Avoiding Ballot Stuffing in eBay-like Reputation Systems

Rajat Bhattacharjee * Ashish Goel †
Stanford University Stanford University

December 8, 2004

Abstract

We study the robustness of binary feedback reputation systems (e.g. eBay) to ballot stuffing and bad mouthing. In a feedback based reputation system, a seller can collude with other buyers to undertake fake transactions in order to enhance her reputation. This problem is referred to as ballot stuffing. A seller can also be targeted by a group of buyers to deliberately lower her reputation. This problem is referred to as bad mouthing. For the reputations to be meaningful, any practical reputation system needs to be resistant to these problems. We give an explicit relation between the reputation premium and the transaction cost that needs to hold in order to avoid ballot stuffing. Thus we draw attention to the necessity of transaction costs for a well functioning reputation system.

1 Introduction

While reputation systems have long been studied in the economics literature, recently there has been a renewed surge of interest due to their immense application in electronic commerce. Websites/services like eBay, Epinions, Amazon.com, etc. have been successfully using reputation systems. Indeed reputation systems have always been implicitly used in business transactions. However, in transactions on the Internet where there is a lack of other traditional indicators of trustworthiness, the importance of feedback from previous customers is greatly enhanced. We refer the reader to [4, 8] for a more detailed introduction to reputation systems. Since reputations are essential for trust in electronic commerce, it is important that the purported reputation of an agent is an actual indicator of its trustworthiness. There are numerous fraudulent behaviors possible in a reputation system. In fact, it has been pointed out [3, 8] that it is surprising that these systems actually work with such success.

In this paper we concentrate on two strategies that can lead to reputations which are not reflective of the trustworthiness of a seller. A seller can collude with other buyers to undertake fake transactions resulting in positive ratings. These fake transactions have the effect of inflating a seller’s reputation. This is known as ballot stuffing [1, 2]. On the other hand, a group of buyers can collude (may be at the behest of a rival seller) to deliberately give negative feedback to a particular seller and hence lower her reputation. This is known as bad mouthing. For reputations to be
meaningful, any practical reputation system needs to be resistant to these behaviors. Else, agents would game the system to inflate/deflate reputations, rendering the system useless. Mechanisms based on anonymity and clustering based on similarity of buyer profiles have been suggested in previous work [1][2]. In this paper, our aim is to give restrictions on reputation premiums such that there is no incentive for agents to indulge in either ballot stuffing or bad mouthing. We give explicit relations between the reputation premium and the transaction costs that need to hold in order to avoid ballot stuffing. Thus we draw attention to the necessity of transaction costs for a well functioning reputation system. We also point out how similar ideas can be extended to bad mouthing in a restricted setting.

2 Related work

Resnick et al. [9] report a controlled experiment conducted on eBay. The experiment involved a high-reputation established eBay dealer selling matched pairs of lots (batches of vintage postcards in this case). They observed that, in consonance with the expected, the established identity fared better. In this experiment, the difference in buyer’s willingness-to-pay was 8.1% of the selling price. In the case of relatively new sellers, one or two negative feedbacks did not affect buyer’s willingness-to-pay. Another feature of eBay is the rarity of negative feedback. Resnick and Zeckhauser [7] observe that sellers received negative feedback only 1% of the time. Eliciting feedback from entities is another problem. Many other empirical studies have been done on eBay; we refer the reader to [4][9] for pointers to these studies.

The reputation systems for Internet based transactions are different from conventional business transactions as the sellers who have got negative feedback can leave the system and join under a new name. Friedman and Resnick [5] study this scenario and conclude that buyers need to impose some disadvantage on sellers with no feedback at all. Among the suggestions they make, they discuss the possibility of charging for name changes, that is, a newcomer has to make some payment to the system before joining the system.

Dellarocas [3] introduced a formal model for binary reputation mechanisms like the one used by eBay. Here a buyer gives a rating of +1 or −1 (i.e. “good” or “bad”) after a transaction is over. It is shown that leniency of buyer while rating sellers is important for the system to settle down to steady-state levels avoiding oscillations. He argues that the ratio of negative feedback to total feedback is a good indicator of reputation. Also, Dellarocas [1][2] considers the problems of ballot stuffing and bad mouthing. The methods suggested in these papers are based on anonymity and cluster filtering. Anonymising buyers and sellers works for bad mouthing. However it might not work for ballot stuffing as the seller might give hidden indications of its identity to its colluders. For example, the seller might signal its colluders by pricing its products at a price having a specific decimal point. The method of cluster filtering involves profiling buyers based on their other transactions and feedback. Then these profiles are used to give weights to specific feedback. Our work contrasts with these methods as we study the inherent capability of binary feedback based reputation system to deter dishonest behaviour without using additional heuristics.

3 Model

In a binary reputation system, a buyer provides feedback for a seller after a transaction. This feedback is either “positive” or “negative”. This is very similar to eBay where the buyer provides
feedback in the form of \{−1, 0, 1\} and also any additional comments. The reputation of a seller then constitutes the net feedback, i.e., the number of positive feedbacks less the number of negative feedbacks accrued by the seller. Any additional comment made by the system is also displayed. Dellarocas [3] has studied these binary reputation mechanisms. It is worth noting here that binary reputation systems are not the only kind of reputation systems being used. For example, Amazon asks its users to rate sellers in a scale of 5. There are other schemes which use a more complex feedback. We confine ourselves to the important subclass of binary reputation systems. We make the following assumptions in our model:

**Price of goods:** It is assumed for simplicity that there is only one good that the seller wishes to sell and the price of the good is 1.

**Feedback:** Each transaction results in a positive feedback or a negative feedback. It is assumed that the sums of positive and negative feedbacks displayed by the system spans the entire history of the seller. However, it might make sense to only display results for a particular time window. For example, eBay only spans a history of six months. We note here that our results hold for such systems too.

**Intrinsic error rate:** The rate at which a seller commits errors (i.e., unsatisfactory transactions) is denoted by $\epsilon$. The error rate, $\epsilon$, is assumed to be intrinsic and hence fixed to a seller. All the errors result in negative feedback. Also, any genuine transaction which is not an error results in positive feedback. It should be noted that these assumptions are not simple ones as empirical studies have shown that buyers tend to negotiate before giving negative feedback. Moreover, it is difficult to elicit feedback from buyers [7, 8].

**Ballot stuffing rate:** The rate at which fake transactions leading to a positive feedback occur (for a particular seller) is denoted by $\rho$, i.e. for each genuine transaction undertaken by the seller, there are $\rho$ fake transactions on the average which all lead to positive feedback.

**Perceived error rate:** The rate at which the system observes a seller making error is denoted by $\epsilon_p$:

$$\epsilon_p = \frac{\epsilon}{1 + \rho}$$

Thus a seller's perceived error rate can decrease with an increase in $\rho$.

**Central System:** It is assumed that there is a central system that keeps record of the feedback and collects transaction costs.

**Reputation premium:** Buyers use the reputation of a seller in order to make decision on a purchase. We model this effect by means of a function we call reputation premium, $\xi$. It is a function of the perceived error rate $\epsilon_p$. Reputation premium implies that the seller can sell the product at a price $1 + \xi$, i.e. it represents the buyers’ willingness-to-pay for higher reputation. We make the assumption that $\xi$ is continuous and differentiable. Based on the work of Dellarocas [3], we assume that $\xi$ is a function of the perceived error rate, $\epsilon_p$. We also assume that the reputation premium decreases with an increase in perceived error rate, $\epsilon_p$, i.e.,

$$\xi'(\epsilon_p) \leq 0.$$  (1)
Transaction Costs: Each transaction has a fixed seller transaction cost, $\tau$, paid by the seller to the central system.

4 Reputation premiums resistant to ballot stuffing

Definition. An $(\epsilon, \rho)$-Bernoulli seller is a seller for which the intrinsic error rate is $\epsilon$ and the ballot stuffing rate is $\rho$. More precisely, the seller errs in a genuine transaction with probability $\epsilon$ independent of all other transactions. Moreover, the probability that a transaction is fake is $\frac{\rho}{1 + \rho}$ independent of all other transactions.

Definition. The transaction cost per genuine transaction is denoted by $\psi(\rho)$ and is equal to $\tau(1 + \rho)$.

Definition. A reputation premium, $\xi$, is inflation resistant if the following relation holds $\forall \epsilon \in [0, 1]$:

$$\left. \frac{\partial}{\partial \rho} (\xi - \psi) \right|_{\rho=0} \leq 0.$$  \hspace{1cm} (2)

In other words, $\xi$ should be such that the increase in reputation premium due to increase in positive fake transactions is offset by the increase in transaction costs.

Theorem 1 A reputation premium is inflation resistant iff for all $x \in (0, 1)$:

$$-x\xi'(x) \leq \tau.$$  \hspace{1cm} (3)

Proof: Suppose equation (3) holds for $\xi$. From (2), $\xi$ is inflation resistant if, $\forall \epsilon \in [0, 1]$,

$$\left. \frac{\partial}{\partial \rho} (\xi - \psi) \right|_{\rho=0} = 0 \Leftrightarrow \left. \frac{\partial \xi}{\partial \rho} \right|_{\rho=0} \leq \left. \frac{\partial \psi}{\partial \rho} \right|_{\rho=0} = \tau.$$

However,

$$\left. \frac{\partial \xi(x)}{\partial \rho} \right|_{\rho=0} = \left. \frac{\partial}{\partial \rho} \left( \frac{\epsilon}{1 + \rho} \right) \right|_{\rho=0} = -\frac{\epsilon}{(1 + \rho)^2} \left. \xi' \left( \frac{\epsilon}{1 + \rho} \right) \right|_{\rho=0} = -x\xi'(x) \leq \tau.$$

Hence, $-x\xi'(x) \leq \tau$, $\forall x \in (0, 1)$ is a necessary and sufficient condition for $\xi$ to be an inflation resistant reputation premium.

4.1 How high can reputation premiums be?

We now consider the following question: assuming a reputation premium is inflation resistant, how high can a reputation premium be? Suppose $\xi$ is an inflation resistant reputation premium
satisfying equation (4) by equality at each point. We assume that $\xi(1) = 0$, i.e., if the perceived error rate of a seller is 1, then she doesn’t receive any reputation premium. Therefore,

$$\xi(\epsilon_p) = -\int_{\epsilon_p}^{1} \xi'(x) dx = \int_{\epsilon_p}^{1} \frac{\tau}{x} dx = (\tau) \log \frac{1}{\epsilon_p}. \quad (4)$$

Note that $\xi \to \infty$ as $\epsilon_p \to 0$.

### 4.2 A linear reputation premium

Another obvious choice for reputation premium is $\xi(\epsilon_p) = \tau(1 - \epsilon_p)$. The reputation premium is clearly inflation resistant since $\forall x \in [0, 1],$

$$-x\xi'(x) = x\tau \leq \tau$$

![Figure 1: $\tau = 0.1$](image)

### 4.3 Importance of transaction costs

The most important contribution of this paper is that it highlights the necessity of transaction costs for a reputation system to work well. Further, the reputation premium accrued to the seller should be related to the transaction cost. This is in consonance with the empirical results of Resnick et al. [9]. It is demonstrated by experiments that the difference in buyer’s willingness-to-pay is 8.1% of the selling price, when buyers with established reputation and no reputation are compared. We note that the order of the reputation premium is the same as that of the transaction cost.

### 4.4 Bad mouthing

Here we show that our analysis carries over to bad mouthing, although in a restricted setting. We now assume that the rate at which fake transactions leading to a negative feedback occur is denoted
by \( \eta \), i.e., for each genuine transaction undertaken by the seller, there are \( \eta \) fake transactions on the average which all lead to negative feedback. The perceived error rate, \( \epsilon_p \), now becomes \( \frac{\epsilon + \eta}{1 + \eta} \) (we assume for now that there is no ballot stuffing). We assume that each transaction also has a transaction cost for the buyer which we denote by \( \beta \) (this may correspond to shipping and handling costs). It is further assumed that \( \beta > \tau \). We note that this is a fairly restrictive condition, as it is saying that the transaction cost of the buyer is more than the transaction cost of the seller. The transaction cost for the buyers intending to deflate the sellers reputation is denoted by \( \chi(\eta) = \beta \eta \).

A reputation premium, \( \xi \), is deflation resistant if the following relation holds \( \forall \epsilon \in [0, 1]: \)

\[
\frac{\partial}{\partial \eta} (-\xi + \psi - \chi) \bigg|_{\eta=0} \leq 0. \tag{5}
\]

In other words, \( \xi \) should be such that the decrease in reputation premium and increase in transaction costs of the seller should not be more than the transaction costs of the malicious buyers. We claim that a reputation premium is deflation resistant iff, \( \forall x \in [0, 1], \)

\[-(1-x)\xi'(x) \leq \beta - \tau.\]

The proof is similar to the inflation resistant premium. First, observe that

\[
\frac{\partial \xi(x)}{\partial \eta} \bigg|_{\eta=0} = \frac{\partial}{\partial \eta} \xi \left( \frac{\epsilon + \eta}{1 + \eta} \right) \bigg|_{\eta=0} = \left( \frac{1}{1+\eta} - \frac{\epsilon + \eta}{(1+\eta)^2} \right) \xi' \left( \frac{\epsilon + \eta}{1 + \eta} \right) \bigg|_{\eta=0} = (1-x)\xi'(x).
\]

Then, observe that \( \frac{\partial}{\partial \eta}(\chi - \psi) = \beta - \tau \) to obtain the desired condition: \( -(1-x)\xi'(x) \leq \beta - \tau. \) To also avoid ballot stuffing, we still need the condition \( -x\xi'(x) \leq \tau. \)

The maximising \( \xi \) would now be the one which has \( -\xi'(x) = \frac{x}{\tau}, \forall x \in \left[ \frac{x}{\tau}, 1 \right] \) and \( -\xi'(x) = \frac{\beta - \tau}{1-x}, \forall x \in (0, \frac{x}{\tau}) \). Hence, \( \xi(\epsilon_p) = \tau \log \frac{1}{\epsilon_p}, \forall x \in \left[ \frac{x}{\tau}, 1 \right] \) and \( \xi(\epsilon_p) = \tau \log \frac{\beta}{\tau} + (\beta - \tau) \log \frac{1-\epsilon_p}{\beta-\tau}, \forall x \in (0, \frac{x}{\tau}) \). It is easy to verify that the linear reputation premium, \( \xi(\epsilon_p) = \tau(1 - \epsilon_p) \) is also deflation resistant for \( \beta \geq 2\tau \). Figure 2 plots the maximising and the linear reputation premium for \( \beta = 0.2, \tau = 0.1 \).

5 Caveat

In this section, we draw attention to the problem of an initial strategising seller. This section is illustrative rather than formal, and we are sometimes going to rely on informal approximations.

Definition. A \( \rho \)-initial strategising seller is one who does \( \rho k \) positive fake transactions and then undertakes \( k \) genuine transaction with a Bernoulli error rate \( \epsilon \). An honest Bernoulli seller is nothing but the 0-initial strategising seller.

For the purpose of discussion in this section, we assume that the reputation premium used by the system is the linear reputation premium discussed above. The total reputation premium accrued by a \( \rho \)-initial strategising seller is approximately equal to:

\[
\sum_{i=1}^{k} \tau \left( 1 - \frac{\epsilon i}{\rho k} \right) \approx k \int_{0}^{1} \tau \left( 1 - \frac{\epsilon x}{\rho} \right) dx = k \tau \left( 1 - \epsilon \left( 1 - \frac{1+\rho}{\rho} \log \left( \frac{1+\rho}{\rho} \right) \right) \right).
\]
The total reputation premium accrued by a 0-initial strategising seller (honest) is: $\tau(1-\epsilon)$. Hence, the net reputation premium that the seller makes by being a $\rho$-initial strategising seller is:

$$k\tau\epsilon\rho \log \left( \frac{1+\rho}{\rho} \right)$$

On the other hand, the excess transaction cost paid by a $\rho$-initial strategising seller is: $\tau\rho$. Hence in the scenario

$$k\tau\epsilon\rho \log \left( \frac{1+\rho}{\rho} \right) > k\tau\rho$$

$$\rho < \frac{1}{exp\left(\frac{1}{\epsilon}\right) - 1}$$

the initial strategising seller is going to benefit.

The discussion in this section reinforces the need for a special treatment to new sellers. Please see [5] for a detailed treatment of this subject. In our setting, the extra profit that the initial strategising seller can make is unbounded (see appendix A), making the problem even more challenging. Providing disincentives against initial-strategising remains an important open problem.

6 Conclusions and future work

In this paper we study the problem of ballot stuffing. We obtain a relation between reputation premium and transaction cost. This highlights the importance of transaction costs in a reputation system. Our results can be extended to a time window setting. The case of a new seller pointed out in the previous section is a very interesting open problem. We have considered only a restricted class of transaction costs and reputation premiums in this paper. It would be interesting to study other functions for transaction costs and reputation premiums which might alleviate the problem of the cheating newcomers.
References


A Unbounded returns for the initial-strategising seller

Let $\pi(k, \rho)$ denote the extra profit (i.e. the extra reputation premium minus the extra reputation cost) that a $\rho$-initial-strategising seller can make after $k$ genuine transactions. In section 5 we showed that

$$
\pi(k, \rho) \approx k\tau\epsilon\rho \log \left( \frac{1 + \rho}{\rho} \right) - k\tau\rho
$$

The second quantity above is maximized when $\rho = \exp(-1 - 1/\epsilon)$, and with this value of $\rho$, we obtain $\pi(k, \rho) \approx k\rho\tau\epsilon$. Here $\tau$ and $\epsilon$ are intrinsic parameters of the system and the merchant respectively, and $\rho$ depends only on $\epsilon$. Hence, by letting $k$ grow arbitrarily large, the extra profit $\pi(k, \rho)$ becomes unbounded. Interestingly, the rate of return (i.e. the ratio of the extra profit to extra investment) for this choice of parameters is $\approx \epsilon$. 