December 2005

Enhanced Reputation Scoring for Online Auctions

Bonnie Montano
Georgetown University

Dawn Porter
Georgetown University

Ross Malaga
Montclair State University

J. Ord
Georgetown University

ENHANCED REPUTATION SCORING FOR ONLINE AUCTIONS

Bonnie R. Montano
McDonough School of Business
Georgetown University
Washington, DC U.S.A.
brm7@georgetown.edu

Dawn C. Porter
McDonough School of Business
Georgetown University
Washington, DC U.S.A.
dcp22@georgetown.edu

Ross A. Malaga
School of Business
Montclair State University
Montclair, NJ U.S.A.
montclair@ramconsult.com

J. Keith Ord
McDonough School of Business
Georgetown University
Washington, DC U.S.A.
ordk@georgetown.edu

Abstract

To handle the uncertainty inherent in eCommerce transactions, reputation systems have emerged as a way to represent reliability and develop trust between transaction participants. Despite the value added by reputation systems, limitations of existent systems remain (Malaga 2001). We empirically test Porter et al.'s (2004) reputation scoring procedure, which was designed to address the shortcomings of current systems. This study uses computer simulation to replicate the auction process between buyers and sellers with reputation scores calculated using both the Porter et al. model and the eBay auction model. Results from the two models are then analyzed to show that the Porter et al. model more accurately estimates reputation scores.

Keywords: Reputation score, online auction

Introduction

Auctions are marketplaces where information is exchanged for determining a price for delivery of goods or services. With the United States Treasury alone auctioning billions of dollars of bills and notes weekly, the sheer volume of goods and services exchanged by auctions suggests they should be studied (McAfee and McMillan 1987). Even before the Internet there were electronic auctions. Since then, auctions have become a potent phenomenon online with an estimated annual market in the range of tens of billions of dollars. For example, with 88 percent of market share, eBay's valuation exceeds $18 billion (Fischer 2004).

In traditional auctions, participants may see the product, but quality is not necessarily guaranteed as it might be in a retail store. For goods, however, buyers can often see them in person. Sometimes there are only pictures. Online auctions like the ones conducted on Yahoo! and eBay always only have pictures. In either case, auction theory addresses the issue of uncertainty through a concept termed *commitment*. Commitment typically refers to how the seller promises to sell his goods or services, whereby once a seller sets out rules for how its goods or services will be sold, the seller cannot renege. Sometimes the only means for assessing commitment is through a seller's reputation (see McAfee and McMillan 1987). Formally stated, reputation is the "overall quality or character as seen or judged by people in general" (Merriam-Webster 2005). In online auctions, reputation is increasingly used as a proxy for gauging the quality of products or services for sale. Is the product or service accurately described? Will the product or service be provided as promised? What if there is a problem with the product or service?

Given the uncertain nature of online auctions, myriad online reputation systems have been developed to serve as a benchmark for seller reliability and for promoting trust between buyers and sellers. Ba (2003) note the importance of answering the question *How can mechanisms be designed to promote trust in online auctions*? Trust is a key factor in the success of online transactions, including electronic auctions (Brynjolfsson and Smith 2000; Resnick et al. 2000). Despite the critical role of trust in online auctions, problems with reputation systems designed to develop trust persist. Ba and Pavlou (2002) question the reliability of existent feedback mechanisms. Additionally, Ba (2003, p. 276) notes that "the 'reputation' developed through the feedback system may not be accurate and truthful."

The primary purpose of this paper is to test a reputation scoring procedure developed by Porter et al. (2004) that promises to address this issue of accuracy and truthfulness. Reputation scoring is a method that uses feedback to assess the trustworthiness of participants in a transaction. According to the theories of reasoned actions and planned behavior, an individual's intentions can predict actual behavior, but there is a disconnect between the two (Ajzen 1991). Ideally, a reputation score will minimize the difference between an individual's intentions and their actual behavior (deserved reputation) so that the trustworthiness of a seller is accurately represented by the reputation score. Simulation is used to test the Porter et al. model through replication of the auction process between buyers and sellers. Because eBay is the most widely known online auction site, this paper compares the Porter et al. reputation scoring model to the eBay reputation scoring model.

The remainder of the paper is organized as follows. The next section of the paper introduces the theoretical underpinnings of this work. Then the reputation scoring procedures being examined are described and hypotheses are presented. The experimental design is given; the results are presented, and the paper is concluded.

Background and Motivation

Trust

A large body of research has shown that trust plays an important role in a consumer's decision to purchase a product in an electronic market (Jarvenpaa et al. 2000; Pavlou and Gefen 2004). One definition of trust in electronic markets, proposed by Lim et al. (2001), is "the willingness of a consumer to expose himself/herself to the possibility of loss during an Internet shopping transaction, based on the expectation that the merchant will engage in generally acceptable practices, and will be able to deliver the promised products or services."

Many types of trust have been discussed in the literature such as control trust, which refers to embedded protocols and procedures for reducing opportunistic behavior (Tan and Thoen 1998), personality-based trust, which refers to trust developed during childhood in response to experience with care givers (Bowlby 1982), cognition-based trust, which refers to first impressions as the basis of trust instead of personal interactions (Meyerson et al. 1996), and institution-based trust, which refers to institutional structures supporting the development of trust (McKnight et al. 2002; Weiss et al. 1999). Each type of trust is relevant in a different situation.

Institution-based trust is of particular relevance in this paper because it involves trust between parties who do not know each other personally, as is the case in online auctions (Pavlou and Gefen 2004). It reflects the belief a party has about the security of a particular situation because of guarantees (i.e., structural assurances) or safety nets that are in place (Zucker 1985). The importance of institution-based mechanisms is highlighted by the fact that both eBay and Yahoo! Auctions use various institution-based mechanisms including escrow services, dispute resolution policies and procedures, and reputation systems for encouraging online transactions (Pavlou and Gefen 2004). This paper focused on reputation systems, but for such systems to effectively develop trust, they must be reliable and believable (Weiss et al. 1999). Pavlou and Gefen (2004) highlight the importance of subjective *perceptions* of institutional mechanisms for increasing trust in online marketplaces, regardless of whether those mechanisms are legal or market driven. Thus, it is important for buyers to perceive reputation as reliable and believable.

Reputation

Researchers interested in reputation have reached a consensus regarding its operational definition. Wilson (1985 pp. 27-28) states,

> In common usage, reputation is a characteristic or attribute ascribed to one person, industry, etc. by another (e.g. "A has a reputation for courtesy"). Operationally, this is usually represented as a prediction about likely future

> behavior (e.g., "A is likely to be courteous"). It is, however, primarily an empirical statement (e.g., "A has been observed in the past to be courteous"). Its predictive power depends on the supposition that past behavior is indicative of future behavior.

Weiss et al. (1999) refer to the idea of *perceived* reputation. Just as there is a distinction between perceived and actual behaviors (Ajzen 1991; Ajzen and Fishbein 1980; Venkatesh and Davis 2000), reference to a perceived reputation suggests there is a distinction between perceived and actually deserved reputation. One's deserved reputation is that which he/she has earned based on actual behavior instead of intention. We use Ajzen's (1991) theories of reasoned actions and planned behavior to provide a theoretical framework for understanding how to best predict an individual's behavior and thus his/her truly deserved reputation. The theory of reasoned actions, and the later theory of planned Behavior, suggests that the best predictor of an individual's actual behavior is his/her intention to perform the behavior. This intention is, in turn, a function of one's attitude toward the behavior. Intention is the cognitive representation of a person's readiness to perform a given behavior, and it is considered to be the immediate antecedent of behavior (Ajzen 1991). Additionally, individuals' perception of their control, or lack thereof, over their behavior also drives actual behavior (Madden et al. 1992).

Reputation Scoring Models

In many contexts such as offline auctions and stores, buyers have multiple clues that help them determine their level of trust in a seller. For example, if Sotheby's or Christie's holds an auction, the institutional trust of these companies transfers to some extent to the product sold because buyers trust Sotheby's or Christie's to ensure the auction is conducted and concluded as promised. However, in online auctions, there is no such context. In most online auction situations, reputation is the only factor influencing the development of trust. Therefore, it is important that the stated reputation accurately reflect the past actions of the seller so that fair levels of trust can be established by buyers.

Because online auctions are typically characterized by one-time transactions where participants possess neither a shared history nor the promise of future transactions, mechanisms for developing reputation scores are very important in online auction settings. The theoretical basis for reputation scoring models is game theory (Resnick et al. 2000). The traditional prisoner's dilemma indicates that in a consumer-to-consumer online auction, in which no reputation mechanism is present, the participants always have an incentive to cheat each other (Yamamoto et al. 2004). Essentially, reputation scoring systems are designed to extend the relationship between buyer and seller beyond the one-time transaction or to provide information about previous dealings.

Different online auctions, as well as other e-businesses, use different mathematical representations for reputation. Academic papers that attempt to develop these representations are lacking. However, the commercial sector has been active in researching new reputation scoring methods. For example, Open Ratings of Waltham, MA, holds three patents (6,892,178; 6,892,179; 6,895,385) related to reputation score. In addition, the algorithm used by Google (described in patent filing 6,725,259) to rank search results is based, in part, on reputation. eBay's reputation model, one of the most well-known, is described in the subsequent section followed by a description of the Porter et al. model tested in this paper.

eBay's Online Auction System

eBay does not participate in transactions between buyer and seller; rather, eBay provides the platform to enable buyers and sellers to connect directly to each other. As such, eBay does not verify the authenticity of products or guarantee that payments or auction items will be received as promised. Instead, eBay offers services such as authentication, insurance, and escrow through third parties.

In addition to third party services, eBay supports the trust building necessary for the conduct of online transactions (Pavlou and Gefen 2004) through the use of two feedback scales: feedback score and [percent] positive feedback. The eBay model "was founded on the belief that most people are trustworthy and committed to honorable dealings with each other.…We are encouraging our community to think that basically 99 percent of the people out there are doing the right thing " (Bunnell 2000, p. 55).

In the eBay procedure, any completed transaction may be rated through a feedback system by the winning bidder of an item. The feedback scores are +1, representing a positive experience where the seller delivered an item of promised quality in a timely fashion; 0, representing neutral feedback; and -1, meaning the purchasing experience was negative for some reason. These ratings are then used to calculate an overall reputation score (feedback score) as well as a percent positive feedback rating. The overall reputation score is an aggregation of all unique feedback scores and has no upper bound; it has been criticized as being more of an "experience" measure than a measure of true reputation (Kauffman et al. 2000). Further, the feedback score is founded on a belief (that 99 percent of people behave honorably) that is fundamentally flawed. The literature on opportunism indicates dishonorable behavior does, in fact, occur (e.g., Fan et al. 2005; Fukuyama 1995). Reichheld and Schefter (2000) further assert that undesirable behavior is of even greater concern in online transactions than in traditional ones because consumers cannot judge the trustworthiness of a seller face-to-face.

Since it has been shown that sellers can easily change their identities and use pseudonyms, there is no incentive for a seller with a negative score to remain in the market (Friedman and Resnick 2001). Because both feedback score and (percent) positive feedback are presented to potential buyers, there is no upper bound on the feedback score, and the basic premise of the feedback score is fundamentally flawed, we have chosen to use the eBay *percent positive feedback* rating as our comparative measure. Percent positive is the percentage of positively rated transactions from all transactions rated as either positive or negative, which better reflects the notion of reputation, as defined above, than the total feedback score. The calculation for percent positive for n transactions is

$$\% pos = \frac{N_{+n}}{N_{+n} + N_{-n}} \tag{1}$$

where N_{+n} represents the number of unique positive feedback responses at time n, and N_{-n} represents the number of unique negative feedback responses.[1] This calculation disregards any neutral ratings, which essentially drops all 0 values from consideration and thus provides less information to auction participants. Depending on the actual number of neutral feedback ratings, this deletion typically results in an upward bias in the calculation.

The Porter et al. Model

The Porter et al. model uses a parametric statistical model to create a reputation score using the same buyer feedback values that are currently used by online auction sites. The model was specifically developed to address and correct some of the limitations of existent reputation scoring models identified by Malaga (2001).

As such, the Porter et al. model uses the same +1, 0, -1 ratings as the eBay model, but instead of a percent of positive rating, a reputation score is calculated by averaging all feedback scores (i.e., -1, 1, and 0) rather than dropping out the 0 ratings. The probability that a buyer assigns feedback r_i, where $r_i = +1, 0, -1$, is written as Pr(buyer assigns score r_i) = p_i, $i = 1, 2, 3$, and we denote the corresponding set of probabilities by $\{p\} = \{p_1, p_2, p_3 \,|\, p_i \geq 0, \sum p_i = 1\}$. We use the feedback from n unique raters, denoted by $\{x\} = \{x_1, x_2, x_3 \,|\, \sum x_i = n\}$, where x_i represents the total feedback tallied in the i^{th} category, and describe $\{x\}$ by the multinomial distribution

$$\Pr\left(x_1, x_2, x_3 \,\middle|\, n, \{p_i\}\right) = \binom{n}{x_1 x_2 x_3} \prod_i p_i^{x_i} \tag{2}$$

The prior distribution for $\{p\}$is specified as a standard Dirichlet distribution

$$f\left(\{p\} \,\middle|\, \{\alpha\}\right) \propto \prod_i p_i^{\alpha_i - 1}, \quad \sum \alpha_i = A, \quad \alpha_i > 0 \tag{3}$$

with parameters $\{\alpha\} = (\alpha_1, \ldots, \alpha_K$. Equations (2) and (3) lead to the Dirichlet posterior

[1]The eBay system uses unique scores from members only. That is, a buyer can leave only a positive, neutral, or negative rating for a seller, but each of these ratings affects the other member's feedback score only once. For example, if a member leaves two negatives for a seller, only one of them will be counted. However, if the same member leaves two negatives and one positive, the seller's score will reflect one negative and one positive rating. Subsequent negatives or positives from the same member will not affect the feedback score. For the sake of simplicity, the Porter et al. model will be analyzed using the same structure.

$$f\left(p_1, p_2, p_3 \middle| \{x_i\}\right) = \frac{\Gamma\left(A+n\right)}{\prod_i \Gamma\left(\alpha_i + x_i\right)} \prod_i p_i^{\alpha_i + x_i - 1} \tag{4}$$

If we know the population proportions $\{\pi_i\}$ attributed to the three possible ratings, we could use the simple score function $S_i = \sum r_i \pi_i$, where $\pi_i = \lim E\left(p_i \middle| x_i\right)$ as $n \to \infty$, to represent the true reputation of seller *i*. For the Porter et al. model, in the absence of the true proportion values, the posterior distribution is used and the reputation function after *n* transactions is:

$$M_n = E\left(S_n\right) = \sum r_i \, E\left(p_i \middle| \{x_i\}\right) = \frac{\sum_i r_i \left(\alpha_i + x_i\right)}{\left(A+n\right)} \tag{5}$$

This posterior mean is the expected reputation score, representing our current estimate of a seller's true intention value. This parameter estimate is updated after new information arrives, or as each new feedback rating is recorded. We use this posterior mean to represent the seller's reputation score and compare it to the percent positive rating from eBay.

The ultimate goal of reputation scoring is to accurately represent an individual's character or quality and thus the true level of trust attributable to that particular individual. Reputation scoring is a measure external to the individual being scored as it is initiated by transaction partners. In contrast, the true level of trustworthiness is internal to the individual. Hence, the true level of trustworthiness can be conceptualized as an individual's internal *intention* to behave in a particular manner. Examples include the seller's propensity to do such things as (1) represent an item in its true condition, (2) ship a sold item appropriately and in a timely fashion, or (3) fail to do either of the above. Because of the weaknesses previously noted in the eBay reputation scoring system, we seek to test the Porter et al. model and illustrate that it better reflects a seller's true level of trustworthiness than does the eBay model. Per the theories of reasoned action and planned behavior, we use a measure of internal intention as the basis for comparison, assuming that each seller on the auction site has a fixed, internal intention vector that corresponds to his/her auction behavior. We formally state our first hypothesis as

Hypothesis 1: The Porter et al. model more closely approximates a seller's true intention $\{\pi_i\}$ than the eBay percent positive model.

Continuing with the theme of accurately reflecting an individual's true level of trust, we also posit that time may have an impact on an individual's reputation (Fan et al. 2005; Porter et al. 2004). That is, behavior may change over time. For example, a seller may behave honestly until a solid reputation is developed and then exploit the high reputation score for economic gain through dishonest actions. Conversely, if a seller has become more familiar with the auction system and has subsequently performed better, his/her score should reflect the increased reputability. In either case, changes in behavior should be identified quickly to minimize the negative impact on buyers (in the case of a change to worse behavior) or minimize negative impact on sellers (in the case of a change to better behavior). Thus, recent ratings should play a larger role in determining reputation scores than earlier ratings (Fan et al. 2005). When older ratings are weighted less heavily than newer ratings, this is termed time decay because the relevance of the ratings decrease (i.e., decay) over time.

From Porter et al. (2004), we use a time decay approach to capture this property. Let

$$z_i\left(n\right) = w z_i\left(n-1\right) + d_i\left(n\right) \tag{6}$$

and we define the indicator variable

$$d_i\left(n\right) = \begin{cases} 1 \text{ if } n^{th} \text{ rater gives feedback } r_i \\ 0 \text{ otherwise} \end{cases}$$

with the prior distribution given by $z_i\left(0\right) = \alpha_i$.

The coefficient w denotes the discount factor, where $0 \leq w \leq 1$. There is no discounting when $w = 1$, and $w = 0$ implies that only the most recent rater is being used. More general weighting schemes are possible, but the use of a discount factor has proved both popular and successful in exponential smoothing, so that is the formulation we pursue here.

Using similar arguments to those applied above, after n rated transactions we have

$$M_n = E\left(S_n\right) = \sum r_i E\left(p_i \middle| z_i\left(n\right)\right) = \frac{\sum r_i z_i\left(n\right)}{Z_n} \tag{7}$$

where $Z_n = z_1\left(n\right) + \cdots + z_K\left(n\right)$. This expression may be rewritten in the following way, which allows for easy updating without needing to store previous individual scores

$$M_n = M_{n-1} + \frac{\sum r_i d_i(n) - M_{n-1}}{Z_n} \tag{8}$$

Analogous to the general model seen in (5), we use this posterior mean with weighted transactions as our reputation score and compare it to the eBay model. We formally state hypothesis 2 as

Hypothesis 2: Over a series of n transactions, the Porter et al. model, which incorporates time decay will more accurately reflect a given seller's reputation than eBay's percent positive model.

Experimental Design

Simulation

This study uses simulation to verify the two hypotheses drawn from the models in Porter et al. (2004) by comparing them to eBay's percent positive score. Simulation uses a mathematical model of a real system, an online auction in our case, for conducting computer-based experiments. The experiments enable one to describe, explain, and predict the behavior of the real system (Hoover and Perry 1990). Similar simulation approaches have been used by a number of researchers to test reputation systems (Fan et al. 2005; Josang et al. 2003; Yu and Singh 2002). This approach is well suited for our study because we wish to examine how reputation scores reflect true intention under numerous scenarios.

The Experiment

Based on the aforementioned notion of intention, we posit that the score from a reputation system should accurately reflect a participant's true intention since this is the best predictor of behavior (Ajzen and Fishbein 1980). We operationalize the notion of intention through a *true intention vector*. We then use simulation to evaluate the efficacy of the Porter et al. model versus the eBay percent positive model when compared to this true intention vector. The experiment assumes that the true intention is given for each tested scenario in the simulation. We further the experiment by comparing the improved time decay model to both the general Porter et al. and eBay models.

A sample intention vector for seller i could be: π_i = 0.9, 0.05, 0.05)', corresponding to ratings $\{r_i : r_1 = +1, r_2 = 0, r_3 = -1\}$ as described above where each rater, or buyer, chooses one of the three values and casts one vote per completed transaction. This seller has a high intention (90 percent) of honestly representing an item and shipping it quickly. There is a 5 percent chance, however, that this seller will end up choosing to behave in an un-trustworthy fashion and a 5 percent chance the seller will behave in a way that will merit a neutral rating. We have chosen to focus on this particular intention vector for descriptive purposes. It is skewed in the positive direction yet still retains enough variability so as to highlight the differences between the models. Per the theories of reasoned action and planned behavior, there will be fluctuations in the actualizations of these intentions as displayed through actual behaviors. To model this, we create bidder profiles that contain true intention vector values for all possible bidder types. This was done by varying the probability values for each rating by 0.05 and creating vector values for every possible combination where $\sum_j \pi_j = 1$, resulting in 227 different bidder profiles. These intention vectors are then used in a computer simulation to create feedback vectors.[2] Assuming intention is a perfect predictor of behavior, the feedback vector represents the feedback scores (+1, 0, -1) an individual deserves to receive. We then use the feedback vector to calculate both the Porter et al. reputation score from (5) and the eBay percent positive score from (7).

For each profile (i.e., intention vector), 500 potential feedback vectors are generated and the average reputation score is calculated at every transaction point for both the Porter et al. model's score and that of the eBay percent positive score. Using the intention vector described above, we could see a string of values for $t = 10$ such as (+1, +1, +1, -1, +1, 0, +1, -1, +1, +1). The seller's true intention value would then be $S_i = R_i' \cdot x_i = (+1, 0, -1) \cdot (0.9, 0.05, 0.05)' = 0.85$. For each possible seller profile, 250 transactions were simulated. The simulation was replicated 500 times and the results were averaged.

For the time decay model shown in (7), a slightly different approach was necessary. In order to model this, we continue with the profile in which a seller has a 0.9 probability of behaving favorably within the system, a 0.05 probability of behaving poorly, and a remaining 0.05 probability of behaving neutrally for a particular transaction. In this part of the experiment, however, we built in a downward shift in this profile after 50 transactions have occurred. From transactions 51 through 250, the seller's profile has only a 0.5 probability of behaving favorably, a 0.25 probability of behaving poorly, and a remaining 0.25 probability of behaving neutrally. The time decay simulation tests a downward shifting profile because the converse situation is of less concern due to the nature of typical online auctions. That is, it is relatively simple to create a new user name and identity as a seller on these sites (Friedman and Resnick 2001), so there is little real incentive for sellers with poor scores to remain in the system. This split profile was run a total of 20 times and the summary statistics of means, variances, and sum standard of squared error (SSE) values were averaged.

We specifically focused on the results from the two cases of $w = 0.75$ and $w = 0.9$. From Porter et al., it was noted that these two parameter values provide a good tradeoff between focusing on recent feedback scores and creating a relatively smooth function. Lower values of w allow for large fluctuations, and larger values provide curves that rely a great deal on past data, thereby taking a long time to adjust to changing profiles.

Results

Once the simulation was completed, standard sum of squared error (SSE), and, in some cases, mean squared error (MSE) were used to analyze the results. The SSE measure was calculated at the end of the 250 transactions and used to compare the two models to the true intention value.

Figure 1 shows the results of the eBay percent positive ratings, the reputation score obtained using the Porter et al. model, and the true intention value of 0.85 derived from the sample profile described earlier, where the seller has a 0.9 probability of performing in a favorable fashion. This profile was selected because it represents a reasonable example based on anecdotal evidence and random observation.[3] Figure 1 shows that, although both the percent positive score and the Porter et al. reputation score start at the same point, the eBay value is consistently further from the true intention value than the Porter et al. score. Figure 2 displays only the first 20 transactions, when there is more variability for both the eBay and the Porter et al. model.

Figure 2 illustrates that the Porter et al. model better approximates the true intention than the eBay model even during start-up of the scoring system. In order to quantify the differences between the two models' results, we calculated the SSE value. This measure describes how far each model is from the true intention value by aggregating the squared differences between it and each score at every transaction point. Figure 3 shows the change in the SSE value over all 250 simulated transactions.

As shown in Figure 3, it is evident that the eBay procedure has a smaller, and, therefore, a more favorable, SSE value at small numbers of completed transactions. From about the fiftieth transaction on, however, the Porter et al. model results in consistently lower SSE values. At the far right side of the figure, in fact, the discrepancy is quite significant.

---

[2]For the sake of consistency and simplicity, the Porter et al. model assumes that all transactions are completed by unique buyers.

[3]Observe that the eBay score given in (1) fails to distinguish between the profile being studied and (0.947, 0, 0.053),yet this second profile is clearly preferable to the first.

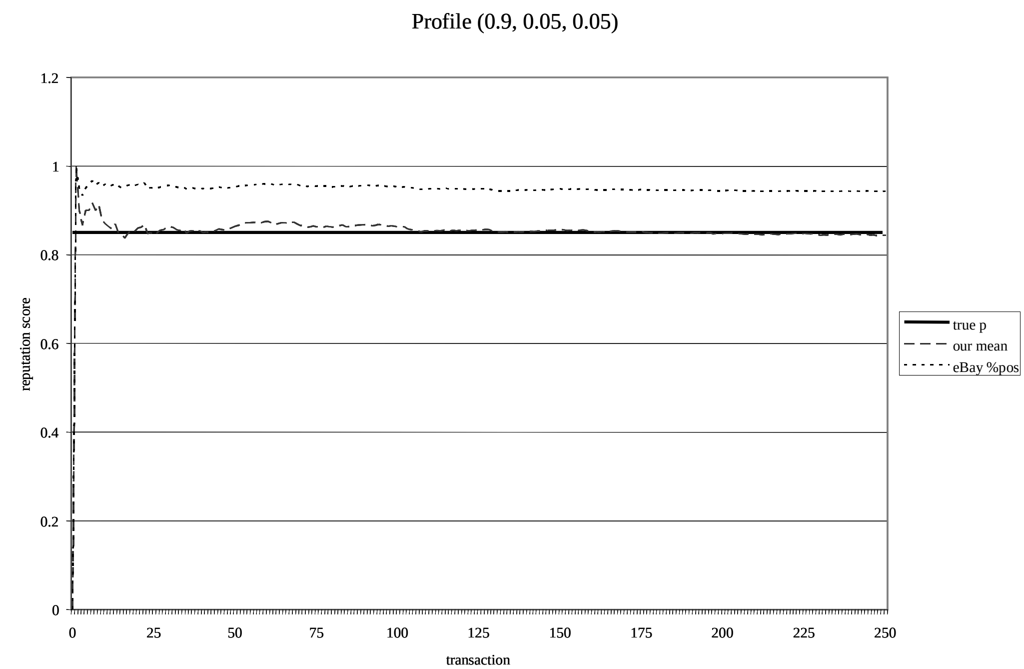

Figure 1. eBay's Percent Positive Ratings and the Porter et al. Scores Compared to True Intention Value

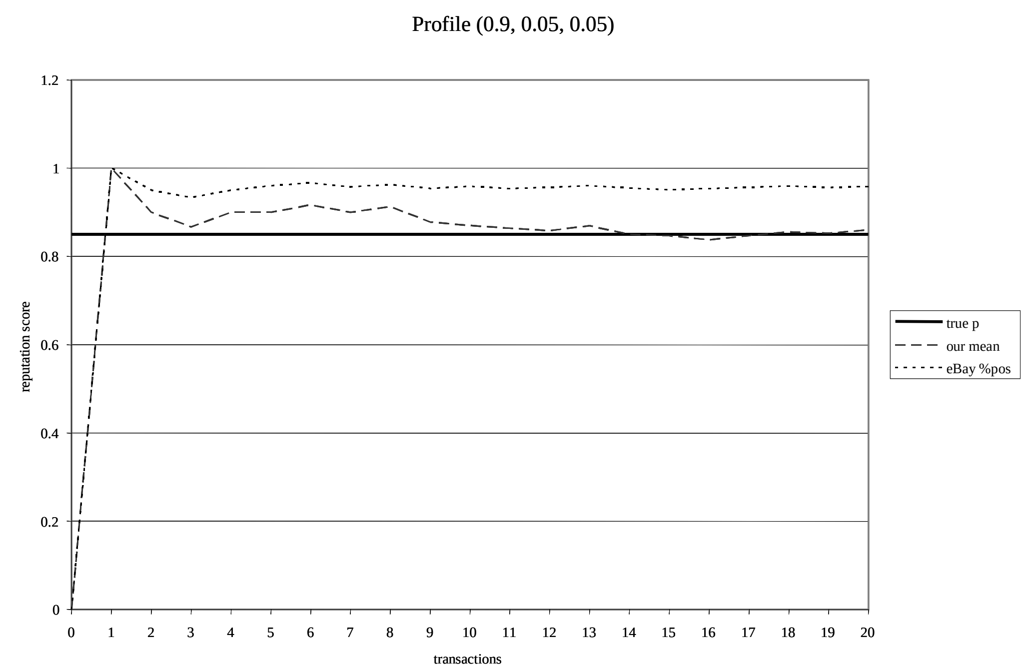

Figure 2. eBay's Percent Positive Ratings and the Porter et al. Scores Compared to True Intention Value, *t* = 20

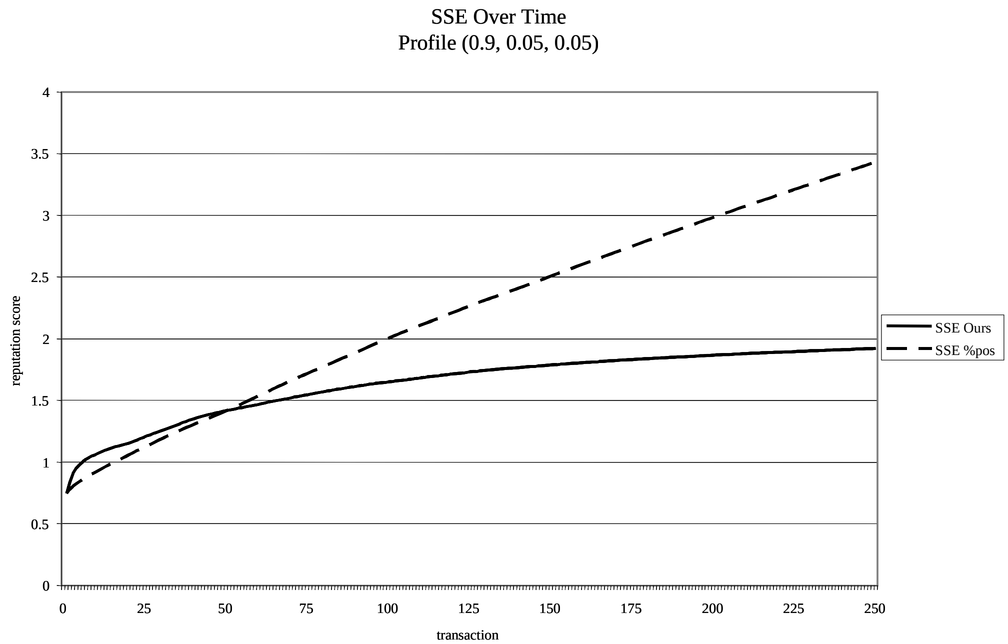

Figure 3. Comparison of SSE for eBay's Percent Positive Ratings and the Porter et al. Model Versus True Intention Values

Table 1. SSE Values for *t* = 250 and 50 Replications

SSE Comparison between Models		
Profile of (+1,0,-1) Feedback	*Porter et al. Score*	*eBay Percent Positive*
(0.50, 0.20, 0.30)	4.43605	43.06294
(0.70, 0.05, 0.25)	5.96090	17.99222
(0.20, 0.20, 0.60)	4.00092	105.5010
(0.25, 0.50, 0.25)	2.64268	63.14386
(0.90, 0.05, 0.05)	1.93063	3.39097

We include Table 1 to demonstrate the difference between the true intention value and each of the two models for a selection of various other profiles, including the one used in the figures above. Even under widely varying profile types, the Porter et al. model has a significantly lower SSE value.

To test the time decay model, we calculated the SSE values and chose to rely on MSE values for comparison. The MSE values display the average squared error at every transaction point. Because of the difference in scale between the SSE and MSE charts, the MSE comparisons show the plot discrepancies more clearly. An SSE chart would still produce the same overall patterns and results.

Using the time decay approach from (7), we should quickly see a shift in a seller's behavior patterns. Figure 4 represents the results for 100 transactions where a downward shift occurred at transaction 50. The figure includes results from our general model from (5), the eBay percent positive values from (1), and two time decay approaches with weights of $w = 0.75$ and $w = 0.9$ from (7).

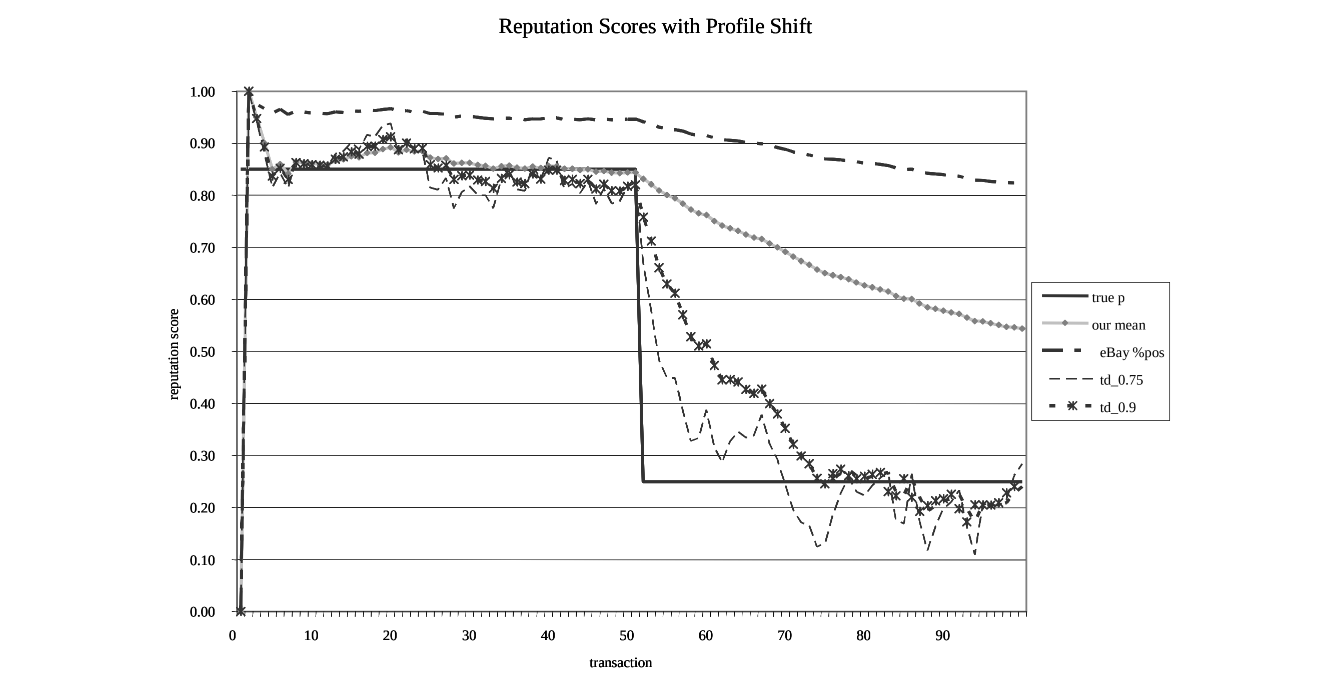

Figure 4. Comparison of the Porter et al. General Model, the eBay Percent Positive Ratings, and the Porter et al. Time Decay Model with *w* = 0.75 and *w* = 0.9

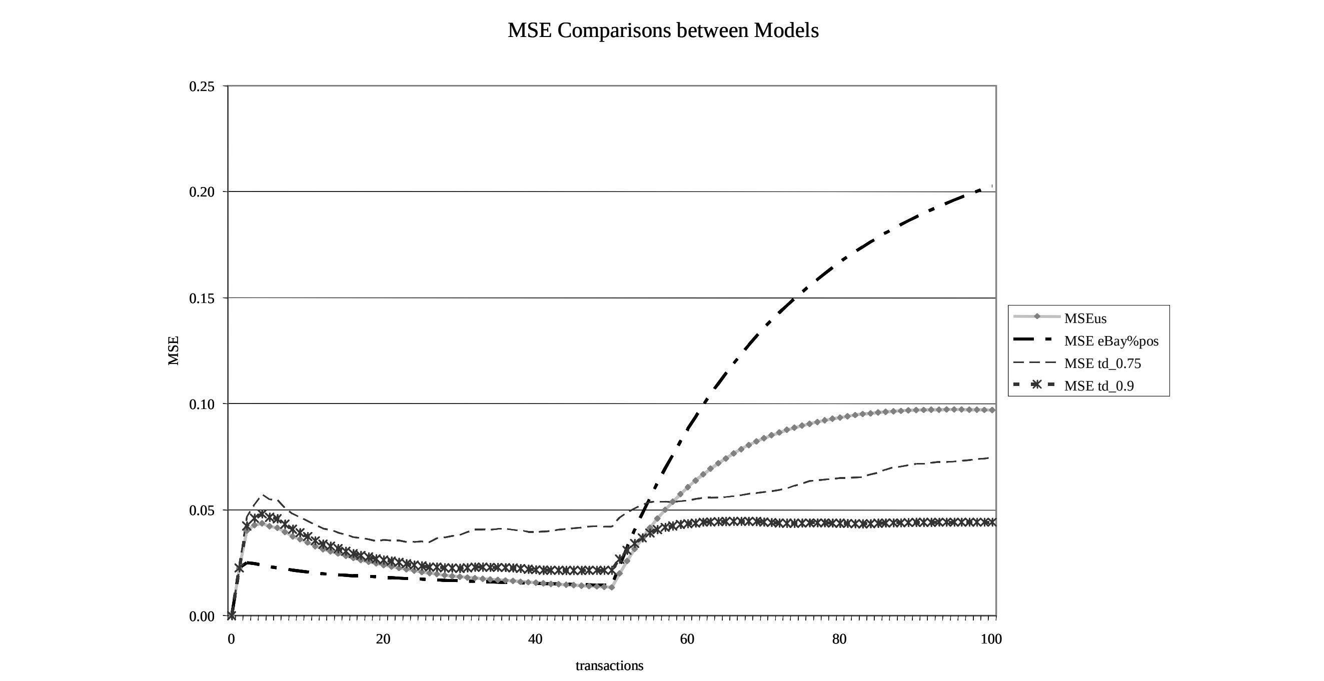

Figure 5. MSE Comparison of the Porter General Model, the eBay Percent Positive Ratings, and the Porter et al. Time Decay Model with *w* = 0.75 and *w* = 0.9

We see that the eBay values are consistently above the true intention value, which is to be expected from Figures 1 and 2. Although the Porter et al. general model picks up the shift and begins to account for it by slowly decreasing the seller's overall reputation score, the time decay plots adjust more quickly to the change. The line representing a weighting parameter of 0.75 has more fluctuation than the 0.9 line. A lower weight will place less emphasis on past transactions more quickly than a plot with a higher value, which tends to decrease more slowly into the past. To further display our results, we include Figure 5, which is a plot of the MSE values for the plots in Figure 4.

As in Figure 3, Figure 5 shows that the eBay model seems to have less average error at the left side of the chart. As the number of transactions increases, however, the Porter et al. models become closer predictors of the true intention value. Especially when the shift occurs, the Porter et al. time decay plots have much less error than even the Porter et al. general model.

Discussion and Conclusions

The results presented above support one underlying conclusion: eBay's model consistently produces scores that are higher than both the Porter et al. reputation score and the true intention. This is shown in Figures 1 and 2 and supported with statistical analyses in Figure 3 and Table 1.

The data in Table 1 shows that the disparity between the SSE values of the two models is greater when the seller profile reflects a higher probability of either neutral or poor behavior. In those circumstances, the Porter et al. model more closely approximates the true intention value. Thus, buyers relying on the eBay-generated reputations have an inflated sense of trust in the seller.

Two specific issues drive the inflated reputation scores of eBay's percent positive model. First, it is bounded below by zero, so that a seller who has received only negative feedback ratings will never acquire a negative reputation score. Second, the presence of neutral scores is not accounted for in the calculation. If a seller receives a large number of neutral, or 0, feedback, there is no change in his/her reputation score. Over time, this will result in the eBay value overestimating a seller's percent positive feedback.

The implication for buyers is twofold. First, buyers may agree to transact with a seller they would not otherwise choose. Second, they may be paying higher prices for goods purchased then they should. A number of studies have shown that higher reputation scores lead to higher prices (Ba and Pavlou 2002; Fan et al. 2005; Melnik and Alm 2002). This is obviously desirable to sellers, but also obviously unfair to buyers.

Inflated reputation scores also have implications for eBay because buyers are agreeing to transact with sellers based on inflated levels of trust. Thus, there is a greater chance their actual experience will not match their expectations than if they enter into transactions with more realistic trustworthiness expectations. Buyers' trust in eBay will decline, and buyers may potentially stop bidding or buying on eBay as expectations are disconfirmed and it becomes clear the reputation score provided by eBay is neither reliable nor believable. Further, Zand (1972) shows a relationship between accurate information and increasing trust in his model of the relationship of trust to information, influence, and control. Thus, it can be argued that accurate reputation information (i.e., reliable and believable) will ultimately increase the trust buyers place in both individual sellers and eBay as a whole.

Furthermore, Bromley (1993) discusses the concept of reputation management. Reputation theory suggests that parties engage in actions designed to sustain or enhance their reputational standing such as very high levels of service or quality of goods. This is important because systems that accurately reflect reputation might encourage more of this reputation management behavior than scoring procedures such as that of the eBay model, where scores are artificially high (and difficult to force down due to the cumulative model used). With less incentive to behave in a way that will enhance one's reputation, seller behavior can negatively affect the buyer experience.

This paper serves to forward our understanding of how to more accurately represent the reputation of members in online auctions. This is important because the trust one party places in another will be misguided if the basis for that trust, in this case reputation scores, is not an accurate reflection of the individual. Despite the contribution of this work, additional research is needed to further improve and test online reputation scoring procedures. In particular, the Porter et al. model includes a feature to handle starting reputation scores which, in turn, may generate (or remove) barriers-to-entry. Future simulations will test this aspect of the model as well as time decay profiles where seller behavior improves over time.

References

Ajzen, I. "The Theory of Planned Behavior," *Organizational Behavior and Human Decision Processes* (50), 1980, pp. 179-211.

Ajzen, I., and Fishbein, M. *Understanding Attitudes and Predicting Social Behavior*, Prentice-Hall, Englewood Cliffs, NJ, 1980.

Ba, S. "Building Trust in Online Auction Markets through an Economic Incentive Mechanism," *Decision Support Systems* (35), 2003, pp. 273-286.

Ba, S., and Pavlou, P.A. "Evidence of the Effect of Trust Building Technology in Electronic Markets: Price Premiums and Buyer Behavior," *MIS Quarterly* (26:3), 2002, pp. 1-26.

Bowlby, J. *Attachment and Loss, Vol. 1: Attachment*, Basic Books, New York 1982.

Bromley, D. B. *Reputation, Image, and Impression Management*, John Wiley & Sons, Chichester, England, 1993.

Brynjolfsson, E., and Smith, M. "Frictionless Commerce? A Comparison of Internet and Conventional Retailers," *Management Science* (46:4), 2000, pp. 563-585.

Bunnell, D. *The eBay Phenomenon*, John Wiley, New York, 2000.

Fan, M., Tan, Y., and Whinston, A. B. "Evaluation and Design of Online Cooperative Feedback Mechanisms for Reputation Management," *IEEE Transactions on Knowledge and Data Engineering* (17:2), 2005, pp. 244-254.

Fischer, J. "eBay (EBAY)," Fourth Quarter Update Released: February 22, 2004, *The Motley Fool* (available online at **http://www.fool.com/research/reports/ebay.htm**).

Friedman, E., and Resnick, P. "The Social Cost of Cheap Pseudonyms," *Journal of Economics and Management Strategy* (10), 2001, pp. 173-199.

Fukuyama, F. *Trust: The Social Virtues and the Creation of Prosperity*, The Free Press, New York, 1995.

Hoover, S. V., and Perry, R. F. *Simulation: A Problem-Solving Approach*, Addison-Wesley, Reading, MA, 1989.

Jarvenpaa, S. L., Tractinsky, N., and and Vitale, M. "Consumer Trust in an Internet Store," *Information Technology & Management* (1:1/2), 2000, pp. 45-71.

Josang, A., Hird, S., and Faccer, E. "Simulating the Effect of Reputation Systems on e-Markets," in *Trust Management: First International Conference*, Springer-Verlag, Berlin 2003, pp. 179-194.

Kauffman, R. J., March, S. T., and Wood, C. A. "Design Principles for Long-Lived Internet Agents," *International Journal of Intelligent Systems in Management* (9:4), 2000, pp. 217-236.

Lim, K. H., Sia, C. L., Lee, M. K. O., and Benbasat, I. "how Do I Trust You Online, and If So, Will I Buy? An Empirical Study on Designing Web Contents to Develop Online Trust," Working Paper, Faculty of Commerce and Business Administration, University of British Columbia, 2001.

Madden, M. J., Ellen, P. S., and Ajzen, I. "A Comparison of the Theory of Planned Behavioral and the Theory of Reasoned Action," *Personality and Social Psychology Bulletin* (18:1), 1992, pp. 3-9.

Malaga, R. A. "Web-Based Reputation Management Systems: Problems and Suggested Solutions," *Electronic Commerce Research* (1), 2001, pp. 403-417.

McAfee, R. P., and McMillan, J. "Auctions and Bidding," *Journal of Economic Literature* (25:2), 1987, pp. 699-738.

McKnight, H., Choudhury, V., and Kacmar, C. "Developing and Validating Trust Measures for e-Commerce: An Integrative Typology," *Information Systems Research* (13:3), 2002, pp. 334-359.

Melnik, M., and Alm, J. "Does a Seller's Ecommerce Reputation Matter? Evidence from eBay Auctions," *Journal of Industrial Economics* (50), 2002, pp. 337-349.

Merriam-Webster. *Merriam-Webster Online Dictionary*, 2005 (available online at **http://www.m-w.com/netdict.htm**).

Meyerson, D., Weick, K. E., and Kramer, R. M. "Swift Trust and Temporary Groups," in *Trust in Organizations: Frontiers of Theory and Research*, R. M. Kramer and T. R. Tyler (Eds.), Sage Publications, Thousand Oaks, CA, 1996, pp. 166-195.

Pavlou, P. A., and Gefen, D. "Building Effective Online Marketplaces with Institution-Based Trust," *Information Systems Research* (15:1), 2004, pp. 37-59.

Porter, D., Ord, J. K., and Montano, B. R. "Scoring Procedures for Online Reputation Systems," Working Paper, Georgetown University, 2004.

Reichheld, F. F., and Schefter, P. "E-Loyalty: Your Secret Weapon on the Web," *Harvard Business Review* (78:4), 2000, pp. 105-113.

Resnick, P., Zeckhauser, R., Friedman, E., and Kuwabara, K. "Reputation Systems: Facilitating Trust in Internet Interactions," *Communications of the ACM* (43:12), 2000, pp. 45-48.

Tan, Y-H., and Thoen, W. "Towards a Generic Model of Trust for Electronic Commerce," *International Journal of Electronic Commerce* (3), 1998, pp. 65-81.

Venkatesh, V., and Davis, F. D. "A Theoretical Extension of the Technology Acceptance Model: Four Longitudinal Field Studies," *Management Science* (46:2), 2000, pp. 186-204.

Weiss, A. M., Anderson, E., and MacInnis, D. J. "Reputation Management as a Motivation for Sales Structure Decisions," *Journal of Marketing* (63:4), 1999, pp. 74-89.

Wilson, R. "Reputations in Games and Markets," in *Game Theoretic Models of Bargaining*, A. Roth (Ed.), Cambridge University Press, Cambridge, England, 1985, pp. 27-62.

Yu, B., and Singh, M. P. "Distributed Reputation Management for Electronic Commerce," *Computational Intelligence* (18:4), 2002, pp. 535-549.

Yamamoto, H., Ishida, K., and Ohta, T. "Modeling Reputation Management System on Online C2C Market," *Computational & Mathematical Organization Theory* (10), 2004, pp. 165-178

Zand, D. E. "Trust and Managerial Problem Solving," *Administrative Science Quarterly* (17:2), 1972, pp. 229-239.

Zucker, L. G. "Production of Trust: Institutional Sources of Economic Structure," in *Research in Organizational Behavior, Volume 8*, B. Staw and L. Cummings (Eds.), Elsevier, Amsterdam, 1985, pp. 53-111.