Evaluation of Reputation Systems in Internet Auctions—Experimental Approach

Hiroyuki Takahashi, Yukihiro Usui†, and Noriaki Yoshikai

Abstract

Creating and evaluating trust in a network community are major issues for an Internet trading system that hopes to operate successfully in an anonymous and asymmetrical environment. We experimentally investigated the effect of reputation systems on Internet auctions and the behavior of participants in auctions with and without reputation systems. We found that a simple reputation system promotes trading.

1. Introduction

The range of people or entities with which an individual or a company can trade has been expanding as a result of advances in information communication services (such as broadband Internet and cellular phone services) and market globalization. Many users around the world participate in eBay or Yahoo! auctions, the pioneers in Internet auctions, and conduct numerous trades. These changes have increased trading opportunities for individuals and for businesses, which are predicted to take advantage of these opportunities in a range of situations.

However, there is a downside to the increased opportunities: the potential for higher risk in trading. For example, in Internet auctions between individuals, it is easy for the seller to falsify his/her identity or send false information about the product for sale, so participants are always faced with the possibility of being deceived and of suffering a loss. This problem is based on the asymmetrical availability of information about the products for sale between sellers and buyers; that is, sellers know all about the products including their flaws while buyers have no way to get any information that sellers have not disclosed. Since buyers cannot bridge this information gap, they can easily be cheated by fraudulent sellers, so they tend to refrain from trading. Even when honest sellers try to trade high-quality products, buyers tend to refrain from trading because they have no information about who is honest or dishonest. This situation, in which honest traders and high-quality goods are driven out of the trade by dishonest traders and low-quality goods, is called a "lemon market" [1]. It is therefore necessary to develop a method that will maintain the benefit of the increased trading opportunities and reduce the risk inherent in trading on the Internet.

One method used in current net auctions for this purpose is a "reputation system". However, there have been few detailed studies or analyses of whether and to what degree such a system is effective. Moreover, there is no established method for designing such a system.

Kollock [2] studied net auctions with a reputation system using trading data from eBay. This was a study based on observing real net auctions, so it involved each participant's preferences and how they valued the products for their own use. Thus, it did not specifically investigate the reputation system itself.

Matsuda et al. [3] empirically verified Kollock's results. They eliminated the effect of personal value by using a fictitious trading market on computers. However, their experiment studied direct trading models, not auction models. In direct trading, a seller cannot trade successfully if the buyer does not trust the seller. In auctions, on the other hand, a seller can trade successfully if even one of the buyers in the market trusts the seller. This means that there is more

† NTT Information Sharing Platform Laboratories
Musashino-shi, 180-8585 Japan
E-mail: usui.yukihiro@lab.ntt.co.jp

chance of making a successful trade in auctions than in direct trading, even if inferior goods are being traded. To prevent a lemon market we need more efficient reputation systems in auction trading. The papers mentioned above do not provide empirical data about the effectiveness of a reputation system for auctions. We used an experimental system that isolated the effects of reputation on net auctions.

In this paper, we outline the experimental system and experiment flow and present results showing the effects of reputation systems on net auctions.

2. Need to create trust

Conventionally, when people make decisions and take actions, they rely on information from trustworthy people they have contacted before. However, in a net community, the triumph over distance and time has made communication virtually anonymous. In addition, the turnover of community members has weakened any control over other members, increasing the risk of one's trust being betrayed. On the other hand, the network by its nature allows information to be provided by a variety of people. This means that a net community offers great benefits if information sources can be trusted. It is therefore important to foster trust. The development of Linux is a good example. In this case, although the asymmetrical availability of information is not an important issue, mutual trust between members is a reward in itself. The trust cultivated between members encourages cooperation within the community and can lead to the achievement of a secure position in other communities as one interacts with them. In other words, a reputation for trustworthiness not only represents a quality guarantee, but can also be considered the main driving force making an information society function effectively.

3. Problems with existing reputation systems

Reputation systems give people a way to judge the trustworthiness of other people with whom they are not familiar, and they serve as an incentive that encourages people to be honest. However, Kollock [2] shows that reputation systems are not sufficiently effective in auction markets, and that there are cases where the reputation of the seller does not necessarily correspond to his/her behavior. We believe that this result can be attributed to flaws in the reputation systems used. A reputation system needs (a) a way of acquiring reputations from participants, (b) a method of expressing reputation, and (c) a method of sharing reputation information [4].

To acquire reputation, the designer must consider the fact that participants in the auction generally have no incentive to supply accurate reputation information. For example, eBay users are requested to enter a reputation value in a feedback form after a successful bid, but they seldom take the time and trouble to do so. To solve this problem, the designer should prepare an incentive system that pays users for entering reputation values. Since people tend to be shy of making negative evaluations, another problem is that net auction sites often abound in positive evaluations.

One problem with sharing information is that a user can have multiple online identities. In many net auction sites, participants remain anonymous and can register under new names as often as they want. If a user gets a bad reputation as a result of fraudulent trading, he/she can participate in a new auction under a new name. Thus, the possibility of having a bad reputation is not an effective deterrent to fraudulent trading. This may be countered by allowing the user to participate only under his/her real name or by prohibiting the use of multiple pseudonyms so that the user will be forced to use the same name. However, as we show in section 6.2, fraudulent trading can be prevented to some extent through careful operation of the reputation system. A related problem is that reputation values currently cannot be shared among different auction sites.

The third issue is the difficulty of expressing reputation. Most existing auction sites use the sum of individual evaluations, each of which is expressed as one of three grades (–1, 0, +1). However, there is no theoretical basis for this, so a reputation evaluation method is still an issue.

As a basic review for these issues, we investigated the effects of reputation systems on net auctions.

4. Experimental auction system

We developed an experimental auction system and used it to investigate the problems mentioned above. First, to study the effect of introducing a reputation system in a lemon market, which is characterized by asymmetrical availability of information, we designed a simplified virtual Internet auction site. As **Fig. 1** shows, the auction flow in this system supports the asymmetrical availability of information about products for sale. **Figure 2** shows the system's main page.

At regular intervals, the system gives participants an opportunity to manufacture a product. It gives the

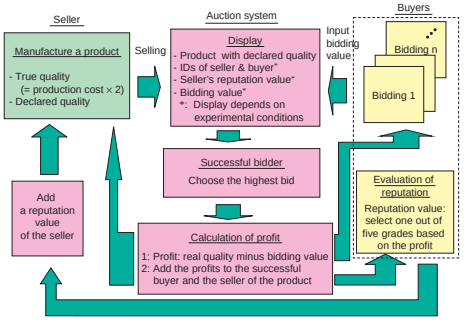

Fig. 1. Auction flow in the experimental auction system.

Fig. 2. Main page of the prototype auction system.

sellers the nominal manufacturing cost, which represents the product's quality level. The manufacturer can choose to produce a cheaper product to save money and pocket the difference. The product in the virtual auction site is an abstract product without any substance. The product's true quality represents its manufacturing cost plus the value added by its manufacturing. For simplicity, in our experiments, the true quality was set to twice the true cost. In real-world trading, the true quality corresponds to the overall quality including any flaws. The product manufactured by a seller is put on an auction market for sale with a price set by the seller. We regard this price as the declared quality. The declared quality is displayed to buyers in the auction market. A seller can set the declared quality freely, independent of the true quality. For example, the case where the declared quality is set higher than the true quality is equal to the case where information about the product for sale does not disclose any flaws.

As buyers, participants can bid for a product during some auction period based on the declared quality that the seller has set, but they do not know the true quality. At the end of the auction period, the highest bidder is successful and receives the product. In this trade, the seller gets the following profit:

Seller's profit = nominal manufacturing cost – true cost + successful bid value.

That is, the higher the successful bid value is above the true cost, the greater the seller's profit. After the auction period has ended, a buyer can get, for the first time, information about the true quality of the product he/she has. This enables the buyer to evaluate this trade based on how the seller has exploited the asymmetrical availability of information about the product, that is, the balance between the true quality and the declared quality the seller set. In this experimental system, the product the buyer has received is resold to the experimenter at the true quality cost immediately after the buyer enters a reputation value for the seller. Therefore, the successful bidder gets the following profit:

Successful bidder's profit = true quality – successful bid value.

In this experimental system, the successful bidder only receives the profit if he/she enters a reputation value. This is an incentive for bidders to provide a reputation evaluation.

The auction system summarizes reputation values for sellers and displays the results with information about products in the auction market. In this experiment, fictitious products were displayed, and each product had associated with it: (1) a declared quality, (2) seller's ID, (3) seller's reputation value, and (4) bidding value. Items (2)–(4) were varied according to the experimental conditions. Participants could use the reputation value to determine whether or how much to bid.

To study the effects of methods of expressing reputation and sharing reputation information, we investigated various experimental conditions, such as with or without a reputation function and with or without a user ID changing capability. We also made it possible to try various reputation algorithms that are intended to promote trading.

5. Experiments

5.1 Participants

We recruited participants through an open invitation on an experimentation website and through advertisements in e-mail magazines. A total of 2030 people across Japan applied and participated in the experiment. 71% were men and 29% women. Half of them were office workers. More than 70% of both the male and female participants had experience in Internet auctions and had been using the Internet for more than five years.

5.2 Procedure

Participants logged in to the experimental Internet auction server through the Internet using their own computers. Subjects in the experiment participated in auctions with others and were rewarded in proportion to the profit they made through the auctions. A number of sessions were defined in the experiments. Different sessions were conducted under different conditions with different parameters.

Participants acted as both sellers and buyers; that is, each subject manufactured and sold products and bid for products in the auction. What subjects did while trading was at their own discretion, and they received a profit or incurred a loss as a result. In the experimental marketplace, there was no information available about the true identity of individual subjects. Instead, IDs that were valid only in this marketplace were used.

The true and declared qualities were expressed using one of ten different grades (10, 20, ..., 100). The seller who manufactured the product set a declared quality independently of the true quality. The reputation value was expressed as one of five grades (i.e., +2: very good, +1: good, 0: neutral, –1: bad, and –2: very bad). The successful bidder's evaluation was added to the reputation information about the seller.

5.3 Conditions

The experimental period was about two weeks. There were 30 to 50 subjects in each experiment group. They could participate in auctions at any time of the day or night via Internet access. The period for each auction was in principle 24 hours, but the finishing time was varied at random by up to plus or minus one hour. The time left until the end of the auction for each product was displayed in 15-minute units. The reason for this was to eliminate the bidding rush in the final minute of the period. The end of the experimental period was announced only 48 hours in advance to prevent any subject knowing the exact finishing time of the experiment, which might allow them to change their attitude and affect the validity of the experiment. In this paper, we report results for the following two cases, where we did not allow users to change IDs.

(1) Case 1: declared quality, seller's ID, and current highest bid were disclosed.
(2) Case 2: in addition to the information in case 1, the seller's reputation score was disclosed.

In case 1, participants could not use reputation information. This condition was necessary to analyze the difference in the participants' behaviors and extract the effects of reputation systems. We assigned 50 subjects in each experiment group with 2 groups for each case. The figures in the following sections show the data for one group in each case.

6. Results and discussion

6.1 Without a reputation system

In case 1, the only information that buyers could get was the declared quality and seller's ID, so they could not use reputation information when selecting products to bid for. **Figure 3** shows the chronological changes in the declared and true qualities of products in one auction. The figure shows the moving average for 50 products. Both qualities decreased with time, and the true quality was much lower than the declared

quality, indicating that goods of inferior quality were offered for sale.

In the absence of a reputation system, the only information the buyer has for determining whether the seller is trustworthy is his/her own experience of trading with that particular seller. If a seller knows that there is no risk of anyone other than the successful bidder learning that the product was of poor quality, he/she may choose to sell a poor quality product (i.e., the declared quality is higher than the true quality) because that is likely to produce more profit in such market conditions. Consequently, the seller has an incentive to sell inferior products. This seems to explain the results shown in Fig. 3

In such an auction market, inferior products prevail. Even if buyers think that they bought products at bargain prices, they may find out that they were cheated. Consequently, they become more reluctant to buy, and products submitted for auction remain unsold. The market gradually shrinks and degenerates into a lemon market.

Figure 4 shows the chronological change in the successful bidder's profit (which is the true quality minus the successful bid value: moving averages for 50 products). It shows that, in a lemon market like this, successful bidders eventually suffer a loss. For buyers, this market is dangerous as well as worthless.

6.2 With a reputation system

We repeated this experiment with a reputation system. **Figure 5** shows the chronological changes in the declared and true qualities of products (moving averages for 50 products). This time, there was no decline in the true quality of products. In other words, products of high true quality continue to be traded. This was because participants felt they could trust this market and they tended to manufacture high-quality products, from which they could expect more profit than low-quality products, and buyers bid higher without unconcern. The market thrived as a result of introducing a reputation system. However, after an announcement that the experiment would end in 48 hours, the true quality began to drop significantly, resulting in a lemon market. At the same time, bar-

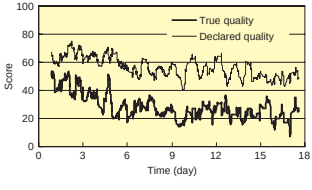

Fig. 3. Chronological changes in the declared and true qualities of products (case 1).

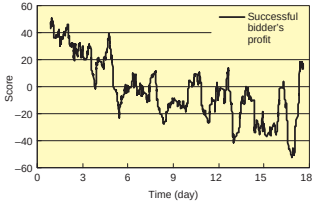

Fig. 4. Chronological change in successful bidder's profit (case 1).

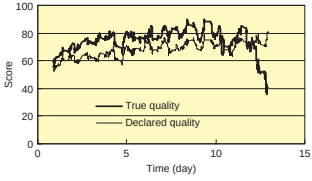

Fig. 5. Chronological changes in the declared and true qualities of products (case 2).

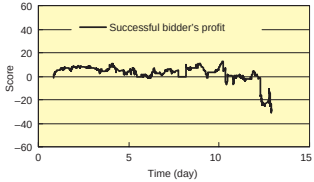

Fig. 6. Chronological change in successful bidder's profit (case 2).

gains, with true quality higher than the declared quality, continued to be offered. This phenomenon was always observed regardless of the different experimental conditions, such as different reputation grades and different algorithms.

Figure 6 shows the chronological change in the profit made by successful bidders. It shows that, unlike the case with no reputation system, successful bidders got a profit on average, except during the period after the termination announcement. This means that the market was useful for buyers to get a profit and they had an incentive to participate in this market.

Figure 7 shows the relationship between the seller's reputation and the balance between the successful bid value and the declared quality. The horizontal axis shows the mean value of the previous evaluations of the seller up to the current product. The vertical axis shows the difference between the successful bid value and the declared quality, which indicates how much the bid value exceeded the declared quality. A regression line has been added to the figure. There is a positive correlation between the reputation score and the balance, and the correlation coefficient is 0.7. We found that, unless the reputation score was high, the successful bid value was rarely above the declared quality.

According to the rules of the experiment, a successful bidder whose bid was higher than the true quality suffered a loss. Through their trades, successful bidders experienced the fact that the true quality was higher than the declared quality for products from reputable sellers. Therefore, it seems that bidders tended to expect a high true quality for a product submitted by a reputable seller. In other words, bidders based their bidding decisions on the reputation system. As a result, a high reputation score assured bidders of the safety of making high bids, which led to greater activity in the auction market.

Figure 8 shows the relationship between the reputation score and the degree of honesty (which corresponds to the true quality minus the declared quality). Although not as prominent as in Fig. 7, the reputation score increased with honesty (correlation coefficient = 0.47). **Figure 9** shows the result of products only by sellers whose final reputation score was more than 1. These figures indicate that sellers with a high final reputation score tended to manufacture products that were not of inferior quality. The percentage of inferior products being submitted increased after an announcement of imminent termination of the auction market. If we look only at data prior to this

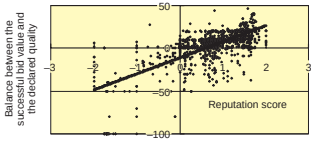

Fig. 7. Relationship between the seller's reputation score and the balance between the successful bid value and the declared quality.

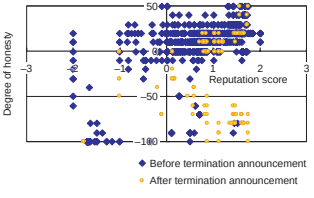

Fig. 8. Relationship between the reputation score and the degree of honesty.

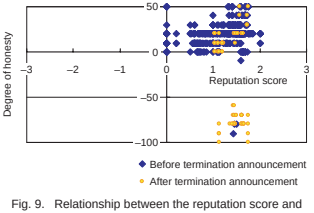

Fig. 9. Relationship between the reputation score and the degree of honesty (high reputation score sellers).

announcement, we can see a stronger correlation between reputation and honesty (correlation coefficient = 0.60).

Since all the sellers also acted as bidders, they knew that bidders are willing to offer a high bid for a product from a reputable seller. So, it can be safely assumed that reputable sellers found it necessary to maintain a high reputation score to have his/her prod-

uct bought at a high price. In other words, we conclude that the existence of the reputation system provides an incentive to refrain from selling inferior products.

7. Conclusion

To measure the effect of a reputation system, we developed a virtual auction system and conducted a number of experiments. They revealed the following.

(1) If the market is characterized by asymmetrical availability of information and anonymity, then an Internet auction is highly like to lead to a lemon market.
(2) When reputation was expressed by the simple mean value of evaluations on a five-grade scale, we observed that, even with such a simple reputation system, the elimination of complete anonymity and the sharing of reputation information could help prevent degeneration into a lemon market.

To refine the above findings, it will be necessary to study various algorithms and situations to gain insight into a desirable reputation system and its effects. Such future issues include: values used to indicate reputation, reputation scores other than a simple average of past evaluations, the effect of a "rogue" subject who deliberately gives poor reputation scores, and the effect of allowing sellers to change IDs or have multiple IDs at the same time.

8. Acknowledgment

We thank Professor Toshio Yamagishi of Hokkaido University for advising us about the Internet auction experiment and its operation from a psychological viewpoint.

References

[1] G. Akerlof, "The Market for "Lemons": Quality Uncertainty and the Market Mechanism," The Quarterly Journal of Economics 84, 1970.
[2] P. Kollock, "The Production of Trust in Online Markets," Advances in Group Processes 16, pp. 99-123, 1999.
[3] M. Matsuda and T. Yamagishi, "Design of reputation system for network trade and experimental study on the effectiveness of the prototype system," Fifth Kobe Forum, Kobe, Japan, 2001 (in Japanese).
[4] P. Resnick, R. Zeckhauser, E. Friedman, and K. Kuwabara, "Reputation systems," Communications of the ACM, Vol. 43, No. 12, pp. 45-48, 2000.

Hiroyuki Takahashi
Research Engineer, Software Architecture Project, NTT Information Sharing Platform Laboratories.
He received the M.S. degree in bioinformatics and biotechnology from the University of Tokyo, Tokyo in 1999. From 1997 until 1999 he was on the research staff at the National Aerospace Laboratory of Japan and Applied Microbiology Laboratory, the University of Tokyo. In 1999, he joined NTT Service Integration Laboratories. Currently he is working on trust engineering and the GRIT Project (Generating Reputation in Internet Trades) for virtual organization and network society. He is interested in the understanding, development, and application of reputation, trust, and human behavior with asymmetrical information availability for the network society in emerging next-generation communications systems.

Yukihiro Usui
Senior Research Engineer, Software Architecture Project, NTT Information Sharing Platform Laboratories.
He received the B.E. and M.E. degrees in applied mathematics and physics from Kyoto University, Kyoto in 1980 and 1982, respectively. In 1982, he joined the Electrical Communication Laboratories, Nippon Telegraph and Telephone Public Corporation (now NTT), Tokyo, Japan. He has been engaged in traffic analysis for telephone services, ATM operation systems, and IP network services. He is a member of the Institute of Electronics, Information and Communication Engineers (IEICE) and the Operations Research Society of Japan.

Noriaki Yoshikai
Senior Research Engineer, Service Development Strategy Projects for Business Promotion, NTT Service Integration Laboratories.
He received the B.E. and M.E. degrees in electronics engineering from Kumamoto University, Kumamoto in 1977 and 1979, respectively, and the Ph.D. degree in electrical engineering from the Tokyo Institute of Technology, Tokyo in 1988. He joined Yokosuka Electrical Communications Laboratories, Nippon Telegraph and Telephone Public Corporation (now NTT) in 1979, and has been engaged in research on line terminal equipment for high-speed digital transmission systems. He is interested in research on multimedia networks and its applications based on the distributed processing technology and computer supported cooperative work. In particular, he is investigating the network and services for intelligent transport systems and network society assessment. He is a member of IEICE and a senior member of IEEE.