Available online at www.sciencedirect.com

ScienceDirect

Journal of Economic Theory 155 (2015) 185–205

JOURNAL OF Economic Theory

www.elsevier.com/locate/jet

From imitation to collusion: Long-run learning in a low-information environment ☆

Daniel Friedman [a], Steffen Huck [b,c], Ryan Oprea [d,*], Simon Weidenholzer [e]

[a] *Economics Department, University of California, Santa Cruz, CA, 95064, United States*
[b] *WZB Berlin, Reichpietschufer 50, D-10785 Berlin, Germany*
[c] *Department of Economics, University College London, Gower Street, London, WC1E 6BT, UK*
[d] *Economics Department, University of California, Santa Barbara, CA, 93106 United States*
[e] *Department of Economics, University of Essex, Wivenhoe Park, Colchester CO4 3SQ, UK*

Received 27 September 2013; final version received 9 October 2014; accepted 24 October 2014

Available online 30 October 2014

Abstract

We explore the stability of imitation in a 1200-period experimental Cournot game where subjects do not know the payoff function but see the output quantities and payoffs of each oligopolist after every period. In line with theoretical predictions and previous experimental findings, our oligopolies reach highly competitive levels within 50 periods. However, already after 100 periods, quantities start to drop and eventually fall deep into collusive territory without pausing at the Nash equilibrium. Our results demonstrate how groups of subjects can learn their way out of dysfunctional heuristics, and suggest elements for a new theory of how cooperation emerges.

☆ We are grateful to the National Science Foundation for support under grant SES-0925039, James Pettit for programming the software, and Luba Petersen for assistance in running the sessions. For helpful comments we are grateful to conference participants at the 2012 American Economic Association Meetings, the 6th Nordic Conference on Behavioral and Experimental Economics in Lund 2011, the 2012 Santa Barbara Conference on Experimental and Behavioral Economics, seminar audiences at Alicante, Baruch College, Cambridge, Carlos III, CERGE-EI, Cologne, Edinburgh, Essex, Frankfurt, Nottingham, UC Merced, Pompeu Fabra, Queen Mary, Stirling, UCL, the University of Zürich and the WZB, and to two referees of this Journal. Weidenholzer acknowledges financial support from the Vienna Science and Technology Fund (WWTF) under project fund MA 09-017.

* Corresponding author.

E-mail addresses: dan@ucsc.edu (D. Friedman), steffen.huck@wzb.eu (S. Huck), roprea@gmail.com (R. Oprea), sweide@essex.ac.uk (S. Weidenholzer).

http://dx.doi.org/10.1016/j.jet.2014.10.006
0022-0531/

JEL classification: C73; C91; D43

Keywords: Cournot oligopoly; Imitation; Learning dynamics; Cooperation

1. Introduction

Imitation is an attractive heuristic when players have little information about the strategic environment but can observe others' choices and success. Compared to popular learning models that focus on own payoffs, imitation makes more comprehensive use of available information — but not necessarily *better* use, as first shown by Vega-Redondo [26] for the case of Cournot games where imitation generates the perfectly competitive Walrasian outcome. Within the broad class of aggregative games (Alós-Ferrer and Ania [2]), Cournot games are, similarly to public good games or common pool resource games, notable for their tension between social efficiency and individual optimization.[1] The efficient collusive profile contrasts with the less efficient Nash equilibrium, and with the still less efficient fully competitive or Walrasian outcome where price is equal to marginal cost. Vega-Redondo showed that imitating quantity choices of the more profitable players leads precisely to that least efficient profile, where economic profits are zero.

Of course, this unfortunate outcome arises from a blind spot in the imitation heuristic — it ignores the fact that prices fall with greater quantities. Nevertheless, the heuristic has been quite descriptive of laboratory behavior in low-information environments where players observe other players' quantity choices and profits but not the underlying payoff function. Most of these studies (including, among others, Huck, Normann, and Oechssler [14], Offerman, Potters, and Sonnemans [19], Apesteguía, Huck, and Oechssler [3] or Apesteguía, Huck, Oechssler, and Weidenholzer [4]), feature what has been considered "long horizon" repeated interaction of around 50 periods.

Our point of departure is to examine a much longer horizon. We employ the new ConG software (Pettit, Friedman, Kephart, and Oprea [20]) which allows for periods to be so short that human subjects perceive action as taking place in continuous time. Here we instead use the software to implement discrete 4-second periods — rather short by recent standards, but perceived by our subjects as comfortable stop-action in discrete time. This enables us to increase the number of periods to 1200.

The results are dramatic — what looked like stable long-run behavior in earlier studies turns out to be transient. In the first 50 periods of our experiment we replicate the very competitive outcomes observed in several previous studies. However, soon thereafter the trend reverses and quantity choices start to drop. Quantities often approach the Cournot–Nash level after 100 periods but they do not halt there, or even pause. Rather they continue to drop until they reach almost fully collusive levels in duopolies and reach, on average, deep into collusive territory in triopolies. These collusive levels are reached despite our use of a hyperbolic demand function that creates much stronger incentives to deviate from collusion than the linear demand functions seen in most Cournot oligopoly experiments.

The primary contribution of the present paper is to document this transition in outcomes — from very competitive to collusive. The transition demonstrates how players can learn to aban-

[1] Since consumers are not considered players in such games, social efficiency refers below to the players' joint payoff maximum at the cartel profile. Of course, if consumers were also regarded as players then the Walrasian outcome would be efficient and the cartel profile would be inefficient.

don dysfunctional heuristics and find better ways to reconcile group interest with self-interest. Interestingly, cooperation does not seem to be supported by Nash reversion or similar strategies; indeed the evidence suggests that our subjects never even learn the myopic best response function. Instead, they appear to gravitate to alternative heuristics that align the players' incentives and enable a form of punishment and forgiveness. We present some suggestive evidence on how these heuristics operate and provide elements for a new theory along with simulations that illustrate players' adjustments to ever more cooperative outcomes.

Several previous Cournot studies with low-information environments have provided empirical support for Vega-Redondo's prediction. In particular, Huck, Normann, and Oechssler [14] have documented the emergence of very competitive outcomes when players can observe each others' choices and outcomes but do not have *a priori* information about the payoff function. However, if players instead have information about the payoff function and receive feedback that makes it easy to play (myopic) best replies, the emergence of Cournot–Nash equilibria is observed. When players have both types of information average production is between Cournot–Nash and Walras. Similar results are obtained in Offerman, Potters, and Sonnemans [19] who also document that when players come very close to Walrasian outcomes they tend to shy away from it soon after — this, however, with information that enables imitation and best replying.

Bosch-Domènech and Vriend [8] also observe more competitive outcomes when players can imitate but information required for (myopic) best replies becomes less salient. However, in their 22-period design even the most competitive markets stay well below the Walrasian benchmark. In treatments where imitation is difficult (because information about others' payoffs is not directly displayed) but best replying is easy, they also observe some partially collusive markets, in particular for $n = 2$. A similar pattern, that is, the partial emergence of collusion in duopolies with information that fosters best replying, is reported in Huck, Normann, and Oechssler [17] who provide a survey on Cournot experiments.[2]

Section 2 describes the basic theory relevant to our investigation, and Section 3 lays out our laboratory procedures including the user interface as well as the treatments and matching protocol. Section 4 summarizes aggregate results. It shows that initially play becomes very competitive, consistent with Vega-Redondo's imitation model, but that eventually behavior changes and overall profits rise towards collusive levels. Section 5 analyzes individual level behavior, and finds that while subjects begin by imitating rivals with higher payoffs, they eventually find different heuristics that lead them in the opposite direction, towards greater cooperation. Although we observe clear end-game effects that demonstrate that subjects are aware of last rounds, and other evidence shows that they are aware of profitable deviations from cooperation, we find that our subjects do not understand Nash reversion. Indeed, they never learn crucial parts of the best-reply correspondence of the stage game, let alone its Nash equilibrium. Nevertheless, subjects enjoy ever-longer spells of collusive play with more effective and shorter "punishment" episodes. Section 6 considers possible theoretical explanations of our data. While existing theories capture early and late phases in our data well, they fail to explain the long middle phase where players move slowly and haltingly towards more cooperative action profiles. Drawing on the "win-continue, lose-reverse" algorithm suggested by Huck, Normann, and Oechssler [15,16]

[2] Besides these studies, we should also note very early work on posted price oligopoly by Friedman and Hoggatt [12] and Alger [1]. Some of their oligopolies lasted over 100 periods, but the results were inconclusive and hard to compare to Cournot oligopolies. See also Berninghaus, Ehrhart, and Keser [6] and the review contained therein for early experimental contributions with long horizons and (almost) continuous time.

Table 1
Static outcomes for payoff function (1).

	Duopoly			Triopoly		
	x_i	P	π_i	x_i	P	π_i
JPM	0.1	600	69	0.1	400	49
CNE	3	20	40	$2.6\bar{6}$	15	$23.\bar{3}$
PCW	6	10	10	4	10	10

we suggest a fairly simple new model that can accommodate some of the key features of the data. Section 7 summarizes our contribution and points out promising directions for future research.

Online Appendix A contains supplemental data analysis, Online Appendix B reproduces instructions to subjects, Online Appendix C collects simulation specifications and supplementary mathematical derivations.

2. Basic theory

We study a repeated Cournot game played by a fixed finite number $n \geq 2$ of strategically identical players with constant marginal cost $c \geq 0$. Each period, each player i chooses a quantity x_i in a finite interval $[x_L, x_U]$. Price P is a decreasing function of the aggregate quantity $X = \sum_{j=1}^{n} x_j$, and player i's profit that period is $\pi_i = a + (P(X) - c)x_i$, including an exogenous additive constant a that captures benefits from other activities net of fixed cost. Our experiment uses $n = 2$ or 3, the interval $[x_L, x_U] = [0.1, \frac{12}{n}]$, $a = c = 10$, and unit elastic demand with $XP(X) = 120$, so

$$\pi_i(x_i, x_{-i}) = 10 + \left(\frac{120}{\sum_j x_j} - 10\right)x_i. \tag{1}$$

Maximal quantity choice $x_i = x_U = \frac{12}{n}$ by every player i yields the minimal price $P = \frac{120}{nx_U} = 10$ equal to marginal cost. Associated minimal profits are $\pi_i^{PCW} = a + 0 = 10$ for every player. We refer to this action profile as the perfectly competitive Walrasian outcome (PCW).

At the other extreme of the action space, minimal quantity choice $x_i = x_L = 0.1$ by every player i yields the maximal price $P = \frac{120}{nx_L} = 1200/n$ and maximal total profits $n\pi_i^{JPM} = 9n + 120$. We call this profile the joint profit maximum (JPM).

The (myopic) best response of player i to $X_{-i} = \sum_{j \neq i} x_j$ is the unique solution $x_i^* = b(X_{-i}) \in [0.1, \frac{12}{n}]$ to the first-order condition

$$0 = \frac{\partial \pi_i}{\partial x_i} = \frac{120}{x_i + X_{-i}} - 10 - \frac{120x_i}{(x_i + X_{-i})^2}, \tag{2}$$

and is given by

$$b(X_{-i}) = 2\sqrt{3X_{-i}} - X_{-i}. \tag{3}$$

Imposing the relevant symmetry condition $x_i + X_{-i} = nx_i$ in (2) and solving for x_i, we obtain the Cournot–Nash equilibrium profile (CNE) as $x_i^{CNE} = \frac{12(n-1)}{n^2}$. The corresponding price is $P^{CNE} = \frac{10n}{n-1}$, and the resulting equilibrium profit for each player is $\pi_i^{CNE} = a + \frac{10}{n-1} \cdot x_i^{CNE} = 10 + \frac{120}{n^2}$. Table 1 summarizes these static predictions for the duopoly ($n = 2$) and triopoly ($n = 3$) cases.

The unit elastic demand embodied in payoff function (1) constitutes a departure from previous Cournot studies. It puts the PCW and the JPM outcomes on the boundaries of the action space and rules out losses. Given subjects' reluctance in many contexts to go to the boundaries, this choice might be problematic if we did not replicate the early emergence of very competitive outcomes close to the PCW. Conversely, given that loss aversion strongly pushes subjects away from the PCW and towards interior choices when Cournot demand is linear, the fact that our chosen profit function eliminates losses might seem problematic. The evidence also alleviates this concern — our subjects do move away from the PCW despite the absence of losses.

We employed unit elastic demand because in our setting it has several advantages relative to the usual linear demand. First, as shown in Table 1, it gives a clean separation between the three static outcomes of interest. Second, it creates a much stronger temptation to defect at the JPM as deviators can reap almost the entire monopoly profit. Finally, for $n < 6$, the payoff function is not as flat around the stage game's best response. See Online Appendix C for supporting details.

When players have little a priori information but can observe their competitors' actions and profits they may resort to imitation; in particular they might simply copy the action of the player who was most successful in the previous period. This "imitate-the-best" heuristic was introduced into the theory literature by Vega-Redondo [26]. Vega-Redondo's model also allows agents from time to time to make mistakes and choose a quantity different from the one prescribed by the imitation rule. He shows that as the error rate goes to zero, the limit of the dynamic process spends almost all time in the PCW profile.

Under imitation, the PCW profile is robust in several senses. It can be reached rapidly: if a single player ever chooses x_U, mistakenly or otherwise, she will immediately be imitated by all players thus achieving the PCW the next period, absent other mistakes. Moreover, once the PCW is achieved, single deviations will never be imitated under Vega-Redondo's [26] rule. Apesteguía, Huck, and Oechssler [3] show that PCW is also the unique stochastically stable outcome for a wide range of other imitation rules, including Schlag's [23] proportional imitation rule, and the imitate-the-best-average rule of Eshel, Samuelson, and Shaked [9].

Alós-Ferrer and Ania [2] show that stochastic stability of the PCW outcome follows also from the fact that it is a strict finite-population ESS in the sense of Schaffer [22]. That is, unilateral deviations from the PCW profile $(x_U, x_U, \ldots, x_U)$ satisfy the strict payoff inequality

$$\pi_i\big(x' | \overbrace{x_U, \ldots, x_U}^{n-1}\big) < \pi_i\big(x_U | x', \overbrace{x_U, \ldots, x_U}^{n-2}\big)$$

for all $x' \neq x_U$, i.e., the deviator earns a lower payoff than the non-deviators.

The intuition behind these stability results is simple. All firms in Cournot oligopoly face the same price, and as long as that price is above marginal cost, the most profitable firm is the one with the largest quantity. Imitation will therefore lead firms to increase quantities, driving price down to marginal cost. (Price below marginal cost is not possible with our restricted strategy space, but even if it were, the firm with the smallest quantity would then be the most profitable, and once again imitation would drive the price back towards marginal cost.) In our game the PCW is the unique profile where price equals marginal cost. At any other feasible profile, a deviation towards the PCW choice x_U will give the deviator higher profits than the non-deviators. Moreover, as just noted in the ESS discussion, any single deviation from PCW earns the deviator smaller profit than the non-deviators. Thus the PCW outcome is the only stochastically stable state, and is relatively robust to mistakes.

Vega-Redondo's theory has been highly predictive in previous experimental studies of low information Cournot oligopoly. Subjects seem attracted to the imitation heuristic and experimental markets do swiftly become very competitive and stay so for the time horizons previously explored. In the process, subjects persistently earn sub-Nash payoffs — though intuitive and sim-

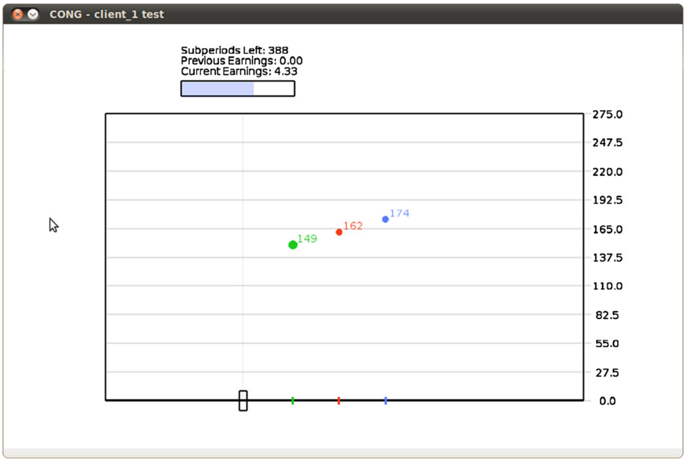

Fig. 1. Screenshot from ConG software. (For interpretation of the colors in this figure, the reader is referred to the web version of this article.)

ple, the imitation heuristic is highly destructive to earnings. The primary question we ask in the present study is whether, given enough experience, subjects will learn to abandon the imitation heuristic and thereby collectively escape from the low earnings PCW trap.

3. Laboratory procedures

The experiment used new ConG software (Pettit, Friedman, Kephart, and Oprea [20]), with the user interface illustrated in Fig. 1. Three key features allow us to run hundreds of Cournot periods in a single session.

- An intuitive graphical interface displays previous-period actions and payoffs, conveying key feedback information in a glance. Color-coded tick marks on the x-axis show each subject's previous-period quantity choice, e.g., the subject's own choice is shown in green. The y-axis measures profit, so the heights of color-matched dots show everyone's previous-period profits; exact amounts can be read from small font text next to each dot.
- Subjects make quantity choices by simply clicking on the screen, or dragging the hollow-box slider at the bottom of the screen. The set of available quantities is nearly continuous, with a granularity of less than 0.007 units over the interval $[0.1, 6]$ in the Duopoly treatment and $[0.1, 4]$ in the Triopoly treatment. A subject who wants to retain the current action into next period can do so simply by not clicking or dragging.
- Periods are time limited at four seconds. A timer bar above the quantity/profit graph fills in over the course of the period; once it is filled the period is over. During the period each subject can adjust her action as often as she likes; the payoff-relevant action is that seen when the period ends. Immediately thereafter subjects see the actions and resulting payoffs achieved in that period by themselves and their fellow oligopolists.

The four second time limit was shown in pilots to steer safely between the twin pitfalls of time pressure and boredom. Subjects did not seem hurried or frantic during game play and, in informal post-experiment interviews, expressed comfort with the pacing of the game. We believe that this comfort arose from the highly visual graphical interface, which allows fine adjustment of actions in a single click and information dissemination in a glance. The default carry-over of previous actions also allowed subjects, at very low cost, to stand still for several periods while thinking about their decisions, further reducing time pressure.

We will see in Section 4 that behavior in the first 50 periods of our experiment is very similar to that seen in previous experiments, reassuring us that our new design features do not drastically reshape behavior. The new features do, however, allow us to run 1200 periods in less than two hours.

We employed 72 subjects in six sessions of twelve subjects each at the LEEPS laboratory at the University of California, Santa Cruz in April 2011. In half of the sessions we matched subjects exclusively into duopolies and in the other half into triopolies, i.e. we ran two treatments using a completely between-subject design. Our matching algorithm grouped subjects into independent "matching groups" of six subjects each. Subjects interacted only with subjects in their own matching groups, thereby giving us six completely independent groups in each treatment. Each 1200 period session is divided into three 400 period blocks. At the beginning of each block, subjects are rematched to new counterparts in their matching group, and no subject interacts more than once with the same counterpart(s). We opted for such rematching in order to make sure that markets would not get stuck forever in some idiosyncratic ways. Rematching, we speculated, would reduce overall noise and, thus, increase the power of our experimental design (even if it means fewer perfectly independent observations).

Because our focus is on adaptation to low-information environments, we told subjects very little about their payoff functions. Using clear but non-technical language, we told them only that the functions were symmetric, time-invariant and determined uniquely by the [quantity] choices of the group members. Subjects were students from all majors and recruited online via ORSEE (Greiner [13]). Instructions, read aloud to the subjects at the beginning of the session, are reproduced in Online Appendix B. Subjects were paid their average earnings in each of the three blocks at the rate of 12 US cents per point in Duopoly and 18 cents in Triopoly. We paid an additional a show-up fee of $5. On average, sessions lasted just under two hours and subjects earned $21.00.

4. Aggregate results

We first examine the early periods of each session to see whether there are any obvious qualitative differences from earlier Cournot experiments in low-information environments. The left-hand panel of Fig. 2 plots median quantities from the Duopoly treatments while the right-hand panel does the same for Triopoly in the first 25 periods of the experiment. In Online Appendix A we plot the evolution of median profits in the same manner.

Markets become very competitive within just a few periods and settle into the competitive region between Cournot–Nash (CNE) and Walras (PCW). This is not only true for the overall medians but also for every single observed oligopoly in both treatments. There are slight differences between duopolies and triopolies with the latter being closer on average to PCW than the former. Indeed, in some triopolies price is equal to marginal cost for sustained periods of time.

To document the initial rise in quantities, note that median quantities increase from the first to the 25th period for each of the six independent matching groups ("silos") in each treatment. The increase is statistically significant at the one percent level in both duopoly and triopoly according

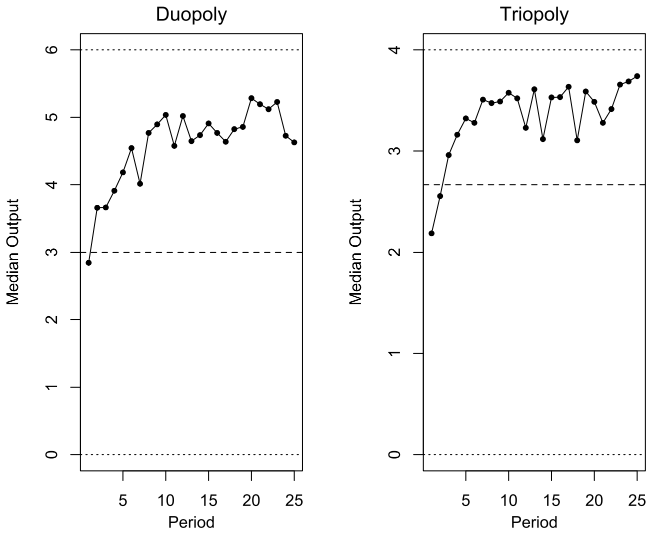

Fig. 2. Median quantities in the first 25 periods.

to a paired Wilcoxon Signed-Rank test. Over the next 25 periods, median quantities continue to fluctuate in the competitive region above CNE with little sign of systematic trend.

Thus, over the first 50 periods we see essentially the same behavior as in earlier studies. This is despite the fact that in those studies it took over an hour to run 50 periods, versus a little over 3 minutes in our experiment.

Result 1. *In both duopolies and triopolies, median output quantities initially trend upwards, and settle between CNE and PCW levels during the first 50 periods.*

The next figure is central to our study, and gives an overview of behavior in the long run. Fig. 3 plots output choice over the full 1200 periods of our experiment, with each dot representing the median quantity in a 25-period bin. The three blocks are demarcated by solid vertical lines. Analogous profit graphs can be found in Online Appendix A.

In Duopoly, there is a stark contrast between the first fifty periods (two dots) and the long-run. Highly competitive outcomes as predicted by Vega-Redondo's imitation model are only observed in those first 50 periods. After that average quantity choices start to drop sharply. Quantities continue to fall even after crossing the Cournot–Nash (CNE) level, and in periods 275–350 are much closer to full collusion (JPM) than to CNE. Of course, the median could hide some interesting heterogeneity. However, inspection of individual groups reveals that none of our matching groups spent any significant time systematically close to the CNE. (More on this below and in Online Appendix A.2.)

In the second Duopoly block, collusion becomes prevalent much more quickly; in some duopolies it is nearly perfect and remarkably stable for long intervals of time. Collusive tendencies are even more pronounced in the third block.[3]

[3] In several pairs we observe spot-on perfect collusion that is stable for long periods. This is probably aided by the fact that the JPM is on the boundary of the action space and we would expect more noise in markets with linear demand, in particular, with very fine action grids as employed here.

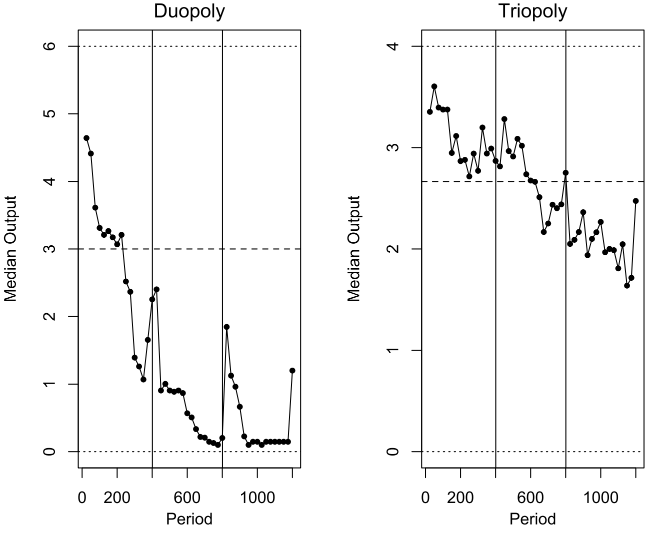

Fig. 3. Median quantities in all periods, plotted in 25 period bins.

Table 2
Median quantities, prices, and profits.

Periods	Duopoly			Triopoly		
	Quantity	Price	Profit	Quantity	Price	Profit
1–50	4.54	13.98	23.74	3.46	12.52	16.52
1–400	3.17	18.43	35.45	3.11	13.59	18.66
401–800	0.57	90.01	63.11	2.74	14.60	21.20
801–1200	0.28	107.36	68.53	2.08	18.70	26.73
1151–1200	0.40	91.30	68.51	2.03	19.44	23.03

In Triopoly, quantities again start to trend downwards after intense competition in the first 50 periods. However, the decline of quantities (and the rise of profits) is much slower than in Duopoly and never approaches full collusion on average (although there is one group of subjects that colludes perfectly in the last block). Also, heterogeneity across groups is much greater than in Duopoly, especially in the last block. Nevertheless there is a systematic trend that takes subjects deep into the collusive territory between CNE and JPM.

To document the drop in quantities, note that the median output choice falls from the first to the final block in each of the six matching groups in each treatment. For each treatment, this decrease is statistically significant at the (two-tail) five percent level according to a paired Wilcoxon Signed-Rank test.

Table 2 summarizes our aggregate results. It shows median quantities, prices, and profits for the three blocks and also for the first and last 50 periods only. An analogous table reporting means can be found in Online Appendix A.

Result 2. *After peaking in the first 50 periods, quantities in both Duopoly and Triopoly begin a long decline towards the collusive JPM level. Median quantities closely approximate JPM by the*

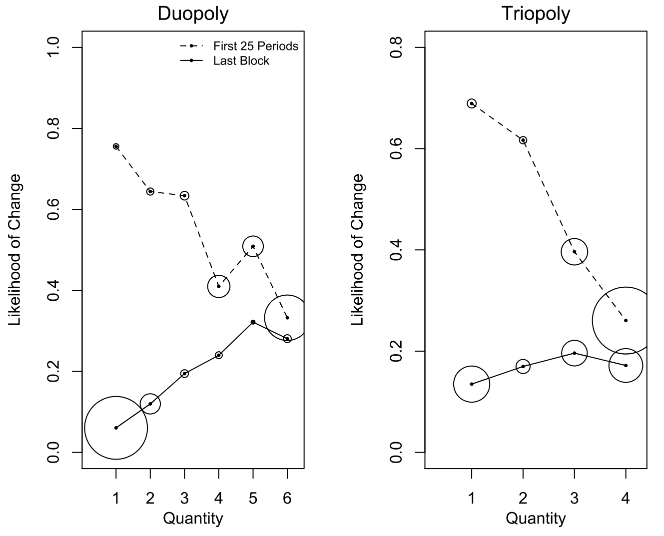

Fig. 4. Stability of quantities.

final block in Duopoly, while in Triopoly median quantities fall nearly by half, and remain well below the CNE level.

Fig. 3 also shows clear end-game effects; evidently subjects are aware of the finite nature of the game. There are also clear restart effects: after rematching, subjects take a while before reducing quantities to the cooperative levels seen in the previous block.

Fig. 4 provides an alternative perspective on aggregate behavior. It plots the likelihood that a subject adjusts quantity in period $t + 1$ as a function of her quantity choice in period t. Each point in the figure represents averages from a quantity bin $[0, 1]$, $(1, 2]$, $(3, 4]$, ... Data from the first 25 periods (dashed line) are plotted separately from data from the final block (solid line).[4] The radius of the circle around each point is proportional to the number of subjects producing that quantity in period t.

The large circles on the dashed lines show that highly competitive output choices are most persistent as well as most common in the first 25 periods in both duopolies and triopolies. This is consistent with the imitation models described earlier. But the pattern changes dramatically by the final block. In Duopoly persistence completely reverses, with collusive quantities becoming most stable and Walrasian quantities least stable. In Triopoly, where groups are more heterogeneous, the relationship becomes almost flat. Still, the change from early period behavior is striking.

Fig. 4 illustrates another important finding. The solid lines lie well beneath the dashed lines, indicating that subjects are considerably less likely to change their quantities later in the experiment. This suggests subjects approach a behavioral equilibrium with experience, particularly in Duopoly where colluding subjects rarely change their quantities.

[4] The last time bin contains strong end-game effects, so we expand the window for capturing behavior late in the session to include the entire 400 period block. The results are qualitatively the same for all plausible specifications, and indeed would be sharper if we excluded the first few time bins, which contain restart effects. See Online Appendix A for finer details.

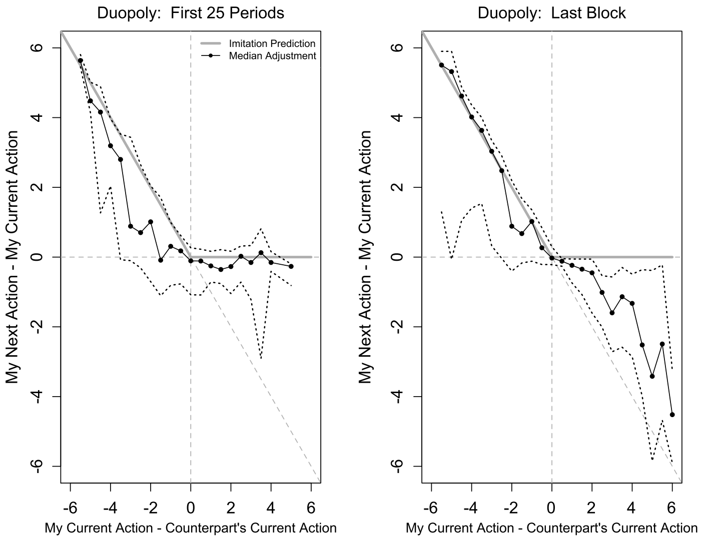

Fig. 5. Median quantities in Duopoly.

Result 3. *In early periods PCW outputs are most stable and JPM outputs least stable. By the final block, this pattern has disappeared, and has reversed in duopoly. Overall, subjects adjust less frequently by the final block.*

Thus the aggregate data suggests that subjects do learn out of the mal-adapted imitation heuristic. Indeed, it suggests that the behavioral change begins just after 50 periods, the number available to previous studies within the usual two-hour session. While earlier studies may have been tantalizingly close to detecting the behavioral change, our data also show that it may take several hundreds of periods to achieve behavioral equilibrium in a low information environment.

5. Individual behavior

What sort of individual adaptation lies behind the change in aggregate behavior? We now compare individual quantity adjustments early on to those in later play, and see whether subjects eventually employ a (myopic) best response, or some other rule, once they gain experience with the payoff function.

Figs. 5 and 6 plot adjustments from period $t-1$ to period t as a function of the difference between own quantity and counterparts' highest quantity in period $t-1$. Dots connected by solid lines show binned medians while dotted lines show the 25th and 75th percentiles, i.e., the central 50 percent confidence interval.[5] The noiseless prediction from Vega-Rodondo's imitation rule is shown in thick gray. That rule prescribes matching higher counterpart output, thus a point on the off-diagonal for negative output differences, and no change when own output is higher, thus following the positive x-axis above zero.

[5] Finer action bins are possible here than in the previous figure, because the difference data are more evenly dispersed than the level data. Bins with extremely small numbers of observations (≤ 5) are omitted from this analysis.

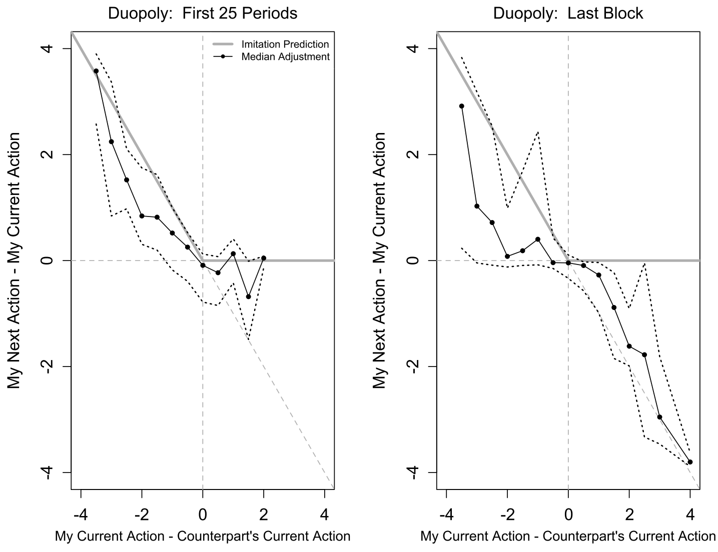

Fig. 6. Median quantities in Triopoly.

The left hand panels of the figures show that output adjustments conform strikingly well to this prediction at the beginning of the session. The prediction nearly lies within the central confidence interval and the empirical noise amplitude generally seems reasonable. We conclude that during the first 25 periods, subjects do indeed tend to "imitate the best."

The right hand panels tell a much different story. The median adjustment is now close to the off-diagonal line regardless of whether subjects have lower or higher quantities than their counterparts. Otherwise put, in the last block the predominant mode of adjustment is to match others' output choices whether or not they earned higher profits last period.

Result 4. *In early periods, subjects tend to imitate the most profitable player (including self). In later periods, by contrast, subjects tend to match counterparts' actions regardless of their profitability.*

We conclude that subjects indeed escape the destructive imitate-the-best heuristic. But do they move towards (myopic) best response? Figs. 7 and 8 plot opponents' (average) quantities in period $t-1$ on the x-axis and own quantity in period t on the y-axis. Again, connected dots show median responses and dotted lines show the 25th and 75th percentiles. The thick gray line shows the best-reply prediction and the main diagonal shows perfect imitation.

The left panels again show data from the first 25 periods. Quantities virtually never coincide with the stage game best response. This is not surprising as subjects are given no initial information about their payoff functions. Consistent with the previous graphs and with the imitation heuristic, quantities tend to roughly follow the diagonal at high quantities and exceed it at low quantities.

The right panels again show data from the final block of 400 periods. Quantities (except at the sparsely populated upper end) are tightly bunched along the diagonal, again showing that subjects tend to indiscriminately match their counterparts' quantities. Moreover, the data show

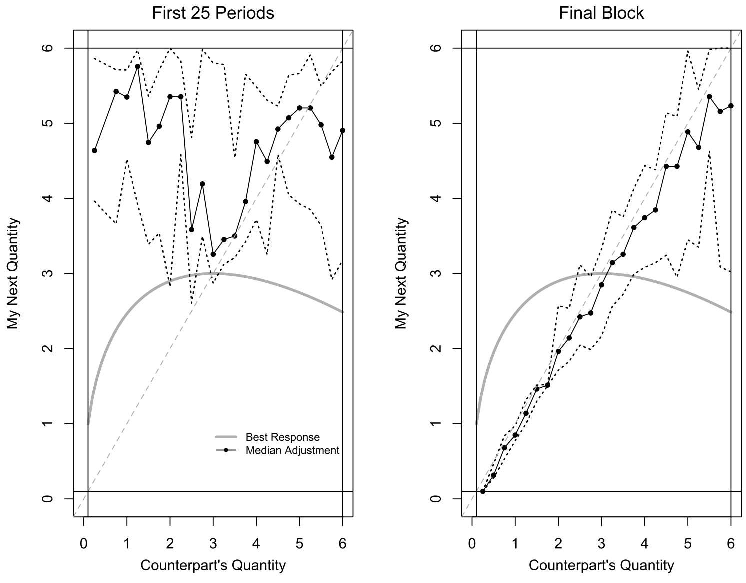

Fig. 7. Median quantities in Duopoly.

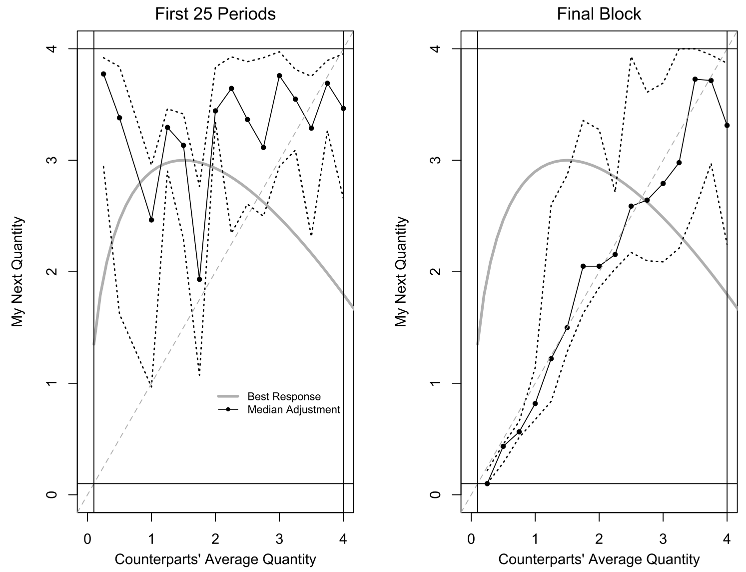

Fig. 8. Median quantities in Triopoly.

no tendency to move towards the stage best response line, except where it intersects the main diagonal.

Additional evidence comes from post-experimental questionnaires, reported in Online Appendix A.3. Fully incentivized elicitations of subjects' beliefs regarding the direction of better replies in the stage game revealed that the vast majority of subjects were aware that one could profitably deviate from the JPM. However, they never acquired systematic knowledge of the

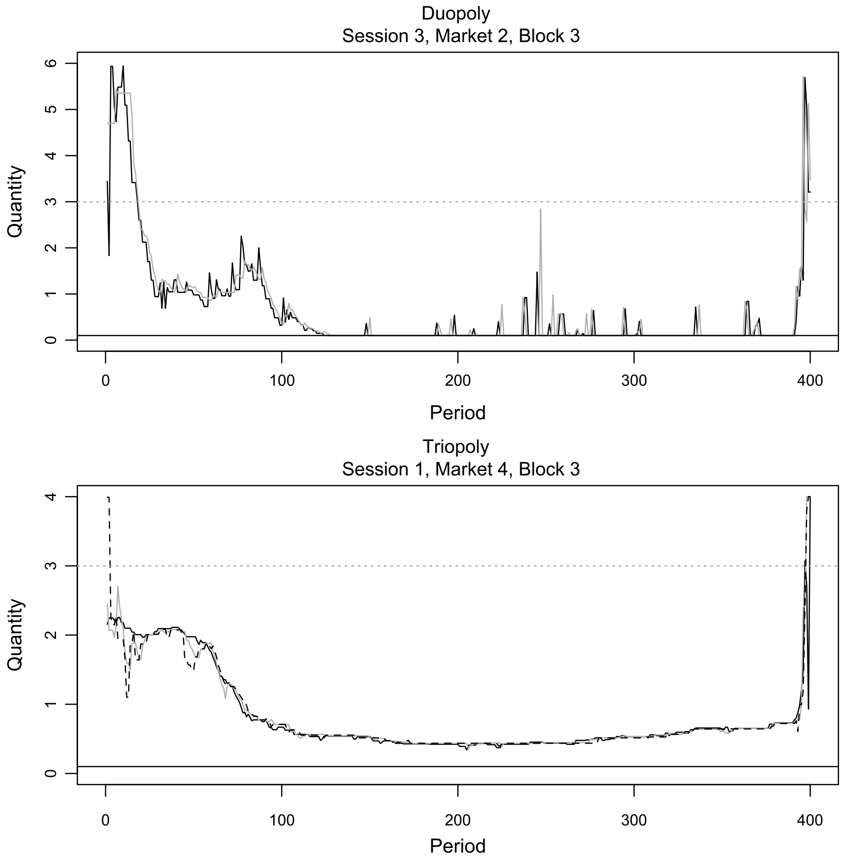

Fig. 9. Examples from the data illustrating matching behavior.

rough shape of the best-reply correspondence. For example, very few subjects realized that the best reply against the CNE profile is the CNE action. Rather they believed that higher quantities would be more profitable. Yet more evidence on the irrelevance of BR in explaining subjects' behavior can be found in Online Appendix A.2.

Result 5. *Subjects show no tendency towards following the myopic best response in either the short run or the long run.*

Granted that, after abandoning the imitate-the-best heuristic, subjects mainly adopt an unconditional strategy of matching others' choices and not an approximate myopic best response, one would then like to know the implications. How does unconditional matching play out over time?

Some insight can be gleaned from inspecting examples of the last 400 period block, as in Fig. 9. Individual subjects' output choices are plotted in solid black and gray and (in triopoly) also dotted black. In the duopoly example, after an initial flurry, subjects closely track each other on a steady decline towards collusion, but at around period 20 (that is, period 820 in aggregate plot numbering) they level out. They then test each other, remaining in the middle of the collusion zone between CNE and JPM until around period 80, at which point they resume a slow downward march with one small interruption. After about 120 periods of their duopoly, they achieve full collusion at the JPM, and remain there for most of the remaining periods. We see a sprinkling of brief episodes in which one duopolist defects for a period or two, and the other immediately retaliates, usually proportionately. Here the gains from unilateral deviations clearly are fleeting, and collusion prevails until the last few periods. Similar patterns can be seen in the triopoly example, but with less exploration of the action space. Perhaps for that reason, quantities stall out slightly above the JPM, until cooperation again fails in the last few periods. Inspecting other example blocks shows that subjects sometimes start reducing their quantities more or less simultaneously, while in other oligopolies a single subject takes the lead, reducing her own quantity to demonstrate to her counterparts that higher payoffs are available.

Online Appendix A includes "bar code" diagrams that compactly summarize behavior over time in all 48 duopoly blocks and all 36 triopoly blocks. The diagrams partition action profiles into three color-coded categories: competitive (all players' payoffs are below the CNE level), collusive (all payoffs above CNE), and other (some earn more and some earn less than in CNE). These figures show that the vast majority of deviations from collusion (to other) pass through the competitive region before returning back to collusion. This is unconditional matching at work which (as discussed below) reminds us of Tit for Tat.

The diagrams also show that the spells in "other" tend to get shorter and the collusive spells get longer. In the first block, the average collusive spell lasts 24.4 periods in duopoly and 2.8 periods in triopoly. This increases to 139.2 and 38 in the second block and, finally, reaches 174.2 and 67.2 in the last block. On the other hand, the average number of consecutive periods spent in non-collusive regions, conditional on a defection from collusion, drops in Duopoly from 68 to 56.6 to 13.8.[6] The pattern is slightly different in Triopoly where the respective figures are 86, 142, and 90.1, illustrating the increased difficulty in coordinating behavior among three players.

6. How might theory account for the data?

In this section we will first discuss how existing theories can account for behavior observed in the early and the late phases of a session, but not in the middle phases where output choices drop from the competitive into the collusive region. We then propose a new model that captures important features of observed adjustment paths. We prove a simple analytical result on the model's global convergence to the JPM, and describe simulations that exhibit patterns parallel to those seen in our data.

Our experiment was originally motivated by the conjecture that, even in a low information environment, the PCW outcome might be unstable. We thought that as players gained experience with the payoff function, they might learn the (myopic) best response and eventually converge to the CNE. Beyond this, we considered it unlikely but possible that players might adopt Nash reversion strategies and move into the collusive region.

What we actually saw was rather different. As documented in previous sections of this paper, abstracting from the idiosyncrasies of particular duopoly and triopoly groups, our data exhibit the following stages of behavior.

1. Group members initially tend to match the most profitable member, pushing the group towards the Walrasian (PCW) outcome.

[6] Besides indicating better coordination in later periods, this finding suggests to us that most of our subjects are not tired or bored even after 1000 periods.

2. After a few dozen periods in the neighborhood of PCW, group members begin to match the average quantity of other members instead of the most profitable, allowing the group to drift towards more jointly profitable profiles.
3. Most groups then begin to ratchet quantities downward while maintaining match; if prolonged, this behavior eventually leads to the joint profit maximum (JPM).
4. Many duopolies and a few triopolies spend considerable time in the vicinity of the JPM, but substantial deviations from the JPM are often seen in the last few periods.

The first stage, of course, is accounted for by well-known theories by Vega-Redondo and others, discussed in Sections 1 and 2 above. Stage 4 can be seen as an instance of an empirical regularity known since Selten and Stoecker [25] — in finitely repeated games, cooperation often disappears in the last few periods but doesn't unravel very far. One theoretical solution is proposed in the famous Gang of Four paper (Kreps, Milgrom, Roberts, and Wilson [18]) where cooperation is maintained as long as rational types have the incentive to mimic cooperative "crazy types," but that incentive eventually evaporates. Another approach generating a similar pattern is due to Radner [21] and is extended in Friedman and Oprea [10]. It shows that cooperation until the last few periods is supported by ϵ-equilibria (indeed, by ϵ-dominant cutoff strategy profiles).

The real surprise, then, lies in stages 2 and 3. Players in stage 2 clearly move away from the strategically naive imitate-the-best heuristic, but not towards the stage game's best response. They move instead to an apparently even more naive heuristic of unconditional matching, irrespective of how profitable the matched strategy was. Perhaps paradoxically, in stage 3 this behavior eventually leads to profits higher than in our most optimistic conjecture.

Our initial intuition slowly changed as we digested our data. We came to realize that, although matching others' quantities unconditionally seems simpleminded, it in fact is very effective in sustaining collusion. The process is something like tit-for-tat. A cooperative move (reducing one's own output) is profitable if it is soon matched by counterparties, i.e., if it invites cooperation on their part. A defection move (increasing output) is not very profitable if counterparties match it next period, i.e., if it provokes others to defect. Indeed, by aligning outputs unconditionally, subjects experience that better collective outcomes are available, and that incentivizes further downward adjustments.

This process seems difficult to formalize using established theory. Standard repeated game theory would tell us that in our finite horizon setting, tit-for-tat and Nash reversion strategies unravel and thus are not consistent with equilibrium. But even in an infinite horizon setting with its plethora of equilibria, there seems to be no reason to select an equilibrium with matching and ratcheting downward, as in stages 2 and 3.

Can a learning model account for what we see? Unfortunately, most learning models predict that players will not ratchet down their quantities below the CNE level. Fictitious play (and related models such as weighted fictitious play) require knowledge of the best-reply correspondence and predict convergence to the CNE. Smoothed (myopic) best response dynamics have similar predictions in our setting. Neither can reinforcement learning explain our subjects' systematic exploration of new (lower) quantities,[7] nor hybrid models like EWA. Selten and Buchta's [24] directional learning model predicts gradual adjustments towards better replies, but on

[7] In Online Appendix C.2 we show that the CNE is the only profile that survives the iterated elimination of strictly dominated strategies. As shown by Beggs [5], a consequence of this is that reinforcement learning also converges to the CNE.

the contrary, once our subjects are below the CNE in step 3, they systematically chose worse replies.

The behavior we observe in stage 3 is reminiscent of the WCLR ("win-continue, lose-reverse") learning algorithm suggested by Huck, Normann, and Oechssler [15,16]. The 2003 paper analyzes a class of dilemma games where agents move on a grid. Each agent determines the direction of the next step on the grid by examining their change in payoff. As long as the payoff increases, a WCLR agent continues to move into the same direction. Once the agent's payoff drops, the direction is reversed.

As a first step towards building a theoretical account of behavior observed in our experiment, we examine an analogous (noisy) WCLR model in a discrete version of our oligopoly setting, combining it with some form of unconditional matching of others' decisions.

Players $i = 1, \dots, n$ in a group choose their output each period $t = 1, 2, ...$ from a grid $\Gamma = \{x_L, x_L + \delta, x_L + 2\delta, \dots, x_L + \mu\delta\}$. We assume that the positive integer μ and the step size $\delta > 0$ are chosen so that Γ contains the strategies that produce the benchmark JPM, CNE, and PCW outcomes. Let x_i^t be player i's quantity, let $x^t = (x_1^t, \dots, x_n^t)$ be the profile of quantities, let $X^t = \{x \in \Gamma \mid x_i^t = x \text{ for some } i \in 1, \dots, n\}$ be the set of group members' quantities, and let $m^t = |X^t| \geq 1$ be the number of distinct quantities actually used in period t. The WCLR direction of player i at period t is $\alpha_i^t = \text{sgn}(x_i^t - x_i^{t-1})(\pi_i^t - \pi_i^{t-1})$, where π is the Cournot profit function specified earlier. Note that α_i^t can take values in the set $A = \{-1, 0, 1\}$. Finally, let $A^t = \{\alpha \in A | \alpha_i^t = \alpha \text{ for some } i \in 1, \dots, n\}$ be the set of WCLR directions present in period t.

The starting configuration in the simulations, x_i^1 and x_i^2 for each player i, are determined as independent draws from the quantity grid Γ. Subsequent output choices are specified iteratively as follows.

a. If $m^t > 1$, then x_i^{t+1} is drawn independently from the uniform distribution on X^t.
b. If $m^t = 1$, then $x_i^{t+1} = x_i^t + \tilde{\alpha}_i^t \delta$, where $\tilde{\alpha}_i^t = \alpha_i^t$ if $\alpha_i^t \neq 0$ and it is drawn independently from the uniform distribution on A^t if $\alpha_i^t = 0$.
c. With probability $\epsilon \geq 0$, the choice x_i^{t+1} specified above is replaced by $x_i^t + \alpha\delta$, where α is an independent uniform random draw from A.

That is, players match randomly if choices in the previous period are heterogeneous. On the other hand, if they are synchronized on the same choice then they use the "win-continue, lose-reverse" algorithm outlined in Huck, Normann, and Oechssler [15,16] — each player will increase her quantity by an incremental step δ if a previous quantity increase (decrease) has led to higher (lower) profits and will decrease her quantity by δ if a previous quantity decrease (increase) has led to higher (lower) profits. If there was no change in profits, she randomizes among the WCLR-steps of all agents. There is also the possibility of noise, consisting here of a one-step tremble.[8]

Proposition. *In the absence of noise ($\epsilon = 0$) the process defined above converges to the JPM with probability one from any state (x^{t-1}, x^t) such that either*

[8] Our specification of local noise follows Huck, Normann, and Oechssler [15]. Alternatively, we could have also modeled global noise by permitting mutations to any quantity in Γ. While such a process shares similar qualitative features with the process studied here we believe local noise to be more realistic in that a tremble does not usually take the whole group far from the previous output profile.

i. $m^t > 1$, *or*
ii. $m^t = 1$ *but* $m^{t-1} > 1$.

Proof. If $m^t > 1$ then by a. players will match each others' outputs. With positive probability all players except one stand still and that player copies one of the other players. In duopoly we have reached a state with $x_j = x_i$. For $n > 2$, we have reached a state with strictly fewer strategies present. By iteration, we will eventually reach a symmetric profile (a state with only one quantity played) in some period τ. Since only one player moved in the previous period we have $|A^\tau| = 2$. In period $\tau + 1$ agent i will choose $x_i^\tau + \alpha_i^\tau \delta$. With positive probability, all players who did not move sample α_i^τ and will start to move in the same direction as i. As all previous transitions occur with probability bounded away from zero, this will eventually happen with probability one. From this point on, all players will choose the same quantity in every period. They will increase (decrease) their quantity if a joint increase (decrease) resulted in higher profits and will reverse the direction otherwise. Note that at symmetric profiles with $x_i^t = x$ for $i = 1, \ldots, n$ profits are given by $10 + \frac{120}{n} - 10x$. As these are decreasing in x our process will move in small increments downwards until it hits the lower bound of the grid x_L, the JPM. This establishes the conclusion in case i.

In case ii, the state in period $t + 1$ either falls in class i or else is monomorphic, i.e., $x_i^{t+1} = z \neq x$ for all $j = 1, \ldots, n$. By the argument in the previous paragraph, in either case the state then converges to the JPM with probability one. □

In order to get a better understanding of the model we also ran simulations. With a moderate amount of noise present, the simulations are able to replicate some of the main stylized facts about our data. By construction we see a process of matching as in stage 2, and there is a clear tendency for groups to ratchet output downward as in stage 3. Crucially, we see much faster and more reliable convergence in duopoly than in triopoly.

To be more specific, we ran 100 simulations using the basic parameters of the laboratory experiment and with random starting conditions, for each chosen value of ϵ and for $n = 2, 3$. The grid Γ ranged from 0.12 to the PCW quantity with step size $\delta = 0.06$ in duopoly and $\delta = 0.04$ in triopoly, so that there were roughly 100 steps for both choices of group size n. We classified a group as convergent in period t if all group members then chose the JPM output for the first time.

With $\epsilon = 0.1$, all 100 duopoly simulations converged to JPM within 400 periods, while only 63% of the triopoly simulations did so. The median number of periods until convergence to JPM was 136.5 in duopoly, versus 266.5 periods among the convergent triopoly simulations. Results were qualitatively similar for the higher noise rate of $\epsilon = 0.15$ in that 98% (21%) of duopoly (triopoly) simulations converged, although the median number of periods to JPM among the convergent simulations was similar (about 200) for both group sizes.

We believe the differences between duopolies and triopolies arise because i) it is more difficult to initially align strategies in triopoly, and ii) if for some reason (such as a mistake or a deliberate deviation) a group is thrown off its joint decline, it takes longer to realign quantities and restart a joint downward movement in triopoly. Our simulation captures these differences. Indeed, when starting an $\epsilon = 0$ simulation from a state where all players use different quantities, in a duopoly the probability of converging in one period to a monomorphic state is $\frac{1}{4} + \frac{1}{4} = \frac{1}{2}$. In triopoly this probability is only $\frac{1}{27} + \frac{1}{27} + \frac{1}{27} = \frac{1}{9}$. Likewise, if one player has deviated (from the joint decline in quantities) the probability of recoordinating in one period is $\frac{1}{4} + \frac{1}{4} = \frac{1}{2}$ in duopoly and $\frac{1}{8} + \frac{1}{8} = \frac{1}{4}$ in triopoly.

Although WCLR – a "backwards looking," adaptive heuristic – captures the main tendencies in the third stage quite well, it does not explain everything in the data. Perhaps the most striking example is the endgame effect, visible especially toward the end of the third block of Fig. 3: subjects tend to increase their quantities in the final periods, evidently in anticipation of the end of the game. This partial unraveling of the game may suggest that subjects also have more sophisticated "forward looking" components to their strategies not fully captured by WCLR.

7. Conclusion

We believe that our study makes three fundamental contributions. First, it shows the relevance of long horizons. It sheds light on the relative importance of the amount of experienced feedback as opposed to the mere passing of time. Previously, 50 periods was generally considered sufficient to observe settled behavior. Now we see that the technical limitations of earlier software (for which implementation of longer horizons was very difficult[9]) meant that important aspects of learning in the long run were simply missed. Interestingly, time as such (providing subjects with the opportunity to analyze the game through cognition) turns out not to be the major bottleneck. Behavior in the first 50 periods of our experiment nicely mirrors behavior observed in earlier studies although in our experiment 50 periods take less than four minutes while in previous studies over an hour would have passed. In other words, multiplying the clock time for consideration by a factor of ten to twenty seems (in the case of Cournot games at least) does not to make a difference. Conversely, increasing the amount of feedback through sheer repetition changes the picture dramatically.

Of course, experiments with hundreds of periods run the risk that subjects get bored and lose focus. Fortunately, subjects in our study appear (e.g., by swift reactions to deviations from collusive play) to retain or even increase focus late in a session. Another potential issue is that four seconds for a decision seems very short. But our graphical interface allows subjects to stand still (and at worst incur minor transient losses) while processing new information, so they don't feel much time pressure. As noted in the previous paragraph, we saw similar outcomes during the first 50 periods for our 4 second periods as for long periods in previous studies. Yet it remains an open question whether long-run dynamics would be different with much longer (or perhaps even shorter) periods. The difficulty in keeping subjects in the laboratory beyond two or three hours discourages us from investigating that question.

Our second contribution is to show how additional repetitions help subjects to learn their way out of a superficially attractive but ultimately fallacious heuristic. Eventually imitation of successful others ceases to be attractive in our Cournot game. Subjects learn that they are hurting themselves and are able to overcome their initial impulse to copy what has made others relatively more successful. Escape is possible even from a devilishly baited trap.

Third, we offer a new perspective on the emergence of cooperation. Subjects replace maladapted imitation by other heuristics. Interestingly, these other heuristics are neither more complicated nor more sophisticated; they are just better suited to the repeated-game setting. Subjects learn that it is in their collective interest to produce small quantities. As we see it, they move into

[9] A notable exception is a Prisoners' Dilemma study by Bigoni, Casari, Skrzypacz, and Spagnolo [7] who manage to reduce period lengths in z-Tree to 0.16 seconds.

collusive territory through alignment of actions ("unconditional matching") and a local ("win-continue, lose-reverse") search heuristic. By matching quantities, subjects teach one another that their actions will be shadowed by others in the future, encouraging search for high collective payoffs (rather than search for individual myopic best responses). This is reminiscent of the old literature on conjectural variations (Friedman [11]). In our experiment, subjects do not merely conjecture that others will match their output adjustments; they actually experience it first hand. Consequently, they learn over time that deviations from cooperation do not pay. The ever increasing length of collusive spells in our data confirms this sort of emerging sophistication.

While we are not able to identify and estimate a precise structural model of the underlying behavior across all 1200 periods of our experiment, we point out that existing theories explain early (rather naive) and late (rather sophisticated) phases of play, and we suggest a new model that captures the essence of the long middle phase where subjects slowly move towards ever more cooperative action profiles. Matching of others' actions and a gradual slow search process are the key ingredients. We prove that such a combined process globally converges to the JPM and run simulations that generate dynamics reminiscent of the lab data. In particular, the differences between duopoly and triopoly treatments in the simulation model parallel the differences in the human subjects data.

Low information Cournot games provide one useful setting for studying the stability of heuristics,[10] but we believe that the same issues arise in a broad range of strategic settings. In larger classes of games and under different informational conditions there will, of course, be other heuristics that subjects might find initially attractive, and other heuristics to which they might eventually converge. The heuristics identified in this paper are particularly suited for symmetric low-information settings with ordered strategies. The oligopoly literature dealing with better *a priori* information about demand and cost functions, for example, has shown that myopic best replies are of immediate attraction to subjects, leading them into Nash equilibrium outcomes. Again one might ask whether, in long-horizon settings like ours, players would learn their way out of such an inefficient heuristic. In general, the set of relevant heuristics might be large and in some games it might be harder to overcome mal-adaption than in others. Studying long-run learning of heuristics in different circumstances may emerge as an attractive new agenda in experimental economics. That agenda would also open new avenues for economic theory.

Appendix A. Supplementary material

Supplementary material related to this article can be found online at http://dx.doi.org/10.1016/j.jet.2014.10.006.

References

[1] D. Alger, Laboratory tests of equilibrium predictions with disequilibrium data, Rev. Econ. Stud. 54 (1987) 105–145.
[2] C. Alós-Ferrer, A.B. Ania, The evolutionary stability of perfectly competitive behavior, Econ. Theory 26 (2005) 179–197.
[3] J. Apesteguía, S. Huck, J. Oechssler, Imitation—theory and experimental evidence, J. Econ. Theory 136 (2007) 217–235.
[4] J. Apesteguía, S. Huck, J. Oechssler, S. Weidenholzer, Imitation and the evolution of Walrasian behavior: theoretically fragile but behaviorally robust, J. Econ. Theory 145 (5) (2010) 1603–1617.

[10] Though our experiment is focused on a game designed for studying oligopoly we do not claim our results are any more (or less) relevant to oligopoly than any other dynamic strategic setting.

[5] A. Beggs, On the convergence of reinforcement learning, J. Econ. Theory 122 (1) (2005) 1–36.
[6] S.K. Berninghaus, K.-M. Ehrhart, C. Keser, Continuous-time strategy selection in linear population games, Exper. Econ. 2 (1) (1999) 41–57.
[7] M. Bigoni, M. Casari, A. Skrzypacz, G. Spagnolo, Time horizon and cooperation in continuous time, Econometrica (2014), forthcoming.
[8] A. Bosch-Domènech, N.J. Vriend, Imitation of successful behaviour in Cournot markets, Econ. J. 113 (487) (2003) 495–524.
[9] I. Eshel, L. Samuelson, A. Shaked, Altruists, egoists, and hooligans in a local interaction model, Amer. Econ. Rev. 88 (1998) 157–179.
[10] D. Friedman, R. Oprea, A continuous dilemma, Amer. Econ. Rev. 102 (1) (2012) 337–363.
[11] J.W. Friedman, Oligopoly and the Theory of Games, North Holland, Amsterdam, New York, 1977.
[12] J.W. Friedman, A.C. Hoggatt, An Experiment in Noncooperative Oligopoly, Jai Press, Greenwich, CT, 1980.
[13] B. Greiner, The Online Recruitment System ORSEE 2.0 – a guide for the organization of experiments in economics, working paper series in economics 10, University of Cologne, Department of Economics (2004).
[14] S. Huck, H.-T. Normann, J. Oechssler, Learning in Cournot oligopoly — an experiment, Econ. J. 109 (1999) C80–C95.
[15] S. Huck, H.-T. Normann, J. Oechssler, Zero-knowledge cooperation in dilemma games, J. Theor. Biol. 220 (1) (2003) 47–54.
[16] S. Huck, H.-T. Normann, J. Oechssler, Through trial and error to collusion, Int. Econ. Rev. 45 (1) (2004) 205–224.
[17] S. Huck, H.-T. Normann, J. Oechssler, Two are few and four are many: number effects in experimental oligopolies, J. Econ. Behav. Organ. 53 (4) (2004) 435–446.
[18] D.M. Kreps, P. Milgrom, J. Roberts, R. Wilson, Rational cooperation in the finitely repeated prisoners' dilemma, J. Econ. Theory 27 (2) (1982) 245–252.
[19] T. Offerman, J. Potters, J. Sonnemans, Imitation and belief learning in an oligopoly experiment, Rev. Econ. Stud. 69 (4) (2002) 973–997.
[20] J. Pettit, D. Friedman, C. Kephart, R. Oprea, Software for continuous game experiments, Exper. Econ. (2014), http://dx.doi.org/10.1007/s10683-013-9387-3, forthcoming.
[21] R. Radner, Can bounded rationality resolve the Prisoner's dilemma?, in: A. Mas-Colell, W. Hildenbrand (Eds.), Contributions to Mathematical Economics, North-Holland, Amsterdam, 1986, pp. 387–399.
[22] M. Schaffer, Evolutionarily stable strategies for a finite population and a variable contest size, J. Theor. Biol. 132 (1988) 469–478.
[23] K. Schlag, Why imitate, and if so, how? A boundedly rational approach to multi-armed bandits, J. Econ. Theory 78 (1998) 130–156.
[24] R. Selten, J. Buchta, Experimental sealed bid first price auctions with directly observed bid functions, in: I.E.I. Budescu, R. Zwick (Eds.), Games and Human Behavior: Essays in Honor of Amnon Rapoport, Lawrence Erlbaum Associates, Mahwah, NJ, 1998.
[25] R. Selten, R. Stoecker, End behavior in sequences of finite Prisoner's Dilemma supergames a learning theory approach, J. Econ. Behav. Organ. 7 (1) (1986) 47–70.
[26] F. Vega-Redondo, The evolution of Walrasian behavior, Econometrica 65 (1997) 375–384.