Strategic Pricebot Dynamics

Amy R. Greenwald, Jeffrey O. Kephart, and Gerald J. Tesauro

IBM Institute for Advanced Commerce
IBM Thomas J. Watson Research Center
Yorktown Heights, New York, USA

amygreen@cs.nyu.edu, {kephart,tesauro}@watson.ibm.com

Abstract

Shopbots are software agents that automatically query multiple sellers on the Internet to gather information about prices and other attributes of consumer goods and services. Rapidly increasing in number and sophistication, shopbots are helping more and more buyers minimize expenditure and maximize satisfaction. In response at least partly to this trend, it is anticipated that sellers will come to rely on *pricebots*, automated agents that employ price-setting algorithms in an attempt to maximize profits. This paper reaches toward an understanding of strategic pricebot dynamics.

More specifically, this paper is a comparative study of four candidate price-setting strategies that differ in informational and computational requirements: game-theoretic pricing (GT), myoptimal pricing (MY), derivative following (DF), and Q-learning (Q). In an effort to gain insights into the tradeoffs between practicality and profitability of pricebot algorithms, the dynamic behavior that arises among homogeneous and heterogeneous collections of pricebots and shopbot-assisted buyers is analyzed and simulated.

In homogeneous settings — when all pricebots use the same pricing algorithm — DFs outperform MYs and GTs. Investigation of heterogeneous collections of pricebots, however, reveals an incentive for individual DFs to deviate to MY or GT. The Q strategy exhibits superior performance to all the others since it learns to predict and account for the long-term consequences of its actions. Although the current implementation of Q is impractically expensive, techniques for achieving similar performance at greatly reduced computational cost are under investigation.

1 Introduction

Shopbots — software agents that automatically query multiple on-line vendors to gather information about prices and other attributes of consumer goods and services — herald a future in which automated agents are an essential component of electronic commerce [2, 5, 11, 16]. Shopbots outperform and out-inform humans, providing extensive product coverage in just a few seconds, far more than a patient and determined human shopper could ever achieve, even after hours of manual search. Rapidly increasing in number and sophistication, shopbots are helping more and more buyers minimize expenditure and maximize satisfaction.

Since the launch of BargainFinder [12], a CD shopbot, on June 30, 1995, the range of products represented by shopbots has expanded dramatically. A shopbot available at `shopper.com` claims to compare 1,000,000 prices on 100,000 computer-oriented products. Another shopbot, `DealPilot.com` (formerly `acses.com`), gathers, collates and sorts prices and expected delivery times of books, CDs, and movies offered for sale on-line. One of the most popular shopbots, `mysimon.com`, compares office supplies, groceries, toys, apparel, and consumer electronics, just to name a few of the items on its product line. As the range of products covered by shopbots expands to include more complex products such as consumer electronics, the level of shopbot sophistication is rising accordingly. On August 16th, 1999, `mysimon.com` incorporated technology that, for products with multiple features such as digital cameras, uses a series of questions to elicit multi-attribute utilities from buyers, and then sorts products according to the buyer's specified utility. Also on that day, `lycos.com` licensed similar technology from `frictionless.com`.

Shopbots are clearly a boon to buyers who use them. Moreover, when shopbots become adopted by a sufficient portion of the buyer population, it seems likely that sellers will be compelled to decrease prices and improve quality, benefiting even those buyers who do not shop with bots. How the widespread utilization of shopbots might affect sellers, on the other hand, is not quite so apparent. Less established sellers may welcome shopbots as an opportunity to attract buyers who might not otherwise have access to information about them, but more established sellers may feel threatened. Some larger players have even been known to deliberately block automated agents from their web sites [4]. This practice seems to be waning, however; today, sellers like `Amazon.com` and `BarnesandNoble.com` tolerate queries from agents such as `DealPilot.com` on the grounds that buyers take brand name and image as well as price into account as they shop.

As more and more buyers are relying on shopbots to increase their awareness about products and prices, it is becoming advantageous for sellers to increase flexibility in their pricing strategies, perhaps by using *pricebots* — automated agents that employ price-setting algorithms in an attempt to maximize profits. A primitive example of a pricebot is available at `books.com`, an on-line bookseller. When a prospective buyer expresses interest in a given book, `books.com` automatically queries `Amazon.com`, `Borders.com`, and `BarnesandNoble.com` to determine the price that is being offered at those sites. `books.com` then slightly undercuts the lowest of the three quoted prices, typically by 1% of the retail price. Such dynamic pricing on millions of titles is virtually impossible to achieve manually, yet can easily be implemented with a modest amount of programming.

As more and more sellers automate price-setting, pricebots are going to interact with one another, yielding unexpected price and profit dynamics. This paper reaches toward an understanding of strategic pricebot dynamics via analysis and simulation of four candidate price-setting algorithms that differ in their informational and computational needs: game-theoretic pricing (GT), myoptimal pricing (MY), derivative following (DF), and Q-learning (Q). Previously, we studied the price and profit dynamics that ensue when shopbot-assisted buyers interact with homogeneous collections of pricebots that utilize these algorithms [9, 10, 15]. In this work, we first establish that our previous results are not significantly altered when buyers' valuations are inhomogeneous rather than identical. Later, we examine the behavior that ensues when pricebots employing different strategies are pitted against one another.

This paper is organized as follows. The next section presents our model of an economy that consists of shopbots and pricebots. This model is analyzed from a game-theoretic point of view in Sec. 3. In Sec. 4, we discuss the price-setting strategies of interest: game-theoretic, myoptimal pricing, derivative following, and Q-learning. Secs. 5 and 6 describe simulations of homogeneous and heterogeneous collections of pricebots that implement these algorithms. Finally, Sec. 7 presents our conclusions and discusses ideas for future work.

2 Model

We study an economy in which there is a single homogeneous good offered for sale by S sellers and of interest to B buyers, with $B \gg S$. Each buyer b generates purchase orders at random times, at rate ρ_b, and each seller s reconsiders (and potentially resets) its price p_s at random times, at rate ρ_s. The value of the good to buyer b is v_b, and the cost of production for seller s is r_s.

A buyer b's utility for a good is a risk-neutral function of its price as follows:

$$u_b(p) = \begin{cases} v_b - p & \text{if } p \le v_b \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

We do not assume that buyers are necessarily utility maximizers. Instead, we assume that they use one of a set of *fixed sample size* search rules in selecting the seller from whom to purchase. A buyer of type i (where $0 \le i \le S$) searches for the lowest price among i sellers chosen at random from the set of S sellers, and purchases the good if that seller's price is less than the buyer's valuation v_b. (Price ties are broken randomly.)

A few special cases are worth mentioning. A buyer of type $i = 0$ simply opts out of the market without checking any prices. Buyers of types $i = 1$, $i = 2$, and $i = S$ have been referred to in previous work [9, 10] as employing the *Any Seller*, *Compare Pair* and *Bargain Hunter* strategies, respectively; the latter corresponds to buyers who take advantage of shopbots. [1] The buyer population is assumed to consist of a mixture of buyers employing one or another of these strategies. Specifically, a fixed, exogenously determined fraction w_i of buyers employs strategy i, and $\sum_i w_i = 1$.

The profit function π_s for seller s per unit time is determined by the price vector $\vec{p}$, which describes all seller's prices: $\pi_s(\vec{p}) = (p_s - r_s)D_s(\vec{p})$, where $D_s(\vec{p})$ is the rate of demand for the good produced by seller s. This rate of demand is determined by the overall buyer rate of demand, the likelihood that the chosen seller's price p_s will not exceed the buyer's valuation v_b, and the likelihood of buyers selecting seller s as their potential seller. If $\rho = \sum_b \rho_b$, and if $h_s(\vec{p})$ denotes the probability that seller s is selected, while $g(p_s)$ denotes the fraction of buyers whose valuations satisfy $v_b \ge p_s$, then $D_s(\vec{p}) = \rho B h_s(\vec{p}) g(p_s)$. Note that the function $g(p)$ can be expressed as $g(p) = \int_p^\infty \gamma(x)dx$, where $\gamma(x)$ is the probability density function describing the likelihood that a given buyer has valuation x.

[1] In this framework, it is also possible to consider buyers as utility maximizers, with the additional cost of searching for the lowest price expressed explicitly in the utility functions. In doing so, the search cost for bargain hunters is taken to be zero, while for those buyers who use the any seller strategy, its value is greater than v_b. The relationship between exogenously determined buyer behavior and the endogenous approach which incorporates the cost of information acquisition and explicitly allows for buyer decision-making is further explored in computational settings in Kephart and Greenwald [10]; in the economics literature, see, for example, Burdett and Judd [1].

Without loss of generality, define the time scale *s.t.* $\rho B = v$. Now $\pi_s(\vec{p})$ is interpreted as the expected profit for seller s per unit sold systemwide. Moreover, seller s's profit is such that $\pi_s(\vec{p}) = v(p_s - r_s)h_s(\vec{p})g(p_s)$. We analyze the functions $h_s(\vec{p})$ and $g(p)$ in the next section.

3 Analysis

We now present a game-theoretic analysis of the prescribed model viewed as a one-shot game. [2] Assuming sellers are utility maximizers, we derive the symmetric mixed strategy Nash equilibrium. A Nash equilibrium is a vector of prices at which sellers maximize their individual profits and from which they have no incentive to deviate [13]. There are no pure strategy Nash equilibria in our economic model whenever $0 < w_A < 1$ [8, 9]. There does, however, exist a symmetric mixed strategy Nash equilibrium, which we derive presently.

Let $f(p)$ denote the probability density function according to which sellers set their equilibrium prices, and let $F(p)$ be the corresponding cumulative distribution function. Following Varian [17], we note that in the range for which it is defined, $F(p)$ has no mass points, since otherwise a seller could decrease its price by an arbitrarily small amount and experience a discontinuous increase in profits. Moreover, there are no gaps in the said distribution, since otherwise prices would not be optimal — a seller charging a price at the low end of the gap could increase its price to fill the gap while retaining its market share, thereby increasing its profits. The cumulative distribution function $F(p)$ is computed in terms of the quantity $h_s(\vec{p})$.

Recall that $h_s(\vec{p})$ represents the probability that buyers select seller s as their potential seller. This function is expressed in terms of the probabilistic demand for seller s by buyers of type i, namely $h_{s,i}(\vec{p})$, as follows:

$$h_s(\vec{p}) = \sum_{i=0}^{S} h_{s,i}(\vec{p}) \tag{2}$$

The first component $h_{s,0}(\vec{p}) = 0$. Consider the next component, $h_{s,1}(\vec{p})$. Buyers of type 1 select sellers at random; thus, the probability that seller s is selected by such buyers is simply $h_{s,1}(\vec{p}) = 1/S$. Now consider buyers of type 2. In order for seller s to be selected by a buyer of type 2, s must be included within the pair of sellers being sampled, which occurs with probability $(S-1)/\binom{S}{2} = 2/S$, and s must be lower in price than the other seller in the pair. Since, by the assumption of symmetry, the other seller's price is drawn from the same distribution, this occurs with probability $1-F(p)$. Hence, $h_{s,2}(\vec{p}) = (2/S)\,[1-F(p)]$. In general, seller s is selected by a buyer of type i with probability $\binom{S-1}{i-1}/\binom{S}{i} = i/S$, and seller s is the lowest-priced among the i sellers selected with probability $[1-F(p)]^{i-1}$, since these are $i-1$ independent events. Therefore, we have derived $h_{s,i}(\vec{p}) = (i/S)[1-F(p)]^{i-1}$, and[3]

$$h_s(p) = \frac{1}{S}\sum_{i=0}^{S} i w_i [1-F(p)]^{i-1} \tag{3}$$

A Nash equilibrium in mixed strategies requires that all prices assigned positive probability yield equal payoffs — otherwise, it would not be optimal to randomize. Thus, assuming $r_s = r$ for all sellers s, the equilibrium payoff $\pi \equiv \pi_s(p) = v(p-r)h_s(p)g(p)$, for all prices p. The precise value of π can be derived by considering the maximum price that sellers are willing to charge, say p_m. At this price, $F(p_m) = 1$, which by Eq. 3 implies that $h_s(p_m) = w_1/S$. Identifying the expression $v(p-r)g(p)$ as the profit function of a monopolist, this function attains its maximal value π_m (the monopolist's profit) at price p_m. Therefore, for all sellers s,

$$\pi = v(p_m - r)h_s(p_m)g(p_m) = \frac{w_1 \pi_m}{S} \tag{4}$$

and hence,

$$v(p-r)g(p) = \frac{w_1 \pi_m}{\sum_{i=0}^{S} i w_i [1-F(p)]^{i-1}} \tag{5}$$

implicitly defines p and $F(p)$ in terms of one another, and in terms of $g(p)$, for all p such that $0 \le F(p) \le 1$.

In previous work, all buyer valuations were taken to be equal (*i.e.*, for all buyers b, $v_b = v$), and hence $g(p) = \Theta(v-p)$ (see [9] and [10]). In this paper, we assume $\gamma(x)$ is a uniform distribution over interval $[0, v]$, with $v > 0$, in which case the integral yields a step function as follows:

$$g(p) = \begin{cases} 1 - p/v & \text{if } p \le v \\ 0 & \text{otherwise} \end{cases} \tag{6}$$

For this uniform distribution of buyer valuations, the monopolist's profit function is simply $(p-r)(v-p)$, for $p \le v$, which is maximized at the price $p_m = (v+r)/2$. At this price, the monopolist's profit $\pi_m = [(v-r)/2]^2$. Inserting these values into Eq. 5 and solving for p in terms of F yields:

$$p(F) = p_m - \sqrt{\pi_m}\left\{1 - \frac{w_1}{\sum_{i=0}^{S} i w_i [1-F(p)]^{i-1}}\right\}^{1/2} \tag{7}$$

[2] The analysis presented in this section applies to the one-shot version of our model, although the simulation results reported in Sec. 5 focus on repeated settings. We consider the Nash equilibrium of the one-shot game, rather than its iterated counterpart, for at least two reasons, including (i) the Nash equilibrium of the stage game played repeatedly is in fact a Nash equilibrium of the repeated game, and (ii) the Folk Theorem of repeated game theory (see, for example, Fudenberg and Tirole [7]) states that virtually all payoffs in a repeated game correspond to a Nash equilibrium, for sufficiently large values of the discount parameter. Thus, we isolate the stage game Nash equilibrium as an equilibrium of particular interest.

[3] In Eq. 3, $h_s(p)$ is expressed as a function of seller s's scalar price p, given that probability distribution $F(p)$ describes the other sellers' expected prices.

Eq. 7 has several important implications. First of all, in a population in which there are no buyers of type 1 (*i.e.*, $w_1 = 0$) the sellers charge the production cost r and earn zero profits; this is the traditional Bertrand equilibrium. On the other hand, if the population consists of just two buyer types, 1 and some $i \neq 1$, then it is possible to invert $p(F)$ to obtain:

$$F(p) = 1 - \left[\left(\frac{w_1}{i w_i} \right) \left(\frac{(p_m - p)^2}{2(v-p)(p-r)} \right) \right]^{\frac{1}{i-1}} \qquad (8)$$

The case in which $i = S$ and $v_b = v$ for all buyers b was studied previously by Varian [17]; in this model, buyers either choose a single seller at random (type 1) or search all sellers and choose the lowest-priced among all sellers (type S) and all buyers have equal valuations.

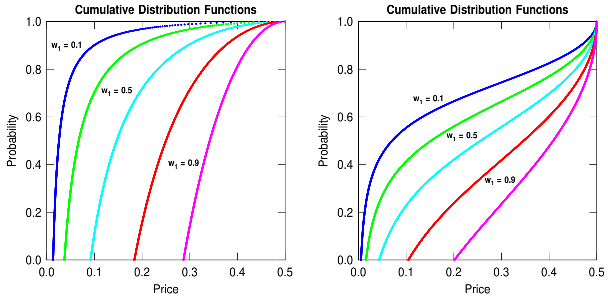
Since $F(p)$ is a cumulative probability distribution, it is only valid in the domain for which its valuation is between 0 and 1. The upper boundary is $p = p_m$, since prices above this threshold leads to decreases in market share that exceed the benefits of increased profits per unit. The lower boundary p^* can be computed by setting $F(p^*) = 0$ in Eq. 7, which yields:

$$p^* = p_m - \sqrt{\pi_m} \left\{ 1 - \frac{w_1}{\sum_{i=0}^{S} i w_i} \right\}^{1/2} \qquad (9)$$

In general, Eq. 7 cannot be inverted to obtain an analytic expression for $F(p)$. It is possible, however, to plot $F(p)$ without resorting to numerical root finding techniques. We use Eq. 7 to evaluate p at equally spaced intervals in $F \in [0, 1]$; this produces unequally spaced values of p ranging from p^* to p_m.

Fig. 1(a) and 1(b) depict the CDFs in the prescribed model under varying distributions of buyer strategies — in particular, $w_1 \in \{0.1, 0.25, 0.5, 0.75, 0.9\}$ — when $S = 2$ and $S = 5$, respectively. In Fig. 1(a), most of the probability density is concentrated just above p^*, where sellers expect low margins but high volume. In contrast, in Fig. 1(b), the probability density is concentrated either just above p^*, where sellers expect low margins but high volume, or just below p_m, where they expect high margins but low volume. In both figures, the lower boundary p^* increases as the fraction of random shoppers increases; in other words, p^* decreases as the fraction of shopbot buyers increases. Moving from $S = 2$ to $S = 5$ further decreases p^*, which is consistent with Eq. 9. These relationships have a straightforward interpretation: shopbots catalyze competition among sellers, and moreover, small fractions of shopbot users induce competition among large numbers of sellers.

Recall that the profit earned by each seller is $\frac{1}{S}\pi_m w_1$, which is strictly positive so long as $w_1 > 0$. It is as though only buyers of type 1 are contributing to sellers' profits, although the actual distribution of contributions



(a) 2 Sellers (b) 5 Sellers

Figure 1: CDFs for $w_1 \in \{0.1, 0.25, 0.5, 0.75, 0.9\}$.

from buyers of type 1 vs. buyers of type $i > 1$ is not as one-sided as it appears. In reality, buyers of type 1 are charged less than p_m on average, and buyers of type $i > 1$ are charged more than r on average, although total profits are equivalent to what they would be if the sellers practiced perfect price discrimination. In effect, buyers of type 1 exert negative externalities on buyers of type $i > 1$, by creating surplus profits for sellers.

4 Pricebot Strategies

When sufficiently widespread adoption of shopbots by buyers forces sellers to become more competitive, it is likely that sellers will respond by creating *pricebots* that automatically set prices in attempt to maximize profitability. It seems unrealistic, however, to expect that pricebots will simply compute a Nash equilibrium and fix prices accordingly. The real business world is fraught with uncertainties, undermining the validity of traditional game-theoretic analyses: sellers lack perfect knowledge of buyer demands, and have an incomplete understanding of competitors' strategies. In order to be deemed profitable, pricebots will need to learn from and adapt to changing market conditions.

We now introduce four pricebot strategies, each of which places different requirements on the type and amount of information available to the agent and upon the agent's computational power.

GT The *game-theoretic* strategy is designed to reproduce the mixed strategy Nash equilibrium that was computed in the previous section. It makes use of full information about the buyer population, and it assumes its competitors utilize game-theoretic pricing as well.

GT is a constant function since it makes no use of historical observations. Nonetheless, it is of interest in our simulation studies in part because there exist learning algorithms that converge to stage game-theoretic equilibria over repeated play (see Foster and Vohra [6] and Greenwald [8]).

MY The myopically optimal, or *myoptimal*, [4] pricing strategy (see, for example, [11]) uses information about all the buyer characteristics that factor into the buyer demand function, as well as competitors' prices, but makes no attempt to account for competitors' pricing strategies. Instead, it is based on the assumption of static expectations: even if one seller is contemplating a price change under myoptimal pricing, this seller does not assume that this will elicit a response from its competitors; it assumes that competitors' prices will remain fixed.

The myoptimal seller s uses all available information and the assumption of static expectations to perform an exhaustive search for the price p_s^* that maximizes its expected profit π_s. The computational demands can be reduced greatly if the price quantum ϵ (the smallest amount by which one seller may undercut another) is sufficiently small. Under such circumstances, the optimal price p_s^* is guaranteed to be either the monopolistic price p_m or ϵ below some competitor's price, limiting the search for p_s^* to S possible values. In our simulations, we choose $\epsilon = 0.002$.

DF The *derivative-following* strategy is less informationally intensive than either the myoptimal or the game-theoretic pricing strategies. In particular, this strategy can be used in the absence of any knowledge or assumptions about one's competitors or the buyer demand function. A derivative follower simply experiments with incremental increases (or decreases) in price, continuing to move its price in the same direction until the observed profitability level falls, at which point the direction of movement is reversed. The price increment δ is chosen randomly from a specified probability distribution; in the simulations described here the distribution was uniform between 0.01 and 0.02.

Q The *Q-learning* price-setting strategy is based on a reinforcement learning procedure called Q-learning [18], which can learn optimal pricing policies for Markov Decision Problems (MDPs). It does so by learning the function $Q(x, a)$ representing the cumulative discounted payoff of taking action a in state x. The discounted payoff is expressed as $\sum_n \gamma^n r_n$, where r_n is the expected reward n time steps in the future, and γ is a constant "discount parameter" lying between 0 and 1. The optimal policy for a given state is the action that maximizes the Q-function for that state. Q-learning yields a *deterministic* policy, and is therefore unable to represent equilibrium play in games where the equilibria are composed solely of randomized strategies. Q-learning finds the optimal policy in cases where the Q-learner's opponents use stationary Markovian strategies. Situations that deviate from this, such as history-dependent opponents or non-stationary learning opponents (*e.g.*, another Q-learner), constitute an interesting and open research topic that is touched upon here and in some of our prior work (see Tesauro and Kephart [15]).

[4] In the game-theoretic literature, this strategy is often referred to as Cournot best-reply dynamics [3]; however, price is being set, rather than quantity.

5 GT, MY, and DF Simulations

We simulated an economy with 1000 buyers and initially 2, and later 5, pricebots employing various mixtures of pricing strategies. In the simulations depicted, buyer valuations are uniformly distributed in the interval $[0, 1]$, and each seller's production cost $r = 0$. The mixture of buyer types is set at $w_1 = 0.25$, $w_2 = 0.25$, and $w_S = 0.5$; when $S = 2$, $w_2 = w_S = 0.75$. The simulation is asynchronous: at each time step, a buyer or seller is randomly selected to either make a purchase or reset a price. The chance that a given agent is selected to act is determined by its rate; the rate ρ_b at which a given buyer b attempts to purchase the good is set to 0.000999, while the rate ρ_s at which a given seller reconsiders its price is 0.00005. Each simulation was iterated for 100 million time steps.

5.1 Homogeneous Simulations

Here, we augment our previous work (see [9]) on simulations of 2 pricebots of matching types by assuming the uniformly distributed buyer valuations introduced in Section 3. If we consider 2 GT pricebots, simulations verify that the cumulative distribution of prices closely resembles the derived $F(p)$, to within statistical error. The time-averaged profits for each seller were 0.0319, nearly the theoretical value of 0.03125. When 2 pricebots use the myoptimal pricing strategy, cyclical price wars result (see [9]). The myoptimal sellers' expected profits averaged over one price-war cycle are given by $\pi_s^{\mathrm{MY}} = 1/vS[p_m - p^*] \int_{p^*}^{p_m} (v-p)(p-r)\, dp$, which is equal to 0.0893. Simulation results closely match this theoretical value with an average profit per time step for MYs of 0.0892, nearly thrice the average profit obtained via the game-theoretic pricing strategy.

In related work (see [9]), assuming all buyer valuations to be equal, the behavior of DF pricebots tended towards what is in effect a collusive state in which all prices approached the monopolistic price. In this study, 2 DF pricebots once again track one another closely; in doing so they accumulate greater profits than either 2 myoptimal or 2 game-theoretic pricebots. The effect of uniformly distributed buyer valuations is nonetheless apparent in substantial fluctuations in price, ranging from approximately 0.15 to 0.55 — the upper bound of which is above the monopolistic price $p_m = 0.5$ — rather than remain in the neighborhood of the monopolistic price. Specifically, DF pricebots achieved time-averaged profits of 0.1127.

5.2 Heterogeneous Simulations

Fig. 2(a) and Fig. 2(b) portray simulations of heterogeneous play. Consider the price dynamics of Fig 2(a), which depicts 1GT vs. 1DF. In this scenario, the game-theoretic pricebot outperforms the derivative follower by more often than not capturing greater market share with its lower price. In response, the derivative follower charges relatively high prices, although it does not oscillate precisely around the monopolistic price $p_m = 0.5$; instead, it prices in a range slightly below this optimum, since in doing so it more frequently finds itself to be the lower-priced of the two pricebots. The average profits of GT were 0.0682, while DF's average profits were less than half this value at 0.0334.

In contrast, Fig 2(b) depicts the price dynamics of 1GT vs. 1MY. Unlike the derivative follower, the myoptimal pricebot outperforms the game-theoretic pricebot; specifically, the time-averaged profits of GT were merely 0.0235, while MY achieved 0.0494. MY pricing has two notable advantages over derivative following: (i) access to full information pertaining to both competitors' prices and buyer demand, and (ii) the ability to change its price discontinuously, if necessary. Accordingly, Fig. 2(b) reveals that MY prices generally just undercut GT prices, unless GT charges p^*, in which case MY charges $p_m = 0.5$. We temporarily defer discussion of competition between MY and DF pricebots.

Table 1 summarizes the results of our 1-on-1 GT, MY, and DF simulations by depicting the time-averaged profits obtained by pricebots that employed the various strategies as indicated. It is interesting to consider this profit matrix as representing the payoffs of a normal form game in which there are three possible strategies, namely MY, DF, and GT. In doing so, we observe that the strategy profiles (1MY, 1MY) and (1DF, 1DF) are both pure strategy Nash equilibria, with the latter as Pareto optimal. Moreover, regardless of the opponents' behavior, it is always preferable to choose strategy MY or DF, rather than behave as prescribed by GT: *i.e.*, the elimination of dominated strategies eliminates the game-theoretic strategists. This outcome is not entirely surprising in view of the fact that GT is rooted in a stage game analysis, and prescribes play with no regard for historical data. In contrast, MY and DF take into account changing environmental conditions, and are thus more apt in asynchronous, repeated game settings.

We now turn to heterogeneous simulations of 5 pricebots. These simulations were conducted assuming 1 individual pricebot vs. 4 pricebots playing the same strategy: *e.g.*, 1MY vs. 4DF. Fig 3 displays a 3×3 matrix depicting the 9 possible combinations of 1 vs. 4 pricebots of strategies MY, DF, and GT, with the cells in the matrix indexed by $1 \leq i, j \leq 3$: *e.g.*, Fig $3_{1,2}$ refers to the cell that depicts 1MY vs. 4DF. We describe two of the more interesting off-diagonal entries.

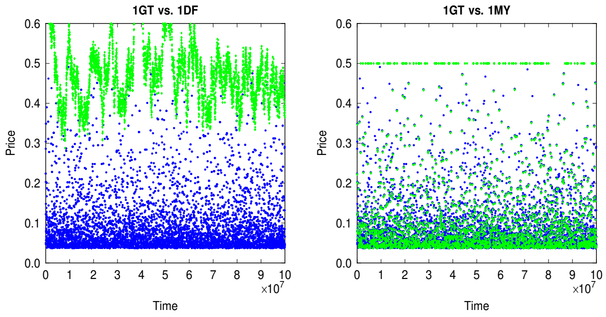


Figure 2: 1-on-1 Pricebot Simulations

	1MY	1DF	1GT
1MY	.0892,.0892	.0932,.0456	.0494,.0235
1DF	.0456,.0932	.1127,.1127	.0334,.0682
1GT	.0235,.0494	.0682,.0334	.0319,.0319

Table 1: 1-on-1 Profit Matrix. Within a given cell, the left-hand profit is that received by an agent employing the strategy corresponding to that cell's row, while the right-hand profit is that received by an agent employing the strategy corresponding to that cell's column.

1MY vs. 4DF

The fact that a homogeneous group of derivative followers is able to extract a larger profit than more clever, better informed, game-theoretic and myoptimal agents (see Sec. 5.1) may seem paradoxical: how is it that it is so smart to be so ignorant? Fig. $3_{1,2}$, in which one myoptimal pricebot is pitted against 4 derivative followers, suggests that in fact ignorance may not be bliss. During the simulation, MY averaged a profit of 0.0690. Hovering close to one another, just below the monopolistic price p_m, 2 of the 4 DFs received average profits of 0.0210 and 0.0227. A third DF never learned that it would be better off charging below p_m; even after 10 million iterations it continued to set its price above p_m and earned only 0.0188. The final DF pricebot, however, had a more interesting experience: it engaged in head-to-head combat with the MY pricebot, managing in the process to do better than its cohorts obtaining an average profit of 0.0275. Towards the end of this simulation though, the prices of the competing DF and MY pricebots approach p_m, and consequently, those of the other DF pricebots. In longer runs, we observed

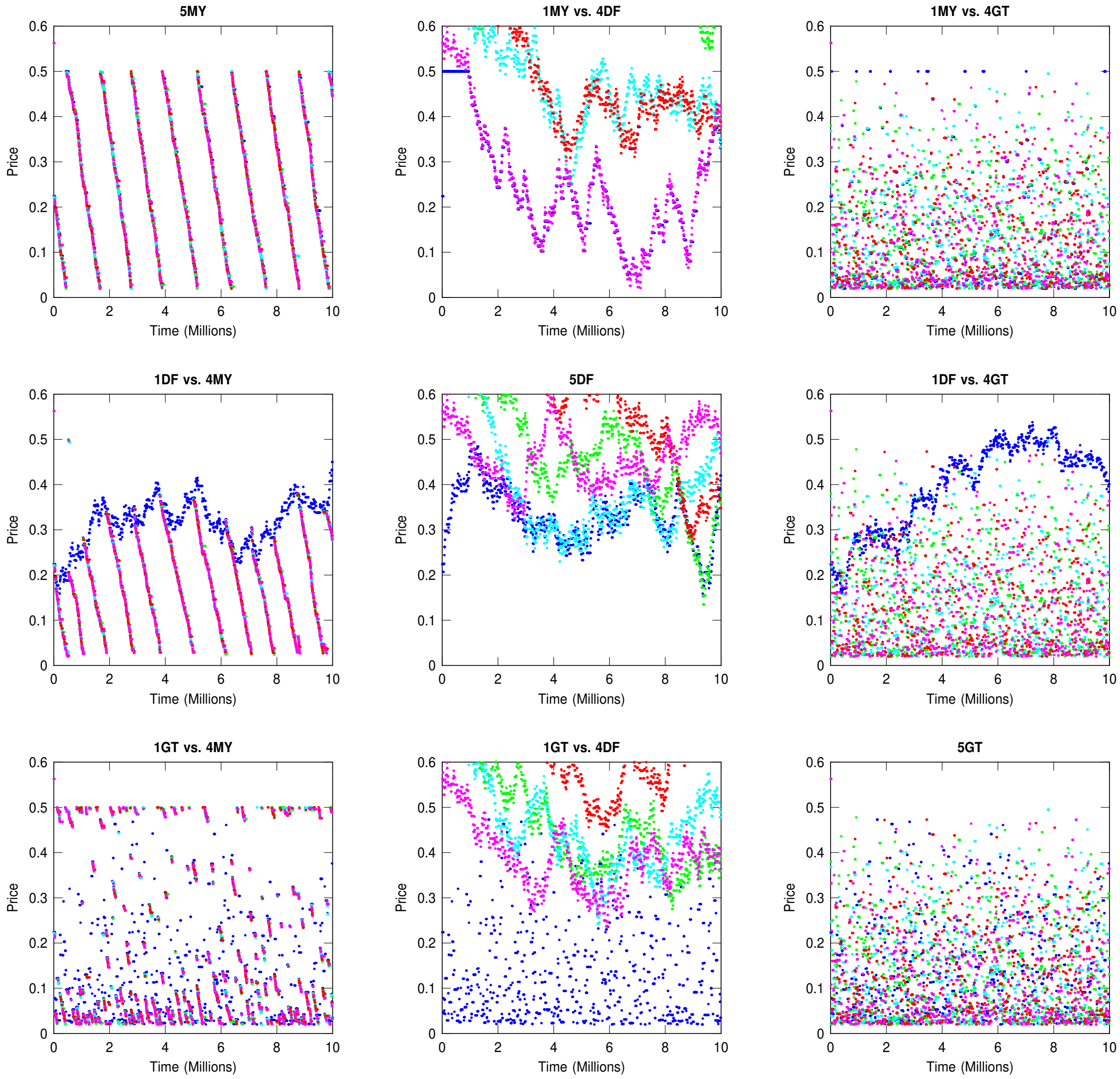


Figure 3: 4-on-1 Pricebot Simulations: Prices vs. Time

other DFs going head-to-head with the MY pricebot suggesting that all DFs ultimately fare equally well.

1DF vs. 4MY

A simulation of 1 derivative-following pricebot competing with 4 myoptimal pricebots is depicted in Fig. $3_{2,1}$. In this simulation, the derivative follower achieved average profits of 0.0136, while the myoptimal pricebots earned 0.0335 on average. Note that the profit obtained by a myoptimal seller in this heterogeneous setting is nearly identical to that obtained when *all* sellers are myoptimal. This equivalence is due to a coincidental cancellation of two opposing effects. First, note in Fig. $3_{2,1}$ that the derivative follower's price forms an upper bound above which no myoptimal seller is willing to price. This upper bound is in general less than p_m, since the derivative follower is continually enticed into half-hearted price wars, but its recovery phase is typically too short to allow it to meander up as high as p_m. These dynamics tend to reduce the profit received by a myoptimal seller. On the other hand, the more savvy myoptimal pricebots rarely permit the derivative follower to undercut them, so the profits earned in the market segment consisting of the more diligent shoppers is shared by just four myoptimal sellers, rather than five. Thus, each myoptimal seller receives a larger piece of a smaller pie. These opposing effects happen to counterbalance one another almost perfectly for the choice of parameters made in this example.

Prisoners' Dilemma

Table 2 summarizes the results of 5 pricebot simulations by depicting the time-averaged profits obtained by pricebots that employed the various strategies as indicated. (This table parallels Fig. 3.) Given 4 opposing pricebots that behave uniformly, the best-response of one additional pricebot is myoptimal pricing; conversely, assuming 1 opposing pricebot behaves according to MY, the best-response for 4 pricebots acting in unison is again myoptimal pricing. Thus, in this framework, the strategy profile (1MY, 4MY) can be viewed as a Nash equilibrium; in fact this is the unique Nash solution in Table 2. This strategy profile is not Pareto optimal, whereas (1DF, 4DF), for example, is Pareto optimal. In effect, this profit matrix has the same basic character as that of the Prisoner's Dilemma: the cooperative strategy is DF, but there is every incentive for individuals to defect to MY.

	4MY	4DF	4GT
1MY	.0337,.0337	.0690,.0225	.0185,.0109
1DF	.0136,.0335	.0387,.0387	.0134,.0159
1GT	.0119,.0169	.0536,.0226	.0129,.0129

Table 2: 4-on-1 Profit Matrix

Consumer Surplus

Table 3 presents the consumer surplus — the benefit achieved beyond the buyers' willingness to pay — obtained by the simulated buyers for the usual combinations of seller strategies. As expected, when all pricebots play DF and seller profits are maximized, consumer surplus is minimized. Notice, however, that the introduction of a single GT pricebot substantially increases consumer surplus; similarly, the introduction of a GT pricebot into a group of otherwise MY pricebots leads to even greater increases in consumer surplus. Apparently, the presence of even a single GT agent substantially benefits buyers. Interestingly, however, the greatest consumer surplus is seen, not when all pricebots are GT strategists, but rather when one pricebot plays MY and the remainder play GT. The deterministic, explicit undercutting by the MY pricebot (as opposed to the probabilistic undercutting by the GT pricebot) not only benefits the MY seller (see Table 2), but it also slightly improves the consumers' lot. What is a gain for MY and the consumers is a loss for the GT sellers.

	4MY	4DF	4GT
1MY	.2907	.3066	.4349
1DF	.3010	.2638	.4144
1GT	.4092	.3238	.4326

Table 3: Consumer Surplus

6 Q-Learning Simulations

The MY, DF, and GT pricing strategies studied in the previous section are all predefined strategies that do not vary during a simulation run. In this section, we study Q-learning, which incorporates a training period during which time Q pricebots adapt to specific opponent strategies. Simulation results of Q-learning against each of the 4 pricebot strategies (including Q-learning itself) are presented below. Due to the lookup table representation of the Q-function, these simulations were limited to 2 pricebots. In future work, we plan to study Q-learning in the case of multiple pricebots using function approximators (*e.g.*, neural networks and decision trees) rather than lookup tables to represent the Q-functions. Details of our Q-learning methodology are presented in Tesauro and Kephart [15].

Q vs. GT

Initially, we trained a Q-learner against a GT pricebot, and obtained a resulting policy that is virtually identical to the myoptimal policy, regardless of the value of the discount parameter γ. This is due to the fact that the game-theoretic strategy is state-independent, which implies the actions of the Q-learner have no impact on its opponent's future course of action. Hence, the best the Q-learner can do is optimize its immediate reward, which amounts to following the myoptimal policy.

Q vs. DF

The DF strategy is history dependent; thus, Q-learning is not guaranteed in theory to find an optimal policy against DF. In practice, however, Q-learning performs quite well. The Q-derived policy, which is once again largely independent of γ, is similar to the myoptimal policy. The primary difference is that the Q-learner tends to undercut by 2ϵ rather than by just ϵ, thereby guaranteeing that it will continue to have the lowest price even after the DF replies, since the DF can only lower its price by ϵ at each given time step. [5]

Q vs. MY

In Tesauro and Kephart [15], Q-learning vs. MY and Q vs. another Q-learner were studied in a variant of the present model of shopbots and pricebots in which all buyers had equal valuations. The results discussed here, assuming a uniform distribution of buyer valuations, are similar to the results reported previously. Q-learning always yielded exact convergence to a stationary, optimal solution (as expected) against a myoptimal opponent. Moreover, the Q-derived policies always outperformed

[5] In the simulations reported here, DF was modified slightly so that the size of its price increments was fixed at 0.01, which matched the price discretization of the Q-learner's lookup tables.

the myoptimal strategy, and average utility increased monotonically with γ. Q-learning also had the side effect of improving the utility of its myopic opponent, because the Q-learner learned to abandon undercutting behavior more readily than MY as the price decreased. As shown in Fig. 4(a), the price-war regime is smaller and is confined to higher average prices, leading to a closer approximation to collusive behavior, with greater expected utilities for both sellers.

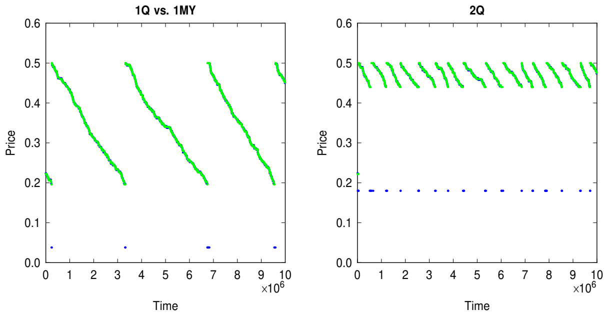


Figure 4: (a) Prices vs. time for Q-learner vs. myoptimal pricebot after training. (Training used a discount parameter $\gamma = 0.5$.) Compared to MY vs MY, the price-war amplitude is diminished, and profitability is increased for both sellers. (b) Prices vs. time for one Q-learner vs. another after training. (Both Q-learners trained with $\gamma = 0.5$.)

Q vs. Q

In simultaneous Q-learning, exact convergence of the Q-functions was only found for small γ. For large γ, there was very good approximate convergence, in which the Q-functions converged to stationary solutions to within small random fluctuations. Different solutions were obtained at each value of γ. For small γ, a symmetric solution was generally obtained (in which the shapes of $p_1(p_2)$ and $p_2(p_1)$ were identical), whereas a broken symmetry solution was obtained at large γ. There was a range of γ values, between 0.1 and 0.3, where either a symmetric or asymmetric solution could be obtained, depending on initial conditions. The asymmetric solution is counter-intuitive because one would expect that symmetric utility functions would lead to symmetric policies. Fig. 4(b) illustrates the results of simultaneous Q-learning at $\gamma = 0.5$. Note that there is a further reduction in the price-war regime as compared to MY vs. MY and Q vs. MY, leading to even greater profits for both sellers. Profits in the Q-learner simulations are summarized in Table 4.

In the cases of Q vs MY and Q vs GT, it appears that the Q-learner has obtained the theoretically optimal strategies against the respective opponents. For

Pricebots	Profits
(Q, GT)	(.0469, .0360)
(Q, DF)	(.2132, .0301)
(Q, MY)	(.1089, .1076)
(Q, Q)	(.1254, .1171)

Table 4: Profits per time step in 2 pricebot simulations where one of the pricebots was a Q-learner (trained at $\gamma = 0.5$) and the other was as indicated.

Q vs DF, the Q-learner does quite well despite the non-Markovian nature of the opponent's strategy, and for simultaneous Q-learning, generally good behavior was found, despite the absence of theoretical guarantees. Exact or very good approximate convergence was obtained to simultaneously self-consistent Q-functions and optimal policies. Spontaneous symmetry breaking is found for large γ, even though the utility functions for both players are symmetric. Understanding how this broken symmetry solution comes about is an active area of research. Also of interest would be to better characterize the solutions that appear to converge approximately but not exactly; such behavior has no analog in ordinary Q-learning.

7 Conclusions and Future Work

This paper examined several approaches to the design and implementation of pricebot algorithms that differ in their informational and computational requirements, and further differ in terms of their merits and limitations when tested against themselves and one another. The simplest algorithm is derivative following (DF) — its price adjustments are trivial to compute, and it requires no knowledge of either the buyer responses or of the other sellers' prices. DF is a reasonable approach in the absence of such knowledge, and performs well when other sellers also use DF. If more knowledge is available, however, other approaches can be designed to yield greater profits. If expected buyer demand is known, it is often possible to compute a game-theoretic equilibrium solution (GT). While this strategy is perhaps appropriate for one-shot games, in iterated games where there is knowledge of other sellers' prices, it is possible to outperform GT. The myoptimal strategy (MY), for example, uses knowledge of expected buyer demand to compute a short-term optimal price (but ignores the expected long-term behavior of its competitors) and outperforms both GT and DF. Computation of MY would appear to require an exhaustive search over all possible prices; however, MY's computational overhead can be reduced to examining only ϵ-undercuts of the other sellers' prices and the monopolistic price, for sufficiently small values of ϵ.

Pricing based on Q-learning is superior to myoptimal pricing since Q anticipates the longer-term consequences of its actions resulting both from competitor responses and its own counter-responses. Thus, among the pricing strategies studied, Q appears to offer the most promising performance; however, its computational requirements are rather substantial. This is true for both off-line and on-line Q-learning. For off-line Q, an extensive simulation is required, including models of both the buyers' demands and the other sellers' strategies. For on-line Q, real competitor prices and buyer responses are used, so that models are not required, but the training process remains long and tedious, and the learner must endure true losses during the training period while exploring suboptimal actions. In future work, we plan to continue investigating adaptive learning approaches such as Q-learning, with an aim towards designing algorithms that are feasible for practical implementation. For example, the Q-function can be represented using function approximators rather than lookup tables. Preliminary (and encouraging) results have been obtained based on neural networks [14], and work based on decision trees is underway. Lastly, while knowledge of competitors' prices seems reasonable (by the very existence of shopbots), accurate models of competitors' strategies or buyer demand may in fact be unavailable in actual agent economies; methods to overcome such limitations remain open to investigation.

References

[1] K. Burdett and K. L. Judd. Equilibrium price dispersion. *Econometrica*, 51(4):955–969, July 1983.

[2] A. Chavez and P. Maes. Kasbah: an agent marketplace for buying and selling goods. In *Proceedings of the First International Conference on the Practical Application of Intelligent Agents and Multi-Agent Technology*, London, U.K., April 1996.

[3] A. Cournot. *Recherches sur les Principes Mathematics de la Theorie de la Richesse.* Hachette, 1838.

[4] J. B. DeLong and A. M. Froomkin. The next economy? In D. Hurley, B. Kahin, and H. Varian, editors, *Internet Publishing and Beyond: The Economics of Digital Information and Intellecutal Property.* MIT Press, Cambridge, Massachusetts, 1998.

[5] J. Eriksson, N. Finne, and S. Janson. Information and interaction in MarketSpace — towards an open agent-based market infrastructure. In *Proceedings of the Second USENIX Workshop on Electronic Commerce*, November 1996.

[6] D. Foster and R. Vohra. Regret in the on-line decision problem. *Games and Economic Behavior*, 21:40–55, 1997.

[7] D. Fudenberg and J. Tirole. *Game Theory.* MIT Press, Cambridge, 1991.

[8] A. Greenwald. *Learning to Play Network Games.* PhD thesis, Courant Institute of Mathematical Sciences, New York University, New York, May 1999.

[9] A. Greenwald and J. Kephart. Shopbots and pricebots. In *Proceedings of Sixteenth International Joint Conference on Artificial Intelligence*, To Appear, August 1999.

[10] J. Kephart and A. Greenwald. Shopbot economics. In *Proceedings of Fifth European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty*, To Appear, July 1999.

[11] J. O. Kephart, J. E. Hanson, D. W. Levine, B. N. Grosof, J. Sairamesh, R. B. Segal, and S. R. White. Dynamics of an information filtering economy. In *Proceedings of the Second International Workshop on Cooperative Information Agents*, 1998.

[12] B. Krulwich. The BargainFinder agent: Comparison price shopping on the Internet. In J. Williams, editor, *Agents, Bots and Other Internet Beasties*, pages 257–263. SAMS.NET publishing (MacMillan), 1996. URLs: http://bf.cstar.ac.com/bf, http://www.geocities.com/ResearchTriangle/9430.

[13] J. Nash. Non-cooperative games. *Annals of Mathematics*, 54:286–295, 1951.

[14] G. Tesauro. Pricing in agent economies using neural networks and multi-agent q-learning. In *Proceedings of IJCAI Workshop on Learning About, From, and With Other Agents*, August 1999.

[15] G. Tesauro and J. Kephart. Pricing in agent economies using multi-agent q-learning. In *Proceedings of Fifth European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty*, To Appear, July 1999.

[16] M. Tsvetovatyy, M. Gini, B. Mobasher, and Z. Wieckowski. MAGMA: an agent-based virtual market for electronic commerce. *Applied Artificial Intelligence*, 1997.

[17] H. Varian. A model of sales. *American Economic Review, Papers and Proceedings*, 70(4):651–659, September 1980.

[18] C. Watkins. *Learning from Delayed Rewards.* PhD thesis, Cambridge University, Cambridge, 1989.