Decision Sciences
Volume 39 Number 1
February 2008

Reverse Auctions with Multiple Reinforcement Learning Agents*

Subhajyoti Bandyopadhyay†
Warrington College of Business Administration, 362 Stuzin Hall, University of Florida, Gainesville, FL 32611-1310, e-mail: shubho.bandyopadhyay@cba.ufl.edu

Jackie Rees
Krannert Graduate School of Management, 4037 Jerry S. Rawls Hall, 100 S. Grant Street, Purdue University, West Lafayette, IN 47907-2076, e-mail: jrees@purdue.edu

John M. Barron
Krannert Graduate School of Management, 464 Krannert Building, 403 West State Street, Purdue University, West Lafayette, IN 47907-2056, e-mail: barron@purdue.edu

ABSTRACT

Reverse auctions in business-to-business (B2B) exchanges provide numerous benefits to participants. Arguably the most notable benefit is that of lowered prices driven by increased competition in such auctions. The competition between sellers in reverse auctions has been analyzed using a game-theoretic framework and equilibria have been established for several scenarios. One finding of note is that, in a setting in which sellers can meet total demand with the highest-bidding seller being able to sell only a fraction of the total capacity, the sellers resort to a mixed-strategy equilibrium. Although price randomization in industrial bidding is an accepted norm, one might argue that in reality managers do not utilize advanced game theory calculations in placing bids. More likely, managers adopt simple learning strategies. In this situation, it remains an open question as to whether the bid prices converge to the theoretical equilibrium over time. To address this question, we model reverse-auction bidding behavior by artificial agents as both two-player and *n*-player games in a simulation environment. The agents begin the game with a minimal understanding of the environment but over time analyze wins and losses for use in determining future bids. To test for convergence, the agents explore the price space and exploit prices where profits are higher, given varying cost and capacity scenarios. In the two-player case, the agents do indeed converge toward the theoretical equilibrium. The *n*-player case provides results that reinforce our understanding of the theoretical equilibria. These results are promising enough to further consider the use of artificial learning mechanisms in reverse auctions and other electronic market

*We are grateful to the editor, the associate editor, and two anonymous reviewers for their insightful comments and suggestions. We also thank Professor Arunava Banerjee from the Department of Computer Science in the University of Florida for his specific suggestions for validating the hypotheses testing. Finally, we would like to acknowledge crucial inputs from Antonio Benecchi, Project Manager at Roland Berger Strategy Consultants, and executives at Indiamarkets.com and MetalJunction.com for sharing some of their live auction data and their observations.

†Corresponding author.

transactions, especially as more sophisticated mechanisms are developed to tackle real-life complexities. We also develop the analytical results when one agent does not behave strategically while the other agent does and show that our simulations for this environment also result in convergence toward the theoretical equilibrium. Because the nature of the best response in the new setting is very different (pure strategy as opposed to mixed), it indicates the robustness of the devised algorithm. The use of artificial agents can also overcome the limitations in rationality demonstrated by human managers. The results thus have interesting implications for designing artificial agents in automating bid responses for large numbers of bids where human intervention might not always be possible.

Subject Areas: B2B E-Commerce, Strategic Decision Making, and Supply Chain-Information Systems Interface.

INTRODUCTION

Online exchanges for business-to-business (B2B) transactions have become ubiquitous in industries ranging from automotive to retailing. Zhu (2004) defines a B2B exchange as "an online platform that creates a trading community linked by the Internet and provides the mechanism for B2B interactions using industry-wide data standards and computer systems." The Wall Street Journal (Renaissance in Cyberspace, 2003) mentioned the remarkable turnaround of the B2B Internet commerce sector and that U.S. businesses spent $482 billion in B2B transactions, up 242% from 2 years earlier. The latest available data (from 2006) calculated that revenue to have increased to nearly $1 trillion in 2004 (Japan Ministry of Internal Affairs and Communications, 2006), and it is estimated to be more than $14 trillion dollars by 2012 (OYO_Media, 2007). It was observed that, as early as 2001, 49% of organizations that buy more than $1 billion per year reported using an online auction, with most of them subsequently increasing their usage of these venues (Dixon-Burger, Howe, Schaeffer, & Freiden, 2001).

An example of a successful exchange is ChemConnect, a consortia-backed exchange connecting buyers and sellers in industries like chemicals, plastics, pharmaceuticals, paper, and more (Chemconnect, 2004). In 2002, its transaction volume exceeded $8.8 billion. Some prominent and successful private exchanges worth mentioning in the present context are GE Global eXchange Services (GXS), with 100,000 trading partners (Tomak & Xia, 2002); Dow Chemical Co., which now conducts 15% of its business online, a fivefold increase over 2000 levels (Renaissance in Cyberspace, 2003), and Wal-Mart, which started its own private exchange in 2001. Some of the more prominent advantages B2B exchanges are expected to bring include lower costs due to automating the procurement process, reverse auctions, interoperability among users, collaborative planning, and collaborative design (Helper & Macduffie, 2003).

The competition between sellers in reverse auctions has been analyzed by Bandyopadhyay, Barron, and Chaturvedi (2005, 2008) using a game-theoretic framework. The Nash equilibria were described for several scenarios. In an environment in which sellers can collectively meet total demand with the highest-priced seller catering to a residual (i.e., selling only a fraction of its capacity), the sellers resort to a mixed-strategy Nash equilibrium in which the players choose strategies randomly according to preassigned probabilities. Although price randomization in industrial bidding is an accepted norm (the intuition behind this is that a seller would not like to signal his bid by deciding on one particular bid price every time, which can then be easily undercut by the competition who is studying his moves), one might argue that in reality managers probably do not utilize advanced game theory calculations in bidding. More likely, managers adopt simple learning mechanisms that incrementally hone their bidding strategies over time. In this situation, it is still an open question as to whether these evolving bid prices using such learning strategies converge to the theoretical (and optimal) equilibrium over time.

This article examines this question by modeling seller behavior using artificial agents in a simulation setting that start bidding randomly and use reinforcement learning (RL)-styled algorithms to learn the ideal strategy over time. If the exercise is successful, it will provide impetus for both researchers and practitioners to study the feasibility of using artificial software agents in industrial bidding environments. It builds upon the work of a previous article (Bandyopadhyay, Rees, & Barron, 2006), where it was found that, when two sellers are identical (in the sense that they have the same production capacities and marginal costs of production), a simple RL strategy does indeed result in the sellers learning the correct mixed-strategy distribution over time, but that algorithm was ineffective when the sellers were nonidentical. In this article, we extend those results to show that our modified algorithm is indeed able to replicate the theoretical optimal behavior predicted in Bandyopadhyay et al. (2005) when the sellers are nonidentical. We extend the results of the theoretical model of Bandyopadhyay et al. (2005) to analyze seller behavior when there are more than two sellers. Finally, we develop new analytical results that find the best response of a seller in reaction to a seller who does not react strategically and check the performance of our learning algorithm in such an environment. This is important because in real-life scenarios it is conceivable that some sellers might fail to react strategically, and we explore this phenomenon. The best response here, in contrast to the extant results, is a pure-strategy Nash equilibrium. And it was interesting, therefore, to check whether our algorithm, which was successful in learning mixed-strategy equilibria, could replicate its success here and learn this strategy over time.

Bidding in reverse auctions is fraught with the risk of either underbidding (thereby "leaving money on the table" that could have been captured given the buyer's reservation prices and the other bidders' bids) or overbidding (and thereby losing out to the other bidders). The results of this article, therefore, have interesting implications for the design of software agents that can aid managers in bidding effectively in reverse auctions.

The use of simulation to test analytical results from an economic setting is well established. Some early examples include Wolf and Shubik (1978), who utilized an experimental duopoly game to test the effects of market structure, opponent behavior, and information about opponent behavior in the game, and Slusher, Sims, and Thiel (1978) who studied labor–management negotiations in the context of a business simulation game.

Reinforcement learning seeks to have an agent or automata "learn" via signals from its environment. Positive actions are positively rewarded, so that the agent might learn that positive actions are worth repeating (the so-called "exploitation" component). There is also an exploration component to reinforcement learning, where agents try new actions in an attempt to discover possibly fruitful new areas of a specific search space. Reinforcement learning procedures are used to solve problems that are Markov decision problems. In many multiagent settings, the Markov property typically does not hold due to colearning, and the agents are thus not guaranteed to converge in such settings. Therefore, our agents use learning algorithms similar to the zero intelligence plus (ZIP) agents demonstrated in Tesauro and Das (2001) and Cliff (1997) with additional memory and intelligence capabilities.

Exploring the usefulness of such learning algorithms is especially important given the growing proliferation of electronic marketplaces. Relatively simple transactions such as those studied in this research will evolve into more sophisticated and complex auction mechanisms (e.g., ones that incorporate limitations on the number of players given quality specifications, product differentiation, etc.). As these marketplaces continue to grow in popularity and importance, monitoring potentially hundreds of such concurrent transactions individually by human agents will conceivably be very difficult and time consuming, if not impossible. Artificial or machine learning agents provide a possible remedy to this problem, given their abilities to mimic and ideally complement human behavior in such environments. By using a relatively simple machine learning mechanism, we demonstrate that complex mixed-strategy Nash equilibria can be successfully assimilated in artificial agent behavior in various reverse-auction scenarios. These results should spur further research into more sophisticated machine learning algorithms for handling actual reverse auction and other B2B automated transactions.

In the following section, we provide the background literature surrounding the nature of the competition used for testing the specific learning algorithms demonstrated in the simulations. The use of artificial software agents in the modeling of such games is also discussed in this section. Next, we discuss the game-theoretical underpinnings of our work. We first summarize the existing theoretical results, and then develop the analytical results for finding the equilibrium when only one of the two sellers reacts strategically. Simulation experiments allow detailed and repeatable study of behaviors exhibited by the players under various treatment conditions. The agents used in this study are provided learning capabilities from previous actions by the use of an RL-styled algorithm described subsequently. The results of the simulations are discussed in the next section, and the final section concludes by providing the managerial implications of our findings and discussing some future research directions.

BACKGROUND LITERATURE

Competition in a Supply Chain Framework

Several articles have looked at the competition between participants in a supply chain framework. Xiao, Xia, and Zhang (2007) consider the quantity competition between two manufacturers producing partially substitutable goods in a game-theoretic framework. Saeed, Malhotra, and Grover (2005) look at the linkages between interorganizational information systems, buyer–supplier relationships, and manufacturing performance. Mahajan, Radas, and Vakharia (2002) analyze the effect of constrained and unconstrained capacities in the supply chain on the retailers' stocking policies. Chiang, Fitzsimmons, Huang, and Li (1994) use game theory to analyze the quantity discount problem in cooperative and noncooperative settings.

There is a wealth of economic literature that analyzes various facets of firms competing on quantity to supply a homogeneous product. Traditional oligopolistic Cournot competition between firms facing a downward-facing demand curve yields pure-strategy equilibria, where both firms sell at the same price point. Kreps and Scheinkman (1983) derive a similar outcome in a Bertrand setting (i.e., competing on price) with capacity constraints, where quantity precommitments and Bertrand competition yield Cournot outcomes that have equilibrium prices above marginal cost. As opposed to both of these results involving pure-strategy Nash equilibria, the competition in reverse-auction environments yields mixed-strategy equilibria under certain conditions that we elaborate upon next.

Bandyopadhyay et al. (2005, 2008) analyzed the nature of the equilibrium in a reverse-auction environment using various assumptions of the sellers' cost, capacities, and market demand. The problem becomes nontrivial when the assumption is that there is no combined capacity constraint on the part of the sellers. In other words, the sellers were able to supply the entire demand before the advent of the exchange and continue to do so after the exchange comes into play. We do not explore the case where there is a superfluous seller, for in that case the competition reduces to a trivial (pure-strategy) Bertrand model in which the suppliers supply at cost. We make the assumption that, when the set of suppliers is limited and all suppliers meet any given reputation (or quality) requirements in the marketplace, the buyer is indifferent to which supplier(s) fulfills its order. When the time horizon for decision making is extended, this is not necessarily true. In a series of procurement auctions run by IBM for Mars Candy, Mars wanted to ensure that certain high-price sellers received a token amount of business. This ensured that the long-standing good relations with suppliers were maintained, and a variety of suppliers were sustained for future procurement needs (Hohner et al., 2003). This implies that, while competition exists between the firms to be the low-price bidder, this competition is not as extreme as a Bertrand game resulting in prices equal to marginal cost. However, from the supplier point of view, an incentive exists to be the low-price bidder and be extended the first invitation to supply a requirement.

Most B2B auctions are used to transact huge quantities of homogeneous goods (Katok & Roth, 2004). This competition is exacerbated in open exchanges, as Dai and Kauffman (2003) point out that, unlike closed extranets, open electronic markets expose sellers to increased competition from diverse competitors. The environment we consider in this article has several special characteristics. First, the demand for materials posted on the exchange will tend to be inelastic. The buyer decides on the quantity and the reservation price below which the buyer is ready to buy the entire quantity. (We contacted Roland Berger, an automotive industry consulting organization, to confirm this assumption. We found out that auto manufacturers indeed have "target cost" structures for components, which is equivalent to the reservation price in our nomenclature.) Further, any seller meeting the basic requirements of the buyer is qualified to bid. Wal-Mart, for example, has a seller certification process and a seller standards program; similar vendor certification programs are common in a variety of industries, and nowadays many organizations detail these programs on the Internet, allowing for any potential seller worldwide to become certified. With the advent of the online exchange, a buyer puts its requirements on the exchange only once instead of contacting the sellers individually. This reduces transaction costs in terms of maintaining dedicated account teams for buyers, sales force for sellers, cost of sending individual request for quotes (RFQs) to the entire universe of sellers, and so on (Kerrigan, Roegner, Swinford, & Zawada, 2001). Sellers look at this entire requirement and then decide to quote their prices. The sellers are also given a date (usually several weeks in the future) by which they are expected to place their bids. Sellers bid their selling prices, all of which are opened at a later date. The seller with the lowest bid price gets the first invitation to cater to the demand, followed by the seller with the second-lowest bid price, and so on, until the entire demand is met. (A confirmation of this mechanism was obtained from IndiaMarkets.com, India's largest B2B portal, and MetalJunction.com, a B2B portal for metals and minerals [and the world's largest e-marketplace for steel]; the software used at IndiaMarkets.com is licensed from one of the pioneer B2B portals, Ariba.) The use of multiple vendors for the same component is very much a standard practice. Other than the benefits of competition, the practice also reduces the possibility that the production is disrupted due to some problems of a single vendor (Rubin & Benton, 1993).

In such an environment, sellers would prefer to have limited production capacities. An unlimited capacity of the sellers would result in a Bertrand competition with sellers supplying to cost, a result that is clearly not advantageous to the sellers. Kreps and Scheinkman (1983) (and several variants of the original model, such as Allen, Deneckere, Faith, & Kovenock, 2000) show that, if sellers limit capacity, then a quantity precommitment and Bertrand competition yield Cournot equilibrium with prices above marginal cost.

Learning in Artificial Software Agents

Multiple studies exist in which artificial agents have been used to simulate human agents or players in a number of different settings. For example, Oliver (1996) used artificial agents to conduct automated negotiations in a highly structured e-commerce transaction environment. The use of dynamic pricing by automated *pricebots* (automated software agents that decide on pricing mechanisms) was examined in Greenwald and Kephart (2001) and Kephart, Hanson, and Greenwald (2000). Stone, Schapire, Littman, Csirik, and McAllester (2003) used artificial agents in the interacting auction format. Epstein and Axtell (1996) advocates the use of artificial agents to study systems and structures from the bottom up, which proves especially useful when no closed-form solution exists to the problem being addressed (or it is difficult to obtain).

RL is a machine learning technique shown to be useful in situations in which agents must learn the outcomes from previous actions in order to carry out a task. RL agents typically have a goal, receive feedback from the environment, and can make decisions or undertake some actions in response to the feedback from the environment. Uncertainty is also often incorporated into the RL agent environment (Sutton & Barto, 1998). As an example related to this research, an RL agent's goal might be to win a simple auction by making the highest bid (within a specific bound). The agent would receive feedback from the environment of whether or not it won the auction with the bid that was tendered. The next bid tendered in the following round would then be adjusted based on the information received from the previous round. Moreover, in the earlier rounds of the auction, the RL agent would operate under much uncertainty as it learns the bidding behavior of other RL agents participating in the auction. This uncertainty could decrease as the agent learns successful bidding strategies.

RL has been used in modeling competitive scenarios such as sealed bid k-double auctions under asymmetric and incomplete information dynamics (Rapoport, Daniel, & Searle, 1998), market entry games (Erev & Rapoport, 1998), and rule learning in repeated games (Bell, 2001). Tesauro and Das (2001) compared various types of learning agents in the setting of continuous double auctions and reported that relatively simple learning agents performed fairly well when trading prices were available to all participants. As the winning bid might not be publicly available in the reverse-auction scenario, the agents in our research know only their own win/loss experience and incorporate that experience into their bidding behavior.

THE MODEL

To understand the setting of the game, it is useful to first consider the dynamics of the two-player model. We consider the case of two sellers S_i, $i = H, L$, who have different (but fixed) marginal costs of production, c_i. They also have unequal capacities k_i that are more than the requirements of the buyers when combined, but is less than the total requirements of buyers Q (i.e., $(k_H + k_L) - Q > 0$, $k_i < Q$) when taken individually. In such a setting, the lower-priced seller is invited first to sell the required quantity, and after he has supplied his total capacity k_L, the other seller can then sell the residual demand $Q - k_L$. Depending on whether both sellers react strategically or only one seller does, the nature of the Nash equilibrium is different.

It has been established before that, when both sellers react to each other strategically, the resultant equilibrium is in mixed strategies (Bandyopadhyay et al., 2005, 2008), the results of which are summarized in the following subsection. In the subsequent subsection, we establish the nature of the equilibrium when only one seller reacts strategically to another nonstrategic seller (who randomizes his actions according to a known distribution, but does not modify this strategy based on the other seller's actions). The analysis shows that the best response in this environment is in pure strategies.

The Existence of a Mixed-Strategy Equilibrium

It has been shown that a mixed-strategy equilibrium of prices exists with the sellers randomizing bids between a range of prices (Bandyopadhyay et al.,

2008) that is common to both sellers. A similar analysis with n identical players, with $((n-1)k < Q < nk)$, yields a mixed-strategy equilibrium $F_n(p) = [\frac{(p-c)k-(Q-(n-1)k)(r-c)}{(p-c)(nk-Q)}]^{1/n-1}$, with the support of the strategy given by (p_n^1, r), where $p_1^n = \frac{(r-c)(Q-(n-1)k)}{k} + c$.

When the sellers are dissimilar, a crucial parameter is what is called the *indifference price*, the price at which a seller is indifferent between winning with certainty (by supplying to capacity) or losing with certainty (by supplying the residual) at the reservation price (Bandyopadhyay et al., 2005). It can be shown that, if a mixed-strategy equilibrium exists, the pricing strategy of the seller with the lower indifference price (who we call S_L) is defined by a continuous distribution over the range $[p^*, r]$ (p^* is defined below in equation (2)), with no mass at any price. The other seller (or S_H) has a pricing strategy that is also defined by a continuous distribution over the same range $[p^*, r]$, but S_H now places a positive mass at the upper bound of the distribution (r). In the equations that follow, the subscripts H and L refer to sellers S_H and S_L, respectively. It can be shown that (for details of the analysis for the following equations, see Bandyopadhyay et al., 2005)

$$p^* = \frac{(r-c_H)(Q-k_L)}{k_H} + c_H. \tag{1}$$

Further, the distributions of S_L and S_H are given by

$$F_L(p) = \frac{(p-c_H)k_H - (r-c_H)(Q-k_L)}{(p-c_H)(k_L+k_H-Q)}, \tag{2}$$

and

$$F_H(p) = \frac{\Pi_{sL}(p) - \Pi_{fL}(r) - (1-F_H(r))(\Pi_{sL}(r) - \Pi_{fL}(r))}{\Pi_{sL}(p) - \Pi_{fL}(p)}, \quad p < r, \tag{3}$$

respectively, where

$$F_H(r) = \frac{\left(r - p_H^*\right)}{\left(r - p_L^*\right)} \tag{4}$$

with $p_H^* = p^*$, and $p_L^* = \frac{(r-c_L)(Q-k_H)}{k_L} + c_L$ and the symbol Π with the subscripts s and l refer to the profits made by the sellers S_H and S_L at their points of success (i.e., supplying to capacity) and failure (i.e., supplying the residual).

A unique feature about this equilibrium is how the seller with the higher indifference price places a positive probability of pricing his bid at the reservation price (i.e., the distribution is continuous at all points on the support except r), and the magnitude of this probability is given by equation (4).

The Existence of a Pure-Strategy Equilibrium

The above results show the nature of the equilibrium where both players behave strategically. In real life, however, many managers might not think of learning from the responses of the competing seller. Such managers might continue to respond myopically by randomizing their bids, but not altering the distribution of the randomization in subsequent bids. In the context of our two-seller problem, the nonstrategizing seller does not react strategically to the responses of the other (strategizing) seller, who in fact continues to learn from the behavior of the first seller. In such an environment, we prove that the following proposition holds:

Proposition 1. *The best response to a seller who randomizes her bids across a price interval is a pure-strategy response that maximizes the expected profit for the strategizing seller. In particular, if the bid randomization by the nonstrategizing seller is uniform in the interval* $[c, r]$*, the pure-strategy response is given by the price* $p' = \frac{k(r+c) - 2c(Q-k)}{2(2k-Q)}$*, when both sellers have capacity* k *and marginal cost* c*.*

Proof. Let us denote the nonstrategizing seller as Seller 1 and the strategizing seller as Seller 2.

Consider the best response of Seller 2. Because Seller 1 is behaving nonstrategically (in the sense that the bid does not depend on Seller 2's bid), one can calculate the one bidding price that maximizes the expected profit of Seller 2 in response. In other words, the best response to this strategy (by Seller 2) is now not a mixed strategy (because nothing is gained by employing any randomization in the response), but a pure strategy that maximizes Seller 2's expected profit.

Specifically, consider the theoretical equilibrium if Seller 1, irrespective of the response of Seller 2, continues to randomize its bids uniformly between the prices c and r (the nature of the solution would hold true regardless of the specifics of the distribution followed by Seller 1). Let Seller 1 randomize his bids between c and r by following a uniform strategy U$[c, r]$ (i.e., Seller 1 chooses any price in the range $[c, r]$ with the same probability). If the response to this strategy by Seller 2 is given by a bid price p, the probability of Seller 2 losing with this bid price (or in other words, for Seller 1 to have a bid price lower than p) is given by $(\frac{p-c}{r-c})$ (as Seller 1 could have chosen any price with uniform probability in the range $[c, r]$), while the probability of winning with this bid price (or for Seller 1 to have a bid price higher than p) is given by $1 - (\frac{p-c}{r-c}) = (\frac{r-p}{r-c})$. Let the profit of the seller when he picks a higher bid price by choosing a price p be given by Π_l, and the profit when he picks a lower bid price be given by Π_w (the subscripts l and w signify a loss and win, respectively). When Seller 1 loses a bid, his profit with bid price p is given by $\Pi_l = (p-c)(Q-k)$ (as by losing, he gets to supply only the residual $(Q-k)$), while his profit when he wins at this price is given by $\Pi_w = (p-c)k$ (as he then gets to supply his capacity k). Thus, the expected profit of Seller 1 by bidding the price p is given by:

$$\Pi = \left(\frac{p-c}{r-c}\right)\Pi_l + \left(\frac{r-p}{r-c}\right)\Pi_w$$

$$\text{that is,}\quad \Pi = \left(\frac{p-c}{r-c}\right)(p-c)(Q-k) + \left(\frac{r-p}{r-c}\right)(p-c)k. \tag{5}$$

From first-order conditions, we then verify that the best response of the second strategizing seller is given by choosing a single price p':

$$p' = \frac{k(r+c) - 2c(Q-k)}{2(2k-Q)}. \tag{6}$$

□

While the first seller's behavior is nonstrategic in nature, it is entirely conceivable that many sellers in real life might not react strategically and would respond in this fashion, and it is, therefore, necessary to check for the performance of our learning agents in this sort of an environment.

Although price randomization in industrial bidding is an accepted norm, it is doubtful that many managers execute advanced game theory calculations to determine their bids. In any case, real-life situations are far more varied and sophisticated than the simplified model scenarios, making any game theory analysis extremely complex. Therefore, the research question becomes whether a heuristic learning technique utilized by the seller agents results in a bid distribution approaching the predicted theoretical equilibrium. The results provide some confidence in carrying out the learning technique with more than two sellers, thus providing insights about the strategies employed by the sellers in an environment that has not been analyzed so far in literature.

THE ALGORITHM

We model the competing sellers as agents and examine both the two-player and the n-player cases. Similar to humans, we propose that these agents understand the following (without resorting to explicit knowledge of game theory):

i. There are two opposing forces in the pricing strategy: A higher price (toward r) means greater per-unit profit, but also brings about a higher probability of losing to the competition (and, therefore, not supplying to capacity).
ii. It does not make any sense to price below their respective indifference prices, as is clear from the prior discussion.
iii. Provided the need to balance between a higher probability of winning and a higher per-unit profit, there is no a priori reason to rule out any price between the indifference price and r, and further, there is reason to believe that price randomizing might be the ideal (or equilibrium) solution.

Note that, according to the theoretical distribution, the two agents should learn over time that (i) it makes no sense for the agent with the lower indifference price to price below the higher indifference price and (ii) the agent with the higher indifference price should price his bid with positive probability at the seller's reservation price.

Tesauro and Kephart (2002) provide insights into a particular type of reinforcement learning, called Q-learning, which does not require a model of its environment. They studied single-agent and multiagent learning in a similar competitive setting to the one utilized in this article. In single-agent Q-learning, there is one Q-learning agent, which faces another agent that follows a fixed strategy. This scenario guarantees that the Q-learner will eventually find the optimal pricing policy in a repeated game. In the multiagent setting, as utilized here, due to the nonstationary environment, it is much more difficult to reach convergence. Tesauro and Kephart (2002) report success in the multiagent setting when the state space is small and the discount parameter for future rewards is kept small. In the setting

Figure 1: Basic learning algorithm for multiple agents.

For a given set of values for Q, k_i, c_i and r:

1. Determine the indifference price. The range between the indifference price and r determines the range from which the sellers select their prices.
2. Choose a price randomly for either seller.
3. The winner is the seller that chooses the lower price. Note the price, whether the seller won or lost, and the profit of each seller in every transaction. Profit in each round = (p-c_i)**Qty. sold*, where *Qty. sold* is either k_i or $Q-k_i$ depending on whether the seller wins or loses.
4. If seller wins then explore higher prices. The seller explores higher prices by incrementing the bid by α. If seller loses, then lower prices are systematically explored, again by decreasing the bid by α until the indifference price is reached; otherwise, the price is raised to r.
5. Repeat steps 3 and 4 for 1500 games.
6. Discard the results of the first 500 games and analyze the results of games 501 through 1500.

used in this article, a larger state space is used, therefore the traditional lookup tables for the action-state pairs for the agents is not as feasible. Therefore, a static estimation approach is employed, in which the agents make trade-offs between bid pricing and maintaining higher levels of average profit. The agents learn that, in certain price intervals, winning actually can lead to lower levels of profit and thus should be avoided.

Figure 1 describes the algorithm that the agents employ in a two-player game to determine their prices. The algorithm is essentially the same for the n-player model, except that we use p_n^* to determine the support of prices and compare the experimental distribution with $F_n(p)$ rather than $F(p)$. We note that, while it is known that convergence can result in relatively simple environments with multiple agents, there is no proof currently for theoretical convergence (Tesauro, 2001).

The sellers thus start off initially with a random pricing strategy (i.e., the price distribution is uniform in its support), with the anticipation that over time they will discover alternative strategies that enhance their profits. This is the same assumption that Erev and Roth (1998) employ in their experiments and refer to as the “initial propensities” (p. 859) of the players for their pure strategies.

Thus, we attempt to find out whether, through our learning algorithm, the players can finally converge on the arguably more sophisticated theoretical equilibrium. The rationale for the algorithm is as follows: Because the players a priori have no reason to believe that some prices are more likely than others, they start off by selecting any price in the range (p_i^*, r) ($i = H$, L) (step 1) with uniform probability (step 2). The relative magnitude of the prices (bids) determines the profit of either seller (step 3). However, by means of the learning component of the algorithm, they ensure that they are aware of any emerging pattern of wins and losses. This process was deemed the *exploration* component of RL by Sutton and Barto (1998). The agents learn or explore the problem space during this component of learning. The *exploitation* process comes into play (step 4), such that, if an agent wins at a certain price, after a certain number of consecutive wins at this price the agent increases its bid, hoping to obtain higher profits. Choices or decisions that led to good outcomes in the past are more likely to be repeated in the future (Thorndike, 2000). The game is then repeated a number of times so that the players can learn to converge to an equilibrium distribution. The experiment can be repeated with other values of Q, k_i, c_i, and r.

HYPOTHESIS TESTING, RESULTS, AND DISCUSSION

The Two-Player Simulations

For experimental purposes, we selected various values of k_i and c_i, while keeping Q fixed at 100 units and r fixed at \$80. While it is customary to consider the parameter values in a simulation to be random variables that follow certain probability distributions, we decided to constrain our choices of the parameter values due to three considerations. First, the combinations of k_i and c_i were such that the equilibrium of the game remained nontrivial. For example, if we chose the capacity k_i to be too large such that a single supplier can cater to the entire demand, the equilibrium of the game is trivial; if this large supplier has a lower cost of production, it prices the other supplier out of the market by supplying at a price that is just lower than the other supplier's cost. If this large supplier has a higher cost of production, the other supplier would then supply to capacity, while this large supplier will cater to the residual demand. At the other extreme, if the capacity is chosen to be too low such that the two suppliers combined cannot cater to the entire demand, then either supplier will get to supply their entire capacity at the buyer's reservation price.

Another factor that we considered while choosing the parameter values is that the range of the support of prices remained more or less unchanged. Note that a seller with a lower indifference price should, after effective learning, not place any of his bid randomizations in the interval between the two indifference prices (placing a bid in that interval would amount to choosing suboptimally because this seller should have learned from past observations that the other seller *never* places his bids this low). Thus, theoretically, the *effective* interval of choice is the range $[p^*, r]$. In order to study the effectiveness of the learning algorithms with different parameter choices, we wanted to keep this interval more or less invariant, with the sellers choosing their bids in more or less the same price bands. Thus, if the learned strategy distribution in some particular subinterval in any of the simulations deviated significantly from the theoretical distribution, we could investigate the causes by looking at the results in the same subinterval with the other simulations. Therefore, these choices of parameter values made it effectively possible to compare the performance of the algorithms in different settings (in other words, we could test our hypotheses about the artificial agents learning effectively in more or less the same price bands in the different sets of simulations). Thus, all our choices of parameter values except one, the reason for which is explained in the next paragraph, had a price support interval that was roughly in the range (\$50, \$80), and this allowed us to have the bins for the hypothesis testing in the ranges 50–55, 55–60, 60–65, 65–70, 70–75, and 75–80.

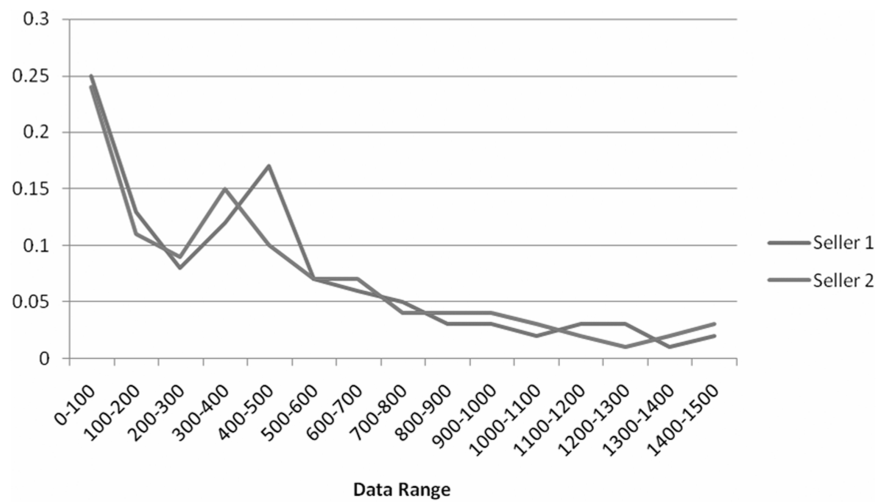
The lone choice of parameter values that had a significantly different support for prices was dictated by our third consideration. This final factor for consideration, the fact that whether a seller had a lower indifference price (and therefore more pricing power) than the other, is determined not only by his cost of production but also his capacity. *Ceteris paribus*, the theoretical results show that a seller with a lower marginal cost of production has an advantage over the other seller by having a lower indifference price. However, the other seller can negate this advantage by having a higher capacity and thereby get to have the lower indifference price. The intuition behind this result is as follows: A seller with a lower marginal cost can price his bids lower more often than the other seller, as the loss in per-unit profit is compensated for by supplying to capacity. However, a competing seller with a higher capacity deters such behavior as now his competitor realizes that a loss will leave him (the competitor) with a very little amount to sell. With this in mind, we wanted to explore different parameter choices such that both the marginal cost and the capacity can play a role in determining the lower of the two indifference prices. Thus, when we chose to have the two agents with the (k_i, c_i) parameter choices of (45, 30) and (85, 35), respectively (see Table A4), the first agent has the benefit of having a slightly lower marginal cost of production, but the second agent more than compensates for that relative disadvantage by having a much larger relative capacity, and this translates into a lower indifference price for the second agent.

The simulations were run on a standard Intel Pentium 4, 3.2 GHz workstation running Microsoft Windows XP SP2 with 1.0 GB RAM. The simulations were each run for 1,500 rounds. The run time for each simulation varied from under 1 second for two-seller simulations to under 3 seconds for five-seller simulations.

The nature of the algorithm meant that the initial rounds of the simulations would differ widely from the theoretical distribution because the players start off by choosing uniformly in the interval (p_i^*, r) $(i = H, L)$. We therefore had to decide on the number of the initial rounds that we would not consider when comparing the expected learned distribution with the theoretical distribution. To do that, we found out the maximum vertical deviation between the experimental cumulative distribution and the theoretical cumulative distribution (in other words, the statistic D in the Kolmogorov-Smirnov test) using a sliding window of 100 observations from the simulations. Thus, we found out the cumulative frequency distributions for the first 100 observations, the second 100 observations, and so on. The result is shown in Figure 2, which shows the variation of the statistic D for the different ranges of observations for the first set of parameters (Table A1). As is apparent from Figure 2, the graph initially (and expectedly) shows a high deviation between the experimental and theoretical distributions, but settled down at around 500 observations. The results were similar with other parameter values, with the distributions settling around roughly in the (500, 600) range. Thus, the first 500 rounds of each simulation were set aside and the final 1,000 rounds were analyzed according to the theoretical distribution.

The learning step parameter, α, essentially controls the rate of exploration of the search space. For the simulations, α was kept constant and operationalized as the bid increment and decrement in the reverse auction, fixed at \$5.00. This figure was arrived at after experimentally varying the step size against the theoretical distribution. Future payoffs are not discounted (i.e., the discount rate $\gamma = 0$) for

Figure 2: Distribution of the Kolmogorov-Smirnov statistic D for the two sellers with $k_1 = 80$, $c_1 = \$20$, $k_2 = 65$, and $c_2 = \$40$.



the simulation, given the short-term time horizon utilized in the reverse-auction model, similar to Kutschinski, Ulthmann, and Polani (2003). The only information available to each seller agent is her own cost, win/loss history, profit, and quantity sold, and the capacities of the other agents (as would be realistic to obtain in the real world). Therefore, the agents in this simulation operate under information asymmetry conditions.

The results of the simulations are shown in Tables A1–A5. We conducted the Kolmogorov-Smirnov test (the critical value at 95% significance level being .043) for testing the goodness-of-fit of the experimentally determined distribution with that of the analytical predicted distribution:

H0: $F(p) = F^*(p)$ where $F^*(p)$ is the experimentally generated distribution.
H1: $F(p) \neq F^*(p)$

We explain the results within Table A1 to illustrate our findings. Table A1 shows the frequency distribution of the bid prices placed by the two sellers with capacities 80 and 65 units, and marginal cost of production of \$20 and \$40, respectively, after 500 rounds of bidding have taken place. The quantity demanded by the seller is 100 units, and the seller's reservation price is \$80. The two indifference prices can be computed from these data and are found to be \$38.46 and \$52.50. The theoretical results then dictate that neither seller would place a bid lower than the higher indifference price that is, \$52.50. The simulations indicate, however, that Seller 1, nevertheless, places some bids (40 out of 1,000) in the interval 45–50, indicating that, while the seller has learned that it does not make sense to bid many times in the interval 45–50, the learning is not perfect. We then proceed to compute the theoretical cumulative frequencies of the bids that would be placed in the intervals 45–50, 50–55, . . . 70–75, and 75–80 from the computed theoretical distribution. These are shown in the fifth column of the tables. These cumulative values in turn give us the theoretical frequency distributions in these ranges (column 4 in the tables). Note that as the sellers are dissimilar, the theoretical frequency distributions are different: Seller 2, with his higher indifference price, bids considerably less in the lower price bands as compared to Seller 1 (and thereby loses more often). In the simulated learning environment, we then find out the number of times the bids have been placed by the two sellers in the above intervals, yielding the values in column 2 of the tables. Column 3 then computes the cumulative frequency distributions from column 2. The relative absolute difference of the simulated cumulative frequency distribution (column 2) and the corresponding theoretical cumulative frequency distribution (column 5) forms the basis of the cumulative difference column (column 6 of the tables) and is used for the Kolmogorov-Smirnov test. These values indicate that we cannot reject our null hypothesis.

We see that the simulation results in Tables A1–A5 show that the learning algorithm manages to reach the theoretical equilibrium over time, insofar as in all the five testable two-seller simulations we cannot reject the null hypothesis. We indicate this graphically in Figures 3 and 4, which depict the goodness-of-fit curves for the first two two-seller cases (the goodness-of-fit curves with the other parameter values are very similar). Note the change of notation in the graphs: The first seller is denoted by the subscript 1 and the second seller by the subscript 2. This is because a priori we are not aware which seller is the seller S_H (the seller

Figure 3: The two–heterogeneous-seller case comparing simulated results with theoretical results with $k_1 = 80$, $c_1 = \$20$, $k_2 = 65$, and $c_2 = \$40$.

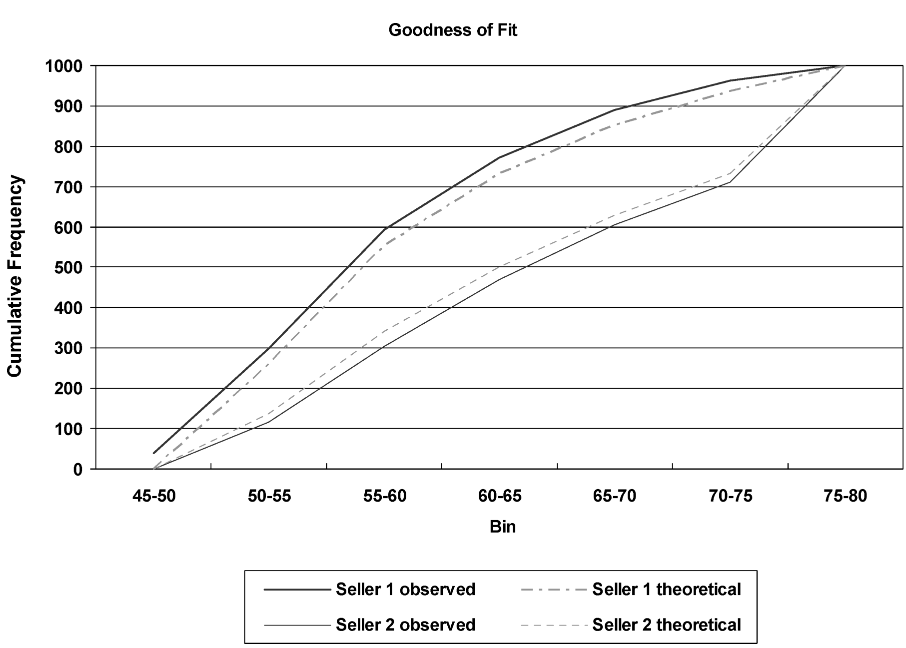
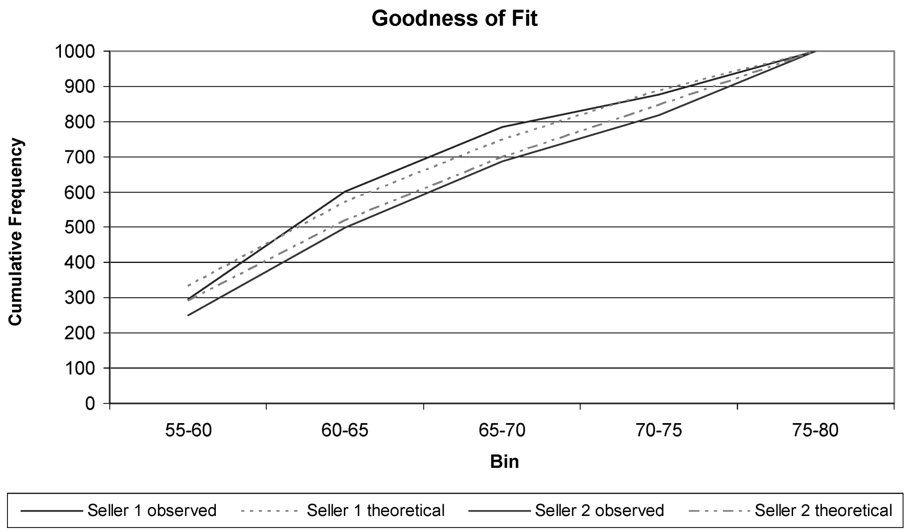


Figure 4: The two–heterogeneous-seller case comparing simulated results with theoretical results with $k_1 = 70$, $c_1 = \$20$, $k_2 = 60$, and $c_2 = \$30$.



with the higher indifference price) and which one is the seller S_L (the seller with the lower indifference price). These are found after computing the values p_L^* and p_H^* from equations (7) and (8), respectively.

We recall that there were two main learning objectives for the agents of the simulation exercises: The seller with the lower indifference price should learn not to price below the other seller's indifference price, and the seller with the higher indifference price should price his bid with positive probability at the seller's reservation price.

We note that the algorithm is able to distinguish between the two different sellers, and the seller with the higher indifference price ends up pricing his bid at the reservation price with positive probability (e.g., for the two-seller case depicted in Figure 3, the seller with the higher indifference price ends up choosing the residual price of \$80 in 46 simulation runs). Further, as both agents very rarely price their bids below the lower of the two indifference prices, it becomes apparent that they realize quickly that such behavior is strictly dominated by other strategies. Both these results are significant because it shows that even complex mixed-strategy equilibrium strategies, where the probability distribution is discontinuous at certain price points, can be effectively assimilated as learned behavior through artificial learning algorithms. The promise of using artificial agents in online transactions would be deemed credible if it could be demonstrated that these agents were effectively learning the behavior of complex but known theoretical equilibria, and the above algorithm demonstrates just that.

The results have interesting ramifications in real-world scenarios. Managers might not have the luxury of learning over a large number of observations themselves as in these simulations, but they can utilize the *organizational memory* (i.e., the experiences of them as well as their predecessors) to effectively build the learning capability over time. Managers also have their own intuition, which these artificial agents lack that might result in accelerated learning toward the equilibrium (and therefore optimal) behavior. If this learning process is considered to be the analogue of the process by which managers analyze their past actions, it becomes easy to understand how a mixed-strategy equilibrium can develop as an emerging behavior without the players actually resorting to game-theoretic calculations. Of course, real-life competition would be significantly more complex than these simple symmetric equilibria, and we wish to explore these considerations in our future work.

The *n*-Player Simulations

We first tested our algorithm for symmetric sellers using the known result of equation (4). The results of simulation with three and five identical players are summarized in Tables A6–A12. We carried out simulations for five sets of parameter values for the three-player case and four sets of parameter values for the five-player case. For considerations of space, we show the results for two of the five-player simulation exercises, as the results of the other two exercises (with $k = 23$, $c =$ \$20 and $k = 24$, $c =$ \$20, respectively) were very similar. To explain the results in greater detail, consider Table A6 that shows the frequency distribution (simulated and theoretical) of the three sellers who have identical capacities (40 units) and marginal costs (\$20). Just as in the two-seller cases, we compute the theoretical cumulative frequency distributions in the price ranges 50–55, 55–60, . . . 70–75, and 75–80. These are shown in column 5. The simulation cumulative frequency distributions are shown in column 3. As before, the absolute difference of the two cumulative frequency distributions forms the basis of the hypothesis testing. Significantly, the Kolmogorov-Smirnov tests do not reject the null hypothesis in each of these cases.

No general analytical results exist for a theoretical equilibrium with three or more nonidentical sellers. But as we obtained promising results using artificial learning mechanisms in the (nonidentical) two- and (identical) *n*-seller models, we decided to explore the algorithm using three and five dissimilar sellers and see whether we get any meaningful insights. In each of the cases, we ensured that no seller was superfluous (i.e., any seller would sell a positive quantity as long as his price was less than or equal to the reservation price), as otherwise the equilibrium settles down into a cost-cutting Bertrand equilibrium.

The results with more than two players are summarized in Tables A13 and A14. As noted before, we do not have the means of comparing the results with any theoretical results (the various columns just indicate the observed frequency distributions of the different sellers in the simulations), but the results are consistent enough to comment on certain general characteristics:

i. Most importantly, the simulation results show that the concept of the indifference point as a strategic variable for competition still holds for multiple players. The indifference price for the *i*th seller is given by $p_i^* = \frac{(r-c_i)(Q-\sum_{j=1,}^{n} k_{j,j\neq i})}{k_i} + c_i$. In both the simulations, if one compares any two sellers, the seller with the lower indifference price on average bids lower than the seller with the higher indifference price and thus supplies to capacity more often.

ii. We also note that, when there are more than two sellers, only the seller with the highest price supplies the residual under the setting of this model, and therefore the best response of a seller is to try to be the seller with the second-highest bid.

The significance of the indifference price is of particular interest. The formulation of the indifference price shows us that a seller with a lower marginal cost of production or a higher production capacity has an advantage relative to its competition. What does this result mean operationally? In an industrial reverse-auction bid setting, where the product is relatively homogeneous (e.g., memory chips), a seller has two choices to increase its chances of winning: either lower its marginal cost of production or increase its production capacity. While lowering marginal costs might involve reengineering one's processes to extract better efficiencies, the choice of increasing production capacity is more direct: build more plants and/or acquire its competition. Perhaps as a result we might observe consolidation in industries that have homogeneous products, driven either through merging of competitors or the exit of relatively inefficient producers.

The Nonstrategic Player Simulations

The simulations so far consider the behavior of the sellers in environments where both sellers are expected to respond (theoretically) with price randomizations. However, the analytical results derived earlier in this article indicate that, when one seller behaves nonstrategically, the best response of the other (strategizing) seller is a pure strategy. In other words, even if the strategizing seller might start off with price randomizations, it should learn over time to bid around a single price that maximizes his expected profit (note that ideally the strategy would be a degenerate distribution placing the entire mass on a single price, but as the algorithm—by the very nature of RL algorithms—responds to the bidding behavior of the other seller, the result shows some perturbations even after learning the ideal behavior). Because this pure-strategy response is materially very different from the prior mixed-strategy results, the response of the RL algorithm is, therefore, of much interest.

Thus, the final set of simulations with our algorithm was conducted to test the behavior of a strategic agent who has to learn to bid only a particular price given that the other seller continues to bid uniformly across the interval of $[c, r]$. The theoretical result indicates the single price to be given by the expression in equation (6). We conducted different sets of simulations with two homogeneous sellers having the capacity and marginal cost combinations of [90, \$35], [80, \$20], [75, \$40], [70, \$30], and [60, \$60], respectively. As before, the buyer demand was kept at 100 units and his reservation price was fixed at \$80.

Once again, we carried out 1,500 simulation runs with each of the parameter values, and in order to discount the learning phase of the strategic agent, we discounted the first 500 observations and used the remaining 1,000 observations to verify whether the strategic agent had effectively learned the theoretically predicted behavior. The Kolmogorov-Smirnov test could not reject the null hypothesis. The strategic seller nearly always placed his bid at around one fixed value, which was very nearly the optimal bid. Because the learning process and the results of the five simulations are very similar, we just show the simulation results with the first two set of parameter values in Tables A15 and A16. The theoretical probability distribution in strategies in this case is just a degenerate distribution at the price predicted by equation (6). Note that we do not indicate the bid distributions of the nonstrategic seller as his bid probability distribution in these simulations is uniform by design.

CONCLUSIONS

This study explores the potential of using artificial agents to automate transactions in electronic marketplaces, particularly reverse auctions. The theoretical results of a mixed-strategy equilibrium in capacity-constrained reverse auctions involving two dissimilar competitors are successfully replicated. We also successfully replicate the mixed-strategy equilibria in capacity-constrained reverse auctions with more than two similar sellers through the simulation results. Importantly, useful insights were gathered through the three– and five–dissimilar-player games that were previously unavailable given the difficulty in obtaining closed-form solutions to the analytical model. Finally, we develop new analytical results that determine the best response of a seller responding to another seller behaving nonstrategically and show that the resultant response is a pure strategy. Encouragingly, the RL algorithm is able to successfully converge on the pure strategy even though the agents start off initially by placing their bids according to a uniform distribution.

All of these results provide further support for the design and incorporation of artificial learning mechanisms that will enable organizations to take greater part in the increasing number of electronic marketplaces. Both the marketplaces and the managers involved in the decision making for placing the bids stand to benefit from such advances in technology.

As we indicated in the introduction, with such e-marketplaces becoming more and more prevalent, the number of transactions in such environments has exploded. To take one typical and well-suited example in the current context, the B2B portal MetalJunction.com, which is the world's largest e-marketplace for steel, has a buyer community of over 5,400 who buy an average of over 150,000 tons of steel per month through the portal. With such large volumes involved, it is imperative that managers in charge of bidding on such platforms bid efficiently. While it would be presumptuous to assume that our algorithms can replace other bidding mechanisms in these environments, the results of this article do indicate an interesting direction of research whereby increasingly sophisticated artificial learning algorithms can at least assist experienced managers in deciding on a bid level in a competitive, reverse-auction environment. With transactions in this electronic auction environment that is now estimated to grow to over $14 trillion (OYO_Media, 2007), this is an issue of increasing importance to businesses worldwide.

There are several advantages of having an efficient artificial agent in these environments. For example, such agents can effectively take out any emotional factors that might color a human agent's decisions. Further, especially if the equilibrium strategy is complicated, the agents can compute the necessary frequency distributions of the bids and enforce that behavior without errors. We do note that it presumably would take much more sophisticated algorithms that can credibly replicate bidding behavior in live environments, but we hope that our results provide some impetus for further research in this direction.

A possible prescription for implementation of these algorithms might be to use them initially in mock environments (with the data of past auctions being used as training data, which are the analogue of discarding the first 500 simulation runs in our research). The artificial agents can then be gradually introduced in a monitored environment in routine procurement activities (e.g., office supplies), with regular audit to compare performance of the agents as compared to human bidders.

In our future research, we intend to utilize more sophisticated algorithms to gain further insights into seller agent behavior in the *n*–dissimilar-player model. Varying the value of the discount parameter, γ, especially toward smaller discount values greater than 0 has been shown to impact convergence, at least for single RL agents using lookup tables (Sridharan & Tesauro, 2000). Given the commoditized nature of the products under consideration, lookup tables are not entirely practical given the large number of price values when continuous prices are used, but research into this area is ongoing. We also wish to consider more complex models of competition. The complexity of the e-marketplace could be better modeled by increasing the number of buyers, sellers, and commodities that constitute the market. Furthermore, the artificial agents employed in this simulation could be enhanced to capture a wider range of behaviors exhibited by managers participating in B2B exchanges. [Received: May 2006. Accepted: October 2007.]

REFERENCES

Allen, B., Deneckere, R., Faith, T., & Kovenock, D. (2000). Capacity precommitment as barrier A to entry: A Bertrand-Edgeworth approach. *Economic Theory, 15*, 501–530.

Bandyopadhyay, S., Barron, J. M., & Chaturvedi, A. R. (2005). Competition among sellers in online exchanges. *Information Systems Research, 16*, 47–60.

Bandyopadhyay, S., Barron, J. M., & Chaturvedi, A. R. (2008). Capacity and entry issues in online exchanges. *European Journal of Operational Research, 185*, 849–863.

Bandyopadhyay, S., Rees, J., & Barron, J. M. (2006). Simulating sellers in online exchanges. *Decision Support Systems, 41*, 500–513.

Bell, A. M. (2001). Reinforcement learning rules in a repeated game. *Computational Economics, 18*, 89–111.

Chemconnect. (2004). Chemconnect fact sheet. Accessed February, 2004 at http://www.chemconnect.com/pdfs/FactSheet.pdf.

Chiang, W.-C., Fitzsimmons, J., Huang, Z., & Li, S. X. (1994). A game-theoretic approach to quantity discount problems. *Decision Sciences, 25*, 153–168.

Cliff, D. (1997). *Minimal-intelligence agents for bargaining behaviors in market-based environments*. Research reports, Hewlett Packard Labs, Palo Alto, CA.

Dai, Q., & Kauffman, R. J. (2003). *To B2B or not to B2B*. Working paper, Carlson School of Management, University of Minnesota, Twin Cities, MN.

Dixon-Burger, M., Howe, C. D., Schaeffer, J., & Freiden, S. (2001). *Sellers make tough BTO choices*. Research report, Forrester Research, Cambridge, MA.

Epstein, J. M., & Axtell, R. (1996). *Growing artificial societies: Social science from the bottom up*. Washington, DC: Brookings Institution.

Erev, I., & Rapoport, A. (1998). Coordination, 'magic,' and reinforcement learning in a market entry game. *Games and Economic Behavior, 23*, 146–175.

Erev, I., & Roth, A. E. (1998). Predicting how people play games: Reinforcement learning in experimental games with unique, mixed strategy equilibria. *American Economic Review, 88*, 848–881.

Greenwald, A. R., & Kephart, J. O. (2001). Probabilistic pricebots. *Fifth International Conference on Autonomous Agents (AGENTS '01)*. Montreal, Canada: ACM, 560–567.

Helper, S., & Macduffie, J. P. (2003). B2B and modes of exchange: Evolutionary and transformative effects. In B. Kogut (Ed.), *The global internet economy.* Cambridge, MA: MIT Press, 536.

Hohner, G., Rich, J., Ng, E., Reid, G., Davenport, A. J., Kalagnanam, J. R., et al. (2003). Combinatorial and quantity-discount procurement auctions benefit Mars, Incorporated and its suppliers. *Interfaces, 33*, 23–35.

Japan Ministry of Internal Affairs and Communications. (2006). *Information and communications in Japan*. Tokyo, Japan: Author.

Katok, E., & Roth, A. E. (2004). Auctions of homogeneous goods with increasing returns: Experimental comparison of alternative Dutch auctions. *Management Science, 50*, 1044–1063.

Kephart, J. O., Hanson, J. E., & Greenwald, A. R. (2000). Dynamic pricing by software agents. *Computer Networks, 32*, 731–752.

Kerrigan, R., Roegner, E. V., Swinford, D. D., & Zawada, C. C. (2001). B2basics. *The McKinsey Quarterly*, February, 45–53.

Kreps, D., & Scheinkman, D. (1983). Quantity precommitment and Bertrand competition yield Cournot outcomes. *The Rand Journal of Economics, 14*, 326–337.

Kutschinski, E., Ulthmann, T., & Polani, D. (2003). Learning competitive pricing strategies by multi-agent reinforcement learning. *Journal of Economic Dynamics and Control, 27*, 2207–2218.

Mahajan, J., Radas, S., & Vakharia, A. J. (2002). Channel strategies and stocking policies in uncapacitated and capacitated supply chains. *Decision Sciences, 33*, 191–222.

Oliver, J. (1996). A machine learning approach to automated negotiation and prospects for electronic commerce. *Journal of Management Information Systems, 13*(3), 83–112.

OYO_Media. (2007). B2B software. Accessed April 25, 2007, http://www.oyo-media.com/static.php?page=B2B_SOFTWARE&menuID=38.

Rapoport, A., Daniel, T. E., & Searle, D. A. (1998). Reinforcement-based adaptive learning in asymmetric two-person bargaining with incomplete information. *Experimental Economics, 1*, 221–253.

Renaissance in cyberspace. (2003). *The Wall Street Journal*, November 20, D1.

Rubin, P. A., & Benton, W. C. (1993). Jointly constrained order quantities with all-units discounts. *Naval Research Logistics, 40*, 255–278.

Saeed, K. A., Malhotra, M. K., & Grover, V. (2005). Examining the impact of interorganizational systems on process efficiency and sourcing leverage in buyer-supplier dyads. *Decision Sciences, 36*, 365–396.

Slusher, E. A., Sims, H. P., & Thiel, J. (1978). Bargaining behavior in a business simulation game. *Decision Sciences, 9*, 310–321.

Sridharan, M., & Tesauro, G. J. (2000). Multi-agent Q-learning and regression trees for automated pricing decisions. *Seventeenth International Conference on Machine Learning (ICML '00)*. Stanford, CA, 927–934.

Stone, P., Schapire, R. E., Littman, M. L., Csirik, J. A., & McAllester, D. (2003). Decision-theoretic bidding based on learned density models in simultaneous, interacting auctions. *Journal of Artificial Intelligence Research, 19*, 209–242.

Sutton, R. S., & Barto, A. G. (1998). *Reinforcement learning: An introduction*. Cambridge, MA: MIT Press.

Tesauro, G. (2001). Pricing in agent economies using neural networks and multi-agent Q-learning. *Lecture Notes in Computer Science*, 288–307.

Tesauro, G., & Das, R. (2001). High-performance bidding agents for the continuous double auction. *3rd ACM Conference on Electronic Commerce (EC-2001)*. Tampa, FL: ACM, 206–209.

Tesauro, G. J., & Kephart, J. O. (2002). Pricing in agent economies using multi-agent Q-learning. *Autonomous Agents and Multi-Agent Systems, 5*, 289–304.

Thorndike, E. L. (2000). *Animal intelligence: Experimental studies*. New Brunswick, NJ: Transaction Publishers.

Tomak, K., & Xia, M. (2002). Evolution of B2B marketplaces. *Electronic Markets, 12*(2), 84–91.

Wolf, G., & Shubik, M. (1978). Market structure, opponent behavior, and information in a market game. *Decision Sciences, 9*, 421–428.

Xiao, T., Xia, Y., & Zhang, G. P. (2007). Strategic outsourcing decisions for manufacturers that produce partially substitutable products in a quantity-setting duopoly situation. *Decision Sciences, 38*, 81–106.

Zhu, K. (2004). Information transparency of business-to-business electronic markets: A game-theoretic analysis. *Management Science, 50*, 670–685.

APPENDIX: SIMULATION RESULTS

In the following tables, all theoretical frequencies have been rounded to the nearest whole numbers. As discussed in the main text, the first 500 rounds of simulations are not considered for analysis.

Table A1: Two-seller simulation run results with $k_1 = 80$, $c_1 = \$20$, $k_2 = 65$, and $c_2 = \$40$. Number of rounds = 1,500, $Q = 100$, and $r = \$80$.

Bin	Seller 1 Freq.	Seller 1 Cum. Freq.	Seller 1 Th. Freq.	Seller 1 Th. Cum. Freq.	Cumulative Difference
45–50	40	40	0	0	.040
50–55	259	299	259	259	.040
55–60	295	594	296	555	.038
60–65	179	773	178	733	.039
65–70	117	890	119	852	.038
70–75	73	963	85	937	.026
75–80	37	1,000	63	1,000	0
Bin	**Seller 2 Freq.**	**Seller 2 Cum. Freq.**	**Seller 2 Th. Freq.**	**Seller 2 Th. Cum. Freq.**	**Cumulative Difference**
45–50	0	0	0	0	0
50–55	116	116	137	137	.020
55–60	188	304	205	342	.038
60–65	165	469	159	501	.032
65–70	136	605	128	629	.024
70–75	107	712	104	733	.021
75–80	288	1,000	267	1,000	0

Table A2: Two-seller simulation run results with $k_1 = 70$, $c_1 = \$20$, $k_2 = 60$, and $c_2 = \$30$. Number of rounds = 1,500, $Q = 100$, and $r = \$80$.

Bin	Seller 1 Freq.	Seller 1 Cum. Freq.	Seller 1 Th. Freq.	Seller 1 Th. Cum. Freq.	Cumulative Difference
55–60	296	296	333	333	.037
60–65	306	602	238	571	.031
65–70	184	786	179	750	.036
70–75	91	877	139	889	.012
75–80	123	1,000	111	1,000	0
Bin	**Seller 2 Freq.**	**Seller 2 Cum. Freq.**	**Seller 2 Th. Freq.**	**Seller 2 Th. Cum. Freq.**	**Cumulative Difference**
55–60	251	251	292	292	.041
60–65	247	498	227	519	.021
65–70	189	687	181	700	.013
70–75	132	819	148	848	.029
75–80	181	1,000	152	1,000	0

Table A3: Two-seller simulation run results with $k_1 = 65$, $c_1 = \$40$, $k_2 = 75$, and $c_2 = \$20$. Number of rounds = 1,500, $Q = 100$, and $r = \$80$.

Bin	Seller 1 Freq.	Seller 1 Cum. Freq.	Seller 1 Th. Freq.	Seller 1 Th. Cum. Freq.	Cumulative Difference
50–55	103	103	130	130	.027
55–60	199	302	204	334	.032
60–65	172	474	158	492	.018
65–70	151	625	122	614	.011
70–75	76	701	108	722	.021
75–80	299	1,000	278	1,000	0
Bin	**Seller 2 Freq.**	**Seller 2 Cum. Freq.**	**Seller 2 Th. Freq.**	**Seller 2 Th. Cum. Freq.**	**Cumulative Difference**
50–55	286	286	245	245	.041
55–60	306	578	302	547	.031
60–65	178	760	179	726	.034
65–70	100	870	120	846	.024
70–75	83	953	86	932	.021
75–80	47	1,000	68	1,000	0

Table A4: Two-seller simulation run results with $k_1 = 45$, $c_1 = \$30$, $k_2 = 85$, and $c_2 = \$35$. Number of rounds = 1,500, $Q = 100$, and $r = \$80$.

Bin	Seller 1 Freq.	Seller 1 cum. Freq.	Seller 1 Th. Freq.	Seller 1 Th. Cum. Freq.	Cumulative Difference
55–60	0	0	0	0	0
60–65	801	801	786	786	.015
65–70	92	893	89	875	.018
70–75	71	964	58	933	.031
75–80	36	1,000	67	1,000	0
Bin	**Seller 2 Freq.**	**Seller 2 Cum. Freq.**	**Seller 2 Th. Freq.**	**Seller 2 Th. Cum. Freq.**	**Cumulative Difference**
55–60	32	32		0	.032
60–65	46	78		93	.015
65–70	176	254		285	.031
70–75	476	730		766	.036
75–80	270	1,000		1,000	0

Table A5: Two-seller simulation run results with $k_1 = 65$, $c_1 = \$30$, $k_2 = 80$, and $c_2 = \$25$. Number of rounds = 1,500, $Q = 100$, and $r = \$80$.

Bin	Seller 1 Freq.	Seller 1 Cum. Freq.	Seller 1 Th. Freq.	Seller 1 Th. Cum. Freq.	Cumulative Difference
55–60	270	270	289	289	.019
60–65	222	492	227	516	.024
65–70	225	717	178	694	.023
70–75	116	833	158	852	.019
75–80	167	1,000	148	1,000	0
Bin	**Seller 2 Freq.**	**Seller 2 Cum. Freq.**	**Seller 2 Th. Freq.**	**Seller 2 Th. Cum. Freq.**	**Cumulative Difference**
55–60	297	297	327	327	.030
60–65	305	602	253	580	.022
65–70	165	767	164	744	.023
70–75	105	872	153	897	.025
75–80	128	1,000	103	1,000	0

Table A6: Symmetric three-seller simulation runs with $k = 40$, $c = \$20$. Number of rounds = 1,500, $Q = 100$, $r = \$80$.

Bin	Seller 1 Freq.	Seller 1 Cum. Freq.	Seller 1 Th. Freq.	Seller 1 Th. Cum. Freq.	Cumulative Difference
50–55	363	363	383	383	.02
55–60	269	632	285	668	.036
60–65	256	888	202	870	.018
65–70	88	976	93	963	.013
70–75	16	992	25	988	.004
75–80	8	1,000	12	1,000	0
Bin	**Seller 2 Freq.**	**Seller 2 Cum. Freq.**	**Seller 2 Th. Freq.**	**Seller 2 Th. Cum. Freq.**	**Cumulative Difference**
50–55	402	402	383	383	.019
55–60	303	705	285	668	.037
60–65	193	898	202	870	.028
65–70	90	988	93	963	.025
70–75	9	997	25	988	.009
75–80	3	1,000	12	1,000	0
Bin	**Seller 3 Freq.**	**Seller 3 Cum. Freq.**	**Seller 3 Th. Freq.**	**Seller 3 Th. Cum. Freq.**	**Cumulative Difference**
50–55	342	342	383	383	.039
55–60	298	640	285	668	.028
60–65	193	833	202	870	.037
65–70	103	936	93	963	.036
70–75	34	970	25	988	.018
75–80	30	1,000	12	1,000	0

Table A7: Symmetric three-seller simulation runs with $k = 35$, $c = \$20$. Number of rounds = 1,500, $Q = 100$, $r = \$80$.

Bin	Seller 1 Freq.	Seller 1 Cum. Freq.	Seller 1 Th. Freq.	Seller 1 Th. Cum. Freq.	Cumulative Difference
70–75	412	412	435	435	.023
75–80	588	1,000	565	1,000	.023
Bin	**Seller 2 Freq.**	**Seller 2 Cum. Freq.**	**Seller 2 Th. Freq.**	**Seller 2 Th. Cum. Freq.**	**Cumulative Difference**
70–75	424	424	435	435	.011
75–80	576	1,000	565	1,000	.011
Bin	**Seller 3 Freq.**	**Seller 3 Cum. Freq.**	**Seller 3 Th. Freq.**	**Seller 3 Th. Cum. Freq.**	**Cumulative Difference**
70–75	401	401	435	435	.034
75–80	599	1,000	565	1,000	.034

Table A8: Symmetric three-seller simulation runs with $k = 45$, $c = \$20$. Number of rounds = 1,500, $Q = 100$, $r = \$80$.

Bin	Seller 1 Freq.	Seller 1 Cum. Freq.	Seller 1 Th. Freq.	Seller 1 Th. Cum. Freq.	Cumulative Difference
30–35	32	32	49	49	.017
35–40	85	117	63	112	.005
40–45	52	169	85	197	.028
45–50	103	272	88	285	.013
50–55	130	402	128	413	.011
55–60	135	537	152	565	.028
60–65	155	692	165	730	.038
65–70	121	813	122	852	.039
70–75	109	922	81	933	.011
75–80	78	1,000	67	1,000	0
Bin	**Seller 2 Freq.**	**Seller 2 Cum. Freq.**	**Seller 2 Th. Freq.**	**Seller 2 Th. Cum. Freq.**	**Cumulative Difference**
30–35	38	38	49	49	.011
35–40	86	124	63	112	.012
40–45	62	186	85	197	.011
45–50	81	267	88	285	.018
50–55	131	398	128	413	.015
55–60	144	542	152	565	.023
60–65	159	701	165	730	.029
65–70	113	814	122	852	.038
70–75	103	917	81	933	.016
75–80	83	1,000	67	1,000	0

Continued

Table A8: (Continued)

Bin	Seller 3 Freq.	Seller 3 Cum. Freq.	Seller 3 Th. Freq.	Seller 3 Th. Cum. Freq.	Cumulative Difference
30–35	42	42	49	49	.007
35–40	88	130	63	112	.018
40–45	61	191	85	197	.006
45–50	88	279	88	285	.006
50–55	129	408	128	413	.005
55–60	146	554	152	565	.011
60–65	156	710	165	730	.02
65–70	116	826	122	852	.026
70–75	95	921	81	933	.012
75–80	79	1,000	67	1,000	0

Table A9: Symmetric three-seller simulation runs with $k = 42$, $c = \$20$. Number of rounds = 1,500, $Q = 100$, $r = \$80$.

Bin	Seller 1 Freq.	Seller 1 Cum. Freq.	Seller 1 Th. Freq.	Seller 1 Th. Cum. Freq.	Cumulative Difference
40–45	67	67	78	78	.011
45–50	107	174	104	182	.008
50–55	111	285	121	303	.018
55–60	131	416	132	435	.019
60–65	115	531	135	570	.039
65–70	172	703	153	723	.02
70–75	190	893	154	877	.016
75–80	107	1,000	123	1,000	0

Bin	Seller 2 Freq.	Seller 2 Cum. Freq.	Seller 2 Th. Freq.	Seller 2 Th. Cum. Freq.	Cumulative Difference
40–45	73	73	78	78	.005
45–50	107	180	104	182	.002
50–55	111	291	121	303	.012
55–60	110	401	132	435	.034
60–65	127	528	135	570	.042
65–70	163	691	153	723	.032
70–75	196	887	154	877	.01
75–80	113	1,000	123	1,000	0

Bin	Seller 3 Freq.	Seller 3 Cum. Freq.	Seller 3 Th. Freq.	Seller 3 Th. Cum. Freq.	Cumulative Difference
40–45	84	84	78	78	.006
45–50	115	199	104	182	.017
50–55	133	332	121	303	.029
55–60	137	469	132	435	.034
60–65	142	611	135	570	.041
65–70	89	700	153	723	.023
70–75	203	903	154	877	.026
75–80	97	1,000	123	1,000	0

Table A10: Symmetric three-seller simulation runs with $k = 38$, $c = \$20$. Number of rounds = 1,500, $Q = 100$, $r = \$80$.

Bin	Seller 1 Freq.	Seller 1 Cum. Freq.	Seller 1 Th. Freq.	Seller 1 Th. Cum. Freq.	Cumulative Difference
55–60	91	91	102	102	.011
60–65	113	204	128	230	.026
65–70	313	517	304	534	.017
70–75	295	812	255	789	.023
75–80	188	1,000	211	1,000	0
Bin	**Seller 2 Freq.**	**Seller 2 Cum. Freq.**	**Seller 2 Th. Freq.**	**Seller 2 Th. Cum. Freq.**	**Cumulative Difference**
55–60	88	88	102	102	.014
60–65	179	267	128	230	.037
65–70	301	568	304	534	.034
70–75	263	831	255	789	.042
75–80	169	1,000	211	1,000	0
Bin	**Seller 3 Freq.**	**Seller 3 Cum. Freq.**	**Seller 3 Th. Freq.**	**Seller 3 Th. Cum. Freq.**	**Cumulative Difference**
55–60	111	111	102	102	.009
60–65	132	243	128	230	.013
65–70	327	570	304	534	.036
70–75	251	821	255	789	.032
75–80	179	1,000	211	1,000	0

Table A11: Symmetric five-seller simulation runs with $k = 22$, $c = \$20$. Number of rounds = 1,500, $Q = 100$, $r = \$80$.

Bin	Seller 1 Freq.	Seller 1 Cum. Freq.	Seller 1 Th. Freq.	Seller 1 Th. Cum. Freq.	Cumulative Difference
50–55	619	619	615	615	.004
55–60	213	832	180	795	.037
60–65	61	893	85	880	.013
65–70	5	898	54	934	.036
70–75	44	942	38	972	.030
75–80	58	1,000	28	1,000	0
Bin	**Seller 2 Freq.**	**Seller 2 Cum. Freq.**	**Seller 2 Th. Freq.**	**Seller 2 Th. Cum. Freq.**	**Cumulative Difference**
50–55	621	621	615	615	.006
55–60	204	825	180	795	.030
60–65	53	878	85	880	.002
65–70	18	896	54	934	.038
70–75	60	956	38	972	.016
75–80	44	1,000	28	1,000	0

Continued

Table A11: (Continued)

Bin	Seller 3 Freq.	Seller 3 Cum. Freq.	Seller 3 Th. Freq.	Seller 3 Th. Cum. Freq.	Cumulative Difference
50–55	622	622	615	615	.007
55–60	213	835	180	795	.040
60–65	46	881	85	880	.001
65–70	12	893	54	934	.041
70–75	39	932	38	972	.040
75–80	68	1,000	28	1,000	0

Bin	Seller 4 Freq.	Seller 4 Cum. Freq.	Seller 4 Th. Freq.	Seller 4 Th. Cum. Freq.	Cumulative Difference
50–55	628	628	615	615	.013
55–60	209	837	180	795	.042
60–65	65	902	85	880	.022
65–70	1	903	54	934	.031
70–75	40	943	38	972	.029
75–80	57	1,000	28	1,000	0

Bin	Seller 5 Freq.	Seller 5 Cum. Freq.	Seller 5 Th. Freq.	Seller 5 Th. Cum. Freq.	Cumulative Difference
50–55	626	626	615	615	.011
55–60	200	826	180	795	.031
60–65	64	890	85	880	.010
65–70	4	894	54	934	.040
70–75	42	936	38	972	.036
75–80	64	1,000	28	1,000	0

Table A12: Symmetric five-seller simulation runs with $k = 21$, $c = \$20$. Number of rounds = 1,500, $Q = 100$, $r = \$80$.

Bin	Seller 1 Freq.	Seller 1 Cum. Freq.	Seller 1 Th. Freq.	Seller 1 Th. Cum. Freq.	Cumulative Difference
65–70	333	333	322	322	.011
70–75	317	650	358	680	.03
75–80	350	1,000	320	1,000	0

Bin	Seller 2 Freq.	Seller 2 Cum. Freq.	Seller 2 Th. Freq.	Seller 2 Th. Cum. Freq.	Cumulative Difference
65–70	307	307	322	322	.015
70–75	349	656	358	680	.024
75–80	344	1,000	320	1,000	0

Bin	Seller 3 Freq.	Seller 3 Cum. Freq.	Seller 3 Th. Freq.	Seller 3 Th. Cum. Freq.	Cumulative Difference
65–70	312	312	322	322	.01
70–75	389	701	358	680	.021
75–80	299	1,000	320	1,000	0

Continued

Table A12: (Continued)

Bin	Seller 4 Freq.	Seller 4 Cum. Freq.	Seller 4 Th. Freq.	Seller 4 Th. Cum. Freq.	Cumulative Difference
65–70	290	290	322	322	.032
70–75	353	643	358	680	.037
75–80	357	1,000	320	1,000	0

Bin	Seller 5 Freq.	Seller 5 Cum. Freq.	Seller 5 Th. Freq.	Seller 5 Th. Cum. Freq.	Cumulative Difference
65–70	330	330	322	322	.008
70–75	341	671	358	680	.009
75–80	329	1,000	320	1,000	0

Table A13: Three-seller simulation run results with $k_1 = 40$, $c_1 = \$15$, $k_2 = 40$, $c_2 = \$20$, $k_3 = 35$, and $c_3 = \$25$. Number of rounds = 1,500, $Q = 100$, and $r = \$80$.

Bin	Seller 1 Freq.	Seller 2 Freq.	Seller 3 Freq.
55–60	259	228	240
60–65	375	363	357
65–70	251	279	266
70–75	80	92	92
75–80	35	38	45

Table A14: Five-seller simulation run results with $k_1 = 35$, $c_1 = \$15$, $k_2 = 15$, $c_2 = \$15$, $k_3 = 30$, $c_3 = \$20$, $k_4 = 15$, $c_4 = \$25$, $k_5 = 10$, and $c_5 = \$30$. Number of rounds = 1,500, $Q = 100$, and $r = \$80$.

Bin	Seller 1 Freq.	Seller 2 Freq.	Seller 3 Freq.	Seller 4 Freq.	Seller 5 Freq.
65–70	0	5	0	45	69
70–75	593	594	613	594	520
75–80	407	401	387	361	411

Table A15: Two homogenous-seller simulation run with one nonstrategic seller with $k = 90$, $c = \$30$. Number of rounds = 1,500, $Q = 100$, and $r = \$80$. Results shown for the strategic seller.

Bin	Strat, Seller Freq.	Strat. Seller Cum. Freq.	Strat. Seller Th. Freq.	Strat. Seller Th. Cum. Freq.	Cumulative Difference
60–65	965	965	1,000	1,000	.035
65–70	28	993	0	1,000	.007
70–75	6	999	0	1,000	.001
75–80	1	1,000	0	1,000	0

Table A16: Two homogenous-seller simulation run with one nonstrategic seller with $k = 80$, $c = \$20$. Number of rounds = 1,500, $Q = 100$, and $r = \$80$. Results shown for the strategic seller.

Bin	Strat, Seller Freq.	Strat. Seller Cum. Freq.	Strat. Seller Th. Freq.	Strat. Seller Th. Cum. Freq.	Cumulative Difference
60–65	960	960	1,000	1,000	.040
65–70	27	987	0	1,000	.013
70–75	9	996	0	1,000	.004
75–80	4	1,000	0	1,000	0

Subhajyoti Bandyopadhyay is currently an assistant professor in the Department of Information Sciences and Operations Management in the University of Florida, Gainesville. He received his PhD in management information systems from Purdue University in 2002. His work has been published or is set to appear in *Information Systems Research, Journal of Operations Management, European Journal of Operational Research, Decision Support Systems, Communications of the ACM, International Journal of Electronic Commerce, Journal of Electronic Commerce Research, Journal of Organizational Computing and Electronic Commerce, Lecture Notes in Computer Science*, and the *Database for Advances in Information Systems*. His current research interests include economics of information systems, electronic commerce strategies, and information systems policy issues, especially in the area of net neutrality.

Jackie Rees is an associate professor in the Krannert School of Management at Purdue University. She earned her doctorate from the University of Florida in 1998. She has published in *Decision Support Systems, INFORMS Journal of Computing, Communications of the ACM, European Journal of Operational Research*, and *Information Technology and Management*. Her current research interests are in the intersection of information security risk management and machine learning.

John M. Barron is the Loeb Professor of Economics in the Department of Economics at the Krannert School of Management at Purdue University. He earned his doctorate from Brown University in 1976. He has published widely in economics journals. His current research interests cover a variety of topics in labor economics and industrial organization; these papers typically have in common elements of the economics of information and uncertainty.