *reinforcement learning* (Lupi and Sbriglia, 2003; Huck et al., 1999), and several alternative hypotheses on *belief learning*, ranging from basic best reply dynamics to sophisticated *fictitious play* rules (Rassenti et al., 2000; Lupi and Sbriglia, 2003; Huck et al., 1999). The results reached in this area are however still ambiguous: the evidence in favour of *belief*-based behaviours seems to prevail, but *imitation* and simple low rationality routines, such as reinforcement

☆ Financial support from the Italian National Council of Research and MURST is acknowledged by Patrizia Sbriglia, and from the Nuffield Foundation by Huw Dixon.

* Corresponding author. Tel.: +39 (0)823274023; fax: +39 (0)823274042.
E-mail addresses: hdd1@york.ac.uk (H.D. Dixon), patrizia.sbriglia@unina2.it (P. Sbriglia), es10@york.ac.uk (E. Somma).

1090-9443/$ - see front matter © 2006 Published by Elsevier Ltd on behalf of University of Venice.
doi:10.1016/j.rie.2006.08.001

rules, also find significant support in the data. Because of the difficulty in identifying a unique model of behaviour in oligopoly games, the analysis of equilibrium selection in oligopoly is still uncompleted. Between imitative behaviour and simple best reply rules, for instance, the difference is sometimes more conceptual than observational; however, the two rules lead to different selection processes.[1]

The ambiguity of the results suggests the need for more investigation, and the aim of our work is in fact to provide an alternative route to study experimental oligopolies. Our specific aim is to test whether the strategic interaction among "firms", taking place in experimental oligopolies, might lead to collusive outcomes, or, in other words, whether there is a natural tendency for markets to become collusive over time.

Following Dixon (2000), Dixon et al. (2002), and Moss et al. (1995), we focus on the economy-wide interaction of firms competing in different markets, rather than the single-market dynamics. The linkage among the individual markets will be provided by an average measure of performance, e.g., the industry's average profits for each period. The implication of our assumption is that the individual's behaviour might be affected both by what happens in the market where she/he is directly operating and by the industry in which such a market is located.

Assuming agents follow *aspiration learning rules* (Palomino and Vega-Redondo, 1999; Dixon, 2000; Oechssler, 2002), e.g., they switch to a different strategy every time their pay-off is lower than the population mean, it can be proved[2] that convergence to collusive outcomes is likely to be observed in most experimental duopolies.

Our experimental set-up is aimed at testing such a prediction. We consider a simple oligopoly model in a situation where firms know their own and the average pay-off in the industry. In our experimental designs we developed an environment in which the subjects chose the rules by which the oligopoly game is played. We restricted the oligopoly rules to linear reaction functions with a one-period memory. We gave them a limited set of options: six rules which include the most popular oligopoly strategies. The resultant oligopoly game has three types of equilibria: one set is collusive, one corresponds to the Cournot outcome, and the third set corresponds to the Stackelberg equilibrium. Subjects had limited information on the game to be played; participants had no information on the rival's actions but knew their own profit and the "industry" average profit for the market day.

We arrive at three main results. First, our experiments show that in most cases there was clear evidence of convergence to an equilibrium, and whilst both Cournot and collusive outcomes were selected, the collusive equilibrium was more common.

Second, we compare the evidence with the results of a simulated evolutionary model, in order to assess whether the aggregate learning behaviour might be described by a general model of bounded rationality. The comparison between the experimental evidence and the evolutionary model provides support to our hypothesis.

Finally, the analysis of the individual choices shows that *aspiration rules* where successful codes of behaviour among the experimental subjects, and, in the sessions where the collusive equilibrium was selected, such learning behaviour was predominant among participants.

Our main conclusion is that, though it is generally recognised that the study of learning in social games is an area of great importance, mainly for its normative implications, many aspects still need to be studied in order to draw general results from our analyses. We believe that our model, though not conclusive, can be considered as a step in the right direction, because it proves that simple learning rules may lead to systematic deviations from the Nash equilibrium outcome.

The organisation of the paper is as follows. In Section 1 we outline the oligopoly model underlying the experiments, while Section 2 describes the design and the running of the experiments, and states our working hypotheses. Sections 3 and 4 will present, respectively, the analysis of the experimental results as far as the related issues of convergence and equilibrium selection are concerned, and the exploration of the individual and aggregate learning behaviour in the experiments. In the conclusions, we summarise the main findings of our work.

1. The theory underlying the experiments: The oligopoly model

The theoretical set-up of our experiments closely resembles those of Dixon et al., 2002 and Moss et al., 1995. One of our main concerns was to design an oligopoly experiment that, unlike many of its counterparts present in the

[1] In particular, Huck et al. (2000), show that, unlike belief models, *imitation* tends to lead to the selection of competitive equilibria in Cournot experimental games.

[2] See Section 3.

experimental literature, would be perceived by the players as a "realistic" market environment, with few simple rules and a clear information set-up.

The stage game was a Cournot duopoly with homogeneous goods played by a population of firms matched in pairs.[3] Firms choose output levels that are produced without cost. The selling price is assumed to be a linear function of the industry output $p = a - q_a - q_b$.

Firms are endowed with a decision rule mapping the output produced by the rival onto a choice of output:

$$S_a = f_a : \{q_b\} \Rightarrow [0, 24].$$

The decision rule is linear and is composed by an intercept term and a slope term:

$$q_a = h_0 + h_1 q_b.$$

We restrict our attention to a finite number of decision rules[4]; in particular we allow for six different decision rules each corresponding to a well-defined type of behaviour. These are reported in Table 1.

Let $S = \{(1/2, -1/2), (1, -1), (1/2, 0), (1/3, 0), (0, 1), (1/3, -1/3)\}$ be the strategy space.

The Myopic Cournot firm (MC) behaves like in the standard Cournot model; the Walrasian firm (WF) sells output up to the point where price equals marginal cost. The Stackelberg sticker (SS) produces the monopolist profit maximising level of output, whilst the Cournot sticker (CS) produces the Cournot level of output, irrespective of the output decision of the rival. The Copy cat (CC) sets its output equal to that produced by the opponent.[5] The joint profit maximiser firm (JPM) can be thought of as a firm that expects its rival to be cooperative and consequently sets it decision rule so as to produce the joint profit maximising level of output.[6]

The Cournot and Stackelberg sticker, as well as the Copy cat, depict a specific type of behaviour which, in our opinion, might have some relevance when studying experimental duopolies. In the first two cases (CC and SS), in fact, the strategies describe the behaviour of a player who keeps responding with a specific strategy, regardless of the rival's response. Though short-sighted, such behaviour is often observed in market experiments and, as a matter of fact, in an environment where information is limited, such low rationality rules might even be justified.

The Copy cat strategy depicts a "pure" imitative behaviour, in as much as the individual's responses always match her/his opponent's choice. It is easy to verify that an equilibrium for the stage game always exists and it is stable unless both firms have the slope term equal to $+-1$.[7]

The equilibrium output and profits are given by

$$q_a^* = \frac{h_{0,a} + h_{01a} h_{0,b}}{1 - h_{1,b} h_{1,a}} \cdot 24 \tag{1}$$

[3] The matching mechanism adopted in the experiments is of the "playing the field" type, i.e., each firm, in each period, competes with all rivals in the industry. For a more detailed explanation of the experimental design, see the next section. For a comparison between our theoretical and experimental setting for testing aspiration rules and the alternative random matching setting, see Altavilla et al. (2005).

[4] Dixon et al. (2002) develop a methodology for generating decision rules, or "firm types" (see Dixon et al. (2002), pp. 142–143). In the quoted paper, it is shown that symmetric decision rules, i.e., decision rules that are on the diagonal line in the unit triangle, $A \equiv \{x \in [0, 1]^2 : 1 - x_i - x_j \geq 0\}$, are best responses to themselves, and a given rule generated by a point $a \in A$ is a best response to a' which is the reflection of a on the 45° line. For a given x, a Nash equilibrium in decision rules is found, when firm i chooses as reaction function the tangent to the firm j iso-profit curve at x, and vice versa.

[5] The Copy cat rule resembles Axelrod's "tit for tat" strategy. In the experiments we set the pay-off of two CC firms competing with each other equal to the JPM outcome.

[6] In terms of conjectural variation models with quadratic pay-offs, the JPM firm expects its rival to have a reaction function with slope equal to unity.

[7] We can represent the decision rules as defined by a dynamic system of the form $\underline{q}_t = h + H\underline{q}_{t-1}$ where

$$\underline{h} = \begin{bmatrix} h_{0,a} & h_{0,b} \end{bmatrix}, \quad H = \begin{bmatrix} 0 & h_{1,a} \\ h_{1,b} & 0 \end{bmatrix}.$$

The equilibrium of this dynamic is stable if both eigenvalues are real and less than unity in absolute value. The eigenvalues are the root of $\sqrt{h_{1,a} h_{1,b}}$. When two Walrasian firms meet we get a positive unit root; the quantities produced will depend on the initial output while the profits for both firms are always zero. Similar problems arise when two Copy cats meet; here we adopt the convention that the firms equally share the joint profit maximising output. Finally, when a Walrasian firm meets a Copy cat the dynamics gets into a cycle and an equilibrium is never reached. We make the assumption that both firms serve half of the market and profits are zero.

Table 1
The decision rules for the six firm types

	h_0	h_1
Myopic Cournot MC	1/2	−1/2
Walrasian firm WF	1	−1
Stackelberg sticker SS	1/2	0
Cournot sticker CS	1/3	0
Copy cat CC	0	1
Joint profit maximiser JPM	1/3	−1/3

Table 2
The stage game pay-off matrix

	MC	WF	SS	CS	CC	JPM
MC	64	0	36	64	64	92.6
WF	0	0	0	0	0	0
SS	72	0	0	48	0	96
CS	64	0	32	64	64	85.33
CC	64	0	0	64	72	72
JPM	46.08	0	32	56.88	72	72

Table 3
The pay-offs at the theoretical equilibrium points

collusive	Cournot	Stackelberg
648	576	530.18

and

$$\Pi_a^* = \frac{(1 - h_{1,b}h_{1,a} - h_{0,a} - h_{1,a}h_{0,b} - h_{0,b} - h_{1,b}h_{0,a})(h_{0,a} + h_{1,a}h_{0,b})}{(h_{1,b}h_{1,a} - 1)^2} \cdot (24)^2. \tag{2}$$

From (2) and given S the following pay-off matrix for the stage game is obtained. This is a bimatrix game, and the pay-off given is to the row firm when it plays against the column firm (see Table 2):

The pure strategy two-player NE are: (MC, SS), (SS, MC), (CS, CS), (CC, CC), where the first two are strict NE whilst the other two are weak NE. Moreover MC weakly dominates CS. The mixed strategy equilibria are more relevant to our experimental design, where the p can be interpreted as the proportion of the population playing the strategy. The three types of mixed NE are (with suitable titles):

$$\text{Stackelberg} \quad \left(\frac{9}{11}\text{MC}, \frac{2}{11}\text{SS}\right) \tag{3}$$

$$\text{Cournot} \quad (p\text{MC}, (1-p)\text{CS}), \ldots \quad \text{where:} \ldots 0 \le p < \frac{2}{3} \tag{4}$$

$$\text{collusive} \quad (p\text{CC}, (1-P)\text{JPM}), \ldots \quad \text{where:} \ldots \frac{5}{7} < p \le 1. \tag{5}$$

From Table 2, we can categorise the possible Nash equilibria into groupings depending on the equilibrium pay-off. Firstly we have the "collusive" equilibria, which involve the strategies (CC, JPM) in some proportions. Secondly, we have the "Cournot" equilibria, in which the strategies (CS, MC) occur. Finally, we have the "Stackelberg" equilibrium. The expected pay-offs for these three types of equilibria are (to two d.p.) given in Table 3.

At the equilibrium, the three sets comprise both the "traditional" oligopoly strategies and the low rationality rules (CC, SS and Copy cat), since the levels of equilibrium output are the same. However, we shall distinguish among the types of strategies when analysing the players' decision process.

Table 4
Trading strategies as presented to experimental subjects

	MC	WF	SS	CS	CC	JPM
Myopic Cournot MC	8	9	6	8	8	9.6
Walrasian firm WF	24	12	12	16	12	24
Stackelberg sticker SS	12	12	12	12	12	12
Cournot sticker CS	8	8	8	8	8	8
Copy cat CC	8	12	12	8	6	6
Joint profit max. JPM	4.8	0	4	5.3	6	6

2. Market designs and experimental predictions

The experiments[8] were designed as market games in which a firm type (Myopic Cournot, Walrasian firm, etc.) is understood by the subjects as a "trading strategy" (or production plan). A trading strategy is thought of in terms of the reaction functions.[9] However, rather than using the specialist concept of the reaction function, the subjects were told to think of a trading strategy in terms of a schedule of output levels. There are six trading strategies available, and the subjects were given the amount that each trading strategy leads to when it meets each of the other trading strategies. Both in the instructions and in the practice rounds, which explained how the software operated, they were told that by choosing a trading strategy, they would be responding with an output level to a different output level chosen by the player they were matched in the particular duopoly game. This was illustrated by suggesting that each player was selling his/her good in a market street in which one opponent entered at a time, offering the same good in different amounts and so affecting the market price consequently. In the experiment, these trading strategies were given letters {A, B, C,...} as names: here, however, we present them with their economic label. Except for the names, what the subjects were shown is in Table 4.

The experimental subjects were told that the experiment was divided into "days". At the beginning of each day, each subject had to choose his/her strategy for that day's trading. In the first period (day 1), the strategies were chosen by the experimenter (these were chosen to be close to a "flat" distribution, given that there were six strategies and eight to ten players in each experiment). During the day, each participant meets all of the other participants, and in each encounter his chosen trading strategy plays against the strategy chosen by the person he is playing against. Thus, if there are nine experimental subjects, each one meets with the other eight every day.[10]

At the end of the day, each subject was given two bits of information: (a) his/her own profits for the day — this was expressed as the total accumulated throughout the day, (b) the average profits over all subjects that period. As a rough magnitude, the individual subjects profits can vary (with 10 players) between 0 and 922; the average over all players can vary between 0 and 720. These magnitudes translate easily into monetary terms (they are pennies earned that day, so that 650 = £6.50), and the intercept in the demand curve was chosen to yield the right sort of figures.

In the Instructions,[11] subjects were informed of the costs and the demand parameters; however, they had limited information on the rivals' profits and strategies, and they did not receive information on the pay-off matrix (Table 2).

We realised two experimental treatments. In the baseline treatment (Type 1), the only pieces of information delivered to subjects were the individual and the average per day profit, but no extra information on the rivals' performance was given. In the second treatment we explore the effects of imitative behaviour, by allowing subjects to observe the trading strategy and pay-off of another subject randomly chosen by the computer. In order to "imitate", the subject had to pay a fee: this was set at 24.

Furthermore, in one session 21a we explore the effects of experience: in fact, we had the *same* group of subjects perform the same experiment with the first treatment's design.

[8] The experiments were conducted at the University of York; we thank N. Spivey and the personnel of the EXEC (Centre for Experimental Economics) for technical assistance. The participants were undergraduate students from the same University.

[9] Though with a different underlying model, Martini (2003) presents oligopoly experiments where subjects selected reaction functions, in a similar fashion to in our experimental designs.

[10] As a point of experimental design, we did not have the person meet themselves: whilst this might make economic sense, it is not easy to explain this to experimental subjects in a convincing way!

[11] A copy of the Instructions is available on request.

Table 5
The experiments: Design, duration and endpoints[a]

Experiment	Notes	No. of subjects	No. of days	$\Pi(10)$	$\Pi(5)$	Sw 5
24a	Basic	10	81	565	568	14
24b	Basic	10	80	600	607	13
24c	Basic	10	80	591	596	10
24d	Basic	9	62	622	634	5
24e	Imitation	9	80	628	631	7
24f	Imitation	9	80	625	625	6
24g	Imitation	9	80	618	631	9
24h	Imitation	10	36	627	616	8
24i	Basic	10	80	641	645	4
21a	Imitation: same subjects as 24i	10	36	641*	643*	6

[a] Notice that $\Pi(s)$ gives the average profit over all players in the last s periods ($s = 5, 10$). Sw 5 is the total number of instances of switching in the last five periods of the game. The pay-off for 21a is renormalised to be comparable with those for the other experiments.

Another important point about the experimental design was the *stopping rule*. There were a fixed number of market days: we allowed 45 min for the play of the experiment, and so a fixed term play of 80 periods was one possibility. However, if the experiments converged, then the boredom of sitting in a room pressing the same key might create an incentive for experimentation, which we did not want. We therefore decided on an endogenous stopping rule: if all participants each pressed the same key as they had done in the previous three periods, then the game would be stopped. The maximum length of the game if this stopping rule was not activated was set at 80 periods (in one case 24a, it was left to run for 81 days). It was explained that the monetary pay-off to subjects was going to be the *average over the experiment*: thus the length of the game per se was not important.

We ran a total of ten experiments over five days. The details of the experiments are given in Table 5. At the beginning of the session, the subjects were given a leaflet describing the experiment, and the structure of the game was explained by the organisers. We also had a dry run of three market days. Each experiment had a maximum of ten participants. We had calibrated the model (essentially by choosing the intercept on the demand curve) so that we thought that the subjects would earn about £6–8 pounds in a session: ex post, the subjects earned a little less than we had expected, but most were remunerated within the lower half of this range. In the last experiment 21a, were the same people were participating as in 24i, the payments were scaled down by 30%: we adjusted the intercept term for the demand curve from 24 to 21. Our subjects were mainly students, taken from across all subject areas: none would have had any specialist knowledge about the economic theory underlying the experimental design.

The aim of our study is twofold. As stated in the Introduction, we are interested both in studying the learning dynamics of the experimental subjects in order to assess the relevance of the aspiration rules in the context of social games, and in exploring the issue of equilibrium convergence in the oligopoly model described above. The following claims sum up, therefore, our hypotheses testing.

Claim 1. *The dynamics of individual players is better explained by a general model of bounded rationality. Comparing the relative application of simple reinforcement routines and aspiration-based learning, the behaviour of the subjects is better explained by the latter individual learning rules.*

Claim 2. *If the dynamics of individual play can be explained by an aspiration learning model we observe more collusion in the market experiments and equilibrium (3) is more often selected by players.*

3. Convergence and equilibrium selection

Though it was stated as Claim 2, let us begin with the issue of equilibrium selection. Some immediate observations can made by looking at Table 5. Firstly, except for two experiments (24a and 24c), the duopoly's profits were higher than 600 in the last ten periods of play; in some cases (in particular in the imitation setting) the final value was remarkably close the collusive equilibrium. Table 6 divides the experiments into three different groupings, according to the endpoints.

In no case did we find clear evidence of convergence to the theoretical Stackelberg equilibrium point; experiment 24a was, in fact, the only experiment in which the Stackelberg strategy was more frequently adopted

Table 6
Classification according to end-profits

Classification	Profit range	Experiments
Cournot	550–599	24a, 24c
Intermediate	600–629	24b, 24h
collusive	630–650	24d, 24e, 24f, 24g, 21, 21a

(the relative frequency being over than 10%). Nevertheless, looking at the history of play, we can conclude that the combinations of the Myopic Cournot and the Cournot sticker were the more successful strategies in that specific game, with a relative frequency higher than 50%.

However, the information shown in Tables 5 and 6 is not sufficient to establish the issue of equilibrium selection, since it does not explain the history of play in each session; for this reason, we applied an accelerated decay model to our data, in an attempt to model the dynamic process. The general form of the model is

$$\text{In}(X_t - X^*) = \alpha + \beta t + \varepsilon$$

where X_t is the average industry profit in period t, and X^* is the equilibrium value in the collusive and Cournot regimes (648 in the former, 576 in the latter case).

The basic idea is that, as effects of experience and learning, the individuals adjust their behaviour moving towards equilibrium, and the coefficient β measures the proportionate change in X over time:

$$\beta = [\mathrm{d}(X_t - X^*)/(X_t - X^*)]/\mathrm{d}t. \tag{6}$$

If β is negative, then

$$\lim_{t\to\infty}(X_t - X^*) = 0. \tag{7}$$

Tests for convergence to the Cournot and the collusive equilibrium were both run for each experiment. There are of course four possible outcomes: convergence to neither, convergence to both, and convergence to only one.

As is shown in Table 7, the results are close to the conclusions we reached earlier: all experiments, with the notable exceptions of 24a and 24c, clearly converged towards the collusive equilibrium.

It is interesting to notice that the speed of convergence varies across experiments. In experiments 24d and 21a,[12] the absolute value of the β coefficient is higher than in the remaining cases, showing a faster convergence to the equilibrium.

Experiments 24c and 24a differ from the other experiments in as much as convergence is not easy to establish. The two experiments, however, are not comparable.

As far as 24c is concerned, in fact, the inspection of the data reveals that the individual behaviours were very noisy at the early stages, and tend to settle after some time. For this reason, we estimated the model using the last 40 observations and, as is shown in Table 8, a clearer picture emerged.

Experiment 24a is probably the only one in which convergence cannot be easily established, since the individual behaviour was noisy throughout the experiment. However, looking at the size of the coefficient in Table 6, it appears that convergence to the Cournot equilibrium is much faster. As a further aid to the understanding of the players' behaviour, we looked at the long run value of the industry pay-off, calculated estimating an asymptotic root model, of the form

$$X_t = \alpha + \beta X_{t-1} + \varepsilon \tag{8}$$

so that:

$$\lim_{t\to\infty} X_t = \alpha/(1-\beta). \tag{9}$$

[12] In these two sessions the stopping rule was applied and the experiments were terminated at the 62nd and 36th stages, respectively.

Table 7
Estimates of the accelerated decay model

Experiment	Accelerated decay model convergence to the collusive equilibrium			Accelerated decay model convergence to the Cournot equilibrium		
	β	R^2	DW	β	R^2	DW
24a	−0.006 (−3.304)	0.12	1.68	−0.024	0.10	1.92
24b	−0.026 (−6.016)	0.32	1.84	0.010 (1.257)	0.02	1.26
24c	−0.013 (−9.057)	0.51	1.70	−0.028 (−3.385)	0.12	1.30
24d	−0.069 (−9.126)	0.58	1.45	0.018 (1.863)	0.05	1.15
24e	−0.037 (−9.120)	0.52	1.56	0.019 (2.928)	0.09	1.41
24f	−0.026 (−8.634)	0.49	2.15	0.014 (3.54)	0.13	1.80
24g	−0.040 (−5.382)	0.27	2.02	0.010 (3.502)	0.14	2.32
24h	−0.051 (−8.944)	0.51	2.07	0.014 (2.482)	0.07	2.30
24i	−0.049 (−9.787)	0.55	1.43	0.001 (2.198)	0.06	1.44
21a	−0.241 (−5.609)	0.48	1.92	0.044 (2.659)	0.17	0.96

t-statistics are in brackets.

Table 8
Experiment 24c (40–80)

Accelerated decay model convergence to the collusive equilibrium			Accelerated decay model convergence to the Cournot equilibrium		
β	R^2	DW	β	R^2	DW
−0.011 (−3.967)	0.29	2.00	0.061 (3.441)	0.23	2.28

t-statistics are in brackets.

Such a value (for the last 40 periods) corresponds to 555.50,[13] showing therefore that the long run industry pay-off was between the Stackelberg and the Cournot equilibria, moving towards the latter. An inspection of the strategies played by players in the last 11 periods confirmed this hypothesis, since strategies A and D (respectively, Myopic Cournot and Cournot sticker) were the most frequently played, with an increasing proportion of the players' population switching to the Copy cat strategy, E (which, it has to be remembered, yields a Cournot outcome when playing strategies A and D), and with a survival of the Stackelberg strategy (C), which is played by one to three players in 4 out of 11 of the final stages.

We are now able to sum up the results of the analysis so far.

Result 1. All experiments show that individual play settles down to an equilibrium after an early period of noisy behaviour, and whilst the Stackelberg equilibrium is rarely observed in all sessions, both Cournot and collusive were adopted by the majority of the experimental subjects. However, the analysis of the convergence issue shows that, in all cases but one, the long run profits tend to the level of the Joint profit/Copy cat equilibrium points.

[13] It increases to 561.69 for the last 20 periods.

4. Aggregate and individual learning behaviour

We move now to the analysis of the learning behaviour, both at the aggregate and at the individual level. As stated in Claims 1 and 2, we are interested in (i) assessing whether the individuals' behaviour can be explained by a model of bounded rationality, and, specifically, whether they follow aspiration rules, (ii) and whether, in the experiments in which such a behavioural code was predominant, individual play converged toward collusive outcomes. We tackle these problems in two subsequent steps.

First, we compare the experimental evidence with the results of a simulated evolutionary model. Such a comparison will allow us to assess whether the bounded rationality hypothesis is a correct way of proceeding when analysing the individuals' choices. Second, we analyse the individuals' behaviour in two specific sessions (24d and 24a) in which the aggregate output choices converged towards the Cournot and the collusive equilibrium, respectively. For these two experimental sessions we test whether individuals' learning rules can be better described by the model of *aspiration* learning rather than a simple *reinforcement* model.

As for the first step, we concentrate here on the selection mechanism known as *replicator dynamics*.[14] We assume that a large and finite population of firms is programmed to play pure strategies (decision rules) and that they are randomly matched in pairs to play the Cournot duopoly game described above. We adopt the following discrete time version of the replicator dynamics:

$$z_{i,t} = (1-d)\left(z_{i,t-1}\frac{\Pi_{t-1}}{\overline{\Pi}_{t-1}}\right) + d\frac{1}{n} \tag{10}$$

where $z_{i,t}$ is the fraction of the population playing strategy i at time t and $d \in [0, 1]$. $\Pi_{i,t}$ is the average pay-off of firm type i at time t:

$$\Pi_{i,t} = \sum_{j=1}^{n} z_{j,t}\pi_{i,j} \tag{11}$$

where $\pi_{i,j}$ is the pay-off to firm type i against firm type j as given in Table 2. $\overline{\Pi}_{t-1}$ is the average pay-off of all firms in iteration t:

$$\overline{\Pi}_t = \sum_{i=1}^{n} z_{i,t}\Pi_{i,t}. \tag{12}$$

In each period a fraction $(1-d)$ of the population switches progressively towards those strategies that earn pay-offs above average whilst the remaining fraction d randomly mutates into one of the existing strategies. The parameter d introduces noise into the otherwise deterministic process defined by the evolutionary model. The exact process giving rise to the kind of evolution implied by the *replicator* equation is not modelled here. We can think of a process of imitation, which leads the less successful firms to imitate the more successful ones, or of a process of propagation by which thriving strategies are spread by managers, which move from one firm to another, or (as we assume here) agents using näive learning rules, such as reinforcement rules (Borgers and Sarin, 1995).[15] The *replicator* dynamics is inherently a selection mechanism: the least successful firms are more likely to go bankrupt.

Table 9 gives the results of the simulations of the *replicator dynamics*. We consider both cases of evolutionary dynamics, without noise and with noise.

- Evolution without noise.

The sample considered consisted of six firm types (strategies). The simulations ran until the differences of surviving firm types profits from the population average were smaller than 1e−100. We tried several different random initial distributions and the results consistently reported the survival of strategies CC and JPM. The share of CC type firms is

[14] See Vega Redondo (1996).

[15] There are now a large number of research works which analyse the economic microfoundation of the *replicator dynamics* algorithm, which has, in contrast, an immediate application in biological settings. Amongst the recent surveys, see Ponti (2000).

Table 9
The evolutionary results for different levels of noise

	$d = 0.0001$	$d = 0.001$	$d = 0.01$	$d = 0.1$
Myopic Cournot	1.86e−5	1.86e−5	1.86e−5	0.02
Walrasian firm	9.9e−7	9.9e−7	9.9e−7	0.001
Stackelberg sticker	1.23e−6	1.23e−6	1.23e−6	0.001
Cournot sticker	1.47e−5	1.47e−5	1.47e−5	0.016
Copy cat	0.85	0.85	0.85	0.82
Joint profit maximiser	0.15	0.15	0.15	0.14
Iterations	2309 258	253 116	26 751	2243

always extremely close to unity ranging from 0.999 upwards. This result indicates a strong tendency to cooperation. The average pay-off in equilibrium is in fact equal to 0.125 which corresponds to the collusive outcome.

- Evolution with noise.

We ran several simulations for different levels of noise. When noise is introduced, a fraction d of firms undergo a random mutation which, according to the particular specification adopted, does not introduce new firm types. The mutant firms switch to any other firm type with equal probability ($1/6$) and, consequently, all firms survive with at least a share of (d/n) (the underlying idea is that agents mistakenly reintroduce strategies dismissed in the past). The convergence criterion adopted is that the change in population shares for each firm type in two successive periods is less than 1e−300.

As appears from Table 9, the evolutionary dynamics converges towards the cooperative equilibrium, in both the cases of evolution, with and without noise. However, the introduction of noise does mark an important difference, in as much as the Cournot strategy keeps a small proportion of the population when the noise is at a high level. The introduction of "noise", or "mutations" in a biological sense, is generally interpreted as the effects of human mistakes, such as replaying a strategy which was excluded earlier in the game. In experimental settings, mistakes might be brought about by misunderstandings on the nature and/or the rules of the game, or simply "by hitting the wrong key", and therefore the most plausible way of studying evolutionary dynamics in experiments is concentrating on the noisy case, which is more responsive to the specific context.

For this reason, we compared the final distribution of strategies of the noisy *replicator* ($d = 0.1$) to the observed experimental distributions, averaged over the last ten periods, using a χ^2. In all cases but one, 24a, the test accepted the null hypothesis, e.g., the two distributions of strategies were equal.

There is, however, a relevant difference between the experimental evidence and the theoretical distribution of Table 8, in as much as the imitative strategy (Copy cat) tends to have a higher share of the population (in the last periods of play) than the results of Table 8 would suggest, compared to the JPM strategy. The diffusion and the relative success of the *tit for tat* rule are an indication of the degree and type of rationality that prevailed in the sessions.

As stated before, the study of the evolutionary dynamics does not provide a sufficient explanation of the participants' choices. In other words, we need to explore the individual learning process, in order to establish which rule is more apt to describe the players' behaviour. In this respect we will test whether players adopted simple *aspiration*-based learning rules, or whether a simple *reinforcement* model provides the correct interpretation of the individual play.

For this specific aspect, we restrict our analysis to two experiments (24d and 24a), in which the Cournot and the collusive equilibrium points were selected, so that an immediate comparison between the selection mechanism and the individual rules can be drawn.

The stimulus–response paradigm, or *reinforcement* learning, is the first theory of individual behaviour we consider in our hypotheses testing. In very simple terms, the basic idea of this class of models is that the individual reinforces the probability of playing strategy a, according to the observed pay-off of a achieved in the past periods. This result is known as the *Law of Effect* (Luce, 1959), and it implies that choices that have led to "good" outcomes in the past are more likely to be used in the future. The reinforcement theory has been successfully used to explore human and animal behaviour in experimental settings (Bush and Mostseller, 1955), and it depicts an adaptive process of learning that can be described in the following manner. In a game with two players, player 1 at time $t = 0$ has an initial "propensity"

Table 10
Experiment 24a, individual learning

	Player 2	Player 3	Player 4	Player 5	Player 6	Player 7	Player 9	Player 10
α	0.53	0.46	0.56	0.37	0.64	0.6	0.22	2.00
β	0.87	1.30	0.81	0.62	0.71	0.3	0.35	2.25
Conv.	MC (CS)	CS (MC)	CS (MC)	CS (CC)	MC (CS)	CC (SS)	CS (MC)	CS (CC)

Table 11
Experiment 24d: Individual learning

	Player 1	Player 2	Player 3	Player 4	Player 5	Player 6	Player 7	Player 8	Player 9
α	1.33	1	0.61	1.44	1	0.67	0.79	0.37	0.50
β	1.56	0.27	0.17	1.11	1.17	1.60	0.58	0.44	0.61
Conv.	CC (MC)	CC (MC)	CC (MC)	CC (MC)	CC (JPM)	CC (JPM)	MC (CC)	CC (CS)	CC (CS)

(Roth and Erev, 1995, 1998), $A_k^1(t = 0)$, to play strategy k which yields a pay-off U_k^1, once it is actually used. Propensities are numbers that describe the individual's predisposition to play a specific strategy. Such predisposition may derive from her/his pre-play knowledge of the game (Camerer and Ho, 1999). The subject's propensity for i is updated in the following period according to the rule

$$A_k^1(t = 1) = A_k^1(t = 0) + \beta U_j^1(t = 0);$$

where: $0 < \beta \le 1$ if $k = j$;

$$\beta = 0 \quad \text{if } k \ne j;$$

the probability of playing strategy i at time $t = 1$, for player 1 is equal to

$$p_k^1(t = 1) = \frac{A_k^1(t = 1)}{\sum A_j^1(t = 1)}.$$

In economic analyses the reinforcement algorithm has been modelled in several alternative ways. In Borgers and Sarin (1995), actions may receive a negative reinforcement; whilst in Roth and Erev (1995, 1998) probabilities are updated according to a reinforcement function R, which depends on the relative magnitude of the pay-offs.

The alternative learning paradigm – known as *aspiration* learning (Palomino and Vega-Redondo, 1999; Dixon, 2000; Oechssler, 2002) – states that, when information on the industry's average profitability is provided, markets might converge to more collusive outcomes, if such market signals are perceived by players as *aspiration levels* and they therefore try new strategies whenever their individual profits fall below such a threshold.

In other words, the *aspiration* hypothesis assumes that players' choices are more responsive to market stimuli rather than to the individual profit differential, as stated in the *reinforcement* paradigm. The study of *aspiration* learning has a long historical tradition (Simon, 1955), and there has been extensive research on *aspiration* rules in a number of repeated games (Karandikar et al., 1998; Vega Redondo, 1996; Palomino and Vega-Redondo, 1999; Dixon, 2000; Oechssler, 2002). The general result is that aspiration rules lead to cooperation in bilateral games "where the maximum joint pay-off is attained by a symmetric strategy profile" (Palomino and Vega-Redondo, 1999, p. 486).

In the basic treatment design (Sessions: 24a, 24b, 24c, 24e, 24i), subjects were given two different pieces of information, regarding the current individual pay-off and the average profit the market day, whilst they received no information on their rivals' actions. This specific setting is therefore apt for a direct comparison between the two learning paradigms.

In Tables 10 and 11, we report – for the two sessions under investigation – two different adjustment factors. The first one, labelled α, can be defined as a *reinforcement adjustment factor* and is given by the ratio between the number of times the following inequality holds: $\overline{\pi} < \pi_t^i < \pi_{t-1}^i$ – for individual i, at time t – and the total number of strategies' changes. The second one, labelled β and defined as an *aspiration adjustment factor*, is otherwise given by the ratio between the number of times the following inequality holds: $\pi^i_{t-1} < \pi^i_t < \overline{\pi}$ – for individual i, at time t – and the total number of strategies' changes. Finally, we report, in the third row of both tables, an individual convergence ratio, by indicating the strategies which were most frequently used by each player.[16]

We analysed the behaviour of seventeen subjects[17]: 8 in the first session and 9 in the second one. In Session 24a, we considered the second half (periods 41–81) of the experiment, whilst for Session 24d, we considered the entire length of the experiment, rather than the second half. The reason we adopted different time criteria is related the specific speed of convergence. In the first experiment 24a, play did not converge within the experiment allowed time, and, on the contrary, individual behaviours show high levels of randomness during the first half. The opposite situation can be found for the second experiment under consideration 24d, where individual play converged to an equilibrium (Copy cat–JPM) in the 64th period.

Tables 9 and 10 show some interesting aspects of the individual learning behaviours.

Overall, the coefficient β is greater than α in eleven cases out of seventeen, thus confirming that in the majority of cases subjects' choices were more responsive to market stimuli than to individual profit differentials.

However, we are not able to confirm the final part of our Claim 2, related to the equilibrium selection process. In fact, as arises from both Tables 9 and 10, subjects for whom β was greater than α selected the mixed equilibrium (3) in two cases out of eleven, whilst in the majority of cases, the most frequently adopted strategies corresponded to the equilibrium point (1).

It should also be noticed that in no cases did reinforcement learners converge towards the cooperative equilibrium point, since the most frequently played strategies where MC, CS, CC and – in one instance – SS.

Result 2. The studies of the aggregate and the individual learning behaviours show that players' actions are better explained by a model of bounded rationality and that the individual play was more responsive to the market informational signal than the individual past performance. Whilst this latter finding seems to confirm that aspiration rules were successful codes of behaviour, our analysis does not provide support for the hypothesis of aspiration learners converging towards more cooperative equilibria final points.

5. Concluding remarks

The main aim of our experiments was to explore the issue of aspiration learning behaviour in oligopoly games. To this end, we designed experiments where subjects selected trading strategies, and received feedback information on their own and the average (market) performances.

Our main findings may be summed up as follows.

First, there is clear evidence of convergence in the experiments, and both Cournot and Cooperative equilibrium points have been selected.

Second, the behaviour of individuals is consistent with an aspiration learning model, where *aspirations* are close to the average level of profits. In fact, as is shown in our individual learning analysis, *aspiration* rules seem to be more common than *reinforcement* rules, even though our restricted investigation does not allow us to rule out alternative learning behaviours.

Finally, we would like to underline one direction in which our research could be extended. Much attention has been devoted to the Cournot games in the recent literature, but there is very little research on alternative market structures. The limited interest in studying alternative market institutions makes it impossible to compare human behaviours in different contexts, and draw general conclusions on the learning process in market experiments.

---

[16] We are aware that α and β can only be considered as an indication of the relative importance of the two informational signals (individual and average profits) on the subjects' choices, and that both adjustment factors do not take into account random behaviour which might have affected the total number of switches for each individual. Nevertheless, we believe that – whilst the existence of random behaviour has the same influence on both factors – a *direct* comparison of α and β still gives relevant insights into the relative importance of the two pieces information for the individual learning process.

[17] We excluded from our investigation individuals 1 and 8 in Session 24a, since they did not change their strategy throughout the period we considered.

References

Altavilla, C., Luini, L., Sbriglia, P., 2005. Social learning in market games. Journal of Economic Behaviour and Organization (in press).

Borgers, T., Sarin, R., 1995. Naïve reinforcement learning with endogenous aspirations. Mimeo. University College London.

Bosh-Domènech, A., Vriend, N.J., 2003. Imitation of successful behavior in Cournot markets. Economic Journal 113, 495–524.

Bush, R., Mostseller, F., 1955. Stochastic Models for Learning. Wiley Pub., New York.

Camerer, C., Ho, T.-H., 1999. Experience weighted attraction learning in normal form games. Econometrica 67, 827–874.

Dixon, H.D., 2000. Keeping up with the Joneses: Competition and evolution in collusion. Journal of Economic Behavior and Organization 43, 223–238.

Dixon, H.D., Wallis, D., Moss, S., 2002. Axelrod meets Cournot: Oligopoly and the evolutionary metaphor. Computational Economics 20, 139–156.

Huck, S., Normann, H.T., Oechssler, J., 1999. Learning in Cournot oligopoly—an experiment. Economic Journal 109, C80–C95.

Huck, S., Normann, H.T., Oechssler, J., 2000. Does information about competitors' actions increase or decrease competition in experimental oligopoly markets? International Journal of Industrial Organization 18, 39–57.

Karandikar, R., Mokherjie, D., Ray, D., Vega Redondo, F., 1998. Evolving aspiration and cooperation. Journal of Economic Theory 80, 292–331.

Luce, R.D., 1959. Individual Choice Behavior. A Theoretical Analysis. J. Wiley, Pub., USA.

Lupi, P., Sbriglia, P., 2003. Exploring human behaviour and learning in experimental Cournot settings. Rivista Internazionale di Scienze Sociali CXI, 373–395.

Martini, G., 2003. Complexity and individual rationality in a dynamic duopoly: An experimental study. Research in Economics 57, 345–370.

Moss, S., Dixon, H., Wallis, S., 1995. Evaluating competitive strategies. International Journal of Intelligent Systems in Accounting, Finance and Management 4, 361–392.

Nagel, R., Vriend, N., 1999. An experimental study of adaptive behavior in an oligopolistic market game. Journal of Evolutionary Economics 9, 27–65.

Oechssler, J., 2002. Cooperation as a result of learning with aspiration levels. Journal of Economic Behavior & Organization 49, 405–409.

Palomino, F., Vega-Redondo, F., 1999. Convergence of aspirations and (partial) cooperation in the prisoner's dilemma. International Journal of Game Theory 28, 465–488.

Ponti, G., 2000. Continuous-time evolutionary dynamics: Theory and practice. Research in Economics 54, 187–214.

Rassenti, S., Reynolds, S.S., Smith, V.L., Szidarovszky, F., 2000. Adaptation and convergence of behavior in repeated experimental Cournot games. Journal of Behavior and Organization 41, 117–146.

Roth, A., Erev, I., 1995. Learning in extensive-form games: Experimental data and simple dynamic models in the intermediate term. Games and Economic Behavior 8, 164–212.

Roth, A., Erev, I., 1998. Predicting how people play games: Reinforcement learning in games with unique mixed strategy equilibria. American Economic Review 78.

Simon, H.A, 1955. A behavioral model of rational choice. Quarterly Journal of Economics 69, 129–138.

Vega Redondo, F., 1996. Evolutionary games and economic behavior. Oxford University Press, Oxford.