From Imitation to Collusion: Long-run Learning in a Low-Information Environment*

Daniel Friedman (UCSC), Steffen Huck (WZB and UCL),
Ryan Oprea (UBC), and Simon Weidenholzer (U Essex)

July 19, 2012

Abstract

We study long-run learning in an experimental Cournot game with no explicit information about the payoff function. Subjects see only the quantities and payoffs of each oligopolist after every period. In line with theoretical predictions and previous experimental findings, duopolies and triopolies both reach highly competitive levels, with price approaching marginal cost within 50 periods.

Using the new ConG software, we extend the horizon to 1,200 periods, far beyond that previously investigated. Already after 100 periods we observe a qualitative change in behavior, and quantity choices start to drop. Without pausing at the Cournot-Nash level quantities continue to drop, eventually reaching almost fully collusive levels in duopolies and often reaching deep into collusive territory for triopolies. Fitted models of individual adjustment suggest that subjects switch from imitation of the most profitable rival to other behavior that, intentionally or otherwise, facilitates collusion via effective punishment and forgiveness. Remarkably, subjects never learn the best-reply correspondence of the one-shot game. Our results suggest a new explanation for the emergence of cooperation.

Keywords: Cournot oligopoly, imitation, learning dynamics, cooperation.

JEL Codes: C73, C91, D43.

*We are grateful to the National Science Foundation for support under grant SES-0925039, James Pettit for programming the software, Luba Petersen for assistance in running the sessions, 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, Cambridge, Edinburgh, Frankfurt, Carlos III, CERGE-EI, Pompeu Fabra, Queen Mary, Stirling, UCL, Baruch College and the WZB. Weidenholzer acknowledges financial support from the Vienna Science and Technology Fund (WWTF) under project fund MA 09-017.

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 (1997) for the case of Cournot games where imitation generates the perfectly competitive Walrasian outcome.

In this paper we show how subjects, although initially attracted to imitation, learn their way out of it, and eventually acquire heuristics that not only avoid the pitfalls of extreme competition but also enable them to cooperate. The cooperation we see emerges only in the long run, beyond the horizons previously investigated. 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 best response function. Instead, they eventually gravitate to heuristics that are no more complicated than imitation, but that align incentives and enable a form of punishment and forgiveness.

We study Cournot games as they are of special interest not only to industrial organization theorists but also from a wider view. Within the broad class of aggregative games (Alós-Ferrer and Ania 2005), Cournot games are notable for a tension between social efficiency and individual optimization. (Since consumers are not considered players in such games, social efficiency refers to the players' joint payoff maximum at the cartel profile.) The efficient profile contrasts to the Cournot-Nash equilibrium, but there is an even less efficient profile of interest — the fully competitive or Walrasian outcome. 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 reported by several authors in low-information environments where players observe other players' quantity choices and profits but not the underlying payoff function. Most of these studies feature what has been considered "long horizon" repeated interaction of around 50 periods (see, for example, Huck, Normann, and Oechssler 1999, Offerman, Potters, and Sonnemans 2002, or Apesteguía, Huck, and Oechssler 2007).

Our point of departure is to examine a much longer horizon. We employ the new ConG software (Pettit, Friedman, Kephart, and Oprea 2012) 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 1,200.

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 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.

The primary contribution of the present paper is to document that transition in outcomes. The transition illustrates the importance of very long horizons in low information environments, and it also sheds light on an important general question: how can players learn to abandon dysfunctional heuristics and find different rules that better reconcile group interest with self interest?

With that general question in mind, we also examine individual behavior before and after the transition, using known models of dynamic adjustment. We show how subjects abandon the tempting but fallacious imitation heuristic and instead combine two other modes of behavior: simple matching of others' quantities, and a basic algorithm ("win-continue, lose-reverse") that responds to the profitability of the previous quantity adjustment. Together, these two heuristics move quantities towards the collusive region. They effectively make subjects more mindful of future consequences of their current actions: subjects learn that deviations from cooperation are not very profitable as they are quickly matched (an effective punishment) and that repentance brings forgiveness (as a return to cooperation is also matched).

The next section reviews static theory relevant to our experiment, and computes the three benchmark outcomes: the joint profit maximum (JPM), the Cournot Nash equilibrium (CNE), and the perfectly competitive Walrasian outcome (PCW). The section then discusses several well known heuristics intended to describe individual adaptation in low-information environments like ours. The heuristics include Vega-Redondo's imitate-the-best-max (IMIT), matching the average action played by others (MATCH) and win-continue lose-reverse (WCLR). The section also notes some standard adaptation models that require more information, such as myopic best response (BR), and "forward-looking" repeated game strategies such as Tit-for-Tat (TFT). The section concludes with a discussion of observable consequences and hypotheses that motivate our experiment.

Section 3 lays out our laboratory procedures. It describes the ConG user interface we used as well as matching procedures and treatments. Section 4 summarizes results. It shows that initially play

becomes very competitive, consistent with Vega-Redondo's IMIT, but that subjects subsequently change behavior and obtain much more efficient outcomes. Section 5 analyzes the transition in terms of the individual adaptation models.

Section 6 discusses our findings, which do not trivially vindicate standard repeated game theory. Although we observe clear end-game effects that demonstrate that subjects are aware of last rounds and are aware of (stage game) 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 do become more sophisticated over time in that they enjoy ever-longer spells of collusive play with more effective and shorter punishment phases.

Appendix A contains supplemental data analysis, Appendix B uses simulations to examine identification issues in the data, and Appendix C reproduces instructions to subjects. On-line Appendix D collects supplementary mathematical derivations.

2 Perspectives from Received 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

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

2.1 Static predictions

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 indeed maximal total profits $n\pi_i^{JPM} = 9n + 120$. We call this profile the joint profit maximum (JPM).

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

The 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

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

and is given by

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

Imposing the relevant symmetry condition $x_i + X_{-i} = n x_i$ in (2) and solving for x_i, we obtain the Cournot-Nash equilibrium profile 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.

Compared to a linear demand specification, the unit elastic demand embodied in payoff function (1) has three important advantages for experimental work. 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. Finally, for $n < 6$, the payoff function is not as flat around the best response. See on-line Appendix D for details on the limitations arising from a linear specification of the demand function.

2.2 Dynamic adjustment models

The existing theoretical literature offers a variety of predictions regarding which outcomes will emerge as players react to other players' choices and profits in Cournot games like (1). Among these, the best supported in previous low information experiments is the simple heuristic of imitating the choice of the player who earned the highest payoff among all players last period. This heuristic, first analyzed by Vega-Redondo (1997), is often referred to as "imitate-the-best-max". Below we will refer to it as IMIT. 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.

Indeed, convergence is global and rapid: from any other profile, a single player choosing x_U one period will, except perhaps for a few transitory mistakes, immediately be imitated by all players, and single deviations from the PCW profile will never be imitated under Vega-Redondo's (1997) rule. Apesteguía, Huck, and Oechssler (2007) show that PCW is also the unique stochastically stable outcome for a wide range of other imitation rules, including Schlag's (1998) proportional imitation rule, and the imitate-the-best-average rule of Eshel, Samuelson, and Shaked (1998).

Alós-Ferrer and Ania (2005) 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 (1988). That is, unilateral deviations from the PCW profile $(x_U, x_U, \ldots, x_U)$ satisfy the strict payoff inequality

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

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 be the most profitable, and once again imitation would drive the price back towards marginal cost.) In our game 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.

Though IMIT is theoretically prominent and empirically successful, many competing adjustment rules are available. One such alternative is an even simpler variety of imitation: simply match average actions independently of payoffs. There are a number of possible reasons for such behavior. A player's utility might be subject to conformism biases (Bernheim 1994), or players might use popularity weighting (Ellison and Fudenberg 1993) in their decision process. Forward looking subjects may even strategically adopt such a rule as a simple (and salient) way of committing to matching movements towards collusion and deterring deviations in the other direction. Variations on the theme include moving towards the action of a randomly chosen other player, or towards the average action among all other players. For specificity, we will refer to that last variation as MATCH, and include it in our list of heuristics.

For $n > 2$, the stationary points of MATCH, like other unconditional imitation processes, are

symmetric (monomorphic) states, where all players make the same choice. The set of all symmetric states is connected via (a chain of) single deviations. Thus, under MATCH, every symmetric state, including PCW, is stochastically stable for $n > 2$. For the $n = 2$ (duopoly) case, there are also stochastically stable periodic states, which we call "blinkers," in which the two players players swap actions every period.

The heuristics discussed so far respond to the current choices and payoffs of all players, but there are other heuristics that use even less information. Players might respond only to their own past strategies and payoffs, and ignore (or be unaware of) those of other players. For instance, they might simply adjust their actions based on the success of the last adjustment. Huck, Normann, and Oechssler (2003, 2004) propose an adaptive process that we shall call "win-continue lose reverse" (WCLR): agents keep moving in the direction of a previous quantity change if it resulted in an increase in earnings and move in the opposite direction otherwise. WCLR is related to learning direction theory, as first proposed by Selten and Buchta (1998). The distinction is that learning direction theory has players move in the direction of a better reply, which requires counterfactual knowledge. By contrast, WCLR relies only on direct personal experience in the current period plus memory of one's own action and profit in the previous period; it does not require any knowledge of the best reply function.

Huck, Normann, and Oechssler (2003, 2004) show that players following WCLR converge to the JPM profile if they move simultaneously (and to the CNE profile if they move in an alternating fashion). The intuition for JPM convergence is easy to understand in duopoly, where it prevents players from systematically increasing their separation. As opposite movements keep the price unchanged, it is impossible that the large firm increases its profit through a higher quantity if the smaller firm increases its profits through a smaller quantity. Moreover, when firms move in the same direction, it will be the leading firm (the larger firm for upward movements, the smaller firm for downward movements) that is going to reverse first. Hence, over time quantity separation shrinks. Eventually players choose the same quantities, and thereafter any (joint) movement towards the JPM outcome will continue while any movement away from it will result in lower profits and hence reverse direction. On the other hand, if the protocol calls for players to alternate moves, then this sort of entrainment cannot occur. WCLR then leads players towards their best response, eventually resulting in convergence to the CNE profile.

Cournot (1830) assumed that each player always knows enough about the payoff function to best respond to the current profile of other players' choices. Standard learning models generalize this approach mainly by allowing for strategic uncertainty. For example, fictitious play (Brown 1951,

Robinson 1951) postulates that next period's choice is a best response, not to the profile experienced this period, but rather to the time-average action profile over the current and all previous periods. Thus it requires memory as well as counterfactual knowledge of payoffs.

It is well known in games similar to ours that the CNE is the unique point in the serially undominated set, i.e., the set of strategies that survives the elimination of strictly dominated strategies. On-line Appendix D extends that result to our game. Milgrom and Roberts (1991) show that any process of adaptive learning (including fictitious play and Cournot's best response process) converges to the serially undominated set, i.e. to the CNE. Thus in our game, the CNE is a global attractor of such processes, or (in stochastic versions of the models with small noise amplitude) a neighborhood of CNE is stochastically stable. As shown by Beggs (2005), a further consequence is that reinforcement learning (Erev and Roth 1998) which, of course, requires no counterfactual knowledge of the payoff function, also converges to the CNE.

2.3 Questions and Hypotheses

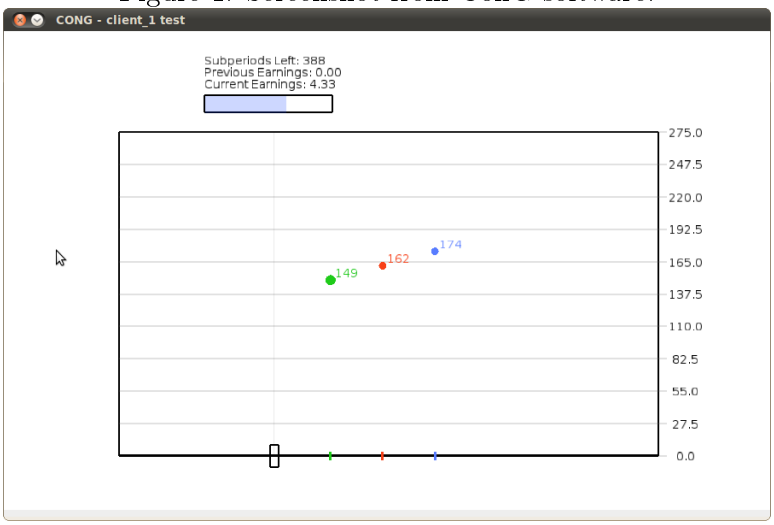
Which adjustment rule best describes actual behavior in these games? From our perspective, this may be the wrong question to ask. Though we often conceive of agents in models (and subjects in experiments) as followers of consistent adaptive rules, there is no reason in principle that this need be the case. Adaptive rules that seem natural to the inexperienced may lead to undesirable outcomes. In such cases, including ours, it seems reasonable to conjecture that moderately sophisticated subjects will eventually abandon such rules, once their pathological consequences become apparent. Such transitions have seldom been considered in the literature.

Low information Cournot games are prime examples of environments that inspire usage of destructive heuristics. The literature contains robust evidence that IMIT-like rules govern behavior in experiments lasting 40-60 periods. Increasing adherence to the IMIT heuristic results in steadily rising quantities and substantial reductions in profits. These games are therefore an ideal testbed for our conjecture.

Our central question is whether, given enough experience, subjects will learn to abandon IMIT and thereby collectively escape from the low earnings at the PCW outcome. Our primary empirical hypothesis is that, though subjects will initially trend towards Walrasian quantities as in previous studies, they will eventually move towards CNE or even JPM quantities, increasing their profits in the process.

Were this to occur, a second question the arises: what do subjects do instead, once they abandon

Figure 1: Screenshot from ConG software.



IMIT? One possibility is that they will develop a crude sense of their best response functions with experience, switch to the BR heuristic, and thus converge to the more profitable CNE profile. If subjects are more sophisticated yet, they might use their newfound knowledge of the best response function to implement repeated game strategies like tit-for-tat, leading to even higher earnings at the JPM.

On the other hand, subjects might not learn their best response function at all, switching instead to simpler heuristics. Subjects might learn to ignore their counterparts entirely, adopting WCLR, leading to eventual collusion (or to CNE). Or subjects could adopt MATCH, perhaps leading to some degree of collusion.

In order to detect these sorts of changes in heuristics, we will estimate simple econometric models on early, intermediate and late data and study whether these reveal significantly different patterns of adjustment. Appendix B provides evidence that such econometric exercises, conducted on simulations of agents following IMIT, BR, MATCH and WCLR heuristics, are capable of identifying and differentiating these behaviors to a reasonable degree. Our second main hypothesis, then, is that we will observe evidence of significantly different rules governing period-to-period adjustment in earlier versus later periods of play.

3 Laboratory procedures

In order to provide a window into long-run behavior, our subjects play Cournot games for hundreds of periods, far more than in any previous experiment on this topic of which we are aware. To make

this feasible in the span of a two-hour session, we introduced several new features using the ConG software package (Pettit et al. 2012). Figure 1 shows a screenshot that illustrates three key features:

- In order to allow subjects to instantly process information, previous-period actions and payoffs are shown to subjects via an intuitive graphical interface. The x-axis represents quantity choices and color coded tick marks show each subject's previous-period quantity choice, e.g., the subject's own choices are shown in green. The y-axis represents profits: the heights of dots (color matched to x-axis ticks) represents previous-period profits, and small font text next to the dot gives the exact amount.

- Second, all 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 subjects can adjust their actions as often as they like; the payoff-relevant actions are those seen when the period ends. Immediately thereafter subjects see the actions and payoffs achieved in that period by themselves and their fellow oligopolists.

- Third, in order to allow subjects to register decisions instantly, 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. When subjects choose not to adjust quantities they can maintain the status quo simply by not clicking on the screen at all.

Several comments are worth making about the design choices. First, the 4 second time limit was shown in extensive piloting 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. Second, this comfort is supported by a carefully constructed graphical interface designed to make information dissemination and action quicker than in standard implementations. Visualization makes subjects almost instantly aware of the actions and payoffs of all members of their group while the point-and-click interface allows subjects to register decisions in a fraction of a second. Third, subjects were not forced to make a new decision in each period as previous quantity decisions were automatically maintained each period unless changed. Thus, at very low cost, subjects could stand still for several periods while thinking about their decisions, reducing time pressure significantly. Finally, the evidence in section 4 shows that during the first 50 periods of play, our experiment yields data very similar to that seen previous experiments. This similarity reassures us that our design choices do not drastically

reshape behavior.

The new design features, however, allowed us to run 1,200 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 "silos" of six subjects each. Subjects interacted only with subjects in their own silos, thereby giving us six completely independent groups in each treatment. Each 1,200 period session is divided into three 400 period blocks. At the beginning of each block, subjects are rematched to new counterparts in their silo, and no subject interacts more than once with the same counterpart(s).

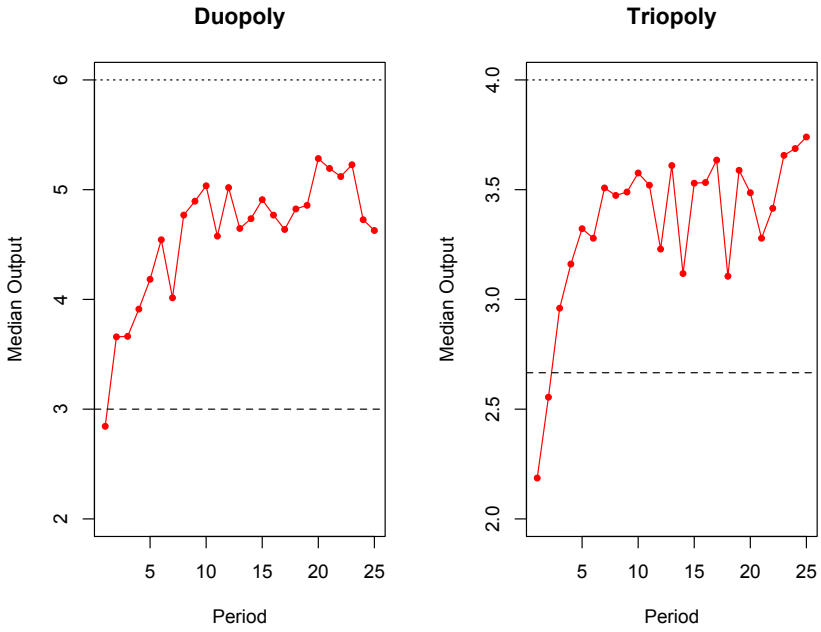
Because our focus is on adaptation to low-information environments, we told subjects very little about their payoff functions. Using clear but non-technical language, we told them only that the functions were symmetric, time-invariant and determined uniquely by the [quantity] choices of the group members. Subjects were students from all fields and recruited online via ORSEE (Greiner 2004). Instructions, read aloud to the subjects at the beginning of the session, are reproduced in Appendix C. 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 analyze the very beginning of the experiment in order to see whether there are any qualitative differences between our data and earlier Cournot experiments in low-information environments. The left-hand panel of Figure 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 Appendix A we plot the evolution of median profits in the same manner.

Markets become very competitive within just a few periods and stabilise in the competitive region between Cournot-Nash and Walras. This is not only true for the overall medians but for all market groups in both treatments. There are slight differences between duopolies and triopolies with the latter being even closer to Walras than the former. There are some triopoly markets where price is equal to marginal costs for sustained periods of time. To statistically establish the initial rise in quantities, note that median quantities rise from the first to the 20th period for each of the

Figure 2: Median quantities in the first 25 periods



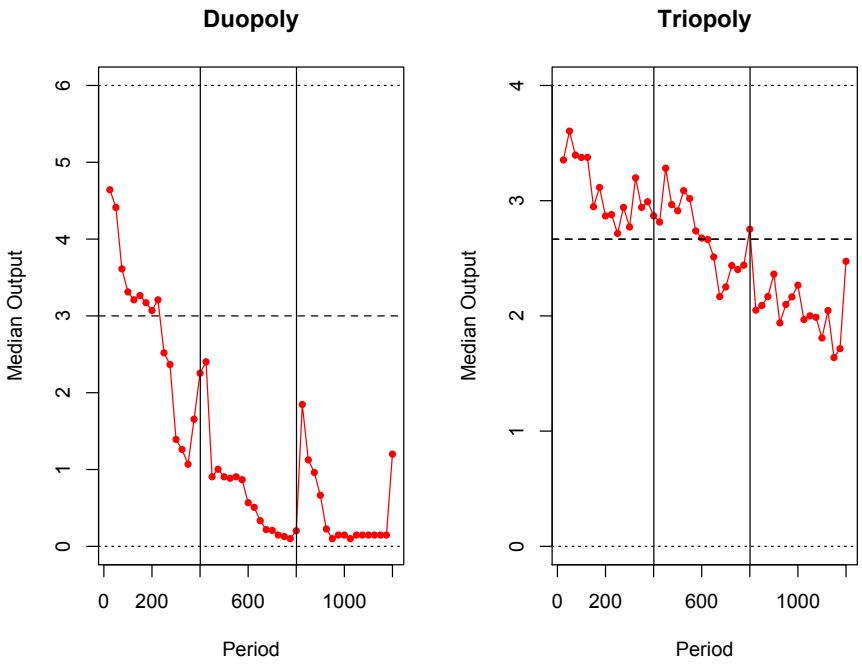
six independent matching groups in each treatment. This difference is statistically significant at the one percent level in both cases by a paired Wilcoxon Signed-Rank test. Over the next 25 periods, median quantities continue to fluctuate in the Walrasian region above Cournot-Nash. Thus, over the first 50 periods we see essentially the same trends as in earlier studies. This is despite the fact that in those studies it took over an hour to run 50 periods, versus about 3 minutes in our experiment.

Result 1 *Median action choices initially trend upwards, and within 25 periods are much closer to PCW levels than to CNE levels in both duopolies and triopolies.*

Figure 3 plots median quantities over the full 1,200 periods of our experiment. Each dot represents the median from a 25-period window. The three blocks are indicated through dotted lines. Analogous profit series can be found in Appendix A.

In Duopoly, there is a stark contrast between the first fifty periods and the long-run. Highly competitive outcomes as predicted by Vega-Redondo's IMIT are only observed in the first 50 periods. After that average quantity choices start to drop sharply. Quantities continue to fall even after crossing the Cournot-Nash level, and in periods 300-380 are much closer to full collusion than to Cournot-Nash. Of course, the median could hide some interesting heterogeneity. However,

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



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 Appendix A.2.

In the second block, collusion becomes prevalent much more quickly; in some duopolies it is nearly perfect and remarkably stable for long intervals of time. Collusion is even more pronounced in the third block.

In Triopoly, quantities again start to trend downwards after the intense competition of 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 Cournot-Nash and the joint profit maximum.

To statistically establish the secular drop in quantities, note that the median quantities fall from the first to the final block in each of the six independent matching groups in each treatment. This difference is statistically significant at the five percent level in both cases by a paired Wilcoxon Signed-Rank test.

Table 2 summarizes our aggregate results. It shows median quantities, prices, and profits for

Table 2: Median quantities, prices, and profits

	Duopoly			**Triopoly**		
Periods	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
$801-1,200$	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

the three blocks and also for the first and last 50 periods only. An analogous table reporting means can be found in Appendix A.

Result 2 *After peaking in the first 25-50 periods, quantities in both Duopoly and Triopoly begin a long decline towards the collusive JPM level. Median quantities closely approximate JPM by the final block in Duopoly, while in Triopoly median quantities fall nearly by half, and remain well below the CNE level.*

Figure 3 also shows both end-game and restart effects. Subjects collude despite being aware of the finite nature of the game. And subjects do not revert to the same level of cooperation once they have been rematched. Rather they enter with more cautious quantities and engage in an adaptive process that leads them gradually towards more cooperative outcomes.

Figure 4 reveals another aspect of the aggregate data. It plots the likelihood that a subject adjusts quantity in period $t+1$ as a function of her quantity choice in period t. A separate series is plotted using data from the first 25 periods (in red) and using data from the final block (in black). Circles around points are scaled in size to the proportion of subjects holding the corresponding quantity in period t.

Highly competitive choices are most persistent in the first 25 periods, consistent with IMIT, in both duopolies and triopolies. The pattern changes dramatically by the final block. In Duopoly persistence completely reverses, with collusive quantities becoming most persistent and Walrasian quantities least persistent. In Triopoly, where group heterogeneity is more challenging, the relationship becomes almost flat, with slightly stronger persistentce near cartel levels than near Walrasian levels. Still, the change from early period behavior is striking.

Figure 4: Stability of quantities.

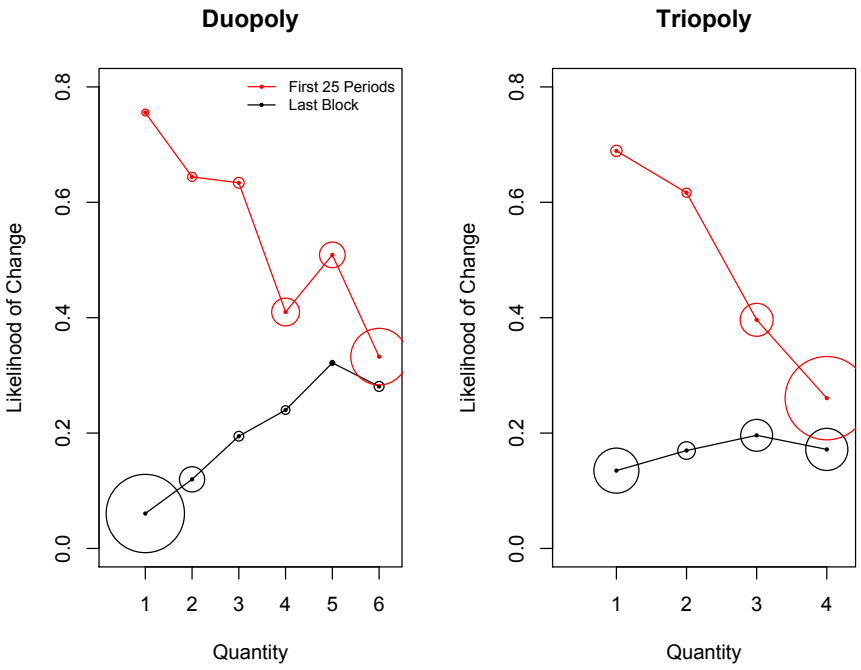


Figure 4 illustrates a second important trend in the data. Subjects are considerably less likely to change their quantities later in the experiment than earlier. This suggests subjects approach a behavioral equilibrium with experience, particularly in Duopoly where colluding subjects rarely change their quantities.

Result 3 *Quantities are less variable later in the session. Early in the session, quantities near the PCW level are the most persistent, but late in the session, quantities near JPM are the most persistent.*

5 Analyzing the dynamics

5.1 Individual adjustments in the experiment

To better understand the transition in aggregate behavior, we investigate the underlying individual heuristics, beginning with BR and MATCH. Figure 5 plots counterpart quantity in period $t-1$ on the x-axis and own quantity in period t on the y-axis. Median responses are plotted in solid red

Figure 5: Median quantities in Duopoly

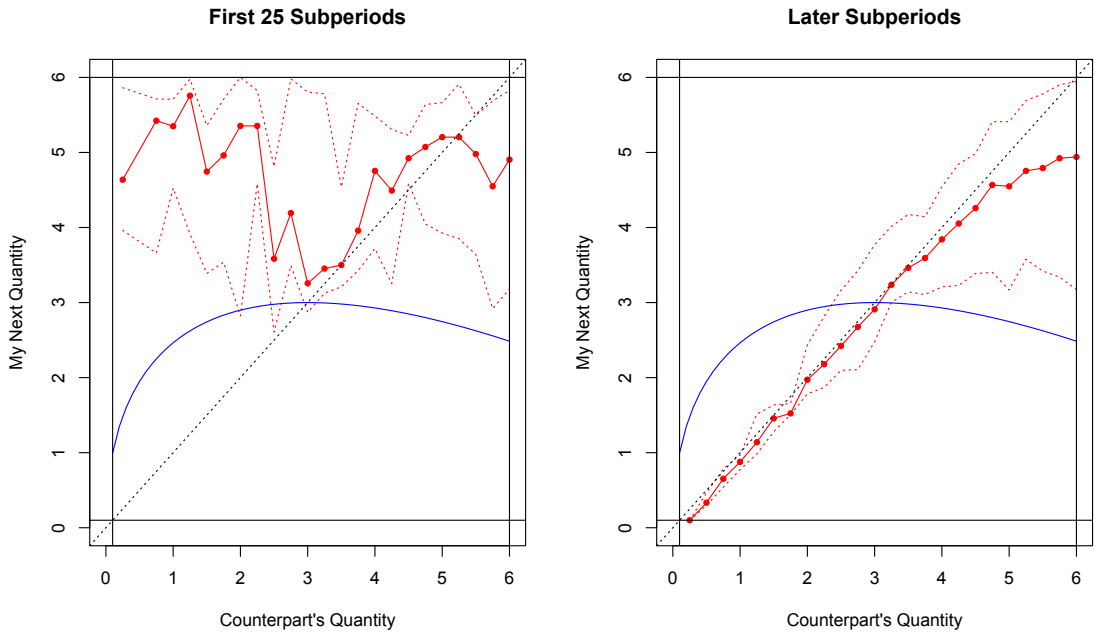
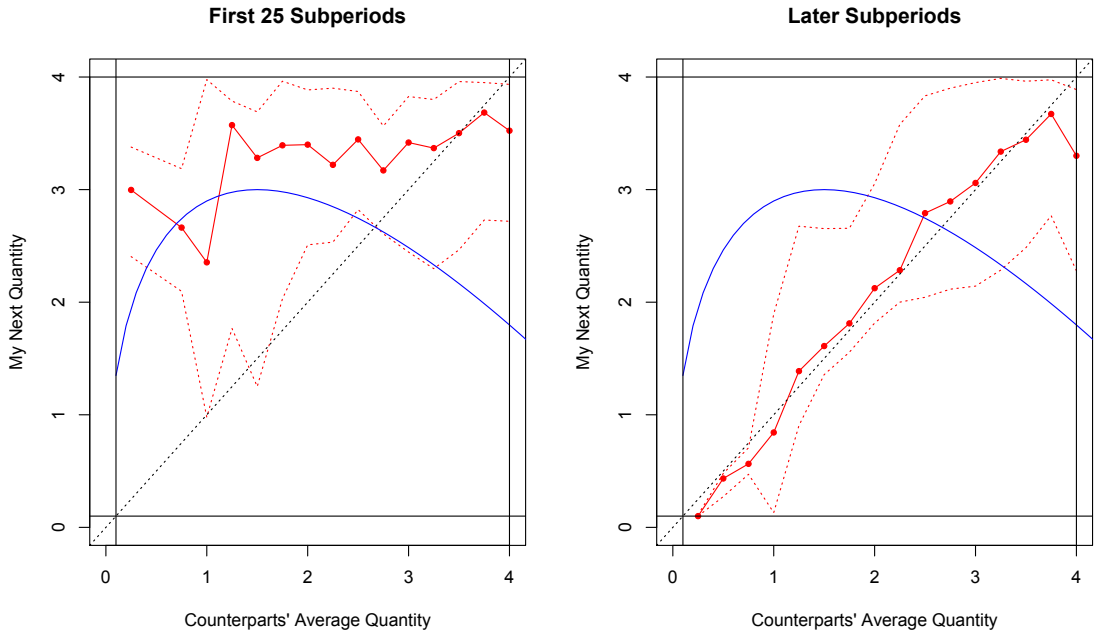


Figure 6: Median quantities in Triopoly



and 25th and 75th percentile in dotted red. For reference we plot the BR prediction in blue and the MATCH prediction as the dotted main diagonal.

The left panel shows data from the first 25 periods. Quantities virtually never coincide with best response. This is not surprising as subjects are given no initial information about their payoff functions. Instead, quantities tend to roughly follow the diagonal at high quantities and exceed it at low quantities. This is roughly consistent with IMIT, where players imitate high-quantity counterparts, and out-produce low-quantity counterparts.

The right panel aggregates all later data. Although best response fares no better in the long run, there is a clear change in behavior. Quantities (except at the sparsely populated upper end) are tightly bunched along the diagonal, consistent with MATCH.

Figure 6 provides analogous graphs for Triopoly, with average counterpart $t-1$ quantities on the x-axis. The pattern for both early and later responses are similar to (though noisier than) that observed in Duopoly.

The complete absence of BR in the adjustment data is mirrored in data from post-experimental questionnaires, reported in Appendix A.3. Fully incentivized elicitation 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 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. Appendix A.2 reports further evidence on the irrelevance of BR in explaining subjects' behavior.

Result 4 *Subjects both in the short and long run show no tendency towards best response.*

The aggregate data suggest that subjects begin some sort of behavioral transition after around 40-60 periods. To investigate, we collected all[1] individual quantity choices x_t^i and, using OLS with robust standard errors clustered at the subject level, fitted them to the following equation:

$$\begin{aligned} x_t^i - x_{t-1}^i = \alpha &+ \beta_1(\overline{x}_{t-1}^{-i} - x_{t-1}^i) + \beta_2(h_{t-1} - x_{t-1}^i) \\ &+ \beta_3 sign(x_{t-1}^i - x_{t-2}^i) sign(\pi_{t-1}^i - \pi_{t-2}^i) + \epsilon_{it} \end{aligned} \tag{4}$$

The explanatory variables correspond to heuristics defined in section 2:

[1] Coefficient estimates are qualitatively similar but larger if we examine only observations with $x_{t+1}^i \neq x_t^i$.

- The intercept α reflects a possible overall trend in the quantity chosen.
- MATCH refers to the player's adjustment towards the average action $\overline{x}_{t-1}^{-i}$ by other players, so perfect adherence to that model would yield coefficient $\beta_1 = 1$.
- Coefficient β_2 picks up the *additional weight* placed on imitation (IMIT) of the highest action $h_{t-1} = \max_{j=1,\ldots,n} x_{t-1}^{j}$, beyond its contribution to the average.
- WCLR is 1 or -1 or 0 depending on whether the most recent change in profit has the same sign as the corresponding change in quantity, or the opposite sign, or there was no change in profit or quantity. The coefficient β_3 estimates the step size.

Can statistical models of this form actually identify and differentiate underlying adjustment heuristics? Some skepticism is warranted because specifications like this raise endogeneity issues, and because behavior at the boundaries of the action space can also derail identification. To deal with such questions, we report in Appendix B estimates using simulated data. The simulations use agents programmed to follow specific heuristics including MATCH, IMIT, WCLR and BR, and have the same number of observations (with the same clustering structure) as in our experiment with human subjects. The results suggest that in practice the estimation actually can do a pretty good job of discriminating among the underlying heuristics. Most importantly, the simulation exercise suggests that we can expect to strongly identify and differentiate the MATCH and IMIT rules from one another.[2]

We estimate equation (4) on the first 50 periods of data and then separately on the final 50 periods of data to examine changes in the dynamic adjustment process. We also include estimates on the 1100 periods between these two endpoints. Under the hypothesis that subjects rely mainly on IMIT at the beginning of the session, we expect β_1 to be indistinguishable from zero and β_2 to be significantly positive. Under the hypothesis that subjects move to MATCH we expect the reverse: β_1 should be significantly positive and β_2 should drop to zero.

Results are presented in Table 3. In both Duopoly and Triopoly, we observe the same pattern: β_2 is significant in the first 50 periods, replaced by β_1 in the final 50. Between, results are intermediate but show a movement from IMIT to MATCH. Coefficients are below 1 in most cases, but this is likely due to an inevitable censoring problem facing this type of model. The subjects who are most

[2]The one heuristic that is difficult to identify in simulations is BR, the best response rule. The simulations suggest that the BR variable will artificially show up as statistically significant *regardless* of the actual rule agents follow. Happily, robust evidence from the data and from incentivized quizzes allow us to rule out BR prior to estimation. This allows us to avoid these identification issues by excluding the BR variable from our empirical specifications.

likely to imitate are also the most likely in any given period to hold quantities identical to their counterparts. These subjects contribute nothing to the estimate of these slope terms, so we expect these estimates to be systematically downward biased. This problem is attenuated in Triopoly where imitation of average quantities does not necessarily entail holding quantities identical to all other players. As a result, we see larger estimates in Triopoly than in Duopoly.

Two other patterns are worth noting. First, there is evidence that subjects follow the WCLR heuristic to some degree. This shows up in all estimates but is stronger later on in both Duopoly and Triopoly. In late-period data the term is quite robust to the window used for the estimate in Duopoly but is somewhat fragile in Triopoly. Our interpretation is that WCLR is stronger and less noisy in Duopoly than in Triopoly. Finally, there is a significant negative trend in the early-period data in both Duopoly and Triopoly. We suspect that this also arises from a boundary problem. Subjects tend to adjust their quantities frequently at the beginning of the experiment but quantities are located, on average, quite close to the Walrasian equilibrium, a boundary in the action space. This combination will tend to drive the mean adjustment below zero when subjects hold identical quantities. Indeed, regressions on simulated IMIT agents, reported in Appendix B, also show significant negative trends even though these agents are not programmed to move autonomously in either direction.

Result 5 *While subjects tend to follow IMIT in the first 50 periods, they switch to MATCH later. There is also evidence that subjects follow WCLR to some degree throughout.*

This seemingly innocuous result makes a powerful difference in evolution of behavior. By ridding themselves from the tempting but fallacious IMIT heuristic, subjects are able to avoid extreme competition and begin to increase profits. Indeed, dissatisfaction with low payoffs and an awareness of an outcome region with collectively much higher payoffs probably is the reason that subjects abandon IMIT. Inspection of many Duopoly and Triopoly matches over 400 periods suggests that in some matches this abandonment happens more or less simultaneously, while in other matches a single player takes the lead, reducing her own quantity to demonstrate to her counterparts that higher payoffs are available.

Result 5 also shows the importance of long horizons in low information environments. With previous technology 40-60 periods were considered sufficient to observe long-run behavior. We now see that that horizon coincided with the turning point, where the allure of IMIT wears off.

The MATCH rule to which our subjects switch may seem excessively simpleminded, but it turns

Table 3: Coefficient estimates from equation (4).

Duopoly				
Coefficient	**Measures**	First 50 Periods	Between	Final 50 Periods
α	Intercept	-0.35	-0.03	0
		(0.10)***	(0.01)***	(0.03)
β_1	MATCH	0.01	0.24	0.32
		(0.044)	(0.03)***	(0.043)***
β_2	IMIT	0.59	0.10	0.10
		(0.10)***	(0.04)***	(0.07)
β_3	WCLR	0.18	0.09	0.18
		(0.05)***	(0.02)***	(0.07)***
Triopoly				
Coefficient	**Measures**	First 50 Periods	Between	Final 50 Periods
α	Intercept	-0.42	-0.09	0.01
		(0.06)***	(0.02)***	(0.05)
β_1	MATCH	-0.05	0.25	0.56
		(0.05)	(0.06)***	(0.16)***
β_2	IMIT	0.66	0.14	-0.00
		(0.069)***	(0.045)***	(0.083)
β_3	WCLR	0.13	0.05	0.09
		(0.04)***	(0.01)***	(0.05)*

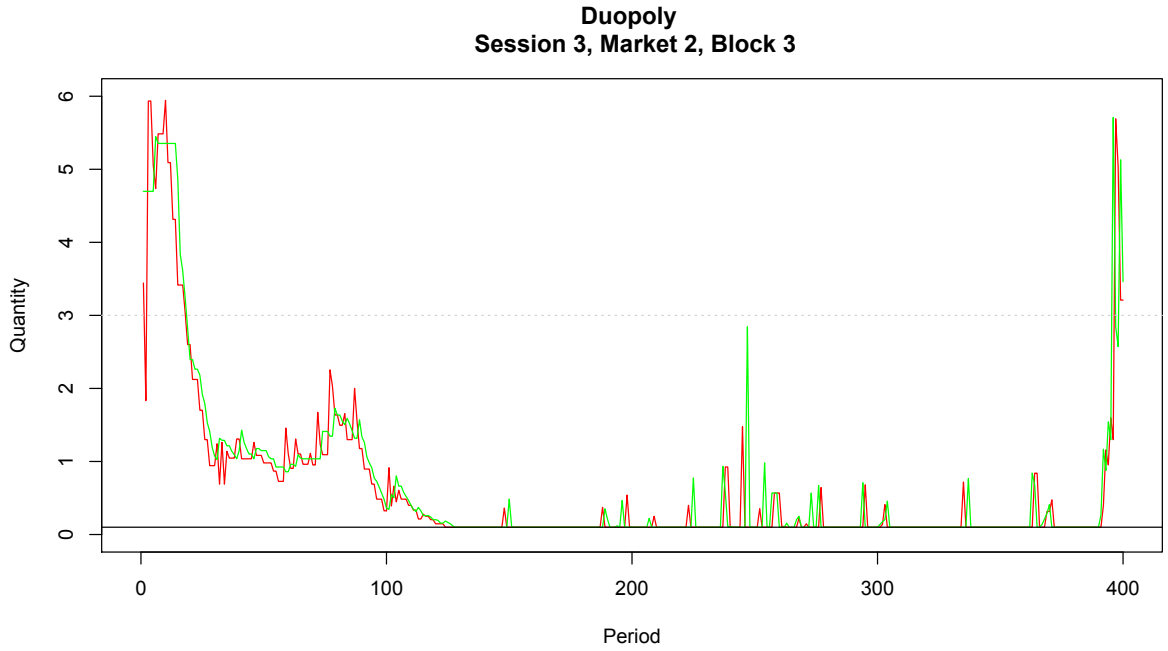
out to be remarkably effective in sustaining collusion. First, as MATCH assures that players are close together, the WCLR component helps them to transverse the quantity region from outcomes close to PCW straight through the CNE into the collusive region. Reducing quantities in parallel benefits everybody, so in WCLR everybody continues to reduce quantity. Second, once players are in the collusive region, simple MATCH also ensure an alignment of incentives that helps keep them there. Indeed, MATCH effectively justifies conjectural variations of $+1$ — the belief that others will quickly match any change in quantity — and thus players have the incentive to reduce quantity whenever it exceeds the JPM. Third, MATCH provides effective punishment for deviations, similar to Tit-for-Tat, even with no knowledge of the best response.

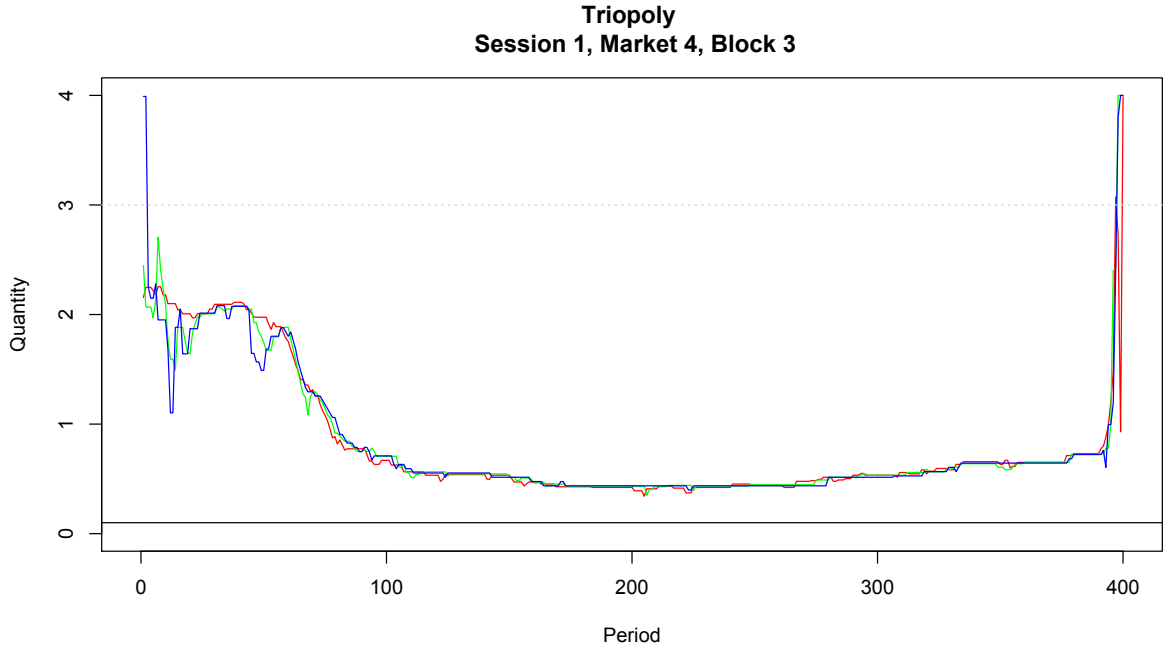
Figure 7 illustrates these points using data from particular matched groups in the final 400 period block. Individual subjects' actions are plotted in red, green and (in Triopoly) blue. In the Duopoly example, subjects closely track one another through a slow decline towards collusion. At around period 20 actions level out but after a few dozen periods of testing the earnings impact of increasing or decreasing joint quantities, the subjects begin moving decisively downwards, reaching collusion after over 100 periods of play. After achieving full collusion, the subjects occasionally defect but the MATCH rule leads to automatic punishment, reminding subjects that gains from unilateral deviations will be fleeting. Collusion collapses only in the last few periods of the block. Similar patterns emerge in Triopoly, but with less exploration of the action space and, perhaps as a result, quantities stall out above fully collusive levels.

Appendix A shows a more comprehensive view of such patterns for all of the subjects in the dataset. It contains bar-code like diagrams for which we partitioned the state space into three regions: competitive (if all players' payoffs are below the CNE payoff), collusive (if all players' payoffs are above the CNE payoff), and other (where some earn more and some earn less than the CNE payoff). Every period is represented by a single color-coded bar with red representing the competitive region, green collusive and black other. These figures show one of the more remarkable features of the data — namely how after a deviation from the collusive region occurs (that is after a change from green to black) play almost always moves into the competitive region (that is into the red) before returning back to collusive play. This is MATCH in action, providing punishment and forgiveness, very much like tit-for-tat.

The plots also clearly demonstrate that subjects do not get tired despite the large number of repetitions. On the contrary, their reaction time decreases and alertness to deviations increases, rendering play ever more efficient as time goes by. One can see that punishment phases get shorter and collusive spells get longer. In the first block, the average collusive spell lasts 24.4 periods in

Figure 7: Examples from the data illustrating MATCH behavior.





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 consecutive time spent in non-collusive regions, conditional on a defection from collusion, drops in Duopoly from 68 to 56.6 to 13.8. This pattern is slightly different in Triopoly where the respective figures are 86, 142, and 90.1, illustrating how much more complicated punishment and subsequent aligned return to collusion is with three players.

5.2 Simulating a mixed process

In order to see whether a simple mixed process consisting of IMIT, MATCH, and WCLR can indeed generate the data patterns we observe we run simulations based on (4) using the various behavioral weights from the experimental data. With probability $1-e$ the program chose the action

$$x_t^i = x_{t-1}^i + \beta_1(\overline{x}_{t-1}^{-i} - x_{t-1}^i) + \beta_2(h_{t-1} - x_{t-1}^i) + \beta_3 sign(x_{t-1}^i - x_{t-2}^i) sign(\pi_{t-1}^i - \pi_{t-2}^i).$$

With the remaining probability e the program chose an action at random according to a uniform distribution with support on the interval $[x_L, x_H]$. The first two periods were also randomly generated. We have simulated early and late behavior using the coefficients for the first and the last 50 periods taken from Table 3.

Table 4 reports the resulting medians and means of this simulation exercise along with the experimentally observed values. Overall the simulated process tracks the experimental data quite well. Note, however, that the noise rate required to replicate the experimental data is rather low in the late Duopoly simulations, as compared to the other simulations. This can be seen as evidence for more settled behavior in the Duopoly experiments, where many duopolies have managed to sustain perfect collusion in later periods. In contrast, early behavior in duopolies and triopolies and late behavior in triopolies appears to be still rather noisy.

6 Discussion and Conclusion

We believe our study makes three fundamental contributions. First, it shows the relevance of long horizons in a low information environment. Thus it sheds light on the relative importance of the amount of experienced feedback as opposed to the mere passing of time. Previously, behavior after some 50 periods was generally considered sufficient experience to observe settled behavior. Now we see that the technical limitations of earlier software (for which implementation of longer horizons was impractical) meant that important aspects of learning in the long run were simply

Table 4: Simulation results for Duopoly and Triopoly

Coefficients from first 50 periods				
	Duopoly		Triopoly	
First 50 periods	Exp. Data	Sim. Data	Exp. Data	Sim. Data
Noise e		0.1		0.1
Median	4.54	4.3	3.46	3.44
Mean	4.22	4.11	3.07	3.21
Coefficients from last 50 periods				
Last block	Exp. Data	Sim. Data	Exp. Data	Sim. Data
Noise e		0.02		0.1
Median	0.28	0.37	2.08	2.06
Mean	1.34	1.05	2.01	2.04

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 just under 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) not to make a difference. Conversely, increasing the amount of feedback or sheer repetition changes the picture dramatically.

Second, we see how additional repetitions with moderately informative feedback help subjects to learn their way out of a superficially attractive but ultimately fallacious heuristic. Eventually imitation of successful others ceases to be attractive. 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 mal-adapted imitation by other heuristics. Interestingly, these other heuristics are neither more complicated nor more obviously sophisticated: they are just better suited to the repeated-game setting. Subjects learn that it is in their collective interest to produce small quantities. They move into collusive territory through alignment of actions and a local (win-continue, lose-reverse) search heuristic. By mutually 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 best response). This is reminiscent of the old literature on conjectural variations (Friedman 1977). In

our experiment, subjects do not just have to believe that others will match their action adjustments; they actually experience it first hand. Consequently, they learn over time that deviations from cooperation do not pay and the ever increasing length of collusive spells in our data impressively confirms this emerging sophistication.

While the heuristics we model and simulate are purely backward-looking, it is clear that the improvements over the three blocks of 400 periods — the longer spells of cooperation and the shorter length of punishment cycles — point to a significant element of forward-looking behavior. However, subjects do not acquire the rationality assumed in folk theorems. In fact, they never learn to best reply (not even for the most relevant of strategy profiles). In some sense, of course, this does not matter. Subjects do not play the one-shot game; they play a repeated game. And what they learn about the repeated game is just enough for achieving collectively rational outcomes that are, from an "as-if" perspective, seductively similar to the predictions of the theory of repeated games.

References

Alós-Ferrer, C., and A. B. Ania (2005): "The Evolutionary Stability of Perfectly Competitive Behavior," *Economic Theory*, 26, 179–197.

Apesteguía, J., S. Huck, and J. Oechssler (2007): "Imitation—theory and experimental evidence," *Journal of Economic Theory*, 136, 217–235.

Beggs, A. (2005): "On the convergence of reinforcement learning," *Journal of Economic Theory*, 122(1), 1–36.

Bernheim, B. D. (1994): "A Theory of Conformity," *Journal of Political Economy*, 102(5), 841–77.

Brown, G. (1951): "Iterative Solutions of Games by Fictitious Play," in *Activity Analysis of Production and Allocation*, ed. by T. Koopmans, pp. 374–376. Wiley,New York.

Ellison, G., and D. Fudenberg (1993): "Rules of Thumb for Social Learning," *Journal of Political Economy*, 101(4), 612–43.

Erev, I., and A. E. Roth (1998): "Predicting How People Play Games: Reinforcement Learning in Experimental Games with Unique, Mixed Strategy Equilibria," *American Economic Review*, 88, 848–81.

Eshel, I., L. Samuelson, and A. Shaked (1998): "Altruists, Egoists, and Hooligans in a Local Interaction Model," *The American Economic Review*, 88, 157–179.

Friedman, J. W. (1977): *Oligopoly and the Theory of Games*. North Holland, Amsterdam, New York.

Greiner, B. (2004): "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.

Huck, S., H.-T. Normann, and J. Oechssler (1999): "Learning in Cournot Oligopoly - An Experiment," *Economic Journal*, 109, C80–C95.

——— (2003): "Zero-knowledge cooperation in dilemma games," *Journal of Theoretical Biology*, 220(1), 47–54.

——— (2004): "Through Trial and Error to Collusion," *International Economic Review*, 45(1), 205–224.

Milgrom, P., and J. Roberts (1991): "Adaptive and sophisticated learning in normal form games," *Games and Economic Behavior*, 3(1), 82–100.

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

Pettit, J., D. Friedman, C. Kephart, and R. Oprea (2012): "Continuous Game Experiments," .

Robinson, J. (1951): "An Iterative Method of Solving a Game," *The Annals of Mathematics*, 54(2), 296–301.

Schaffer, M. (1988): "Evolutionarily Stable strategies for a Finite Population and a Variable Contest Size," *Journal of Theoretical Biology*, 132, 469–478.

Schlag, K. (1998): "Why Imitate, and if so, how? A Boundedly Rational Approach to Multi-armed Bandits," *Journal of Economic Theory*, 78, 130–156.

Selten, R., and J. Buchta (1998): "Experimental Sealed Bid First Price Auctions with Directly Observed Bid Functions," in *Games and Human Behavior: Essays in Honor of Amnon Rapoport*, ed. by I. E. I. Budescu, and R. Zwick. Lawrence Erlbaum Associates, Mahwah, NJ.

Vega-Redondo, F. (1997): "The Evolution of Walrasian Behavior," *Econometrica*, 65, 375–384.

Table 5: Mean quantities, prices, and profits

	Duopoly			Triopoly		
Periods	Quantity	Price	Profit	Quantity	Price	Profit
$1-50$	4.22	17.26	27.81	3.07	13.97	19.32
$1-400$	2.95	82.58	40.53	2.80	18.84	22.05
$401-800$	1.54	259.33	54.57	2.60	33.57	23.98
$801-1,200$	1.34	286.50	56.61	2.01	74.66	29.92
$1151-1200$	1.48	276.75	55.16	2.03	85.51	29.71

A Additional Analysis

A.1 Profit Time Series

Figures 8 and 9 plot profits over time and are analogous to Figures 2 and 3. Top, middle and bottom dotted horizontal lines represent Cartel, Nash and Walrasian profit levels, respectively. The plots suggest that subjects' profits fall well below Nash levels in the first 50 periods and rise above Nash levels in the long run.

A.2 Mean quantities, prices, and profits

Table 5, in analogy to Table 2, shows mean quantities, profits, and prices each of our three blocks and for the first and last 50 periods.

A.3 Failure of Best Response Over Time

In this section, we provide evidence that subjects never in the aggregate experience a period of consistent best response. Figure 10 provides 6 panels. Each corresponds to a 1-point range of counterparts' previous period average quantity (ranges are listed above each plot). In each range the range of best responses is demarcated by dashed horizontal blue lines. Dashed horizontal red lines provide the bounds for imitating average quantity. The x-axis of each panel plots period. Data is binned into 50 period intervals and the black line plots medians. Figure 11 provides analogous data for Triopoly.

It is evident from these figures that median quantities very seldom enter the blue bounds of best

Figure 8: Median profits in early periods

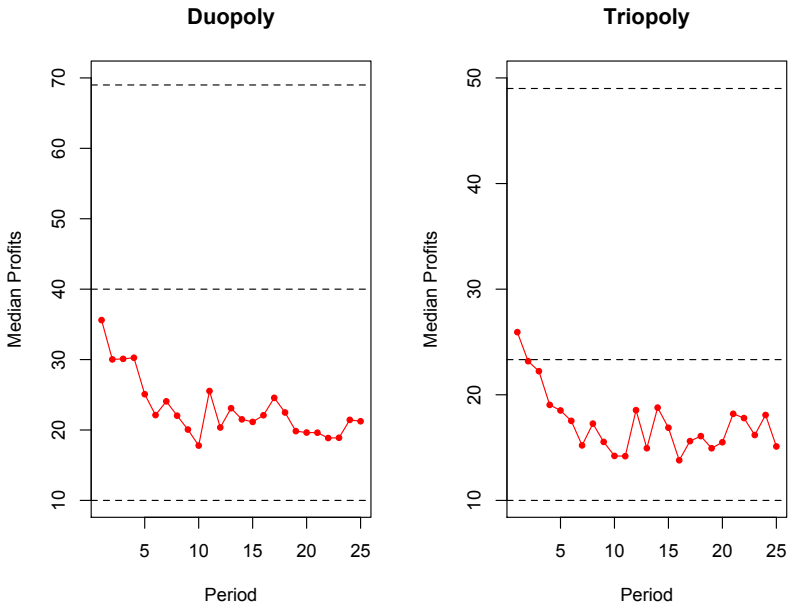
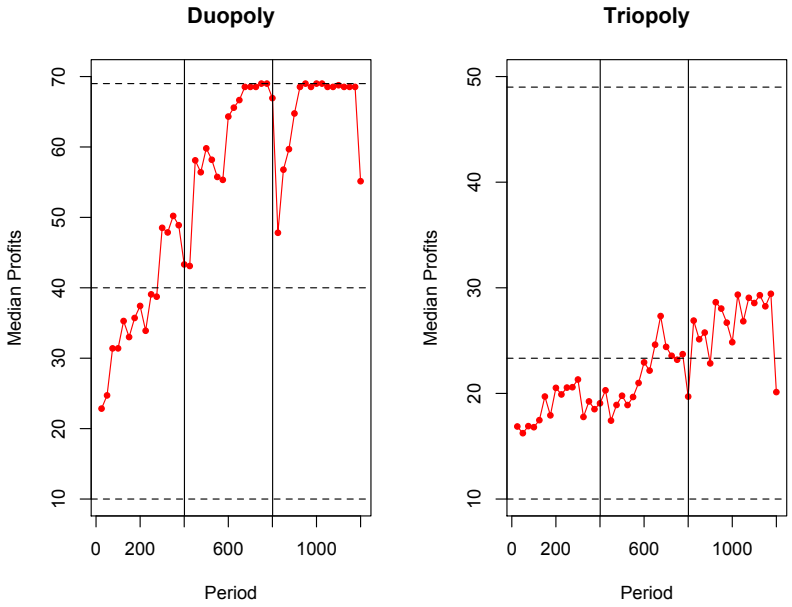


Figure 9: Median profits in all periods, plotted in 20 period bins.



Duopoly				
	Choice other	Alternatives		
Q.	x_2	x_1^1	x_1^2	correct
D1	3	1	3*	12/12
D2	3	3*	6	0/12
D3	1.15	1.15	2.31*	12/12
D4	6	2.49*	6	6/12
Triopoly				
	Choices others		Alternatives	
Q.	x_2	x_3	x_1^1	x_1^2	correct
T1	4	4	1.8*	4	4/12
T2	0.1	0.1	0.1	1.35*	9/12
T3	0.1	0.1	1.35*	4	5/12
T4	2.66	2.66	0.75	2.66*	11/12
T5	2.66	2.66	2.66*	4	0/12

Table 6: Best response quiz. Correct answers are denoted by an asterisk.

response, and that the exceptions are isolated, not bunched. The data therefore are inconsistent with subjects entering a phase of best response at the aggregate level. Instead, plotted datatend to increase from panel to panel after early periods, consistent with the MATCH heuristic.

A.4 Incentivized Quiz Results

At the end of some of the later sessions, subjects were shown printouts of screens similar to the ones used in the experiment. Markers denoted the counterparts' strategies and two slider positions indicated two possible strategies available. Subjects were asked to circle the slider that would earn the higher payoff in the one-shot game given the counterparts' strategies, and they received a cash payment of \$0.50 for each correct answer. Table 6 summarizes the questions and reports on the fraction of correct answers.

Questions D1 and T4 asked whether the CNE quantity or a lower quantity gives a higher profit against the other(s) choosing the CNE quantity. Almost everybody had this question correct, indicating that subjects are aware that downward deviations from the CNE are not profitable. D2 and T5 asked a similar question: is an upward deviation from the CNE profit increasing? Strikingly,

Figure 10: Response to counterpart actions over time in Duopoly.

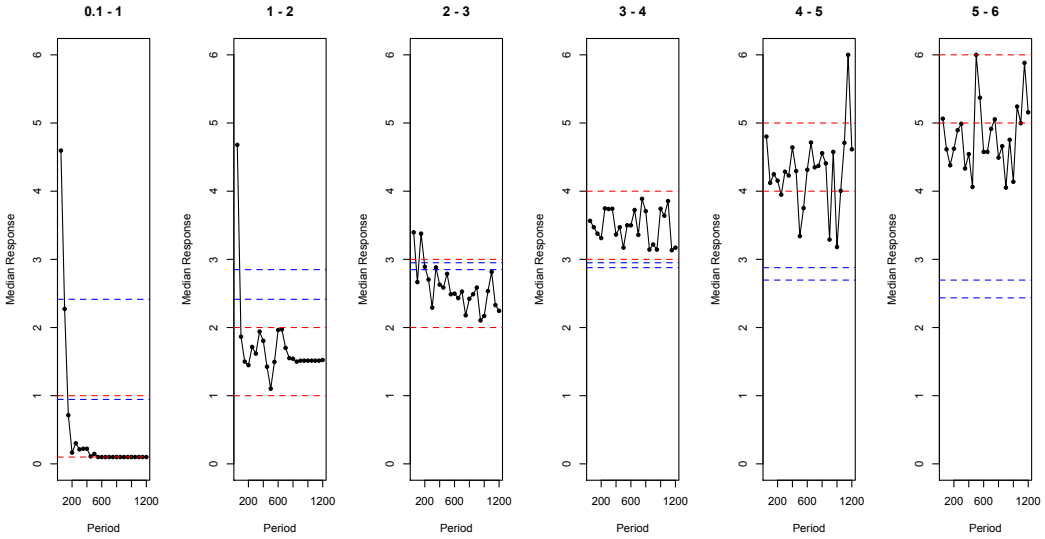
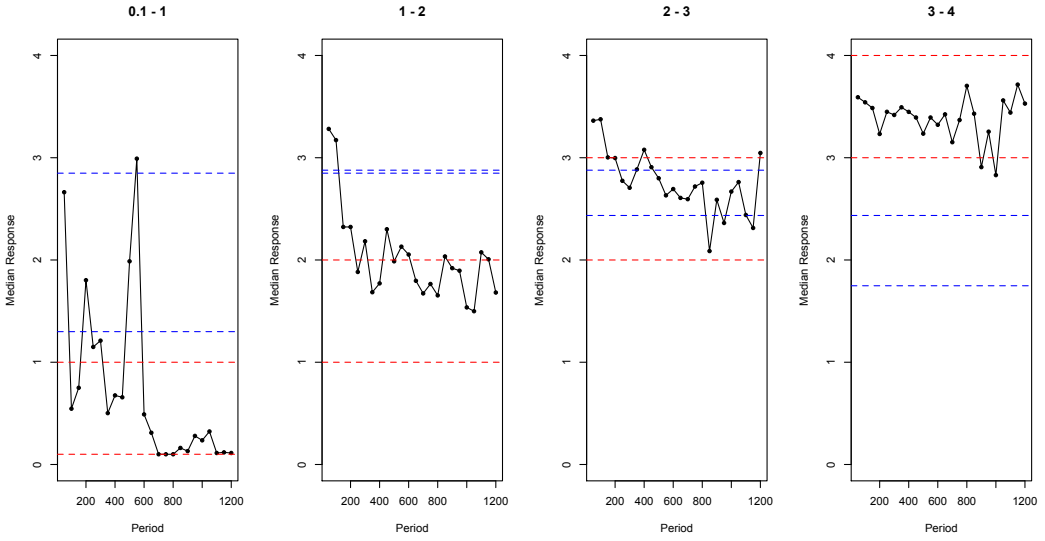


Figure 11: Response to counterpart actions over time in Triopoly.



nobody had this question correct. D4 and T1 asked whether subjects would choose a best response to the PCW-outcome or would go for the PCW outcome themselves. The message that emerges is somehow mixed: in Duopoly half of the subjects belief that the PCW-quantity earns higher profits than the best response and in Triopoly 3/4 of the subjects held this belief. D3 and T2 asked whether individual profits are higher at a (rather) collusive outcome or when deviating to a higher quantity. Everybody had this answer correct in Duopoly and 3/4 had this answer right in Triopoly. Thus, almost everybody was aware that it pays off to deviate from the collusive outcome. Finally, T3 asked whether subjects think that the PCW outcome gives a high payoff than the best response when the others collude. 7/12 subjects had this question wrong. The overall message that emerges from this exercise is that subjects at best have a rather blurred picture of the game and their optimal strategy choice.

A.5 Bar codes and Punishment

We partition the state space into three regions: competitive (if all players' payoffs are below the CNE payoff), collusive (if all players' payoffs are above the CNE payoff), and other (where some earn more and some earn less than the CNE payoff). We color-code these regions red (competitive), green (collusive) and black (other). Figures 5 and 6 plot transition probabilities over time for movements between these regions. Figures 7 to 12 show bar codes where every period is represented by a single color-coded bar indicating in which region subjects stayed in every period. These figures show one of the more remarkable features of the data — namely how, after a deviation from the collusive region occurs (that is after a change from green to black), play almost always moves into the competitive region (that is into the red) before returning back to collusive play.

Subjects' reaction speeds get faster from block to block and punishment phases get shorter and shorter in duopolies. For triopolies, we see how this process is nosier and slower, refecting the more difficult coordination problem.

The transition probabilities demonstrate several features of the data set: They show the increasing stability of collusion for both duopolies and triopolies. And they show how rare are direct transitions from collusive to competitive and vice versa. Almost all changes occur via "other", reflecting individual defections (rather than common dissatisfaction with collusive outcomes) and demonstrating that forgiveness and repentance occur subsequently rather than simultaneously.

Figure 12: Transition probabilities, Duopoly.

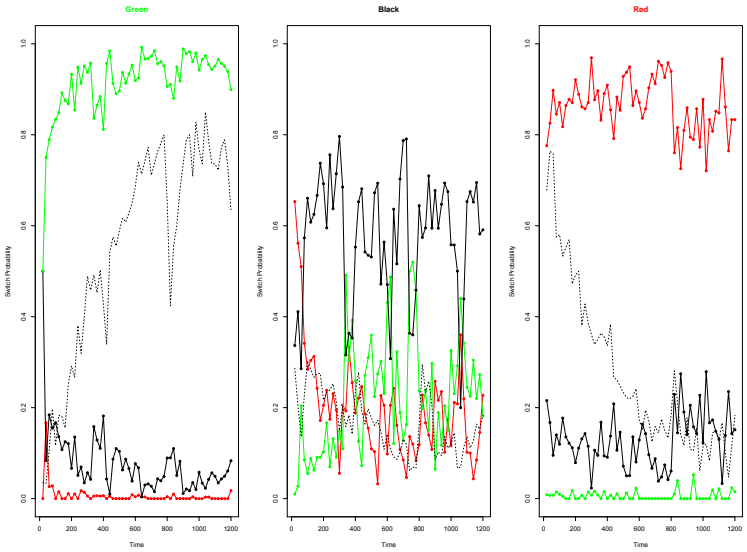


Figure 13: Transition probabilities, Triopoly.

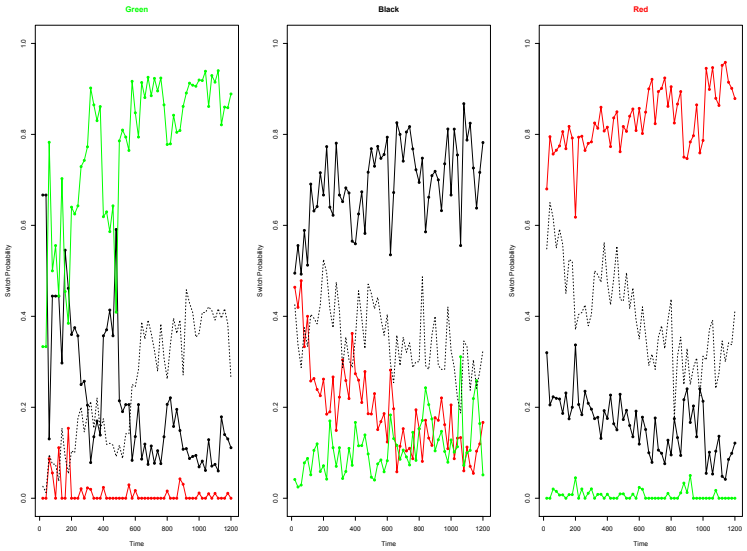


Figure 14: Bar codes from Block 1, Duopoly.

Figure 15: Bar codes from Block 2, Duopoly.

Figure 16: Bar codes from Block 3, Duopoly.

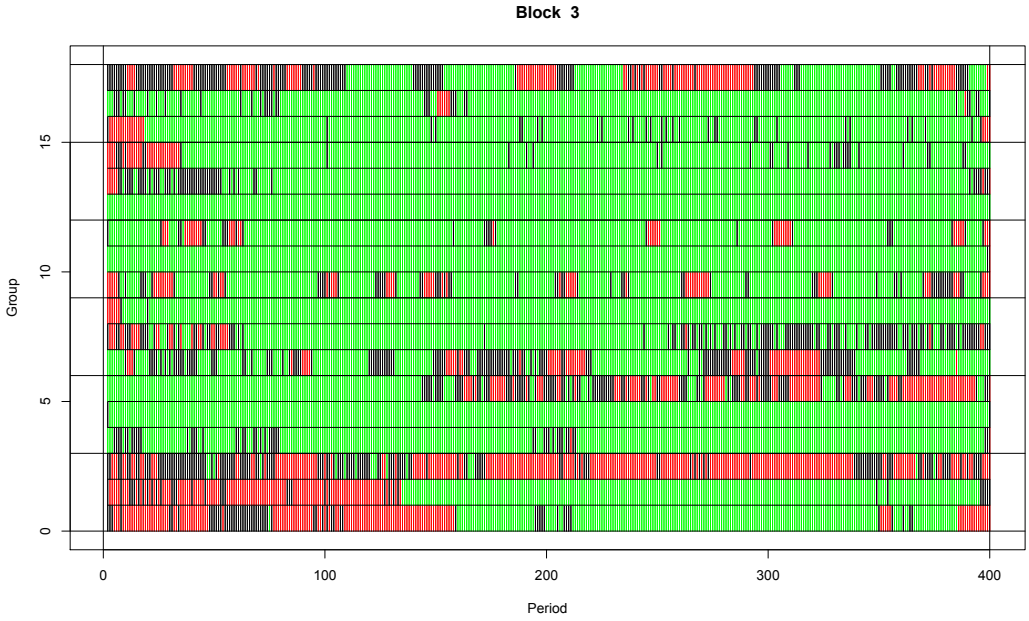


Figure 17: Bar codes from Block 1, Triopoly.

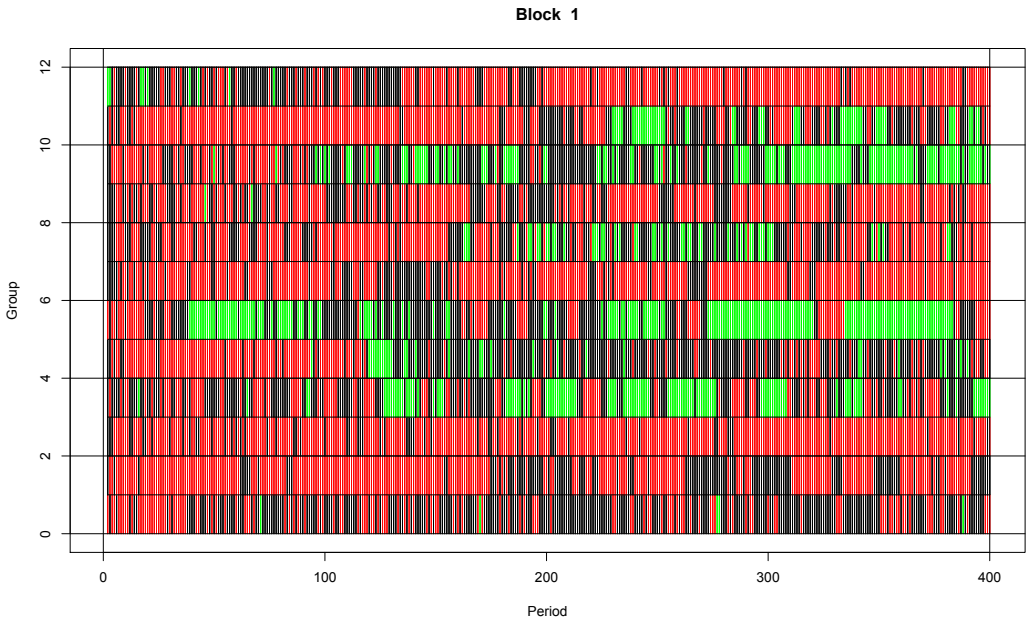


Figure 18: Bar codes from Block 2, Triopoly.

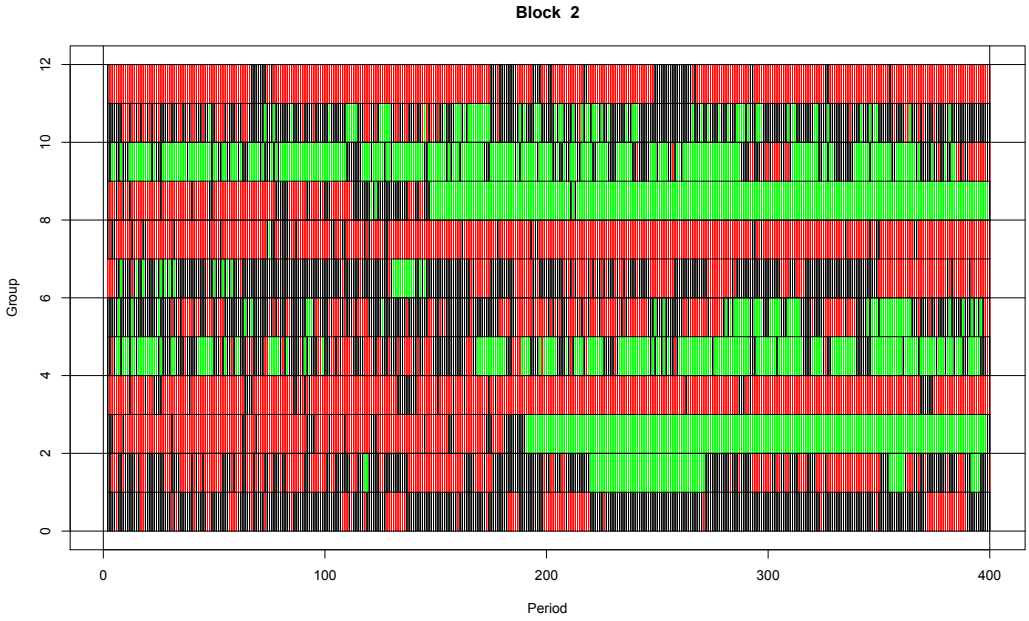
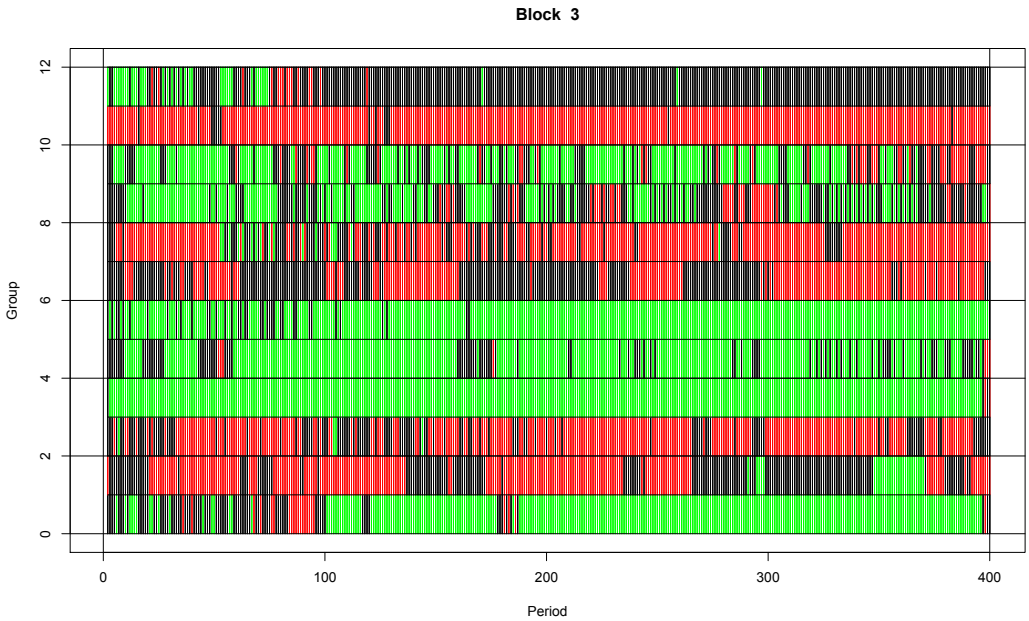


Figure 19: Bar codes from Block 3, Triopoly.



B Observable Implications of Dynamic Models

Can one actually tell the difference between, say, IMIT behavior and MATCH behavior in data like ours using standard econometric techniques as in section 5.1? Empirical models examining intertemporal adjustment are vulnerable to a host of empirical problems, particularly endogeneity, that have the potential to derail identification of underlying adjustment rules. To better understand what we can expect to infer from actual data, we simulate each of the models of interest, and then use standard econometric techniques (the same ones used in section 5.1) to try to recover the true parameters used to generate the simulated data.

For example, our IMIT simulation initializes each player's first period choice using an independent draw from the uniform distribution over $[x_L, x_U]$. With probability 0.05 each subsequent choice for that player is drawn independently from the same distribution, but with probability 0.95 is instead set equal to the highest quantity chosen by any player (including self) in the previous period. We simulate 18 sets of 400 period duopolies and 12 sets of triopolies, the same size and shape as the data set collected in the laboratory, and analyze that simulated data in the second pair of columns in Table B.

We run simulations of the same size using each of the other three adaptive rules discussed above. The first pair of columns in Table B replaces IMIT with 0.95 probability with BR (to the last period's profile) with 0.95 probability. A literal interpretation of MATCH leads to blinkers in Duopoly (players alternate out of phase between two arbitrary actions), so we apply the 0.95 probability instead to an equal mix of no-change and MATCH; results are reported in the third pair of columns. Finally, the last pair of columns applies the 0.95 probability to the WCLR algorithm with fixed step size 0.10. Of course, the simulation truncates the action at the endpoints x_L, x_U.[3]

The table reports coefficient estimates from the OLS regression (with clustered standard errors)

$$x_t^i - x_{t-1}^i = \alpha + \beta_1(\overline{x}_{t-1}^{-i} - x_{t-1}^i) + \beta_2(h_{t-1} - x_{t-1}^i) + \tag{5}$$
$$\beta_3 sign(x_{t-1}^i - x_{t-2}^i) sign(\pi_{t-1}^i - \pi_{t-2}^i) + \beta_4(b(X_{t-1}^{-i}) - x_{t-1}^i) + \epsilon_{it}$$

of quantity adjustments $x_t^i - x_{t-1}^i$ by each player i in periods $t = 2, ..., 400$. Thus the specification is identical to that in equation (4) with the inclusion of the additional explanatory variable BR, the player's adjustment towards the best response $b(X_{t-1}^{-i})$ to the other players' actions last period.

[3] Results are qualitatively similar (with obvious quantitative changes) for other values e of the Noise level than $e = 0.05$, and for other noise distributions than uniform iid, e.g., additive. It may be worth noting that the difference of two iid uniform realizations is triangular, unimodal and symmetric around 0.

	Duopoly Simulations ALL DATA							
Variable	**BR Sim**		**IMIT Sim**		**MATCH Sim**		**WCLR Sim**	
Intercept	0.01	0.02	-0.02	-0.11	0.03	0.00	0.03	0.08
	(0.003)***	(0.004)***	(0.011)	(0.007)***	(0.004)***	(0.004)	(0.008)***	(0.007)***
IMIT	-0.01	-0.32	0.97	1.000	-0.02	-0.02	-0.02	-0.06
	(0.016)	(0.025)***	(0.01)***	(0.01)***	(0.027)	(0.028)	(0.009)**	(0.008)***
MATCH	0.01	0.66	0.01	0.01	0.49	0.52	0.01	0.05
	(0.014)	(0.018)***	(0.008)	(0.008)	(0.021)***	(0.021)***	(0.006)	(0.005)***
WCLR	0.002	0.177	0.006	0.018	0.016	0.05	0.10	0.11
	(0.007)	(0.008)***	(0.021)	(0.021)	(0.031)	(0.03)*	(0.004)***	(0.004)***
BR	1.00		0.03		0.06		0.05	
	(0.012)***		(0.004)***		(0.004)***		(0.005)***	
	Triopoly Simulations							
Variable	**BR Sim**		**IMIT Sim**		**MATCH Sim**		**WCLR Sim**	
Intercept	-0.03	-0.03	-0.02	-0.09	-0.03	-0.01	-0.04	0.03
	(0.005)***	(0.01)***	(0.012)**	(0.004)***	(0.004)***	(0.004)	(0.01)***	(0.007)***
IMIT	-0.00	0.51	0.94	0.99	0.02	0.02	0.002	-0.02
	(0.016)	(0.029)***	(0.02)***	(0.016)***	(0.021)	(0.022)	(0.008)	(0.008)**
MATCH	0.00	0.19	0.03	0.01	0.45	0.48	0.01	0.05
	(0.013)	(0.02)***	(0.014)**	(0.014)	(0.018)***	(0.019)***	(0.007)	(0.006)***
WCLR	0.00	0.01	0.01	0.01	-0.00	0.01	0.09	0.10
	(0.005)	(0.01)	(0.012)	(0.012)	(0.006)	(0.006)	(0.003)***	(0.003)***
BR	0.99		0.03		0.05		0.04	
	(0.012)***		(0.007)***		(0.003)***		(0.004)***	

Table 7: Recovering the Data Generating Process. Simulations all include 0.05 noise. MATCH steps are taken with probability 0.50, and WCLR step size is 0.10. Standard errors are in parentheses. One asterisk indicates significance at the 0.05 level, two at the 0.01 level and three at the 0.001 level.

Notice that the column pairs of the table report two fits to the simulation data, one with the augmented specification just mentioned and the other with specification (4) excluding BR.

In most cases we recover reasonable approximations of the true generating process. For example, the first column reports estimates of the BR response coefficient that are quite precise and not too far from the true value of 0.95, while other coefficient estimates are insignificantly different from the true value of zero. The main discrepancy is the statistically significant (but rather small) artifactual trends implied by the intercepts. This is an artifact arising in most specifications.

Perhaps the most serious problem disclosed by the table is the consistently positive and highly significant coefficient estimates for the BR variable in simulations where the true data generating process does not involve any BR in the mix. Running additional simulations and looking at histograms discloses that these artifacts are larger when the data have large modes near an endpoint. Then random (or other) moves towards the interior are picked up by the BR variable, which always lies in the interior of the action space. This problem persists even if we estimate Tobit instead of OLS models, suggesting that the BR variable is robustly difficult to properly identify using adjustment models of this sort. Note that simply dropping the problematic BR variable (as we do in the second specification for each model) leads to wildly artificial parameter estimates when BR is the true data generating process.

As it turns out, these issues are largely moot. As we document in the body of the paper, we find overwhelming evidence both from non-parametric analysis of the data and from diagnostic incentivized quizzes that subjects do not employ BR. We therefore will drop the BR variable and employ our second specification in our main empirical analysis.

A final issue highlighted by the simulations is that WCLR generates statistically significant results for every coefficient in the model. However, notice that the step size of the WCLR process is recovered with almost perfect accuracy (0.10) whereas all other coefficients are very small (in particular in comparison to the simulations that are based on them) and the IMIT coefficient is even negative.

Although this simulation exercise shows that we can get reasonable parameter recovery from this empirical approach, it is most useful for demonstrating that we can meaningfully distinguish adaptive rules from another to a great degree. Importantly:

- IMIT and MATCH can be clearly distinguished from one another; IMIT simulations do not generate significant MATCH coefficients and vice versa.

- Unlike IMIT and MATCH, WCLR generates significant variables for both IMIT and MATCH. Thus pure WCLR can be distinguished from either of these two adjustment rules.

These results give us confidence that our straightforward econometric techniques can distinguish among the relevant adjustment processes in our experimental data.

C Instructions

These are the instructions used in both Duopoly and Triopoly sessions. In the instructions we used the term "period" to refer to what the paper calls "blocks" and "subperiods" to refer to what the paper calls "periods."

Instructions

Welcome! This is an economics experiment. If you pay close attention to these instructions, you can earn a significant sum of money, which will be paid to you in cash at the end of the last period. Please remain silent and do not look at other participants' screens. If you have any questions, or need assistance of any kind, please raise your hand and we will come to you. If you disrupt the experiment by talking, laughing, etc., you may be asked to leave and may not be paid. We expect and appreciate your cooperation today.

The Basic Idea

The experiment will be divided into a number of periods and in each period you will be anonymously matched with one or two other players via the computer. Each period will be further divided into a number of subperiods. In each subperiod you and your counterparts will secretly select strategies and at the end of the subperiod the combination of your and your counterparts' strategies will determine your earnings for the subperiod.

We will not tell you exactly how earnings are determined but here are a few facts:

- Your earnings in each subperiod depend entirely on your strategy and your counterparts' strategies, and nothing else.

- The function that determines your earnings will not change over the course of the experiment. That is, if you and your counterparts use the same strategies at time A as at time B, you both will all have the same earnings at time A as at time B.
- Your earnings are symmetric with your counterparts'. In particular, if you and your counterparts all choose the same strategy, then you all will earn the same amount.

The screen display

Figure 1 [*identical to Figure 1 in the paper*] shows the computer display you will use to make decisions and interact with your counterpart. At the top of the screen is a bar showing elapsed time in the current subperiod. When the bar fills up the subperiod is over and a new subperiod will immediately begin. Your strategy is the location (from left to right) of the black square slider at the bottom of the screen. During each subperiod you can freely adjust your tentative strategy by clicking on the screen or dragging the slider. Your actual strategy for the subperiod is the location of your slider **at the end** of the subperiod.

When the subperiod is over you will be shown a **green dot** visualizing your payoff rate from that subperiod. The higher the dot, the higher the payoff earned. The precise payoff number is shown floating next to the dot. You will also be shown **blue and red hash marks** at the bottom of the screen showing the location of your counterparts' strategies in the last subperiod and **blue and red dots** representing your counterparts' payoffs from the subperiod that just ended (if you are matched with only one other participant you will only see blue hash marks and dots).

It is important to keep in mind that your counterparts' strategies, your payoff dot and your counterparts' payoff dots always display **outcomes from last subperiod**. You will not learn payoffs or your counterpart's strategy from the current subperiod until after the subperiod is over.

Earnings

Your earnings will be given in points. Point totals reported after each subperiod are given as payoff **rates**, i.e., the payoff you would receive for the entire period if you acted the same way each subperiod. Your actual point earnings for a single subperiod can be calculated by dividing the payoff number reported by the number of subperiods in the current period. For example, if the period contains 50 subperiods and your payoff dot shows earnings rate of 200 in the last subperiod, then you actually earned $200/50 = 4$ points in that subperiod.

Table 8: Static outcomes for the linear payoff function

	Duopoly			Triopoly		
	x_i	P	π_i	x_i	P	π_i
JPM	3	6	28	2	6	22
CNE	4	4	26	3	3	19
PCW	6	0	10	4	0	10

Your points will accumulate over the course of the experiment. The screen will always display your "Current Earnings" during the period so far and "Previous Earnings" accumulated over previous periods. You will be paid cash for points earned at a rate written on the white board at the front of the room.

Frequently asked questions

Q1. Is this some kind of psychological experiment with an agenda you haven't told us?

Answer. No. It is an economics experiment. If we do anything deceptive or don't pay you cash as described then you can complain to the campus Human Subjects Committee and we will be in serious trouble. These instructions are meant to clarify the game and show you how you earn money; our interest is simply in seeing how people make decisions.

D On-line Appendix

D.1 Comparison to linear demand

To document the comparison to linear demand consider the inverse demand function $P = 12 - n\bar{x}$. We summarize the relevant benchmarks for this case in Table 8.

Under our unit elastic demand function, switching to the best response to the JPM-quantity of the other player yields an increase of profits by 58.9% in Duopoly. In Triopoly this temptation is even higher, as the best response to the JPM quantities increases profits by 106.2% Note that the temptations to deviate from the JPM-outcome are much lower in the corresponding linear demand case where a deviator can expect only a 8% rise in profits in Duopoly and a 18.2% increase in Triopoly.

To see that for the unit elastic demand function the payoff function is not as flat around the

best response as in the case of a linear demand function for $n < 6$ note the following. Under linear demand the FOC is $0 = \frac{d\phi_i}{dx_i} = 12-(n-1)\bar{x}_{-i}-2x_i$ and payoff curvature is determined by $\frac{d^2\phi_i}{dx_i^2} = -2$. By contrast, for our constant elasticity specification, FOC is $0 = \frac{d\pi_i}{dx_i} = \frac{120}{\sum_j x_j} - 10 - \frac{120x_i}{(\sum_j x_j)^2}$, and payoff curvature is determined by $\frac{d^2\pi_i}{dx_i^2} = \frac{-240}{(\sum_j x_j)^2} + \frac{240x_i}{(\sum_j x_j)^3}$. Substituting for the last term from the FOC and simplifying yields $\frac{d^2\pi_i}{dx_i^2} = \frac{-20}{nx^*}$, where the symmetric NE quantity is $x^* = 12\frac{n-1}{n^2}$. Hence for $n = 6$ we have $\frac{d^2\pi_i}{dx_i^2} = \frac{-20}{(12)\frac{5}{6}} = -2$, the same as for $\frac{d^2\phi_i}{dx_i^2}$, but for lesser n we have $\frac{d^2\pi_i}{dx_i^2} < -2 = \frac{d^2\phi_i}{dx_i^2}$.

D.2 Iterated Elimination of Strictly Dominated Strategies

To show that the CNE is the unique point in the serially undominated set, let us first consider the derivative of the profit function. If this derivative is positive a higher quantity will lead to higher profits and if it is negative decreasing one's quantity is profit increasing. We have

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

We have $\frac{\partial \pi_i(x_i, X_{-i})}{\partial x_i} > 0$ if $0 < x_i < 3$ and

$$\underline{X}_{-i}(x_i) < X_{-i} < \bar{X}_{-i}(x_i) \tag{6}$$

where $\underline{X}_{-i}(x_i) = 6-2\sqrt{3}\sqrt{3-x_i}-x_i$ and $\bar{X}_{-i}(x_i) = 6+2\sqrt{3}\sqrt{3-x_i}-x_i$. Note that (6) represents the set of quantities of the other players for which a quantity increase pays off. Likewise, we have $\frac{\partial \pi_i(x_i, X_{-i})}{\partial x_i} < 0$ if

$$x_i > 3 \tag{7}$$

or if $0 < x_i \leq 3$ and

$$X_{-i} > \bar{X}_{-i}(x_i) \tag{8}$$

The previous two inequalities capture cases where, depending on the own quantity and the quantity chosen by the others, a quantity decrease results in higher profits.

Duopoly: Consider an interval of the form $[x_L, \hat{x}_U]$. Note that by (7) we know that a slight quantity decrease will earn strictly higher profits (regardless of the quantity X_{-i} chosen by the other) if $x_i > 3$. Since at an upper bound no quantity increase is possible all upper bounds $3 < \hat{x}_U \leq 6$ are strictly dominated by a lower quantity. Iteratively applying this argument, starting from $x_U = 6$, shows that all upper bounds $3 < \hat{x}_U \leq 6$ are iteratively strictly dominated. Now consider any interval of the form $[\hat{x}_L, 3]$. The set of quantities of the other player for which an increase in the own quantity results in higher profits is given by: $\underline{X}_{-i}(\hat{x}_L) < X_{-i} < \bar{X}_{-i}(\hat{x}_L)$. We have $X_{-i} \geq \hat{x}_L$ and $X_{-i} \leq \hat{x}_U$. Thus, if $\underline{X}_{-i}(\hat{x}_L) \leq \hat{x}_L$ and $\bar{X}_{-i}(\hat{x}_L) \geq \hat{x}_U$ it pays off to increase one's quantity for

any quantity chosen by the other player. Both inequalities hold for $\hat{x}_L < 3$. Thus, for any interval of the form $[\hat{x}_L, 3]$ the lower bound is strictly dominated by a higher quantity, showing that the CNE quantity $x_i = 3$ is the only serially undominated strategy.

Triopoly: Again, (8) reveals that as in duopoly all quantities $x_i > 3$ are iteratively strictly dominated by some lower quantity. Thus, we have obtained a new undominated upper bound $x_U^0 = 3$.

Now consider intervals of the form $[x'_L, 3)$. Consider (6) and note that we have $X_{-i} \geq 2x'_L$ and $X_{-i} \leq 6$. We have $\underline{X}_{-i}(x'_L) < 2x'_L$ whenever $x'_L \leq 3$ and we have $\bar{X}_{-i}(x'_L) > 2x_U^0 = 6$ whenever $x'_L < 6(\sqrt{2}-1) = x_L^0$. Thus, for all lower bounds $x'_L < x_L^0$ we can find a profit increasing deviation if the others choose their quantities in the interval $[x'_L, 3)$. Thus, we have obtained a new lower bound $x_L^0 = 6(\sqrt{2}-1)$.

Consider now an interval $[\hat{x}_L, \hat{x}_U]$ with lower bound $\hat{x}_L$ and upper bound $\hat{x}_H$ with $\frac{3}{2} < \hat{x}_L < \hat{x}_U \leq 3$. By (8), it pays off to further reduce one's quantity for each upper bound $\hat{x}'_U$ that satisfies $X_{-i} > \bar{X}_{-i}(\hat{x}'_U)$. We know that $X_{-i} \geq 2\hat{x}_L$. Thus, it pays off to further reduce one's quantity if $2\hat{x}_L > \bar{X}_{-i}(\hat{x}'_U)$. Provided that $\hat{x}_L > \frac{3}{2}$, this can be written as $\hat{x}'_U > f(\hat{x}_L)$ where

$$f(x) = 2\sqrt{6x} - 2x.$$

Thus, we have found a new upper bound $\hat{x}''_U = f(\hat{x}_L)$.

By (6) it pays off to further increase one's quantity for each lower bound $\hat{x}'_L$ if $\underline{X}_{-i}(\hat{x}'_L) < X_{-i} < \bar{X}_{-i}(\hat{x}'_L)$. Since $X_{-i}(\hat{x}'_L) \geq 2x'_L$, the first inequality holds whenever $\hat{x}'_L < 3$. Further, we have $\bar{X}_{-i}(\hat{x}'_L) \leq \hat{2}x_L$ if $\hat{x}_U > \frac{3}{2}$ and $\hat{x}'_L < f(\hat{x}_U)$. Hence, we have found a new lower bound $\hat{x}''_L = f(\hat{x}_U)$.

The previous argument establishes that, for $\frac{3}{2} < \hat{x}_L < 3$ and $\frac{3}{2} < \hat{x}_U \leq 3$, given an undominated interval $[\hat{x}_L, \hat{x}_U]$ we can obtain a new undominated interval $[f(\hat{x}_L), f(\hat{x}_U)]$. We can now iterate the function $f(\cdot)$ on this interval. By the intermediate value theorem, a sufficient condition for the function f to be a contraction mapping is that $|f'(x)| < 1$ which is the case whenever $\frac{2}{3} < x < 6$. Thus f is a contraction mapping which, by the Banach fixed point theorem, assures convergence to the unique fixed point $x = f(x) = \frac{8}{3}$. This, together with the previous observations that $x_L^0 = 6(\sqrt{2}-1)$ and $x_U^0 = 3$, shows that the CNE is the only quantity in the serially undominated set.