CERIAS Tech Report 2005-118

SIMULATING SELLERS IN ONLINE EXCHANGES

by Subhajyoti Bandyopadhyay, Jackie Rees, John M. Barron

Center for Education and Research in
Information Assurance and Security,
Purdue University, West Lafayette, IN 47907-2086

Available online at www.sciencedirect.com
SCIENCE DIRECT®
Decision Support Systems 41 (2006) 500–513
Decision Support Systems
www.elsevier.com/locate/dsw

Simulating sellers in online exchanges☆

Subhajyoti Bandyopadhyay[a,*], Jackie Rees[b], John M. Barron[c]

[a]University of Florida, Gainesville, FL, 364 Stuzin Hall, 32611-1310, United States
[b]Krannert Graduate School of Management, Purdue University, West Lafayette, IN, United States
[c]Krannert Graduate School of Management, Purdue University, West Lafayette, IN, United States

Received 30 December 2003; received in revised form 26 August 2004; accepted 27 August 2004
Available online 2 October 2004

Abstract

Business-to-business (B2B) exchanges are expected to bring about lower prices for buyers through reverse auctions. Analysis of such settings for seller pricing behavior often points to mixed-strategy equilibria. In real life, it is plausible that managers learn this complex ideal behavior over time. We modeled the two-seller game in a synthetic environment, where two agents use a reinforcement learning (RL) algorithm to change their pricing strategy over time. We find that the agents do indeed converge towards the theoretical Nash equilibrium. The results are promising enough to consider the use of artificial learning mechanisms in electronic marketplace transactions.

Keywords: B2B marketplaces; Reinforcement learning; Experimental economics; Game theory; Mixed-strategy equilibrium

1. Introduction

Online exchanges for business-to-business (or B2B) transactions have become ubiquitous in industries ranging from automotive to retailing. The Wall Street Journal [17] recently mentioned the remarkable turnaround of the B2B Internet commerce sector, and that US businesses spent $482 billion in B2B transactions, up 242% from 2 years before. The online research and consulting firm Jupiter Media Metrix predicts that $5.4 trillion in goods and services transacted online among businesses by 2006, while a more optimistic Gartner Group forecast estimates worldwide B2B commerce to swell to $5.9 trillion by the end of 2004. Forrester Research indicates that in Q3 2001, 49% of organizations that buy more than $1 billion per year reported using an online auction, with most of them increasing their usage of these venues.

One of the more prominent advantages of B2B exchanges is lower costs of buyers due to automation of the procurement process, reverse auctions, interoperability among users, collaborative planning and col-

☆ This research is partially funded by the National Science Foundation Grant No. DMI-0122214.
* Corresponding author. Tel.: +1 3523925946; fax: +1 3523925438.
E-mail addresses: shubho.bandyopadhyay@cba.ufl.edu (S. Bandyopadhyay), jrees@mgmt.purdue.edu (J. Rees), barron@mgmt.purdue.edu (J.M. Barron).

0167-9236/$ - see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.dss.2004.08.015

laborative design [8]. For example, Ford announced in July 2001 that it had saved $70 million through Covisint (the online automotive exchange by the Big Three automakers) in terms of reduced paperwork and lower seller prices, which is more than its initial investment in the exchange [8]. Carrier, the world's largest air-conditioning company, realized savings of over 15% in the cost of components by putting its requirements to a larger universe of sellers through an online exchange [7].

As B2B marketplaces evolve, one of the emerging roles of these marketplaces is seen as a demand aggregator for the buyers as well as a specialist in qualifying sellers [21]. Ref. [2] analyzed the competition between sellers in reverse auctions in a game-theoretic framework, and established the Nash equilibria in several scenarios. It was found that in an environment where sellers can collectively cater to the total demand, with the final (i.e. the highest-bidding) seller catering to a residual, the sellers resort to a mixed-strategy Nash equilibrium. While price randomization in industrial bids is an accepted norm, it may be argued that managers in reality do not resort to advanced game theory calculations to bid for an order. What is more likely is that managers learn that strategy over time and finally converge towards the theoretic equilibrium. This paper tests that assertion by modeling the sellers' behavior with artificial software agents that start bidding randomly, and use a simple reinforcement learning (RL) algorithm to "learn" the ideal strategy over time.

The importance of such learning algorithms is potentially very large. As electronic marketplaces proliferate among organizations, transactions such as bidding for buyer requirements in reverse auctions become ubiquitous. While arguably simple transactions like those analyzed in this research will make way for more complex auction mechanisms (for example, those which limit the number of sellers in terms of quality, product differentiation, etc.), it is undeniable that transactions in online marketplaces are here to stay. Monitoring potentially hundreds (or even thousands) of such concurrent transactions individually by human agents will conceivably be very difficult and time-consuming, if not impossible. One promising solution might be to look at artificial agents and whether they can mimic human behavior in such environments. While our experimental setting and the particular RL algorithm used might be too simplistic for real-life scenarios (and is in fact found inadequate for generalized settings), it nevertheless shows promise that even complex mixed-strategy Nash equilibria can be assimilated in artificial agent behavior through simple reinforcement learning mechanisms. Successful learning in this environment should spur further research with more complex algorithms to handle real world transactions.

The remainder of the paper is organized as follows: Section 2 presents the background literature surrounding the nature of the competition that we use for testing our learning algorithms. The use of artificial software agents and reinforcement learning in modeling such games is also discussed in Section 2. The model of the reverse auction for both two and n-sellers is provided in Section 3. Section 5 states the research assertions and hypotheses that are tested in the simulation experiment. Section 4 presents the RL algorithm deployed in the simulation. Results of the experiments are provided in Section 5 and the conclusions and future research directions are discussed in Section 6.

2. Background literature

2.1. Analysis of the market

The competition among sellers in the environment mentioned above is different from the traditional oligopolistic Cournot competition between firms facing a downward facing demand curve, where both firms sell at the same price point. It is also distinct from a capacity-constrained Bertrand model that has been analyzed extensively in the literature, where a quantity precommitment and Bertrand competition yield Cournot outcomes that have equilibrium prices above marginal cost [11].

The analysis of Ref. [2] also established the nature of the equilibrium under various assumptions of the sellers' cost, capacities and the market demand. It is to be noted that the two-seller model has been analyzed by Ref. [13], but the method of analysis in Ref. [2] provides a way to generalize the results to the n-seller model, which is what we use in our simulations. The problem is most interesting when we assume that there is no *combined* capacity constraint as such: the

sellers were supplying to the entire demand before the birth of the exchange, and continue to do so after it comes into play. However, it is conceivable that the firms individually cannot supply to the entire market. In fact, as shown in Ref. [2], the fear of being stranded as the highest priced seller who does not supply anything essentially reduces the competition to Bertrand, with all sellers supplying at cost. Since the set of sellers is limited and all are reputed in the marketplace, the buyers would not mind getting their orders fulfilled by any one or several of these sellers. This means that while there is a competition between the firms to be the low-price bidder, it is not as extreme as a Bertrand game that results in prices equal to marginal cost. However, there remains an incentive to be the low-price bidder and have the "first invitation" to supply a requirement. We note that in practice the bidder with the lowest cost is not always the winning bidder—however, under the assumptions of the model, with the buyer having no other preference than price for a homogeneous good, we make the assumption that a low-price seller is invited to supply to any residual demand before a high-price seller.

Let us suppose that there are two buyers who bought from two sellers (one from each) before the advent of the exchange.[1] What may have prevented buyers from establishing contact with both the sellers (and vice versa) are the search costs and the ongoing cost of establishing relationships within a large organization. Some of these costs are dedicated account management teams for buyers, sales force for sellers, cost of sending individual RFQs to the entire universe of sellers, etc. [10].[2] With the lack of competition, the sellers could afford to sell the required quantities to the buyers at their reservation price, which we assume to be the same for both buyers at r. With the advent of the exchange, the buyers put forward their requirements to the exchange, and the sellers can then bid for the total requirement from both buyers.

[1] The example is just illustrative, and is not crucial to the analysis. There can in fact be only a single large buyer, whose requirements cannot be met by one seller; however, the two sellers together have a combined capacity that is more than the buyer's requirement.

[2] It has been estimated that in terms of reduction of paperwork alone, B2B exchanges can bring down costs per purchase order from $75–$150 to $10–$30 [8].

Next, we discuss how the above market can be simulated by the use of intelligent artificial software agents. The use of simulation allows repeated and detailed study of the behaviors exhibited by the sellers in the market under various experimental treatment conditions. These artificial agents are endowed the ability to learn from previous actions by the use of a type of Reinforcement Learning algorithm described below.

2.2. Artificial software agents and reinforcement learning

Artificial agents have been used to simulate human agents or sellers in a number of different settings. For example, Ref. [15] used artificial software agents to conduct automated negotiations in an e-commerce environment. The use of artificial agents is advocated by Ref. [4] to study systems and structures from the "bottom up", which is especially useful when it is difficult to obtain a closed form solution to the problem at hand.

Reinforcement learning is a machine learning technique that is quite useful in situations where artificial agents need to "learn" from previous actions in order to carry out their functions. RL agents typically have a goal, receive feedback or input from the environment, can make a decision or undertake some action in response to the feedback from the environment. Additionally, a great deal of uncertainty is usually incorporated into the RL agent environment. By incorporating this uncertainty, a more realistic model of the problem is created [19]. For example, an RL agent might have a goal to win a simple auction by making the highest bid (within a specific bound). The agent would receive as feedback from its environment indicating whether the agent won the auction with the bid that was tendered. The bid tendered in the next round would 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 agents participating in the auction, just as a human agent would.

RL has been used to examine various competitive scenarios such as sealed bid k-double auction under asymmetric and incomplete information dynamics [16], market entry games [5] and rule learning in

repeated games [3]. RL is an appropriate choice for the application presented in this paper due to the ability of RL agents to incorporate previous experience (either reward or no reward) into action. The model under which the artificial agents operate in this research is discussed in the next section.

3. The model

The generalized n-seller model derives much of its intuition from the basic two-seller model, and therefore it is instructive to first consider the two-seller model in detail. We consider the case when both sellers have equal capacities K that is less than the respective individual requirements of the buyers, but their combined capacity is lesser than the total requirement of both buyers Q (i.e. $2K-Q>0$). 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, the other seller can then sell the residual demand $Q-K$. Both sellers have a common fixed marginal cost of production, c.

From the modeling point of view, it is important to note is that the entire requirement Q is auctioned to the sellers, and for any unfulfilled demand, a lower priced bidder is invited before a higher priced bidder to satisfy the unfulfilled demand. It is readily apparent that with unlimited capacity, the sellers respond with a Bertrand competition in prices with the seller or sellers with the lowest marginal cost outbidding the others.[3] This is not to the advantage of the sellers. Ref. [11] (and several variants of the original model, such as Ref. [1]) shows that if sellers could limit capacity, then a quantity precommitment and Bertrand competition yield Cournot outcomes that have equilibrium prices above marginal cost. At the other end of the spectrum, if the total capacity of the sellers is so limited as to be less than the total demand, it is easy to see that the sellers can sell their entire capacities at the buyer's reservation price.[4]

It is realistic to think of sellers having limited capacities so that any one seller cannot meet market demand. Further, keeping in mind the discussion of the previous paragraph, we stipulate that the aggregate output of the sellers exceeds total quantity demanded and that a firm sells all it can produce only if it is the low-price seller. That is, the lowest priced seller sells to capacity, but a higher priced seller only sells to a residual demand. Sellers therefore are pulled by two opposing "forces"—on one hand, higher prices fetch higher margins, but on the other, higher prices bring about increased chances of being underbid by competition.

The analysis shows that there exists a mixed-strategy equilibrium of prices where the sellers randomize between a range of prices. The intuition behind such an equilibrium is as follows: with two similar sellers, there cannot be any equilibrium in pure strategies with the sellers settling on different prices. Settling on the same price is also ruled out, since the best response to any price is to set a price that is an infinitesimal amount ε lower than that price. Thus, if any Nash equilibrium exists, it has to be a mixed-strategy equilibrium. It can further be shown that the support of the strategy lies between p_1 and r, where r is the reservation price for the buyer and p_1 is given by

$$p_1 = \frac{(r-c)(Q-K)}{K} + c \tag{1}$$

An intuitive way of looking at p_1 is that below this price, a seller makes less profit by "winning" (supply to capacity) than by "losing" and supplying the residual at the highest possible price r (which is the best price the seller can supply the residual, since he is losing anyway).

The equilibrium strategy for either seller can be expressed in terms of their price randomizing cumulative probability density function $F(p)$:

$$F(p) = \frac{(p-c)K - (r-c)(Q-K)}{(p-c)(2K-Q)} \tag{2}$$

This is a continuous probability distribution within the range (p_1, r), and effectively defines the symmetric Nash equilibrium strategy of the two sellers (i.e. the sellers). The sellers randomize their bids within this interval, such that their randomizing has a probability distribution given by $F(p)$ in Eq. (2).

[3] If seller 1 knows that seller 2 can supply to the entire demand, he responds by charging seller 2's marginal price, since the best response of seller 2 at any higher price is to undercut it by an infinitesimal amount.

[4] Since either seller can sell to capacity, there is no incentive for either to undercut competition.

By definition, the Nash equilibrium maximizes the expected return of the sellers.

The analysis is similar for the n-seller model, where the highest bidder supplies the residual, and the rest supply to capacity ($(n-1)K<Q<nK$). The support of the strategy for the sellers is given by (p_1^n, r), where

$$p_1^n = \frac{(r-c)(Q-(n-1)K)}{K} + c \tag{3}$$

and the expression for the distribution function is given by

$$F_n(p) = \left[\frac{(p-c)K-(Q-(n-1)K)(r-c)}{(p-c)(nK-Q)}\right]^{\frac{1}{n-1}} \tag{4}$$

While price randomization in industrial bids is an accepted norm, it might be difficult to accept that managers go through advanced game theory calculations (and in any case, the real-life situations are far more varied than the simplified model scenarios that make any game theory analysis extremely complex) to determine their bids. It is conceivable that sellers learn from their past experiences to bid in a fashion that maximizes their surplus. It is this assertion that we test in the remainder of this paper.

4. The algorithm

To test our assertion, we model the competing sellers as artificial software agents. We examine both the two-seller and the n-seller cases. Like human subjects, we propose that these agents "understand" the following (without resorting to explicit "knowledge" of game theory):

1. There are two opposing forces in the pricing strategy—a higher price (towards r) means greater per-unit profit, but also brings about a higher probability of "losing" to the competition (in terms of being the first invited bidder to supply the demand).
2. It does not make any sense to price below p_1, as is clear from the above analysis.
3. Since there is a need to balance between higher probability of winning and higher per-unit profit, there is no a priori reason to rule out any price between p_1 and r, and therefore, there is reason not to rule out a price-randomizing solution (at least initially).

Fig. 1 describes the simple RL algorithm that the agents employ in a two-seller game to determine their prices. The algorithm is essentially the same for the n-seller model, except that we use p_1^n to determine the support of prices, and compare the experimental distribution with $F_n(p)$ rather than F(p).

Stated formally, let us denote the average profit for subdivision i as of time t as $\bar{P}_{it}$ and the average profit across generalized n subdivisions as

$$\bar{P}_t = \frac{\left(\sum_{i=1}^{n} \bar{P}_{it}\right)}{n} \tag{5}$$

In this case, the probability of choosing a price from subdivision i is given by

$$W_{it} = \frac{\bar{P}_{it}}{n\bar{P}_t} = \frac{\sum_{i=1}^{n} \bar{P}_{it}}{n} \tag{6}$$

Note that $\sum_{i=1}^{n} w_{it} = 1$, as is required of a probability distribution.

The "sellers" thus start off initially with a totally random pricing strategy (i.e. the price distribution is uniform in its support), with the hope of learning over time about the ideal nature of the randomization. This is the same assumption as employed by Ref. [6] in their experiments and referred to as the "initial propensities" of the sellers for their pure strategies.

Thus, we attempt to find out whether through a relatively simple reinforcement learning algorithm, the sellers can finally converge on the arguably more sophisticated theoretical equilibrium. The rationale for the algorithm is as follows: since the sellers 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_1, r) with uniform probability. However, by subdividing the support and noting in which subdivision each winning or losing price falls, they ensure that they are aware of any emerging pattern of wins and losses. The stipulation of choosing at least 10 prices in each subdivision for either seller is to ensure that when the sellers start making any judgment regarding which price ranges should be favored more over others, they have some amount of experience to

For a given set of values for Q, K, c and r:

1. Determine p_1, from Eq. 1. The price range (p_1, r) determines the range from which the sellers select their prices.
2. Divide (p_1, r) into 10 subdivisions, D1, D2…D10.
3. Choose a price randomly for either seller. The winner is the seller that chooses the lower price. Note in which subdivision each of these prices belongs to, and the profit of each seller in every transaction. Profit in each round = (p-c)*Qty. *sold*, where Qty. *sold* is either K or Q-K depending on whether the seller wins or loses. Continue choosing the prices randomly until there are at least 10 selections of prices in each subdivision for either seller.
4. For each seller, maintain the average profit in each subdivision. This average profit updates itself after each game.
5. Using the average profit in each subdivision as weights, choose prices once again for either seller. Once again, note down the profit, and update the average profit in each subdivision for either seller.
6. Repeat steps 4 and 5 for 1500 games.
7. Use the results of games 501 through 1500 to ascertain the fit with the ideal strategy.

Fig. 1. Algorithm for sellers on B2B exchange.

go by. This process is referred to as the *exploration* component of reinforcement learning in Ref. [19]. The sellers are learning the landscape of the problem space during this component. The exploitation component of the reinforcement learning algorithm then comes into play (Step 5), and each subsequent win with a price within a subdivision ensuring higher probability to that subdivision being picked up in the next game. This is essentially the Law of Effect in action—choices that have led to good outcomes in the past are more likely to be repeated in the future [20].

If we denote the average profit of seller j (j=1, 2) in division k (k=1,…10) in simulation round t as $\Pi_{jk}(t)$, then the probability $p_{jk}(t+1)$ of choosing that division in round t+1 is given by

$$p_{jk}(t+1) = \frac{\prod_{jk}(t)}{\sum_{k=1}^{10} \prod_{jk}(t)} \quad (7)$$

The game is then repeated a sufficient number of times so that the sellers can hopefully learn sufficiently to converge to the ideal distribution. The experiment can be repeated with other values of Q, K, c and r.

The algorithm described above finds support in the work by Ref. [14] for the proof of the existence theorem. Ref. [14] uses a scenario in which sellers adjust their strategies to give greater weight to those pure strategies that are currently best against the strategies of the remaining sellers [12].

5. Hypothesis testing, results and discussion

5.1. The two-seller simulations

For testing the assertion, we selected various values of Q, K, c and r. In the two-seller model

Table 1
Various values of capacity (K) and costs (c) (with Q and r fixed)

Q=100 units, r=$80	
K	C
65	20
65	40
65	60
80	20
80	40
80	60
51	20
51	40
51	60

Table 2
Simulation run results with K=65 units, c=$20

Bin	Frequency (O_i)	Th. Cum. Fr.	Th. Freq. dist. (E_i)	Chi-sq.	Chi-sq. value (p=0.05)
55.07692308	185	171.0526316	171.0526316	1.137247	
57.84615385	156	317.0731707	146.0205392	0.682025	
60.61538462	140	443.1818182	126.1086475	1.530186	
63.38461538	126	553.1914894	110.0096712	2.324256	
66.15384615	93	650	96.80851064	0.149829	
68.92307692	76	735.8490566	85.8490566	1.129936	
71.69230769	67	812.5	76.6509434	1.215128	
74.46153846	59	881.3559322	68.8559322	1.410763	
77.23076923	52	943.5483871	62.19245489	1.670398	
80	46	1000	56.4516129	1.935041	
	1000		1000	13.18481	16.92

equilibrium, we can see from the expressions of p_1 and $F(p)$ that the drivers of interest are the values $(2K-Q)$ and $(r-c)$. If the combined capacity $(2K)$ is barely more than the demand (Q), the sellers have little incentive to lower prices, while if there is a large amount of overcapacity, the sellers would greatly reduce prices, since losing would mean catering to a very small residual demand. The difference $(r-c)$ on the other hand would determine the range of the support of prices. We keep Q (the total quantity demanded) fixed at 100 units and r (the reservation price) fixed at $80. The values of K chosen reflect the amount of overcapacity: at K=65, we have moderate overcapacity, prompting moderate competition; at K=80, the possibility of supplying only a small fraction of the demand (i.e. if the seller bids the higher price, he ends up supplying the residual of only 20 units) should prompt more severe competition; and finally, for K=51, the competition would be very limited, since the seller knows that even by bidding a higher price, he will end up supplying 49 units out of his total capacity of 51 units. For each of these values of K, we choose three values of c, the marginal cost, $20, $40 and $60. Thus, there are a total of nine simulations that are run for the purposes of this experiment. The various variable combinations are summarized in Table 1.

The results are shown in Tables 2–10. For each of the pairs of values of K and c in Table 1, we calculate p_1 and corresponding subdivision limits, which are shown in the Bin column. After running the simulation as described above, we find out the number of times the prices are picked in each subdivision, and this is given in the Frequency (O_i) column. The

Table 3
Simulation run results with K=65 units, c=$40

Bin	Frequency (O_i)	Th. Cum. Fr.	Th. Freq. dist. (E_i)	Chi-sq.	Chi-sq. value (p=0.05)
63.386	190	171.1707859	171	2.071260584	
65.232	157	317.1633904	146	0.829923933	
67.078	139	443.2503631	126	1.32247027	
68.924	111	553.2429816	110	0.009226233	
70.77	102	650.0379157	97	0.279898025	
72.616	77	735.8760527	86	0.909999532	
74.462	72	812.5181359	77	0.281163248	
76.308	54	881.3668246	69	3.202436519	
78.154	51	943.553319	62	2.01229638	
80	47	1000	56	1.580957113	
	1000		1000	12.49963184	16.92

Table 4
Simulation run results with K=65 units, c=$60

Bin	Frequency (O_i)	Th. Cum. Fr.	Th. Freq. dist. (E_i)	Chi-sq.	Chi-sq. value (p=0.05)
71.69230769	176	171.0526316	171	0.143093117	
72.61538462	152	317.0731707	146	0.244855636	
73.53846154	133	443.1818182	126	0.376585912	
74.46153846	109	553.1914894	110	0.009266784	
75.38461538	86	650	97	1.206752397	
76.30769231	78	735.8490566	86	0.717628032	
77.23076923	75	812.5	77	0.035558781	
78.15384615	67	881.3559322	69	0.050024511	
79.07692308	68	943.5483871	62	0.052534014	
80	56	1000	56	0.351041475	
	1000		1000	3.187340659	16.92

theoretical cumulative distribution gives us the theoretically expected number of observations in each subdivision, and this is presented in the E_i column. We compute the χ^2 statistic as $\sum_{i-1}^{10} \frac{(O_i-E_i)^2}{E_i}$, and this is compared with the corresponding chi-square value with p=0.05 (16.92). Formally stated, we would reject the null hypothesis that the data follows the distribution specified in Eq. (2), if the calculated χ^2 exceeded the corresponding χ^2value with a significance level α of 0.05:

H_0: $F_n(p)=F_n^*(p)$ where $F_n^*(p)$ is the experimentally generated distribution.
H_1: $F_n(p)\neq F_n^*(p)$

The simulation results show that as expected from the theoretical results, lower prices are preferred over higher prices. The only slight discrepancy is seen in the case of K=51 units, but that result is easily explained within limits of experimental error. When K=51 units, the winner gets to sell to capacity at 51 units, while the loser gets to sell 49 units—which is almost as good as being the winner. In fact, thanks to the higher price, the difference between the profits between the winner and the loser is very small. However, note that by design, the winner of the auction always makes slightly more money than the loser. Thus, the frequency distribution of the theoretical distribution shows that the frequency by which the prices in the lowest subdivision gets selected is almost the same as the frequency by which prices in the highest subdivision gets selected. The lack of the uniformly falling frequencies in the experimental results can therefore be attributed to the randomness in the process.

Table 5
Simulation run results with K=80 units, c=$20

Bin	Frequency (O_i)	Th. Cum. Fr.	Th. Freq. dist. (E_i)	Chi-sq.	Chi-sq. value (p=0.05)
39.5	310	307.6923077	308	0.017307692	
44	201	500	192	0.392892308	
48.5	140	631.5789474	132	0.538947368	
53	92	727.2727273	96	0.142579904	
57.5	77	800	73	0.251022727	
62	56	857.1428571	57	0.022857143	
66.5	41	903.2258065	46	0.560649309	
71	35	941.1764706	38	0.229414137	
75.5	28	972.972973	32	0.453302385	
80	20	1000	27	1.827027027	
	1000		1000	4.436	16.92

Table 6
Simulation run results with K=80 units, c=$40

Bin	Frequency (O_i)	Th. Cum. Fr.	Th. Freq. dist. (E_i)	Chi-sq.	Chi-sq. value (p=0.05)
53	312	307.6923077	308	0.060307692	
56	190	500	192	0.027692308	
59	144	631.5789474	132	1.172547368	
62	108	727.2727273	96	1.582579904	
65	78	800	73	0.382272727	
68	49	857.1428571	57	1.160357143	
71	40	903.2258065	46	0.802949309	
74	36	941.1764706	38	0.100264137	
77	25	972.972973	32	1.452752385	
80	18	1000	27	3.015027027	
	1000		1000	9.75675	16.92

We also observe the effect of relative overcapacity in the results. If we compare the results of when K is 65 units with those of when K is 80 units (for example, the results in Tables 2 and 5, where the c is same at $20), the observed frequencies in the lowest three bins is much higher when K is 80 units, as compared to when K is 65 units. Correspondingly, the observed frequencies in the other bins are lower when K is 80 units, as compared to when K is 65 units. In other words, when there is relatively more overcapacity, sellers have the risk of losing more by bidding higher, and therefore end up choosing lower prices with higher frequencies. The situation is reversed when we compare the results between K=65 and K=51 (Tables 2–4 and Tables 8–10). For K=51, the sellers end up choosing the lower bins with lower frequencies as compared to when K=65. In fact, if we look at the results for K=51 (Tables 8–10), we see that the sellers hardly discriminate between the lower and higher prices. The sellers know that even if they end up supplying the residual, they still sell most of their capacity. Further, the impact is reduced since he extracts higher per-unit profits as compared to the winning bidder.

We run a chi-square goodness of fit test with each of the simulation settings. The 'Chi-sq.' column in the tables compute the χ^2 statistic, whose sum is shown in the final row, and this value is compared to the corresponding chi-square value with p=0.05 which is shown in the final column. As the results show, the fit with the theoretical distribution is always very good. In all the nine cases, we do not reject the null hypothesis that the experimental frequency distribution follows the theoretical probability distribution. In other words, the simulation run results in the agents

Table 7
Simulation run results with K=80 units, c=$60

Bin	Frequency (O_i)	Th. Cum. Fr.	Th. Freq. dist. (E_i)	Chi-sq.	Chi-sq. value (p=0.05)
66.5	284	307.6923077	308	1.824307692	
68	197	500	192	0.114492308	
69.5	140	631.5789474	132	0.538947368	
71	99	727.2727273	96	0.114229904	
72.5	79	800	73	0.541022727	
74	61	857.1428571	57	0.260357143	
75.5	48	903.2258065	46	0.079749309	
77	36	941.1764706	38	0.100264137	
78.5	35	972.972973	32	0.322752385	
80	21	1000	27	1.344027027	
	1000		1000	5.24015	16.92

Table 8
Simulation run results with K=51 units, c=$20

Bin	Frequency (O_i)	Th. Cum. Fr.	Th. Freq. dist. (E_i)	Chi-sq.	Chi-sq. value (p=0.05)
77.88235294	103	103.6585366	104	0.004183644	
78.11764706	98	206.4777328	103	0.225878561	
78.35294118	101	308.4677419	102	0.009609942	
78.58823529	102	409.6385542	101	0.006795955	
78.82352941	100	510	100	0.001301726	
79.05882353	97	609.561753	100	0.065914653	
79.29411765	102	708.3333333	99	0.105523202	
79.52941176	101	806.3241107	98	0.092410952	
79.76470588	99	903.5433071	97	0.032619704	
80	97	1000	96	0.00306026	
	1000		1000	0.547298599	16.92

learning over time to come very close to the ideal distribution with every set of values of the parameters K and c.

It is therefore observed that the artificial software agents start off selecting their prices uniformly throughout the interval of (p_1, r), but gradually learn over time to select lower prices with monotonically higher probabilities except in the case of K=51. In fact, the final frequency distributions show that the learning is 'perfect' within margins of statistical error. 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 him as well as his 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 towards 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 sellers 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.

5.2. The n-seller simulations

The success of the algorithm in the two-seller game is unfortunately not replicated for games with n sellers. We conducted similar simulations by using the same sets of values for Q, K, c and r for n=3, 5, 8

Table 9
Simulation run results with K=51 units, c=$40

Bin	Frequency (O_i)	Th. Cum. Fr.	Th. Freq. dist. (E_i)	Chi-sq.	Chi-sq. value (p=0.05)
78.58823529	98	103.6585366	104	0.308889527	
78.74509804	102	206.4777328	103	0.00652682	
78.90196078	98	308.4677419	102	0.156095416	
79.05882353	101	409.6385542	101	0.000288392	
79.21568627	99	510	100	0.018468592	
79.37254902	102	609.561753	100	0.059712172	
79.52941176	99	708.3333333	99	0.000528244	
79.68627451	101	806.3241107	98	0.092410952	
79.84313725	102	903.5433071	97	0.235098455	
80	98	1000	96	0.024692913	
	1000		1000	0.902711485	16.92

Table 10
Simulation run results with K=51 units, c=$60

Bin	Frequency (O_i)	Th. Cum. Fr.	Th. Freq. dist. (E_i)	Chi-sq.	Chi-sq. value (p=0.05)
79.29411765	102	103.6585366	104	0.026536585	
79.37254902	101	206.4777328	103	0.032187325	
79.45098039	99	308.4677419	102	0.087657161	
79.52941176	101	409.6385542	101	0.000288392	
79.60784314	100	510	100	0.001301726	
79.68627451	98	609.561753	100	0.024498086	
79.76470588	99	708.3333333	99	0.000528244	
79.84313725	101	806.3241107	98	0.092410952	
79.92156863	100	903.5433071	97	0.079540552	
80	99	1000	96	0.06706026	
	1000		1000	0.412009284	16.92

and 10, respectively, and in each of the cases the use of the naïve RL algorithm led us to reject the null hypothesis that the experimental data followed the distribution in Eq. (4). In all these simulations, we kept K=20 units, c=$20 and r=$80, while Q=50, 95, 150 and 190 for n=3, 5, 8 and 10, respectively. In all these simulations, there was a general trend to form a unimodal distribution somewhere near the midpoint of the range.

These results are not entirely unexpected given the simplicity of the underlying RL algorithm. As Ref. [6] conjecture with their results in which some simulations lead to quick convergence while others do not, these may primarily have to do with the sometimes complex strategy space in which relatively simple kinds of learning might be going on. In other words, while our algorithm was good enough to capture the interactions between two sellers, it is inadequate to capture the complexity of interactions between more than two sellers. It is possible that the solution space in multi-agent settings has multiple equilibria and the simple RL algorithm is converging to an undesirable (or suboptimal) equilibrium.[5] In a more general multi-agent setting, an agent has to not only learn what effects its actions have, but also learn to align its actions with those of the other agents. For example, consider what happens in a five-seller game, when a certain price yields a relatively "high" payoff for a seller (we will call him Seller 1). The four other sellers react to that result, and as a result, there is a multitude of interactions (each individual reaction starts a chain of reactions from the remaining four sellers and so on, as opposed to a two-seller model, where there is a reaction from only one seller) which makes it difficult for the RL algorithm of Seller 1 to "pin down" the exact cause of the increased payoff. It must be noted that our RL algorithm utilizes just a single parameter, which can be thought of the strength of the initial propensities [6] that influences the rate of change of choice probabilities. Ref. [6] also points to the fact that in many RL scenarios, sellers who start away from equilibrium can end up learning "very different things". For games with more sellers, the complexity of the interactions increases the odds of such results.

Recent research suggests that the problem of finding an equilibrium with multiple agents interacting is yet to be tackled effectively. As Ref. [9] points out, "the issue of what happens when multiple interacting agents simultaneously adapt, using RL or other approaches, is largely an open question" (p. 1). It needs to be noted that while the knowledge of game theory can enable us to analyze the nature of a mixed-strategy equilibrium, the basic underlying assumption in such analysis is that of common knowledge, i.e. the players not only know that all players are rational, they also know that all the players know that all the players are rational, and so on, ad infinitum. This means that every player knows how each of the other players would behave in every possible contingency. In case of continuous strategy spaces, this would essentially mean that the agents have to keep track of unlimited possible contingent behavior of the other agents. To look at the problem in another way, with

[5] We are grateful to an anonymous reviewer in pointing this out.

two agents, there is only one interaction between them. The number of interactions between three players increases to three, and in general, between n players, the number of interactions is $\binom{n}{2}$. In comparison, our agents indulge in very simple behavior, with limited look-ahead capability. While it turned out to be enough in capturing the interaction between two players, it is perhaps not so unexpected that the simple algorithms failed to capture the interactions between more than two players.

Some remedies can be considered to address the shortcomings of our current algorithm. Advanced RL algorithms make regular use of multiple parameters like experimentation and recency [6]. Complex RL algorithms with greater look-ahead capability can decide on subsequent courses of action by analyzing the payoffs of other agents in greater detail. Ref. [18] notes that the choice of initial propensities can have a long-term effect on the learning process. In real-life organizations, managers can conceivably do much better than choosing equal initial propensities by virtue of their years of experience. In fact, the initial bin probabilities can be established from past auctions for similar items.[6] We conclude by noting that currently our reinforcement is based on the updated average profit in each interval. The choice of this particular form of reinforcement was dictated primarily by what we thought would be a "common sense" approach by organizations to tackle such transactions. Faster convergence might result if we choose other reinforcement mechanisms.

While we are currently addressing many of these issues in our ongoing research, the results of our simulations with the two-seller model show considerable promise. The distribution of Eq. (2) is certainly not intuitive, and is vastly different from the uniform distribution that the agents start off with. Still, using some common-sense rules of thumb, the agents finally come to mimic the ideal distribution. We hope that these results spur the interest of using automated agents that will enable organizations to effectively compete the increasing number of electronic transactions. While one of the main attractions of B2B exchanges remains in their ability to automate the processes by which organizations can participate in electronic transactions with each other, the problem of overseeing each and every one of them is still very much an issue. This problem will likely exacerbate in future as more and more organizations start to utilize these electronic services. While the algorithms that need to be used in real-world scenarios will be much more complex than those presented in this research, we think that organizations might over time develop such algorithms of increasing sophistication. At first, very basic transactions having routine processes would be entrusted to such learning mechanisms. As algorithms get more complex, and simultaneously organizations also gain confidence in such mechanisms, more complex transactions would probably be entrusted. Organizations might also develop processes by which unusual procedures set off triggers for either human intervention or even a complete abort.

One issue of interest to researchers and practitioners alike will be the cost of learning involved. While the agents did learn the ideal behavior over repeated simulations, there might be a significant cost to the organization as their behavior starts as being completely random, and therefore differs significantly from the ideal in the initial stages. This is a luxury that organizations might not have in real life—in fact, if the costs are high enough, there might not be any incentive for an organization to utilize such automated agents. In such situations, the importance of having experienced managers will be realized, who can "guide" these artificial agents to much better initial "starting points" that would be closer to the optimal solution, thus reducing the cost to the organization.

A related issue is the rate of convergence of the algorithm towards the theoretical equilibrium. For example, the number of price bands should affect convergence. Quite possibly, the time taken towards convergence would increase linearly with the number of price bands, and exponentially with the number of agents. In such an environment, one interesting idea that we wish to explore in future research is to have an "adaptive" number of bands, whereby we start off initially with a few price bands, and progressively increase their number to refine the search in the later stages.[7]

[6] We thank one anonymous reviewer for the suggestion.

[7] We are grateful to an anonymous reviewer for this idea.

6. Conclusions

Our research shows initial promise in the use of artificial agents to automate transactions in electronic marketplaces. We successfully replicate the theoretical results of mixed-strategy equilibrium in capacity-constrained reverse auctions involving two similar competitors through the use of artificial agents that learn their ideal behavior over time by keeping track of their payoffs. Reinforcement learning was successfully employed as the learning mechanism in this simulation. The encouraging results show promise in the use of artificial learning mechanisms that will enable organizations to take part in the increasing number of transactions in electronic marketplaces. Electronic marketplaces can be enhanced and even specifically designed to accommodate artificial agents working on behalf of managers. Additionally, artificial agents could certainly be used to assist managers in their decision making in such scenarios.

In our future research, we intend to apply RL algorithms of increasing complexity that will hopefully learn the idealized seller behavior in an n-seller model. We also wish to consider more complex models of competition (e.g. different marginal costs and capacities of sellers, increasing the number of buyers and sellers, etc.). 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.

Acknowledgements

The authors would like to thank Professor Alok Chaturvedi of the Krannert Graduate School of Management, Purdue University for the generous use of his Synthetic Environment for Analysis and Simulation (SEAS) laboratory for the simulation experiments.

References

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

[2] S. Bandyopadhyay, J.M. Barron, A Generalized Model of Competition among Sellers in a B2B exchange, presented at the INFORMS 2001, Miami (November 4–7, 2001).

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

[4] J.M. Epstein, R. Axtell, Growing Artificial Societies: Social Science from the Bottom Up, Brookings Institution Press, Washington, DC, 1996.

[5] I. Erev, A. Rapoport, Coordination, 'Magic,' and reinforcement learning in a market entry game, Games and Economic Behavior 23 (1998) 146–175.

[6] I. Erev, A.E. Roth, Predicting how people play games: reinforcement learning in experimental games with unique, mixed-strategy equilibria, The American Economic Review 88 (4) (1998) 848–881.

[7] S. Helper, J.P. MacDuffie, B2B and modes of exchange: evolutionary and transformative effects, in: B. Kogut (Ed.), The Global Internet Economy, The MIT Press, Cambridge, MA, 2003.

[8] How I saved $100 Million on the Web, Fast Company 43 (2001 February).

[9] J.O. Kephart, G.J. Tesauro, Pseudo-convergent Q-learning by competitive pricebots, Proceedings of Seventeenth International Conference on Machine Learning (ICMA-00), Stanford University, Stanford, CA, 2000 (June 29–July 2).

[10] R. Kerrigan, E.V. Roegner, D.D. Swinford, C.C. Zawada, B2Basics, Mckinsey Quarterly 1 (2001) 45–53.

[11] D. Kreps, J. Scheinkman, Quantity precommitment and Bertrand competition yield Cournot outcomes, The Rand Journal of Economics 14 (2) (1983) 326–337.

[12] H.W. Kuhn, S. Nasar, Editor's introduction to Chapters 5, 6 and 7, The Essential John Nash, Princeton University Press, Princeton, NJ, 2001.

[13] R. Levitan, M. Shubik, Price duopoly and capacity constraints, International Economic Review 13 (1) (1972) 111–122.

[14] J. Nash, Non-cooperative games, Annals of Mathematics 54 (1951) 286–295.

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

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

[17] Renaissance in cyberspace, The Wall Street Journal (2003 November 20).

[18] A.E. Roth, I. Erev, Learning in extensive-form games: experimental data and simple dynamic models in the intermediate term, Games and Economic Behavior, Special Issue: Nobel Symposium 8 (1) (1995) 164–212.

[19] R.S. Sutton, A.G. Barto, Reinforcement Learning: An Introduction, The MIT Press, Cambridge, MA, 1998.

[20] E.L. Thorndike, Animal intelligence: an experimental study of the associative processes in animals, Psychological Monographs 2 (8) (1898).

[21] R. Wise, D. Morrison, Beyond the exchange: the future of B2B, Harvard Business Review 78 (6) (2000) 86–96.

Subhajyoti Bandyopadhyay is currently an Assistant Professor in the Department of Decision and Information Sciences in the University of Florida, Gainesville. He earned his Ph.D. in Management Information Systems from Prude University in 2002. His areas of specialization are in Economics of Information Systems, Electronic Commerce Strategies, and security aspects of Electronic Commerce. His work has been published in Communications of the ACM, Journal of Organizational Computing and Electronic Commerce, Lecture Notes in Computer Science and the Database for Advances in Information Systems.

Jackie Rees is an Assistant 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 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 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, but typically have in common elements of the economics of information and uncertainty.