

# Learning to Collude Tacitly on Production Levels by Oligopolistic Agents

Steven O. Kimbrough  
University of Pennsylvania  
kimbrough@wharton.upenn.edu

Frederic H. Murphy  
Temple University  
fmurphy@temple.edu

May 5, 2008

## **Abstract**

Classical oligopoly theory has strong analytical foundations but is weak in capturing the operating environment of oligopolists and the available knowledge they have for making decisions, areas in which the management literature is relevant. We use agent-based models to simulate the impact on firm profitability of policies that oligopolists can pursue when setting production levels. We develop an approach to analyzing simulation results that makes use of nonparametric statistical tests, taking advantage of the large amounts of data generated by simulations, and avoiding the assumption of normality that does not necessarily hold. Our results show that in a quantity game, a simple exploration rule, which we call PROBE AND ADJUST, can find either the Cournot equilibrium or the monopoly solution depending on the measure of success chosen by the firms. These results shed light on how tacit collusion can develop within an oligopoly.

**Please do not quote or circulate without the authors' permission.**

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Overview of the Literature</b>	<b>2</b>
<b>3</b>	<b>Experimental Setup</b>	<b>5</b>
<b>4</b>	<b>Reference Model</b>	<b>7</b>
<b>5</b>	<b>Results</b>	<b>11</b>
5.1	Profit maximization . . . . .	11
5.1.1	Detailed statistical analysis of the model results . . . . .	13
5.2	Maximizing total market profits . . . . .	18
5.2.1	Detailed statistical analysis of the model results . . . . .	20
5.3	Maximizing a mixture of market and own returns . . . . .	22
5.4	Maximizing market returns, constrained by own returns . . . . .	23
5.4.1	Detailed statistical analysis of the model results . . . . .	24
5.5	Number effects . . . . .	26
5.6	Differential costs . . . . .	31
<b>6</b>	<b>Discussion</b>	<b>32</b>
<b>7</b>	<b>Conclusions</b>	<b>33</b>
	<b>References</b>	<b>34</b>

# 1 Introduction

Research on decision making in the firm includes two streams—from the economics literature and from the management literature—having different goals. The management literature looks at how managers should learn about their circumstances and make decisions. The purpose of this literature is to advise and to train managers who have to make decisions in the course of running an organization. When looking at interactions of firms in a market, economists generally look at the firm as an aggregate and posit a theory of full information, ideally rational decision making in which the firm actually maximizes profits. The stylized abstractions of economists have been useful for developing theories of the competitiveness of markets and of public policy remedies for market power, a classic concern of microeconomics. The managerial literature deals with making the myriad decisions managers face in organizations. There is less emphasis on the public policy concerns of market power and more on decisions that improve profits and observations on the behaviors of participants. Thus, these two streams of literature address complementary issues, yet treat the same entity, the firm in its environment.

Models of market power standardly make one of two assumptions. Either firms offer prices and the firm with the lowest price gets the whole market, the Bertrand game, or firms produce quantities under the assumption that the other firms do not adjust their quantities in response to the firm's decisions, the Cournot game. In Kagel and Roth [Kagel and Roth, 1995] and elsewhere, one finds comments to the effect that the behaviors of the agents in these economic models are not verifiable and often run counter to the literature in cognitive psychology and behavioral economics. Keunne [Kuenne, 1998] points out that firms in an oligopoly form a community with community-based norms, values, and hierarchies.

The managerial literature addresses the issue of making decisions without full information because it is aimed at practical application and in practice the information available is always incomplete. A better theory of the firm and markets would take into account how managers make decisions and what they know and when they know it. Managers know less about the parameters for functions such as demand curves and probability distributions but know more about the other players in their market. By using people in simulated markets it is possible to observe the decision making and outcomes that get beyond simple abstractions of the decisions managers face. Selten, Mitzkewitz, and Uhlich [Selten et al., 1997] simulate outcomes in a duopoly market by having students program agents. They give students both full information on the supply and demand parameters, and software in which the students can embed their strategies to play against other students in duopoly games. Key to the clever design of the experiments is the revealing and testing of the decision rules used by the players. The authors find that smart agents can obtain higher profits than those in the classic Cournot equilibrium.

Here are several of their observations on the results. One strategy of the players is to try to forecast the actions of the other agents. This strategy is close to positing what is known

as the conjectural variation of the other player in response to the first player's actions [Kuenne, 1998] and performs badly in the experiments. In the Cournot game the conjectural variation is presumed to be zero. The best performing strategies are the oligopoly equivalent of TIT FOR TAT, the winning strategy in Axelrod's tournament with Prisoner's Dilemma [Axelrod, 1984]. The agents in the Selten et al. experiment punished excessive competitiveness and rewarded cooperation, moving the solution from the non-cooperative (defecting) Cournot equilibrium towards the monopoly outcome, which is the Pareto outcome. In essence they are reinforcing a community norm of cooperation without explicitly setting production quotas. (We note that the Pareto outcome of mutual cooperation also results from repeated play in the Prisoner's Dilemma game with the TIT FOR TAT strategy and that under general conditions of mutual learning the Pareto outcome is often achieved or approximated [Kimbrough and Lu, 2005].)

The behavior of the players in the Selten et al. experiments is strikingly at odds with what would be predicted by the classic Cournot theory. The Selten et al. players learned to collude tacitly and thereby achieved rewards in excess of those available from the Cournot outcomes. They did this, however, using explicitly-given knowledge of the market (e.g., the demand curve) that is in fact not generally available in practice.

The foregoing forcefully raises the question of whether there are effective procedures, using realistically available information, that may be actually used by managers in oligopoly settings and that produce the Cournot-improving outcomes found in Selten's experiments. This is the question we explore in what follows. We begin with an overview of the current literature.

## 2 Overview of the Literature

See Tesfatsion and Judd [Tefatsion and Judd, 2006] for a collection of articles on the current status of and issues in agent-based modeling in economics. Tesfatsion [Tefatsion, 2006] surveys the applications of "agent-based computational economics" (ACE) to specific industries and supply chains. Brenner [Brenner, 1999b, Brenner, 2006] summarizes the literature on learning both from cognitive psychology and artificial intelligence. Brenner [Brenner, 1999a] and Bruun [Bruun, 2006] are useful collections of relevant papers. Duffy [Duffy, 2006] looks at what intelligence is necessary for an agent, describing zero-intelligence agents, reinforcement learning and evolutionary algorithms. Pyka and Fagiolo [Pyka and Fagiolo, 2005] provide an overview of the methodological issues in agent-based economic models.

To deal with the institutional features of electricity markets, researchers have developed agent-based models to simulate the auctions they use. See Marks [Marks, 2006], Bunn and Oliveira [Bunn and Oliveira, 2003], and Entrikan and Wan [Entrikan and Wan, 2005].

The rational expectations literature—discussing situations in which economic agents try to forecast the future—relates to the issues raised by Selten, Mitzkewitz, and Uhlich

[Selten et al., 1997]. Hommes et al. [Hommes et al., 2003] experiment with human subjects to forecast the clearing price for a market. The forecasted price sets the production quantity and the actual price results from the production level, resulting in the cobweb model. They show that people can find the market-clearing price and the stability of the market depends on the price sensitivity of supply.

Arifovic [Arifovic, 1994] shows that in an oligopoly that is modeled with a population of agents that evolves using a genetic algorithm, the solution converges to the competitive solution, not the Cournot equilibrium. Vriend [Vriend, 2000] shows that with social learning, where each player sees the returns of every player and can adopt the strategies of the successful players, the solution converges to the competitive equilibrium, and with individual learning, where the player sees only its returns, the solution converges to the Cournot equilibrium. Riechman [Riechmann, 2002] finds it necessary to have more complicated agents to find a solution different from the competitive solution even in an oligopolistic market. Waltman and Kaymak [Waltman and Kaymak, 2005] use Q-learning and find that under certain circumstances the agents move to solutions between the monopoly and Cournot equilibria. Arifovic and Maschek [Arifovic and Maschek, 2005] find that Vriend's results are parameter-driven and not robust. Alkemade, La Poutr, and Amman [Alkemade et al., 2006] show that modeling the agents as chromosomes (the agent has an assigned strategy) in a genetic algorithm instead of modeling the strategies as chromosomes (the agent can choose its strategy) can lead to premature convergence in a genetic algorithm. The parameter settings determine whether the Cournot or the competitive solution is reached when agents are strategies and the Cournot solution is reached when the agents can choose the strategy. Note that the definitions used in this article are different from the definitions used here in that we define a policy as a choice of measure of success and the price and quantity are operating decisions.

Huck et al. [Huck et al., 2003] have a model of agent behavior that, like ours, requires little intelligence or information about its environment. They find that when agents make simultaneous moves of the same step size in a duopoly, the players maximize total social welfare and divide the market equally even though they see only their own welfare. This is because either both players see either the marginal revenue function of the monopolist when both players move in the same direction or the trial price as the marginal revenue, as if the market were in perfect competition, when the players move in opposite directions. The effect of these perceptions is a sequence of steps that leads to equalizing production levels and movements in the same direction. The players converge to the monopoly solution because that is the marginal revenue function they see. Huck et al. also model the players moving sequentially and monitoring the effects of their own actions, resulting in the Nash/Cournot equilibrium. In this case the players follow the tâtonnement process used to explicate the Cournot equilibrium.

Barr and Saraceno [Barr and Saraceno, 2005] use neural networks to represent learning agents that learn about the environment rather than learn the optimal production level. They show that the agents find the Cournot equilibrium. Agents with simple neural net-

works find the equilibrium faster but more complicated networks develop a better demand representation and find a more accurate solution in the long run.

Marks and Midgley [Marks and Midgley, 2006] build a simulation of a market with an oligopoly of coffee manufacturers and a retailer between the producers and customers that decides which coffee promotion to accept. They then simulate the outcomes of retailer strategies that range from zero information to sophisticated measurement of the market and find that a zero information retailer does quite well. This is an example of bringing the management literature into the economic models of markets. Midgley, Marks, and Cooper [Midgley et al., 1997] used an earlier version of this model to look at breeding profit maximizing retailers to examine the frequency of promoting coffee specials. Their work uses point-of-sale information for comparing the model retailer to the actual retailer. Sallans et al. [Sallans et al., 2003] breed firms that compete on production positioning in a market and the firms have to finance their businesses using agents modeled as financial firms. They are able to replicate many phenomena observed in retail markets.

The business literature on decision making, especially the practitioner books and articles, is relevant to agent-based modeling because the economic models should represent business decisions in the way business people make them. Since the focus of oligopoly models is on setting the price and/or quantity, the most relevant literature is in marketing on pricing and capacity expansion. We focus on pricing here as the same issues arise in setting quantities.

We use Nagle and Hogan [Nagle and Hogan, 2006] to illustrate current managerial thought on best practices in pricing products (whether by setting prices or quantities). The main point this book makes is that a firm should first do everything to avoid a focus on price, for example creating product distinctions that are real or only in the minds of customers. The discussion of demand elasticities in this book covers 3 pages in a 30 page chapter with the main discussion about customer perception of product attributes of the firm's product versus the attributes of competitor products and social norms. That is, the discussion focuses on the position of the product relative to the competition with the goal of segmenting the market to customers who are willing to pay a premium for the product's perceived attributes.

The entire approach advocated presumes—and is incomprehensible without assuming—that customers are not fully rational and/or do not have full information. The following passages are representative.

Unless customers actually recognize the value that you create and ask them to pay for, value-based pricing will fail. . . .

The reality is that customers generally don't know the true value delivered by items they buy unless the seller informs them. That leaves the most differentiated and highest quality supplier vulnerable to competitors who offer a lower-price alternative possessing only those value components the customer recognizes, and who portrays additional value elements. . . [Nagle and Hogan, 2006,

page 81]

Although cost should not drive the prices you charge, your prices can definitely affect the cost-to-serve customers and, therefore, your profitability. Many companies differentiate their offers with bundled services, even when demands on those services are subject to customer discretion and therefore are not proportional to the volume of sales. “Service abusers” can boost your average cost of sales while “service avoiders” drive up your average cost of sales by abandoning you in favor of cheaper, low-service competitors. . . .

. . . The solution is to create “roughly right” cost allocation indexes and use them to build a “roughly right” relative profit index by account or segment. [Nagle and Hogan, 2006, pages 113–4]

Their discussion on elasticities looks at the elasticities of the firm’s products and not the market and they note how elasticities differ depending on market share, because the different products in a market are positioned for different customer segments, and products age. In the chapter on estimating price response they state that “The low accuracy of many numerical estimates makes blind reliance on them very risky.” They conclude the chapter with “Even when actual purchase data cannot provide conclusive answers, they can suggest relationships that can then be measured more reliably with other techniques.”

What should be taken from this short discussion is that the management literature recommends that managers explore rather than optimize. The data are not completely clear and circumstances change.

### 3 Experimental Setup

We simulate agents playing a Cournot game. Using NetLogo<sup>1</sup> as the programming environment, we name our program `oligopolyPutQuantity.nlogo`. It is freely available from the authors for purposes of research and education.

Abstractly, the agent we define has three features: (1) a measure of success, (2) a data stream to measure its success, and (3) the ability to do experiments or to learn how its actions affect its success. These three properties are the minimal set of properties for an economic agent to improve its outcomes when operating in a situation without full information. To add an element of realism, the agents can be made to operate in a noisy environment where the demand parameters are a random walk.

One measure of success we use is the classic measure, firm profitability. We term this measure and the policy of using this measure as the objective function “Own Returns.” Another measure is the profitability of whole the industry, termed “Market Returns.” We allow for both measures because firms operate in a complex institutional environment and

---

<sup>1</sup><http://ccl.northwestern.edu/netlogo/>

the leaders of these companies set up and fund institutions that represent the industry. Examples are the National Petroleum Refiners Association and the Iron and Steel Institute. That is, firms choose when to cooperate and when to compete (see Brandenberger and Naibuff [Brandenberger and Nalebuff, 1996]). A less mainstream example is Cosa Nostra, which acts as a chamber of commerce for crime families that mediates conflicts, reduces killing among the families, and works to protect the profitability of organized crime.

We allow the players to use combinations of objectives as a measure of success. In the first, an agent pursuing the “Mixed” policy, at the end of its epochs, uses a convex combination of both objectives. In the second, an agent pursuing the “Market Returns, Constrained by Own Returns” (MR-COR) policy, at the end of its epochs, looks at the mean quantity it produces versus the mean total quantity produced for the entire market (its and the other player(s)’s production). If its mean quantity produced plus epsilon is lower than the mean quantity produced for the market, the agent raises its baseline production by epsilon; otherwise, it uses the “Market Returns” policy. Here a firm pursues a hierarchical policy where it looks to get its share of the market and then looks to keep the market as profitable as possible. Equal shares is the outcome of the Cournot solution when players have equal costs. In most oligopolies the firms have different sizes because of differences in product attributes, unique access to high-quality resources, or the history of the firms and markets, including acquisitions. We view this more as a stylized form of maintaining a sense of fairness while taking the larger view of the industry as a community norm of an implicit willingness to cooperate up to a point as in Kuenne.

In the Cournot game the players make quantity decisions and the market sets the price. Each player starts with a base quantity that remains fixed for a given number of periods and randomly adjusts the quantity up or down in each period, running experiments to observe the effects of altering the base quantity. The number of periods for which the base quantity remains the same for a player is termed an *epoch*. We use a uniform distribution for the random adjustments around the base quantity. Different players can have different epoch lengths. The player is interested in knowing whether it should increase or decrease its production and records its returns and/or the market returns for the quantity increases and decreases separately. After each epoch, the players assess their returns using their measures of success. If the profits for a player are higher with the increases than with the decreases, then the base quantity is increased and vice versa. This begins a new epoch. Epochs are repeated in the simulation until the pattern of behavior stabilizes. Note that each player knows the outcome only in relation to its decisions and retains no information on the other players’ decisions. We term the search/learning method the agents employ as PROBE AND ADJUST. Technically, the method is in the family of line-search algorithms where the algorithm finds the direction of improvement, takes a step of a certain size in that direction, and then assesses the benefit of that move. See Winston [Winston, 2004] for an introduction to algorithms in this class. More importantly, in our context, this algorithm represents a situation in which managers adjust their production incrementally to learn the consequences of their actions, without making radical changes that could risk the business.

Think of this as a form of muddling through. The algorithm approximates the behaviors of consumer products companies that phase in price increases and the capacity-expansion decisions in commodity businesses such as petroleum refining, where increases or decreases in capacity are incremental because of environmental concerns removing the ability to build a wholly new refinery in the US. It does not reflect the situation where capacity has to be added in large increments, such as a firm building a green-field integrated steel mill.

## 4 Reference Model

For clarity we present the underlying model and resulting key quantities that we refer to throughout, as well as the terminology we use. To begin, we assume a linear inverse demand function:

$$P = a - slope \times Q \quad (1)$$

$P$  is the price realized in the market.  $Q$  is the total quantity of good supplied to the market.  $a$  is the price intercept and  $slope > 0$ , we assume. We also assume that negative prices are not permitted, so (2) is actually what is assumed.

$$P = \max\{a - slope \times Q, 0\} \quad (2)$$

We begin with the duopoly case and then generalize the results. Let the agents have unit costs,  $k_i$ , which can differ. In the duopoly case the profit of firm 1 is then

$$\pi_1 = P \cdot Q_1 - k_1 \cdot Q_1 = (a - slope \cdot (Q_1 + Q_2)) \cdot Q_1 - k_1 \cdot Q_1 \quad (3)$$

For firm 2 we have

$$\pi_2 = P \cdot Q_2 - k_2 \cdot Q_2 = (a - slope \cdot (Q_1 + Q_2)) \cdot Q_2 - k_2 \cdot Q_2 \quad (4)$$

Differentiating we get

$$\frac{d\pi_1}{dQ_1} = a - 2 \cdot slope \cdot Q_1 - slope \cdot Q_1 \cdot \frac{dQ_2}{dQ_1} - slope \cdot Q_2 - k_1 \quad (5)$$

$$\frac{d\pi_2}{dQ_2} = a - 2 \cdot slope \cdot Q_2 - slope \cdot Q_2 \cdot \frac{dQ_1}{dQ_2} - slope \cdot Q_1 - k_2 \quad (6)$$

Setting  $\frac{dQ_2}{dQ_1}$  and  $\frac{dQ_1}{dQ_2}$  to 0 as usual leads to

$$0 = a - 2 \cdot slope \cdot Q_1 - slope \cdot Q_2 - k_1 \quad (7)$$

$$0 = a - 2 \cdot slope \cdot Q_2 - slope \cdot Q_1 - k_2 \quad (8)$$

and then on to

$$Q_1 = \frac{a - slope \cdot Q_2 - k_1}{2 \cdot slope} \quad (9)$$

$$Q_2 = \frac{a - \text{slope} \cdot Q_1 - k_2}{2 \cdot \text{slope}} \quad (10)$$

which when solved yield

$$Q_1^C(2, [k_1, k_2]) = \frac{a - 2k_1 + k_2}{3 \cdot \text{slope}} \quad (11)$$

$$Q_2^C(2, [k_1, k_2]) = \frac{a + k_1 - 2k_2}{3 \cdot \text{slope}} \quad (12)$$

Notice that

$$Q^C(2, [k_1, k_2]) = Q_1^C(2, [k_1, k_2]) + Q_2^C(2, [k_1, k_2]) = \frac{2a - k_1 - k_2}{3 \cdot \text{slope}} \quad (13)$$

The formula generalizes. With  $n$  players having proportional costs  $k_i \in \{1, 2, 3, \dots, n\}$  (total cost = unit cost  $\times$  quantity =  $k_i \cdot Q_i$ ) we have expression (17).

The monopoly quantity,  $Q^M$ , may be arrived at as the special case of (18) when  $n = 1$ :

$$Q^M(k) = \frac{(a - k)}{(2 \cdot \text{slope})} \quad (14)$$

Finally, the rivalrous (or competitive, but we've already used  $C$ ) quantity,  $Q^R$ , obtaining in a fully competitive market occurs when price ( $a - \text{slope} \times Q$ ) equals marginal cost ( $k$ , assuming all firms have the same marginal cost). Equating them and solving yields  $Q^R$ .

$$Q^R(k) = \frac{(a - k)}{\text{slope}} \quad (15)$$

Now assume there are  $n$  firms in the market,  $n \geq 1$ . The quantity supplied by firm  $i$  (in a given round or episode) is  $Q_i$ . We stipulate

$$Q = \sum_{i=1}^n Q_i \quad (16)$$

Each firm  $i$  has a unit (marginal) cost of production of  $k_i \geq 0$ .

Given these conditions, then in the Cournot model the equilibrium Cournot quantity,  $Q^C$ , is the sum of the individual  $Q_i^C$ s, and

$$Q^C(n, \vec{k}) = Q^C = \sum_{i=1}^n Q_i^C = \frac{na - \sum_{i=1}^n k_i}{(n + 1) \cdot \text{slope}} \quad (17)$$

When all  $k_i$  are equal to  $k$  we write  $Q^C(n, k)$  for  $Q^C(n, \vec{k})$ . That is,

$$Q^C(n, k) = Q^C = \sum_{i=1}^n Q_i^C = \frac{na - \sum_{i=1}^n k}{(n + 1) \cdot \text{slope}} \quad (18)$$

and the individual firm Cournot quantities are

$$Q_i^C(n, k) = \frac{(a - k)}{(n + 1) \cdot \text{slope}} \quad (19)$$

1. Set parameters  $\delta$ ,  $\varepsilon$ , `currentQuantity`, `epochLength`  
(Typically,  $\varepsilon < \delta \ll \text{currentQuantity}$  and  $\text{epochLength} \approx 30$ .)
2. `episodeCounter`  $\leftarrow$  0
3. `returnsUp`  $\leftarrow$  [] (Initialize `returnsUp` to an empty list.)
4. `returnsDown`  $\leftarrow$  [] (Initialize `returnsDown` to an empty list.)
5. Do forever:
6. `episodeCounter`  $\leftarrow$  `episodeCounter` + 1
7. `bidQuantity`  $\sim U[\text{currentQuantity} - \delta, \text{currentQuantity} + \delta]$   
(The agent's `bidQuantity` is drawn from the uniform distribution within the range `currentQuantity`  $\pm \delta$ .)
8. `return`  $\leftarrow$  *Return-of* `bidQuantity`  
(The agent receives `return` from bidding `bidQuantity`.)
9. If (`bidQuantity`  $\geq$  `currentQuantity`) then:  
  `returnsUp`  $\leftarrow$  *Append* `return` to `returnsUp`  
  else:  
  `returnsDown`  $\leftarrow$  *Append* `return` to `returnsDown`
10. If (`episodeCounter` mod `epochLength` = 0) then:  
(Epoch is over. Adjust `episodeCounter` and reset accumulators.)
  - (a) If (*mean-of* `returnsUp`  $\geq$  *mean-of* `returnsDown`) then:  
  `currentQuantity`  $\leftarrow$  `currentQuantity` +  $\varepsilon$   
  else:  
  `currentQuantity`  $\leftarrow$  `currentQuantity` -  $\varepsilon$
  - (b) `returnsUp`  $\leftarrow$  []
  - (c) `returnsDown`  $\leftarrow$  []
11. Loop back to step 5.

Figure 1: Pseudo code for basic PROBE AND ADJUST

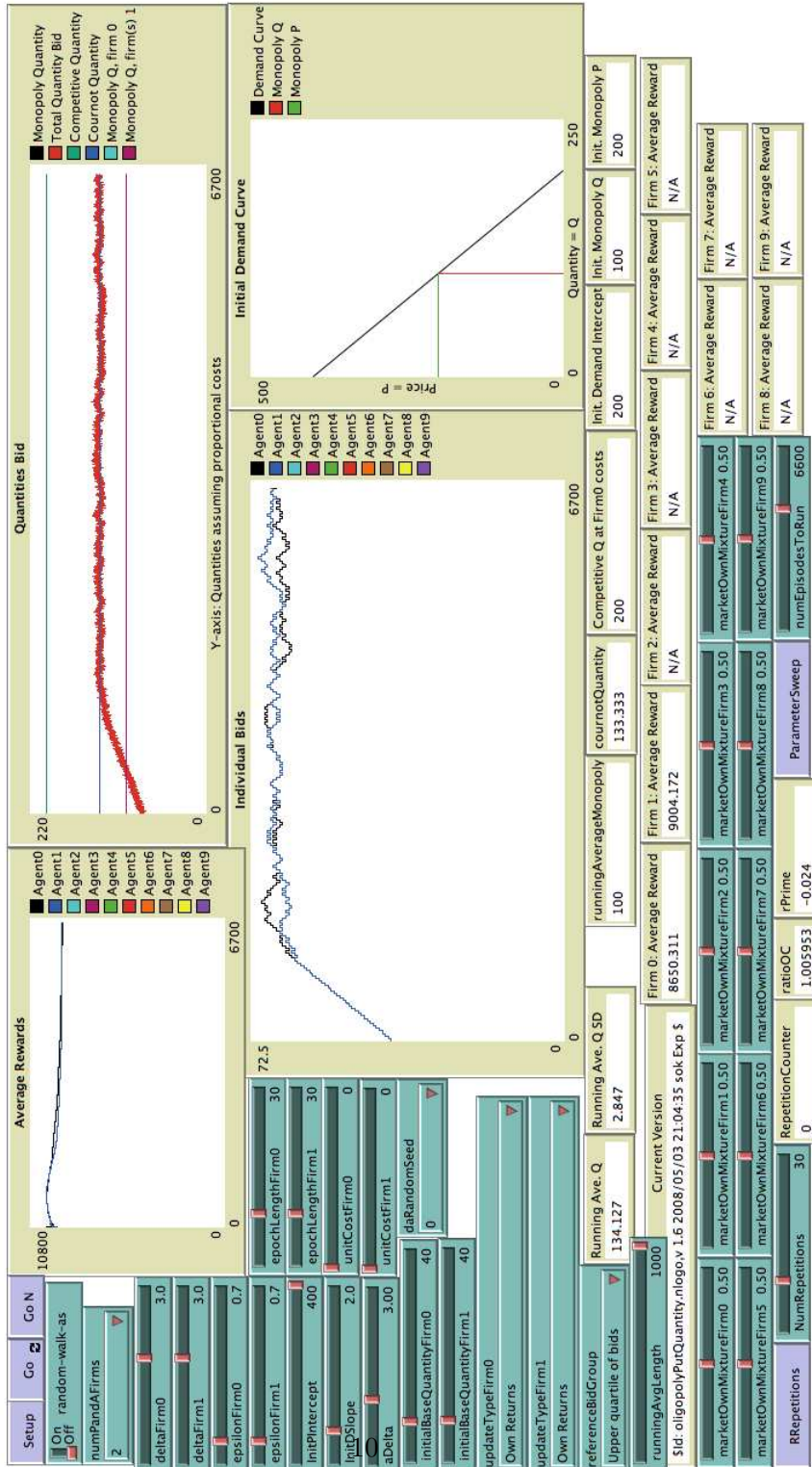


Figure 2: Results from oligopolyPutQuantity.nlogo with two quantity-offering players each using the Own Returns policy of play. File: oligopolyPutQuantity-own-own-r2.jpg.

## 5 Results

Our simulation results are obtained from the program `oligopolyPutQuantity.nlogo`. This program affords simulation of quantity-bidding (or Cournot) agents (“firms”) in an oligopoly. Agents are able to use any of several forms of PROBE AND ADJUST, a learning algorithm suitable for adjusting a continuous parameter, here the quantity an agent offers to the market during a single period (aka: round of play, episode). Under PROBE AND ADJUST, each agent maintains a current base quantity—`currentQuantity`—which it uses as the midpoint of an interval from which it draws uniformly each period to set its production quantity—`bidQuantity`—that period. Each agent maintains its `currentQuantity` for a number of episodes equal to its `epochLength`. When its current epoch is over (by count of periods or episodes played), the agent re-evaluates its `currentQuantity`, adjusting it up or down by its adjustment moiety, its `epsilon`, depending on whether production levels above its `currentQuantity` during the just-completed epoch have or have not been more profitable than those below its `currentQuantity`. Figure 1 presents the basic PROBE AND ADJUST algorithm in pseudo code.

Firms in `oligopolyPutQuantity.nlogo` may use any of several variations (or *policy versions*) of PROBE AND ADJUST. We now present results, focusing especially on the effects of the several policy versions available.

### 5.1 Profit maximization

We begin with the classical objective function of maximizing firm profits. Consider first two agents—Firm0 and Firm1—who independently choose production quantities using PROBE AND ADJUST in the presence of a linear price function,  $P(\text{rice}) = \text{priceIntercept} - d\text{Slope} \times Q(\text{quantity}) = 400 - 2Q$  in our examples below, that is unknown to them. In this section we consider the case in which each firm uses PROBE AND ADJUST and looks only to its own profits (rewards) when adjusting its `currentQuantity`. This is the policy of play labeled “Own Returns” in Figure 2. Specifically, a player using the Own Returns policy observes the market’s price at the end of each period of play, calculates its profits (net of costs) for the quantity it produces (and sells), and uses this as its `return` in step 8 of Figure 1. There, we might label the *Return-of* function as *Own>Returns* in this case. Figure 2 shows simulation results from `oligopolyPutQuantity.nlogo`. Each firm has costs of \$0 per unit and their base quantities—

(`initialBaseQuantityFirm0 = initialBaseQuantityFirm1 = 40`)

—start well below the Cournot equilibrium quantity of 133.33.<sup>2</sup> PROBE AND ADJUST leads them near to the Cournot solution with equal market shares. We see that after 6600 episodes (long after a stable settling has occurred) the average total quantity produced is

---

<sup>2</sup>The results we report are for long after the system has attained stability. Moreover, we find that initial base quantities do not affect either the location of, or the firm shares at, stability. For these two reasons we do not further discuss any starting point effects. There are none, except as noted below.

Firm	Average Reward	Standard Deviation of Average Reward
0	639.8964661413705	70.55142984995368
1	645.3864190695721	67.94933796315856
2	649.0329236965865	62.82872860151167
3	651.7672918703781	68.57175862731928
4	643.1174687232188	64.6236221246588
5	638.8300180138967	61.28529868477505
6	638.2949903490786	66.98362150327631
7	650.5939716430881	79.28907276367299
8	649.4548736132194	62.526625943544516
9	643.725796838893	65.38427951535789

Table 1: Summary of results over 100 replications with 10 firms in the market, all following the Own Returns policy of play

about 134 (134.127 averaged over the last 1000 episodes of play). Each agent is producing about 67.1 at the end (this is a rough estimate, obtained by reading the Individual Bids charts, rather than computed exactly from the data). Firm0 is getting an average profit per play of about 8650 (averaged over the past 1000 episodes), while Firm1 is getting about 9004. The agents have essentially identical average profits overall (see Average Rewards chart). Thus, the cost of learning is quite small. Notice that the quantities fluctuate around the optimal quantity (from the Cournot perspective) because the agents are always randomly varying their base quantity. The standard deviation (`runningAverageBidSD` in the Figure) on total production is 2.847 units. We repeated the experiment 100 times, using the system clock to initialize the random number generator each time. Over the 100 trials the mean of the averages of the total quantity produced is 133.43 with a standard deviation of 1.57. Firm0's average of its average rewards was 8874.72 (standard deviation of 159.08) and Firm1's was 8863.70 (155.99). Note that the standard deviations of profits and production quantities are roughly 2 percent of profits and quantities produced.

We also repeated the experiment 100 times with 10 firms in the market (and using the system clock to initiate the random number generator). The Cournot quantity is now (with 10 firms) 181.818. Averaged across the 100 repetitions, the running average production was 182.08 with a standard deviation of 0.74. Table 1 reports the averaged rewards (and their standard deviations) for the 10 firms in this run of 100 replications.

Thus, we can say the model provides a good approximation to the Cournot solution, in spite of not having the full information assumed in the classical analysis and without each player giving the best response to the other player's plays. A simple explanation of these results is that since there is no correlation or coordination of the player's moves during an epoch, on average the players see the marginal revenue function of a Cournot

player. Unlike Riechmann [Riechmann, 2002] then, we find that a pair of simple searching agents can find the Cournot solution. Figure 2 shows, and we observe in all other runs as well, a certain amount of oscillation about the Cournot quantity. The range of oscillation varies depending on the length of the epoch. Reducing the epoch length to 10 results in an average over five runs of the running average quantity to be 133.85 (using random number seeds 0 through 4), increasing the standard deviation on average to 3.52. After reducing `epochLengthFirm0` and `epochLengthFirm1` further to 5, the running average quantity averaged over five runs (seeds 0 through 4) is 132.8114 and the average standard deviation is 4.2004. At 2 the numbers are 133.7688 and 6.1852. The range on the production levels of the individual agents show more fluctuations with a smaller epoch length than the total because if one firm increases production beyond the optimal quantity due to randomness, the other is more likely to respond with a decrease in the next epoch, a consequence of the shape of the objective function, which is quadratic and has a steeper slope and more curvature the further away from the optimum. With an epoch length of 1 the agents engage in what looks like a random walk with reflecting barriers at production levels of 0 for each player and total production at the competitive equilibrium with no profits for either player. This is because there is either an up or down value but not both in an epoch. Using the parameters of the previous runs, the average of the running averages is 87.45 (= mean of 170.612 22.841 65.555 125.666 52.594) and the average of the standard deviations is 9.75 (= mean of 7.725 8.299 7.784 14.988 9.934). Patience, in the form of a longer epoch length (a tilt towards exploration and away from exploitation), has its rewards for these players.

Increasing the number of players from 2 to 3 but keeping other conditions constant (and returning to epoch lengths of 30) increases the Cournot quantity to 150 and increases the standard deviation of the running average of the total quantity. The numbers are 150.30 (= mean of 148.284 151.984 153.597 149.392 148.221) and 3.6482 (= mean of 3.904 3.591 3.605 3.286 3.855). With 10 firms the Cournot quantity is 181.818 and the numbers are 181.811 (= mean of 180.851 180.708 182.882 182.504 182.11) and 6.33 (= mean of 6.212 6.673 6.233 6.442 6.112). In general, variation in total production increases with the number of firms in the market, but the overall picture remains otherwise accurate.

### 5.1.1 Detailed statistical analysis of the model results

Because it is possible to generate large amounts of data from the simulation results, we are able to avoid making the assumption of normality and use weaker nonparametric tests to examine the statistical validity of the results. Table 2 presents the relevant parameters and their default values (the “Table 2 settings”) for the PROBE AND ADJUST model. We ran 100 repetitions of the `oligopolyPutQuantity.nlogo` model under the conditions described in Table 2. Each repetition produces an ending value of `runningAverageBid`, the total quantity averaged over the past `runningAverageLength` (=1000 in the Table 2 settings) episodes (i.e., in episodes 5601–6600, given the Table 2 settings). Here is a

Agent Parameters	
numPandaFirms	2
epochLengthFirm0	30
epochLengthFirm1	30
initialBaseQuantityFirm0	40
initialBaseQuantityFirm1	40
updateTypeFirm0	Own Returns
updateTypeFirm1	Own Returns
deltaFirm0	3.0
deltaFirm1	3.0
epsilonFirm0	0.7
epsilonFirm1	0.7
unitCostFirm0	0
unitCostFirm1	0
Environment Parameters	
InitPIntercept (Price intercept of demand function)	400
InitDSlope (Negative of the slope of the demand function)	2
numEpisodesToRun (Number of episodes in a run)	6600
runningAvgLength (Running average length)	1000
daRandomSeed (Random number seeded with)	system clock
NumRepetitions (Number of repetitions)	100

Table 2: Default settings of the principal parameters.

summary of the 100 values obtained for `runningAverageBid`.<sup>3</sup>

runningAverageBid summary					
Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
129.5	132.3	133.3	133.3	134.1	138.6

Note that the Cournot value is 133.333. These data are symmetrically and rather tightly centered on or near the Cournot value.

Using the exact binomial test (`binom.test` in R) on the differences between the 100 `runningAverageBid` values and the Cournot value, in this particular run of 100 we actually got 50 values above the Cournot value and 50 below (there were no ties). Under the null hypothesis of  $p = 0.5$ , the p-value is 1 and the 95% confidence interval is [0.3983, 0.6017]. There is no reason to reject the null hypothesis, that PROBE AND ADJUST with Own Returns leads the agents to the Cournot solution.

---

<sup>3</sup>These and all subsequent statistical calculations were made in R, which we gratefully acknowledge [R Development Core Team, 2007]. Console logs and data sets are available from the authors.

At the end of each run `oligopolyPutQuantity.nlogo` reports the standard deviation of the `runningAverageBid` (see Figure 2; in the present case, for the last 1000 episodes). Here is a summary of the 100 values obtained for `runningAverageBidSd`.

runningAverageBidSd summary					
Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
2.596	2.865	3.099	3.146	3.333	4.417

While some asymmetry is apparent, these values are reassuringly regular and concentrated.

Turning now to the rewards obtained by the two individual firms, we would expect them to have no systematic differences. That this is so is certainly suggested by the following summary data.

summary of mean rewards						
	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
<code>meanRewardFirm0</code>	8509	8784	8886	8868	8958	9414
<code>meanRewardFirm1</code>	8366	8766	8900	8895	9006	9276

The nonparametric Wilcoxon rank-sum test (`wilcox.test` in R) yields a p-value of 0.2026 for the null hypothesis of no difference between the two sets of returns. This coheres with the suggestion we drew from the table.

We investigated whether the default number of episodes, 6600, is sufficient for getting beyond any transitory effects from the initialization of the model. As the following three tables of summary information indicate, it is.

summary with 6600 episodes			
<code>runningAverageBid</code>	<code>runningAverageBidSd</code>	<code>meanRewardFirm0</code>	<code>meanRewardFirm1</code>
Min. :129.5	Min. :2.596	Min. :8509	Min. :8366
1st Qu.:132.3	1st Qu.:2.865	1st Qu.:8784	1st Qu.:8766
Median :133.3	Median :3.099	Median :8886	Median :8900
Mean :133.3	Mean :3.146	Mean :8868	Mean :8895
3rd Qu.:134.1	3rd Qu.:3.333	3rd Qu.:8958	3rd Qu.:9006
Max. :138.6	Max. :4.417	Max. :9414	Max. :9276

summary with 6000 episodes			
<code>runningAverageBid</code>	<code>runningAverageBidSd</code>	<code>meanRewardFirm0</code>	<code>meanRewardFirm1</code>
Min. :129.0	Min. :2.583	Min. :8386	Min. :8445
1st Qu.:132.5	1st Qu.:2.911	1st Qu.:8754	1st Qu.:8734
Median :133.6	Median :3.137	Median :8866	Median :8843
Mean :133.6	Mean :3.181	Mean :8878	Mean :8837
3rd Qu.:134.8	3rd Qu.:3.346	3rd Qu.:9002	3rd Qu.:8939
Max. :137.9	Max. :4.529	Max. :9396	Max. :9250

summary with 7200 episodes			
runningAverageBid	runningAverageBidSd	meanRewardFirm0	meanRewardFirm1
Min. :128.5	Min. :2.718	Min. :8347	Min. :8425
1st Qu.:131.9	1st Qu.:2.917	1st Qu.:8777	1st Qu.:8751
Median :133.2	Median :3.069	Median :8870	Median :8880
Mean :133.3	Mean :3.122	Mean :8876	Mean :8876
3rd Qu.:134.9	3rd Qu.:3.281	3rd Qu.:8994	3rd Qu.:9016
Max. :137.4	Max. :3.889	Max. :9570	Max. :9450

We now consider some “parameter sweeping” experiments, first on epoch lengths. We conducted a full factorial experiment using 5 levels of epoch length (24, 26, 30, 34, 38) for each of two variables (`epochLengthFirm0` and `epochLengthFirm1`) with 30 repetitions (for a total of  $750 = 30 \times 5^2$  runs. (Otherwise, the parameter settings are the default settings of Table 2.) From the table below it is evident that on average there is no deviation from our basic findings for the default settings.

summary epoch length sweeps			
meanRewardFirm0	meanRewardFirm1	runningAverageBid	runningAverageBidSD
Min. :8360	Min. :8364	Min. :129.2	Min. :2.503
1st Qu.:8775	1st Qu.:8755	1st Qu.:132.3	1st Qu.:2.889
Median :8887	Median :8872	Median :133.3	Median :3.068
Mean :8883	Mean :8875	Mean :133.3	Mean :3.134
3rd Qu.:9002	3rd Qu.:8990	3rd Qu.:134.3	3rd Qu.:3.308
Max. :9452	Max. :9373	Max. :139.2	Max. :4.622

Looking within these data here is a summary for when Firm0’s epoch length was 24:

summary epoch length sweeps with <code>epochLengthFirm0 = 24</code>			
meanRewardFirm0	meanRewardFirm1	runningAverageBid	runningAverageBidSD
Min. :8365	Min. :8364	Min. :129.2	Min. :2.556
1st Qu.:8790	1st Qu.:8761	1st Qu.:132.4	1st Qu.:2.913
Median :8900	Median :8868	Median :133.2	Median :3.098
Mean :8889	Mean :8863	Mean :133.3	Mean :3.140
3rd Qu.:9005	3rd Qu.:8982	3rd Qu.:134.3	3rd Qu.:3.332
Max. :9435	Max. :9268	Max. :138.2	Max. :4.537

Restricting our attention further, here is a summary for when `epochLengthFirm1 = 38`.

summary with epochLengthFirm0 = 24 & epochLengthFirm1 = 38			
meanRewardFirm0	meanRewardFirm1	runningAverageBid	runningAverageBidSD
Min. :8580	Min. :8590	Min. :129.8	Min. :2.721
1st Qu.:8797	1st Qu.:8673	1st Qu.:132.6	1st Qu.:2.914
Median :8895	Median :8880	Median :133.1	Median :3.060
Mean :8911	Mean :8853	Mean :133.3	Mean :3.129
3rd Qu.:9028	3rd Qu.:8979	3rd Qu.:134.1	3rd Qu.:3.344
Max. :9233	Max. :9172	Max. :136.5	Max. :4.086

Evidently, our basic findings are robust for changes in epoch length between 24 and 38.

Finally, we ran a large full factorial experiment with  $\text{deltaFirm}_i \in \{2.4, 3.0, 3.8\}$ ,  $\text{epsilonFirm}_i \in \{0.4, 0.7, 1.0\}$ ,  $\text{epochLengthFirm}_i \in \{24, 30, 36\}$ , and  $i \in \{0, 1\}$ . There were thus  $3^6$  unique combinations of factors and since we conducted 10 replications, there were  $7290 = 10 \times 3^6$  runs in all.

meanRewardFirm0	meanRewardFirm1	runningAverageBid	runningAverageBidSD
Min. :8069	Min. :8072	Min. :127.5	Min. :2.068
1st Qu.:8756	1st Qu.:8754	1st Qu.:132.3	1st Qu.:2.883
Median :8877	Median :8876	Median :133.4	Median :3.170
Mean :8875	Mean :8872	Mean :133.4	Mean :3.219
3rd Qu.:8994	3rd Qu.:8993	3rd Qu.:134.4	3rd Qu.:3.511
Max. :9538	Max. :9742	Max. :138.8	Max. :6.087

Table 3: Summary data, full factorial own-own experiment

The Wilcoxon signed rank test (`wilcox.test` in R) for the mean of the running average quantities against the null hypothesis of 133.3333 (the Cournot value) produces a p-value of 0.1347, so we do not reject the null hypothesis that the agents are settling on a quantity total equal to the Cournot value. (Recall that this is across a sample of 7,290 data points.) Applying the Wilcoxon test to the mean rewards of Firm0 and Firm1 yields p-value = 0.5189. We cannot reject the null hypothesis that each firm is obtaining, on average, an equal reward.

Given the symmetry of the factorial design, these results are as expected and serve primarily to increase our confidence in the implementation of the model and to provide a baseline for comparison. If we focus on the extreme case in which the factors for Firm0 are at their lowest ( $\text{epochLengthFirm1} = 24$ ,  $\text{deltaFirm0} = 2.4$ , and  $\text{epsilonFirm0} = 0.4$ ) and the factors for Firm1 are at their highest ( $\text{epochLengthFirm1} = 36$ ,  $\text{deltaFirm0} = 3.8$ , and  $\text{epsilonFirm0} = 1.0$ ) we get the summary results in Table 4.

It is evident from these data that the extreme ends of the parameter settings we examined do not perturb the basic findings. If we now use the Wilcoxon (rank sum) test (`wilcox.test` in R) to compare the mean rewards received by the two firms, we get a p-value

Firm0 parameters minimal, Firm1 parameters maximal			
meanRewardFirm0	meanRewardFirm1	runningAverageBid	runningAverageBidSD
Min. :8646	Min. :8667	Min. :130.2	Min. :2.899
1st Qu.:8847	1st Qu.:8875	1st Qu.:131.9	1st Qu.:2.961
Median :8936	Median :8907	Median :132.8	Median :3.195
Mean :8915	Mean :8954	Mean :132.5	Mean :3.182
3rd Qu.:9028	3rd Qu.:9012	3rd Qu.:133.2	3rd Qu.:3.295
Max. :9129	Max. :9389	Max. :134.1	Max. :3.733

Table 4: Summary data, extreme comparison, own-own experiment. `epochLengthFirm0 = 24`, `epochLengthFirm1 = 36`, `deltaFirm0 = 2.4`, `deltaFirm1 = 3.8`, `epsilonFirm0 = 0.4`, `epsilonFirm1 = 1.0`

of 0.9118, prohibiting us from rejecting the null hypothesis that the two firms are, on average, getting the same level of reward, despite using very different parameter values in PROBE AND ADJUST.

Finally, we regressed `meanRewardFirm0` on `epochLengthFirm0`, `epochLengthFirm1`, `deltaFirm0`, `deltaFirm1`, `epsilonFirm0`, and `epsilonFirm1`, including all interaction terms, using OLS. The residuals have a mean of  $-1.965907e-14$  and appear to be quite symmetrical about the mean. The Q-Q plot, Figure 3, indicates a reasonably good match with the normality assumption. Multiple R-Squared for the full model was 0.01412, with a p-value of 0.001058. *None* of the fitted coefficient values had an associated p-value below 0.3.

In summary, for this base case we find that the model settles reliably in the neighborhood of the Cournot outcome, with each player obtaining approximately equal returns. Further, these results are very stable to small to moderate changes in the model’s parameter settings.

## 5.2 Maximizing total market profits

In the previous section we discussed results obtained when all agents use the Own Returns policy. These agents, using PROBE AND ADJUST as their learning regime, set their `return` value at step 8 (Figure 1) to the profit (net of revenue and costs) they individually received during the episode. Agents following the Market Returns policy instead set their `return` values to the total profits of the industry (i.e., all players). This would seem a remarkably unselfish behavior. And so it is, yet it offers certain insights.

When both players maximize total market profits in a duopoly game, the total production settles into an oscillation around the monopoly solution of 100 (given our standard settings; see Figure 2). However, the players can, and normally do, have very different average profits depending on which player has the larger initial quantity. The reason is that

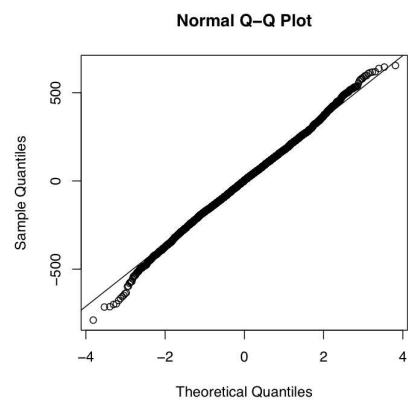


Figure 3: Q-Q plot of residuals from own-own regression

the players see the same signals on total market profitability and tend to move in the same direction. In a simulation where we started one player at 60 and the other at 40, after 1500 events the average profitability of the first player was around 11500 and the second at 8500. Also around 1500 episodes (rounds of play), the first player was producing around 65 and the second 35. At 13500 events the players were much closer, 55 and 45. There is random fluctuation in these runs. If the two players begin with the same initial base quantity, one typically produces more than the other. Which one is random, with small epoch lengths favoring more variation. Production levels can cross, even multiple times. Thus, although the total production stays close to the monopoly solution, the individual-player production levels are not identical. (This variation, as we shall see, averages out and may be deemed random.) A player starting out with a lower production level is at a strong disadvantage relative to the other player and, sometimes, relative to the Cournot solution.

Using the system clock to seed the random number generator we ran 100 repetitions of the standard configuration (above), with both players playing Market Returns. The average (over 100 repetitions) of the concluding running average for 6,600 episodes of play was 100.10 (sd 0.108). Firm0’s overall average reward was 9967.14 with a standard deviation of 628.92, which is very high compared to that for “Own Returns” (above). Firm1’s results were similar: 10019.57 (628.93).

Altruism has its limits. If one player maximizes market profits while the other maximizes its own returns, the total production settles on the monopoly solution. However, the first player is forced to exit the market and the second player settles on the monopoly quantity and reaps the profits of the monopoly solution. From these two simulations it is clear that being a good citizen without any regard to self interest is a weak strategy.

### 5.2.1 Detailed statistical analysis of the model results

We conducted a full factorial experiment with 10 repetitions with the following factors and settings: `numEpisodesToRun` (6600, 7600), `deltaFirm0` (2.4, 3.8), `deltaFirm1` (2.4, 3.8), `epsilonFirm0` (0.4, 1.0), `epsilonFirm1` (0.4, 1.0), `epochLengthFirm0` (24, 36), and `epochLengthFirm1` (24,36). The key summary information appears in Table 5.

meanRewardFirm0	meanRewardFirm1	runningAverageBid	runningAverageBidSD
Min. : 6119	Min. : 6253	Min. : 99.40	Min. :1.927
1st Qu.: 9328	1st Qu.: 9348	1st Qu.: 99.93	1st Qu.:2.365
Median : 9986	Median :10000	Median :100.00	Median :2.675
Mean : 9986	Mean :10000	Mean :100.00	Mean :2.650
3rd Qu.:10639	3rd Qu.:10658	3rd Qu.:100.08	3rd Qu.:2.889
Max. :13728	Max. :13864	Max. :100.62	Max. :3.530

Table 5: Summary data, full factorial market-market experiment

Comparing Tables 5 and 3, we note several points:

1. The `runningAverageBid` in the market-market condition settles tightly and symmetrically about 100, the monopoly quantity, instead of the Cournot quantity of the own-own case (see Table 3).
2. The `runningAverageBidSD` value is discernibly lower in the market-market case.
3. Each firm in the market-market case obtains a mean reward in the neighborhood of 10,000, compared to 8,900 in the own-own case. On average both firms do better by adopting the “cooperative” or “altruistic” policy (acting so as to maximize industry returns, rather than individual returns), than by adopting the “selfish” or “best response” policy.
4. The variation in mean reward obtained is much higher in the market-market case than in the own-own case.

Firm0 parameters minimal, Firm1 parameters maximal			
meanRewardFirm0	meanRewardFirm1	runningAverageBid	runningAverageBidSD
Min. : 7953	Min. : 9597	Min. : 99.75	Min. :2.612
1st Qu.: 9029	1st Qu.:10310	1st Qu.: 99.84	1st Qu.:2.641
Median : 9488	Median :10498	Median : 99.97	Median :2.671
Mean : 9370	Mean :10616	Mean : 99.93	Mean :2.672
3rd Qu.: 9675	3rd Qu.:10956	3rd Qu.:100.04	3rd Qu.:2.701
Max. :10389	Max. :12033	Max. :100.06	Max. :2.730

Table 6: Summary data, extreme comparison, market-market experiment. `epochLengthFirm0 = 24`, `epochLengthFirm1 = 36`, `deltaFirm0 = 2.4`, `deltaFirm1 = 3.8`, `epsilonFirm0 = 0.4`, `epsilonFirm1 = 1.0`

Comparing Table 6 with the analogous results for the own-own (Cournot) case in Table 4, we find some apparent differences for the players, depending on the players’ parameter

settings. In the own-own case, the minimized parameter player (Firm0 in Table 4) is at no apparent disadvantage. Table 6, however, shows Firm0 getting about 10% less on average than Firm1. Note that the mean quantity remains very near to the monopoly level. If we take the negative of these settings (with Firm0’s parameters maximal and Firm1’s minimal), we get the expected result: they settle near the monopoly quantity and Firm0 has the advantage. Evidently, mutual use of PROBE AND ADJUST is (under the current range of conditions) robust to parameter settings with respect to the total quantity, but individuals may gain comparative advantage through parameter settings. The nonparametric Wilcoxon rank-sum test (`wilcox.test` in R) yields a p-value of 0.0004871 for the null hypothesis of no difference between the two sets of returns. This coheres with the suggestion we drew from Table 6: there is a difference resulting from the two treatments.

### 5.3 Maximizing a mixture of market and own returns

Between the extremes of maximizing one’s own returns—Own Returns, §5.1—and maximizing on the market’s overall returns—Market Returns, §5.2—there are an infinite number of weighted combinations. Under the policy of play of Mixture of Market and Own Returns each firm individually has a parameter, `marketOwnMixture`  $\in [0, 1]$ , by which it combines observed market returns and its own returns from each episode of play. Using PROBE AND ADJUST, the firm finds its own reward, `ownReward`, in each episode as well as the average reward for firms in the market, `averageReward`. The firm calculates its `mixedReward` as

$$(1 - \text{marketOwnMixture}) \times \text{ownReward} + \text{marketOwnMixture} \times \text{averageReward}$$

and records this value on the associated up or down list for returns for mixtures, depending on whether its production quantity is above or below its current quantity. (See Figure 1, especially step 9.) Running under the standard conditions, except as noted, we consider the case with two firms in the market each using Mixture of Market and Own Returns as its policy of play, each using 0.5 as its value for `marketOwnMixture`. In a representative run, after 6600 episodes the running average total production is 120.837 with a standard error of 2.983. Recall that the monopoly production is 100 and the Cournot production 133.333. Firm0’s average reward (for episodes 6501–6600) is 9461.997 and Firm1’s is 9651.893. We repeated the experiment 100 times, setting the mixture for both players to 0.1 market (and 0.9 own) returns, and using the system clock to initialize the random number generator each time. Over the 100 trials the mean of the averages of the total quantity produced is 131.31 with a standard deviation of 1.47. Firm0’s average of its average rewards was 9020.88 (standard deviation of 191.22) and Firm1’s was 8994.42 (162.64).

These results are not sensitive to the random seed used, except of course for which firm comes out slightly ahead. The results also extend in the obvious way to more than 2 firms (our program handles up to 10, but this is easily changed to an arbitrary number). By mixing consideration of their own returns and the market’s returns the firms do better than the Cournot outcome but not as well as the monopoly position. The problem

is that the Mixture of Market and Own Returns policy of play is exploitable in the same way as the Market Returns policy. In a representative run, with Firm0 employing Own Returns as its policy of play and Firm1 sticking with Mixture of Market and Own Returns with its `marketOwnMixture` set to 0.5, we got the following outcomes. The running average total production is 128.065 with a standard error of 3.185. Firm0's average reward is 11,024.902 while Firm1's is 7,379.554. (These general results are not sensitive to the random seed used.) Things are a bit more equal if firm 0 uses Mixture of Market and Own Returns but sets its `marketOwnMixture` to 0.4. The running average production comes in at 124.215 with a standard error of 2.946. Firm0's average reward is 9741.348 while Firm1's is 9068.581. (Again, these results are not sensitive to the random seed used.)

At bottom, however, the mixture policy of play is vulnerable to exploitation.

#### 5.4 Maximizing market returns, constrained by own returns

A player pursuing the policy of Market Returns, Constrained by Own Returns (MR-COR) operates as follows. At the end of its epochs it assesses whether the mean of its production quantities during the epoch plus its  $\delta$  is less than the mean of all the production quantities during the epoch. (See Figure 1;  $\delta$  is the search range on each side of the base quantity that a player uses each episode. The program also allows use of  $\varepsilon$  instead of  $\delta$ .) If it is, the player increases its base quantity by  $\varepsilon$ . If it is not, then the player takes the market view and follows the Market Returns policy. The key feature of this policy is that the players have a sense of fair division of the market and if a player does not get what it perceives to be a fair share, it increases production. Two players on average split the market equally, the total production fluctuates around the monopoly total production of 100 and they each make an average profit of 10,000 (under the settings we are discussing).

The Market Returns, Constrained by Own Returns policy may be likened to TIT FOR TAT in Prisoner's dilemma in that if one player tries to take too much market share, the other responds by matching the increase in production. This means that any random increase in production by one is matched by the other and their actual production levels track together, whereas with market returns only, the production levels of the two tend to look like mirror images around 50. It is interesting to see what happens when one player, Firm1, plays MR-COR against the other player, Firm0, playing Own Returns. Typical results under the standard conditions are that Firm1's average profit during the final 1000 episodes is 8833, while Firm0's is 8977. In this type of case, the self interest of Firm0, maximizing its own returns, drives the production to an oscillation around the Cournot equilibrium and has an average episode profit of about 9000. The player that plays Market Returns, Constrained by Own Returns, Firm1, does slightly worse because it almost always produces less than the other player, as it has to be slightly forgiving on share to compensate for noise in the market-share results. Its profits are around 8800, slightly below what it would achieve at the Cournot equilibrium playing myopically. Long-sighted behavior has its risks.

Firm	Mean of Running Averages of Rewards to Firm	SD of Running Averages
0	1994.2025916855077	69.987676171652
1	1988.0676679163748	62.37538356784939
2	1987.651615420313	57.6836838380063
3	1987.8337170456355	68.10393678546707
4	1995.5662569019985	61.40921846900001
5	1995.3897636449321	54.92891914419867
6	1982.9621113765784	48.40642783370098
7	1990.0869668751468	55.82401403170985
8	2004.7913882302219	73.96700484493032
9	1988.9208194082335	56.140148147371825

Table 7: Results for 100 repetitions of 10 firms each playing Market Returns, Constrained by Own Returns (MR-COR) under the standard conditions. The monopoly quantity is 100, the Cournot quantity is 181.818. Across the 100 repetitions the mean (standard deviation) of the running average of the production quantity is 101.9725498548145 (0.6607808254556182).

#### 5.4.1 Detailed statistical analysis of the model results

We conducted a full factorial experiment with 10 repetitions with the following factors and settings: `numEpisodesToRun` (6600, 7600), `deltaFirm0` (2.4, 3.8), `deltaFirm1` (2.4, 3.8), `epsilonFirm0` (0.4, 1.0), `epsilonFirm1` (0.4, 1.0), `epochLengthFirm0` (24, 36), and `epochLengthFirm1` (24,36). Firm0 used the Own Returns update policy throughout, while Firm1 used Market Returns, Constrained by Own Returns. The key summary information appears in Table 8.

meanRewardFirm0	meanRewardFirm1	runningAverageBid	runningAverageBidSD
Min. :8273	Min. :7911	Min. :122.8	Min. :1.831
1st Qu.:8863	1st Qu.:8727	1st Qu.:131.8	1st Qu.:2.830
Median :8960	Median :8823	Median :133.1	Median :3.245
Mean :8979	Mean :8807	Mean :133.0	Mean :3.317
3rd Qu.:9081	3rd Qu.:8906	3rd Qu.:134.3	3rd Qu.:3.670
Max. :9663	Max. :9266	Max. :143.3	Max. :7.650

Table 8: Summary data, full factorial own vs. market-own experiment. Firm0 uses Own Returns and Firm1 uses Market Returns, Constrained by Own Returns.

With this mixture of policies the agents revert to something near, but slightly below, the Cournot solution. The Wilcoxon signed rank test yields a p-value of 2.341e-16 for the null hypothesis that the average value is 133.33333, leading to its convincing rejection. The test fails to reject the null hypothesis that the mean value is the slightly lower number 133.1, showing how close the solution is to the Cournot solution. The Wilcoxon test also rejects the null hypothesis that the two firms are obtaining their rewards from the same distribution (p-value < 2.2e-16). Firm0, using Own Returns, has a discernible advantage over Firm1, using Market Returns, Constrained by Own Returns, but is this Firm0's best policy?

Table 9 presents summary data for the full-factorial experiment (with 30 replications) when both players use the Market Returns, Constrained by Own Returns (MR-COR) policy of play. Notice that both players do very well, with mean rewards fully 1000 higher than in the Own Returns versus Market Returns, Constrained by Own Returns case, summarized in Table 8. Comparing the present case (MR-COR, Table 9) with the Market-Market (both altruistic) case, summarized in Table 5, it would seem that little or nothing is lost by playing Market Returns, Constrained by Own Returns. MR-COR is a robust policy of play. Agents have little to lose by using it and much to gain if everyone uses it.

meanRewardFirm0	meanRewardFirm1	runningAverageBid	runningAverageBidSD
Min. : 9723	Min. : 9739	Min. : 99.42	Min. :1.740
1st Qu.: 9910	1st Qu.: 9912	1st Qu.:100.05	1st Qu.:2.437
Median : 9992	Median : 9994	Median :100.15	Median :2.672
Mean : 9992	Mean : 9994	Mean :100.16	Mean :2.644
3rd Qu.:10072	3rd Qu.:10074	3rd Qu.:100.26	3rd Qu.:3.014
Max. :10246	Max. :10260	Max. :101.00	Max. :3.568

Table 9: Summary data, full factorial, market-own vs. market-own experiment. Both Firm0 and Firm1 use Market Returns, Constrained by Own Returns.

Table 10 presents summary information for extreme parameter settings, with Firm0 as usual having the low settings and Firm1 the high. From the table, it appears that Firm0 does somewhat better than Firm1. This is confirmed by the Wilcoxon rank sum test on `meanRewardFirm0` and `meanRewardFirm1`. The p-value for the null hypothesis of their arising from the same distribution is  $< 2.2e-16$ . In fact, the smallest value of `meanRewardFirm0` exceeds the largest value of `meanRewardFirm1`.

<code>meanRewardFirm0</code>	<code>meanRewardFirm1</code>	<code>runningAverageBid</code>	<code>runningAverageBidSD</code>
Min. :10073	Min. :9773	Min. : 99.86	Min. :2.557
1st Qu.:10130	1st Qu.:9819	1st Qu.:100.07	1st Qu.:2.639
Median :10148	Median :9838	Median :100.16	Median :2.683
Mean :10147	Mean :9839	Mean :100.16	Mean :2.678
3rd Qu.:10168	3rd Qu.:9856	3rd Qu.:100.27	3rd Qu.:2.717
Max. :10212	Max. :9913	Max. :100.44	Max. :2.795

Table 10: Summary data, extreme market-own vs. market-own experiment. Both Firm0 and Firm1 use Market Returns, Constrained by Own Returns. `epochLengthFirm0` = 24, `epochLengthFirm1` = 36, `deltaFirm0` = 2.2, `deltaFirm1` = 3.8, `epsilonFirm0` = 0.4, `epsilonFirm1` = 1.0

## 5.5 Number effects

For the most part, we have focused so far on exploring and establishing the robustness of PROBE AND ADJUST and on presenting results in the case of oligopolies having two firms. In this and the following section we assume the robustness of the basic model and turn our attention to new issues. What happens when the number of firms increases beyond two? The Cournot model from standard economic theory teaches that the Cournot equilibrium will change with increasing numbers of firms, moving asymptotically towards the competitive solution. The theory predicts, however, that the *outcome* reached will continue to be the Cournot solution. That is, the standard theory asserts that there is no *number effect*; the Cournot solution will be the outcome regardless of the number of firms in the market. Considerable experimental work with human subjects, nicely summarized and extended by Huck et al. [Huck et al., 2004], does find, to the contrary, number effects in repeated play by human subjects with fixed counter-players. The title of [Huck et al., 2004] summarizes the experimental findings: “Two are few and four are many.” That is, with four or more firms the Cournot quantity is reached or exceeded, and with two firms there is often evidence of collusion, with  $Q$ , the total production quantity, reduced in the direction of the monopoly level. What happens under PROBE AND ADJUST? Recall the results reported in Table 1 for PROBE AND ADJUST when all players use the Own Returns policy: there are no number effects. In order to facilitate comparison, we now report results with

`InitPIntercept` = 100, `InitDSlope` = 1.0, and `unitCostFirm0` = `unitCostFirm1` = 1.0, which duplicates the demand function and cost structure used in the experiments of Huck et al. (Note that in terms of expression (1),  $a = \text{InitPIntercept} = 100$ ;  $\text{slope} = \text{InitDSlope} = 1.0$ .) The main outcome statistic used in [Huck et al., 2004, page 439] is what they call  $r$ , the ratio of the (mean of the) total production quantity,  $Q$ , to the Cournot solution quantity,  $Q^C$  (or  $Q^N$  in their notation).

$$r = \frac{\bar{Q}}{Q^C(n, k)} \quad (20)$$

(We assume with [Huck et al., 2004] that all firms have the same costs,  $k$ ; cf. expression 18.) Thus,  $r$  values less than 1 indicate a degree of collusion. In `oligopolyPutQuantity.nlogo`,  $r$  is renamed `ratioOfferedCournot` internally and `ratioOC` for display on the user interface panel. Let

$$r_C^M(n, k) = \frac{Q^M(k)}{Q^C(n, k)} \quad (21)$$

Then, for the inverse demand function of [Huck et al., 2004] we have:

$n = \text{number of firms}$	$Q^C(n, 1)$	$Q^M(k)$	$r_C^M(n, 1)$
2	66.00	49.50	0.750
3	74.25	49.50	0.667
4	79.20	49.50	0.625
5	82.50	49.50	0.600
10	90.00	49.50	0.550

We note that  $r$  may be a misleading indicator for our special purposes, since the significance of its value varies with  $n$ . A measure that adjusts for the number of players is

$$r' = \frac{Q^C(n, k) - Q}{Q^C(n, k) - Q^M(k)} = \frac{Q^C - Q}{Q^C - Q^M} \quad (22)$$

Values near 1 indicate very high collusion (with quantities near the monopoly level), while values near 0 would indicate lack of collusion (and quantities near the Cournot level). We will proceed with the discussion in terms of both  $r$  and  $r'$ .

$n$	$Q^C(n, 1)$	$\bar{Q}$	$\bar{r}$	$\bar{r}'$
2	66.00	60.44 (7.05)	0.91 <sup>†</sup>	0.34
3	74.25	72.59 (4.53)	0.98	0.07
4	79.20	80.67 (4.85)	1.02	-0.05
5	82.50	88.43 (8.80)	1.07	-0.18

Table 11: Summary of Huck et al.’s experimental data [Huck et al., 2004, page 441].  $n$  = number of suppliers in the market.  $\bar{Q}$  = average total quantity offered. (Standard deviations in parentheses.) Averages are over episodes 17–25. <sup>†</sup> As reported in [Huck et al., 2004, page 441]. We note that  $60.44/66.0 = 0.9157575\dots$ . For computing  $r'$ ,  $Q^M = 49.5$ .

Huck et al. performed a meta-analysis on the prior human experiments that investigate number effects in Cournot markets with fixed counter-players [Huck et al., 2004]. They found, in aggregate, a modest number effect. While collusion may occur with 2 players, it is reduced or disappears with increasing numbers of players. To this Huck et al. added their own experimental data, which we summarize in Table 11. Subjects offer quantities in 25 rounds of play (episodes). Allowing for some learning, we use the Huck et al. data for rounds 17–25. (Huck et al. also report data for rounds 1–25. The results are not materially different.) Points arising on the Huck et al. data:

1.  $\bar{r}$  increases uniformly with  $n$ . Since at  $Q = Q^C(n, k)$ ,  $r = 1$ , only the result for  $n = 2$  indicates collusion. Huck et al. find the increase statistically significant. Note also that for  $n = 5$  the  $\bar{r}$  and  $\bar{Q}$  values suggest that the subjects systematically offered more than the Cournot amount.
2.  $\bar{r}'$  (not reported by Huck et al.) decreases uniformly with increasing  $n$  and is in apparent broad agreement with  $\bar{r}$ .
3. The standard deviation of  $\bar{Q}$  increases uniformly across  $n = 3, 4, 5$ . It is, however, comparatively high for  $n = 2$ . This suggests (only) that subjects may have used somewhat different decision procedures for  $n = 2$  and  $n \neq 2$ , and that when  $n = 2$  the subjects may have been somewhat more exploratory, perhaps sensing the possibility of collusion.

Table 12 summarizes number-effect results obtained with `oligopolyPutQuantity.nlogo`. We remind the reader that the interpretation of the standard deviations in Table 12 is different from that in Table 11. In the latter case, the given standard deviation is the usual standard deviation of  $Q$ . In Table 12, individual  $Q$  values represent the mean of the total quantities offered for episodes 5601–6600 and  $\bar{Q}$  is the mean (over 30 replications) of these  $Q$  values.  $\bar{Q}$  is, then, a mean of means. The standard deviation values given are the means of the standard deviations of the  $Q$  values.

line no.	$N$	policies	$\bar{Q}$	$\bar{r}$	$\bar{r}'$
1	2	all Own Returns	66.12 (3.24)	1.00	-0.01
2	3	all Own Returns	74.38 (3.86)	1.00	-0.00
3	4	all Own Returns	79.45 (4.35)	1.00	-0.00
4	5	all Own Returns	82.45 (4.72)	1.00	0.00
5	10	all Own Returns	90.90 (6.42)	1.01	-0.02
6	2	all MR-COR	49.39 (2.70)	0.75	0.98
7	3	all MR-COR	49.71 (3.39)	0.67	0.95
8	4	all MR-COR	50.14 (4.01)	0.63	0.93
9	5	all MR-COR	50.58 (4.71)	0.61	0.91
10	10	all MR-COR	52.66 (6.93)	0.59	0.83
11	2	1MR-COR :: 1Mixture 50:50	59.52 (3.25)	0.90	0.38
12	3	1MR-COR :: 2Mixture 50:50	66.37 (3.82)	0.89	0.31
13	4	1MR-COR :: 3Mixture 50:50	70.97 (4.35)	0.90	0.26
14	5	1MR-COR :: 4Mixture 50:50	74.48 (4.81)	0.90	0.23
15	10	1MR-COR :: 9Mixture 50:50	84.20 (6.47)	0.94	0.13
16	2	all 30:70 Mixture	62.57 (3.34)	0.95	0.20
17	2	all 50:50 Mixture	59.35 (3.08)	0.90	0.39
18	3	all 50:50 Mixture	66.09 (3.75)	0.89	0.32
19	4	all 50:50 Mixture	70.92 (4.33)	0.90	0.27
20	5	all 50:50 Mixture	74.46 (4.74)	0.90	0.23
21	10	all 50:50 Mixture	84.28 (6.52)	0.94	0.13

Table 12: Summary of PROBE AND ADJUST data on number effects. (Standard deviations in parentheses.) Averages are over 30 replications.

Points arising on the Table 12 data:

1. When all players use the Own Returns policy (lines 1–5) neither  $\bar{r}$  nor  $\bar{r}'$  shows any evidence of a number effect, further verifying that this policy replicates the Cournot equilibrium.

When all players use the Market Returns, Constrained by Own Returns (MR-COR) policy (lines 6–10), both  $\bar{r}$  and  $\bar{r}'$  are uniformly decreasing in  $n$ . The high value of  $\bar{r}'$  when  $n = 2$  indicates a high degree of collusion. (Note that  $Q^C(2, 1) = 66.0$ , while  $Q^M = 49.50$  and  $\bar{Q} = 49.39$  with  $n = 2$ .) As  $n$  increases,  $\bar{r}'$  also decreases, but even at  $n = 10$  the average quantity on offer is displaced 83% of the distance away from the Cournot quantity and towards the monopoly quantity,  $Q^M$ .

When all players use the Mixture of Market and Own Returns, in a 50:50 combination (lines 17–21),  $\bar{r}$  appears to be constant for  $n = 2, 3, 4, 5$ , but higher for  $n = 10$ .  $\bar{r}'$ , however, is uniformly decreasing, indicating a number effect and reduced collusion. Essentially the same result obtains if Firm0 uses MR-COR and any other firms in the market use Mixture of Market and Own Returns, in a 50:50 combination (lines 11–15).

Interestingly, the relative advantage or disadvantage of the MR-COR policy is also a function of  $n$ . At low values, MR-COR is disadvantaged, but at  $n = 5$  and higher it earns more than the typical firm using Mixture of Market and Own Returns, in a 50:50 combination. See the following table:

Policy Case	Firm0: MR-COR meanRewardFirm0	Firm1: Mixture 50:50 meanRewardFirm1
1MR-COR :: 1Mixture 50:50	1146.0	1193.0
1MR-COR :: 2Mixture 50:50	708.0	717.3
1MR-COR :: 3Mixture 50:50	491.4	498.6
1MR-COR :: 4Mixture 50:50	365.3	359.4
1MR-COR :: 9Mixture 50:50	129.5	121.47

2. Line 11 contains values that are quite close to those reported by Huck et al. for  $n = 2$  (see Table 11). The data in the two tables do not track closely as  $n$  increases. The ordering is in excellent alignment, however, suggesting a good match could be found by a simple transformation.
3.  $\bar{r}'$  decreases uniformly with  $n$  (except in the all Own Returns case), while  $\bar{r}$  is flat for  $n = 2, 3, 4, 5$  (except for the all MR-COR case).  $\bar{r}$ 's flatness is an artifact, since it is sensitive to the absolute level of  $Q$ , which varies with  $n$ .  $\bar{r}'$  is a better indicator.
4. The standard deviation of  $Q$  increases uniformly with  $n$ , without exception. This is hardly surprising, given that PROBE AND ADJUST is a stochastic exploration procedure, undertaken with some independence by the players. A similar explanation

suggests itself for the human data. How much of that stochasticity is systematic (actually part of a learning policy as in PROBE AND ADJUST) and how much is simply error is a fascinating question for future research.

## 5.6 Differential costs

Until now we have made the literature's standard assumption of equal (usually zero) costs for the players. What happens if the players have different costs? We investigate cases in which firms' costs are proportional to the quantities they produce. As usual, we report results from specific, but representative, runs. Here we assume that runs proceed for 6600 episodes of play. This is more than ample for the system to settle.

We revert now to our original setup with  $a = \text{InitPIntercept} = 400.0$  and  $\text{slope} = \text{InitDSlope} = 2.0$ . Let us assume (without loss of generality; the qualitative results are robust to this assumption) that in a duopoly Firm0 has a unit cost of 10 and Firm1 50.  $Q^C(2, [10, 50])$  is then 123.33. (See discussion and formulas in §4.) With these costs the individual monopoly quantities are  $Q_0^M = 97.5$  and  $Q_1^M = 87.5$ . In fact,  $Q$  settles (with some variation;  $\text{SD} \approx 3.2$ ) very near 123.33 when both firms use the Own Returns policy of play. Running for 6,600 episodes of play, in a typical run the last 1,000 episodes of play yield an average  $Q$  of 123.067, with standard deviation 2.895. Firm0, the low cost producer, obtains an average reward of 10,068.879, while Firm1 achieves only 5,498.693.

Similarly, in an experiment with 100 repetitions, using the system clock to seed the random number generator, Firm0's costs were set to 10 and Firm1's to 0. The Cournot quantity is 131.667. The mean (standard deviation) running average production across the 100 repetitions was 131.683 (1.591). Firm0 averaged a reward of 8026.41 (164.09), while Firm1 got 9305.65 (176.60).

Switching to the MR-COR policy for both players, in a typical run, with  $\text{unitCostFirm0}=10$ ,  $\text{unitCostFirm1}=50$ , and  $\text{refereceBidGroup}=\text{All Bids}$ , the last 1,000 episodes of play yield an average  $Q$  of 97.638 (with standard deviation 3.047), which is very close to Firm0's monopoly quantity. Firm0, the low cost producer, obtains an average reward of 9611.584, while Firm1 achieves 7453.02. The two firms have jointly extracted more wealth from the market, but Firm0 has paid a penalty. The example is extreme, however, for it assumes that Firm1's unit costs are five times those of Firm0.

Assume now that Firm1's unit cost is 20 and Firm0's remains at 10, still leaving a substantial cost advantage to Firm0. (Firm1's monopoly  $Q$ ,  $Q_1^M$ , is now 95.0, close to Firm0's of 97.5. The Cournot quantity is  $Q^C(2, [10, 20]) = 128.33$ .) Let both firms use the Own Returns policy of play. In a typical run, using 6,600 episodes of play, the last 1,000 episodes of play yield an average  $Q$  of 129.552, with standard deviation 3.308. Firm0, the low cost producer, obtains an average reward of 8,937.33, while Firm1 achieves 7,386.882. Switching to MR-COR for both players, the last 1,000 episodes of play yield an average  $Q$  of 98.102 (with standard deviation 2.702), which is very close to Firm0's monopoly quantity. Firm0, the low cost producer, obtains an average reward of 9640.072, while Firm1 achieves

8873.914. Switching to Own Returns for Firm0, MR-COR for Firm1, in a typical run the last 1,000 episodes of play yield an average  $Q$  of 131.358 (with standard deviation 3.215). This is very far from Firm0’s monopoly quantity but close to the Cournot quantity. Firm0, still the low cost producer, obtains an average reward of 8431.236, while Firm1 achieves 7617.516.

In sum even with a 2:1 unit cost advantage for one player, the policy pair (MR-COR,MR-COR) Pareto dominates (Own Returns, MR-COR), (MR-COR, Own Returns), and (Own Returns, Own Returns).

Again, we emphasize that these results are typical. They hold up under a broad range of settings and are quite robust to changes in the random number stream.

## 6 Discussion

Recognizing that measuring success by market returns only is not a viable policy because the player is too easily exploited, we compare the firm-only (Own Returns policy) and the firm-and-market (MR-COR policy) measures of success in the following payoff table for a  $2 \times 2$  game in strategic form. Table 13 presents the results approximately, but in strategic form.

	firm-only	firm-and-market
firm-only	(8,900, 8,900)	(9,000, 8,800)
firm-and-market	(8,800, 9,000)	(10,000, 10,000)

Table 13: Payoffs from the strategies as measured by average returns after settlement; both firms have 0 costs

Notice that the relationships among the payoffs across the choice of objective functions is a Stag Hunt game.<sup>4</sup> The payoffs in the Stag Hunt game look like the payoffs in Prisoner’s Dilemma, except for some key differences. The players are each better off if both cooperate (play MR-COR, or hunt stag in the case of the Stag Hunt) versus neither cooperate (play Own Returns, or hunt hare) as in Prisoner’s Dilemma. If one player cooperates and the other does not, both players do worse than if both cooperate. In Prisoner’s Dilemma the noncooperative player is better off if the other player cooperates. The hitch, here and in the Stag Hunt generally, is that players may be led to mutual hunting of hare by considerations of risk. The experimental literature on the Stag Hunt (e.g. [Battalio et al., 2001]; see [Camerer, 2003] for reviews of many experiments) has tended to find that in repeated play subjects are led to the risk-dominant strategy of hunting hare. These findings, however, are based on repeated play by *varying* counter-players, in distinction to *fixed*, counter-players. Our concern is with the latter, with markets containing a fixed number of participants, who

---

<sup>4</sup>See [Skyrms, 2001, Skyrms, 2004] for discussion of other circumstances in which what is seemingly a Prisoner’s Dilemma is transformed into a Stag Hunt game.

produce quantities over time and who have to learn to live with each other. Framing the problem as one of *learning in policy space* (Which policy should I use in setting production? Should it be Own Returns or Market Returns or Market Returns and Own Returns?), it is clear that learning to cooperate, learning to hunt stag, is amply achievable under simple learning regimes [Kimbrough et al., 2005a, Kimbrough et al., 2005b]. Nor is it implausible to think that real players would figure out and implement their collective interests. In short, if we were to allow the players to select the policy as well as search for the maximum return with the given policy, we would expect that the players would choose the firm-and-market strategy, especially if we bring discounting into this repeated game among fixed players. The firm and market policy works only if all players either accept their proportionate share or are willing to accept shares that add up to less than 1. For example, if each player believes it should have 5% above the average production, the firms in the simulation will fight for market share, driving themselves into losses rather than profits. Thus, we do not propose that this policy is actually used or should be used by firms. However, it points to the potential of a more nuanced policy that brings in industry interests and not just firm interests into capacity and production strategies. The policies could be explored using the bargaining frameworks examined by Dawid and Dermietzel [Dawid and Dermietzel, 2006] and Carpenter [Carpenter, 2002].

## 7 Conclusions

In this paper we have developed a model—PROBE AND ADJUST—of an agent that explores its environment and uses that exploration to improve its performance by adjusting a set of continuous parameters. This behavior is an abstraction of typical managerial decision making and is consistent with the notions of continual improvement and of a satisficing player that learns and improves. We emphasize that the required knowledge and computational capabilities for PROBE AND ADJUST are quite credibly available to real agents. We have shown in simulations that this model of an agent reproduces the classic Cournot results in oligopoly theory, under certain assumptions (e.g., use of the Own Returns policy). We have also shown that this model can explain the emergence of tacit collusion (e.g., when players all use MR-COR as their policy of play). Thus, we have a starting point for exploring alternatives to the decision model embedded in classical economic theory and have a more realistic starting point for looking at issues such as market power. Firms operate in complex environments and there are many competing interests within and without the firm. Agents can be given objectives that are more realistic in that organizations both compete with other organizations in some dimensions and cooperate with these same organizations in others (within the bounds of the law) as described in Brandenburger and Nalebuff [Brandenburger and Nalebuff, 1996]. We see that it is possible to engage in tacit collusion by taking into account the interest of the industry as well as the firm while not engaging in explicit price fixing. Thus, being a good corporate citizen can pay. By test-

ing alternative objectives for firms it is possible to represent the richer relationships that managers have to deal with and observe the consequences in the marketplace. This paper opens up several avenues of research. First, players should be able to choose among success measures (e.g., own returns, market returns) to see which ones emerge as most effective for enhancing firm profitability. By doing this, it is possible to see what measures of success emerge in the context of repeated play. Second, the agency problem between the firms' owners and managers can be placed in a larger context, to see how those choices impact the industry as well as the firm.

## References

- [Alkemade et al., 2006] Alkemade, F., La Poutr, H., and Amman, H. M. (2006). Robust evolutionary algorithm design for socio-economic simulation. *Computational Economics*, 28(4):355–370.
- [Arifovic, 1994] Arifovic, J. (1994). Genetic algorithm learning and the cobweb model. *Journal of Economic Dynamic and Control*, 18(1):3–28.
- [Arifovic and Maschek, 2005] Arifovic, J. and Maschek, M. K. (2005). Social vs. individual learning: What makes a difference? Working paper, Simon Fraser University.
- [Axelrod, 1984] Axelrod, R. (1984). *The Evolution of Cooperation*. Basic Books, Inc., New York, NY.
- [Barr and Saraceno, 2005] Barr, J. and Saraceno, F. (2005). Cournot competition, organization and learning. *Journal of Economics Dynamics and Control*, 29(1):277–295.
- [Battalio et al., 2001] Battalio, R., Samuelson, L., and Van Huyck, J. (2001). Optimization incentives and coordination failure in laboratory stag hunt games. *Econometrica*, 69(3):749–764.
- [Brandenberger and Nalebuff, 1996] Brandenberger, A. M. and Nalebuff, B. J. (1996). *Co-opetition: A Revolution Mindset That Combines Competition and Cooperation : The Game Theory Strategy That's Changing the Game of Business*. Doubleday, New York, NY.
- [Brenner, 1999a] Brenner, T., editor (1999a). *Computational Techniques for Modelling Learning in Economics*. Kluwer Academic Publishers, Boston, MA.
- [Brenner, 1999b] Brenner, T. (1999b). *Modelling Learning in Economics*. Edward Elgar, Cheltenham, UK.
- [Brenner, 2006] Brenner, T. (2006). Agent learning representation: Advice on modelling economic learning. In Tesfatsion, L. and Judd, K. L., editors, *Handbook of Computational*

- Economics, Volume 2, Agent-Based Computational Economics*, Handbooks in Economics, pages 895–948. North-Holland, Amsterdam, The Netherlands.
- [Bruun, 2006] Bruun, C., editor (2006). *Advances in Artificial Economics*. Number 584 in Lecture Notes in Economics and Mathematical Systems. Springer, Berlin, Germany.
- [Bunn and Oliveira, 2003] Bunn, D. W. and Oliveira, F. (2003). Evaluating individual market power in electricity markets via agent-based simulation. *Annals of Operations Research*, 121:57–78.
- [Camerer, 2003] Camerer, C. F. (2003). *Behavioral Game Theory: Experiments in Strategic Interaction*. Russell Sage Foundation and Princeton University Press, New York, NY and Princeton, NJ.
- [Carpenter, 2002] Carpenter, J. P. (2002). Evolutionary models of bargaining: Comparing agent-based computational and analytical approaches to understanding convention evolution. *Computational Economics*, 19(1):25–49.
- [Dawid and Dermietzel, 2006] Dawid, H. and Dermietzel, J. (2006). How robust is the equal split norm? On the de-stabilizing effect of responsive strategies. *Computational Economics*, 28:371–397.
- [Duffy, 2006] Duffy, J. (2006). Agent-based models and human subject experiments. In Tesfatsion, L. and Judd, K. L., editors, *Handbook of Computational Economics, Volume 2, Agent-Based Computational Economics*, Handbooks in Economics, pages 949–1012. North-Holland, Amsterdam, The Netherlands.
- [Enriken and Wan, 2005] Enriken, R. and Wan, S. (2005). Agent-based simulation of an automatic mitigation procedure. In *Proceedings of the 38th Hawaii International Conference on System Sciences*.
- [Hommes et al., 2003] Hommes, C. H., Sonnemans, J., Tuinstra, J., and Velden, H. v. (2003). Learning in cobweb experiments. Working paper TI 2003-020/1, University of Amsterdam, Tinbergen Institute.
- [Huck et al., 2003] Huck, S., Normann, H.-T., and Oechssler, J. (2003). Zero-knowledge cooperation in dilemma games. *Journal of Theoretical Biology*, 220:47–54.
- [Huck et al., 2004] Huck, S., Normann, H.-T., and Oechssler, J. (2004). Two are few and four are many: number effects in experimental oligopolies. *Journal of Economic Behavior & Organization*, 53:435–446.
- [Kagel and Roth, 1995] Kagel, J. H. and Roth, A. E., editors (1995). *The Handbook of Experimental Economics*. Princeton University Press, Princeton, NJ.

- [Kimbrough and Lu, 2005] Kimbrough, S. O. and Lu, M. (2005). Simple reinforcement learning agents: Pareto beats Nash in an algorithmic game theory study. *Information Systems and e-Business*, 3(1):1–19. <http://dx.doi.org/10.1007/s10257-003-0024-0>.
- [Kimbrough et al., 2005a] Kimbrough, S. O., Lu, M., and Kuo, A. (2005a). A note on strategic learning in policy space. In Kimbrough, S. O. and Wu, D. J., editors, *Formal Modelling in Electronic Commerce: Representation, Inference, and Strategic Interaction*, pages 463–475. Springer, Berlin, Germany.
- [Kimbrough et al., 2005b] Kimbrough, S. O., Lu, M., and Murphy, F. (2005b). Learning and tacit collusion by artificial agents in Cournot duopoly games. In Kimbrough, S. O. and Wu, D. J., editors, *Formal Modelling in Electronic Commerce*, pages 477–492. Springer, Berlin, Germany.
- [Kuenne, 1998] Kuenne, R. E. (1998). *Price and nonprice rivalry in oligopoly: the integrated battleground*. St. Martins Press, New York, NY.
- [Marks, 2006] Marks, R. (2006). Market design using agent-based models. In Tesfatsion, L. and Judd, K. L., editors, *Handbook of Computational Economics, Volume 2, Agent-Based Computational Economics*, Handbooks in Economics, pages 1339–1380. North-Holland, Amsterdam, The Netherlands.
- [Marks and Midgley, 2006] Marks, R. E. and Midgley, D. F. (2006). Using evolutionary computing to explore social phenomena: Modeling the interactions between consumers, retailers and brands. Working paper, Australian Graduate School of Management.
- [Midgley et al., 1997] Midgley, D. F., Marks, R. E., and Cooper, L. G. (1997). Breeding competitive strategies. *Management Science*, 43(3):257–275.
- [Nagle and Hogan, 2006] Nagle, T. T. and Hogan, J. E. (2006). *The strategy and tactics of pricing: a guide to growing more profitably*. Pearson/Prentice Hall, Upper Saddle River, NJ, 4th edition.
- [Pyka and Fagiolo, 2005] Pyka, A. and Fagiolo, G. (2005). Agent based modeling: A methodology for neo-Shumpeterian economics. Working paper 272, University of Augsburg.
- [R Development Core Team, 2007] R Development Core Team (2007). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.
- [Riechmann, 2002] Riechmann, T. (2002). Cournot or Walras? Agent based learning, rationality, and long run results in oligopoly games. Discussion paper 261, University of Hannover, Faculty of Economics, Königsworther Platz 1, 30 167 Hannover, Germany.

- [Sallans et al., 2003] Sallans, B., Pfister, A., Karatzoglou, A., and Dorffner, G. (2003). Simulation and validation of an integrated markets model. *Journal of Artificial Societies and Social Simulation*, 6(4):<http://jasss.soc.surrey.ac.uk/6/4/2.html>.
- [Selten et al., 1997] Selten, R., Mitzkewitz, M., and Uhlich, G. R. (1997). Duopoly strategies programmed by experienced players. *Econometrica*, 65(3):517–555.
- [Skyrms, 2001] Skyrms, B. (2001). The stag hunt. *Proceedings and Addresses of the American Philosophical Association*, 75(2):31–41.
- [Skyrms, 2004] Skyrms, B. (2004). *The Stag Hunt and the Evolution of Social Structure*. Cambridge University Press, Cambridge, UK.
- [Tsfatsion, 2006] Tsfatsion, L. (2006). Agent-based computational economics: A constructive approach. In Tsfatsion, L. and Judd, K. L., editors, *Handbook of Computational Economics, Volume 2, Agent-Based Computational Economics*, Handbooks in Economics, pages 831–880. North-Holland, Amsterdam, The Netherlands.
- [Tsfatsion and Judd, 2006] Tsfatsion, L. and Judd, K. L., editors (2006). *Handbook of Computational Economics, Volume 2, Agent-Based Computational Economics*. Handbooks in Economics. North-Holland, Amsterdam, The Netherlands.
- [Vriend, 2000] Vriend, N. J. (2000). An illustration of the essential difference between individual learning and social learning and its consequences for computational analysis. *Journal of Economic dynamics and Control*, 24:1–19.
- [Waltman and Kaymak, 2005] Waltman, L. and Kaymak, U. (2005). Q-learning agents in a Cournot oligopoly model. Working paper, Erasmus University, Faculty of Economics, Rotterdam.
- [Winston, 2004] Winston, W. L. (2004). *Operations Research Applications and Algorithms*. Brooks/Cole, Belmont, CA, fourth edition.