An Experimental Study of Reputation with Heterogeneous Goods

Anya C. (Savikhin) Samak*

*School of Human Ecology, University of Wisconsin-Madison
1300 Linden Drive, Madison, WI 53706

October 24, 2012

Abstract

Reputation systems provide decision support for e-commerce. A shortcoming of existing systems is that all transactions are rated equally, and the impact of reputation systems for differently valued goods is not well understood. In an experiment, we study a heterogeneous good market. We find that the reputation system increases surplus by increasing transactions in the high value good. Allowing for heterogeneous goods reduces information, as buyers cannot determine whether the seller previously transacted in low/high value goods. We test a new system, which displays reputation separately for each good. We provide evidence that this additional information is utilized in decisions.

JEL Classifications: D02, D83
Keywords: laboratory experiment, trust, reputation, e-commerce, decision, market

Email Address: asamak@wisc.edu. Phone number: 608-262-5498

1. Introduction[1]

The growing popularity of e-commerce has increased the need for decision support systems that strengthen trust among strangers, reduce information asymmetries, and facilitate transactions when there is an opportunity to cheat [15, 28, 44] .[2] Reputation systems have emerged to fill this role, and have been implemented in several e-commerce sites such as eBay and Amazon Marketplace. The different design aspects of reputation systems, such as the amount and type of information to display, play an important role in determining the success of organizations that conduct business online [35]. The design of these mechanisms has potentially important implications for individual decision-making and for a wide range of management activities, such as brand building, customer acquisition and retention and quality assurance [17]. Effective reputation systems that promote trust and encourage successful transactions will increase the prevalence of e-commerce activities, since Internet consumers' trust impacts the purchasing decision [25, 34, 36].

We contribute to the study of reputation systems through an experimental investigation of how markets function when heterogeneous goods are available to buyers. In practice, there is a substantial difference in value among goods in electronic marketplaces. The popular online electronic exchange site eBay, which holds 80% of the online auction market share, allows sellers to offer goods of widely varying values.[3] For example, with a search on eBay one can see an e-book for sale with a value of less than $1.00, as well as a house for sale with a value above $300,000. Evidence from eBay suggests that when goods of different value are available, sellers engage in strategic reputation building, selling many low value goods to build reputation and then defaulting on high value goods [4, 8]. Yet there is limited

---

[1] SP – Standard preference, MP – Medium preference, HP – High preference

[2] According to a press release by ComScore, a company that tracks the digital world, the popularity of e-commerce continues to increase rapidly. Black Friday (November 27) in 2009 saw $595 million in online sales, an increase of 11% as compared to Black Friday 2008. During the 2009 holiday season, over $10.57 billion dollars was spent online.
http://www.comscore.com/Press_Events/Press_Releases/2009/11/Black_Friday_Boasts_595_Million_in_U.S._Online_Holiday_Spending_Up_11_Percent_Versus_Year_Ago

[3] eBay Inc. reported approximately 647 million listings in the first quarter of 2008 and gross merchandise transaction volume of $16 billion in the third quarter of 2012. Figures are cited from eBay's financial release, available at http://investor.ebay.com/financial_releases.cfm.

understanding of the choices individuals make when faced with heterogeneous goods in an e-commerce market. This necessitates the development of a new framework for understanding reputation systems with heterogeneous goods, which will guide organizational decisions about which reputation systems to employ.

We conducted a laboratory experiment to study consumer-to-consumer transactions in a setting with heterogeneous goods. Laboratory experiments have been promoted as a valuable methodology in decision support systems (DSS) for understanding different market mechanisms and provide more control than data available in the field [26]. In our experiment, sellers choose to offer a high value or a low value good, buyers choose whether to purchase the good, sellers choose whether to send the good to the buyer, and in some treatments the reputation of the seller is automatically updated. We find that the market is less efficient when there is no reputation system because fewer high value goods are traded, but that efficiency is increased with a reputation system. Efficiency is defined as level of surplus that the participants of the market are able to extract relative to the total possible surplus. Successful transactions in high value goods are more profitable than transactions in low value goods, but traders are unable to take advantage of this when there is no reputation system.

The first reputation system we investigate is similar to that used in laboratory experiments with homogeneous goods. In a homogeneous good setting, this reputation system carries complete information about the seller's past behavior. We find that the reputation system increases efficiency, primarily by increasing successful transactions in high-valued goods and decreasing cheating behavior.

However, when heterogeneous goods are available, the information carried by the reputation system is reduced. Specifically, buyers cannot tell from the reputation system whether the seller previously transacted in high or low value goods. Therefore, we investigate a second reputation system that displays reputation separately for each type of good. Previous experimental work has ignored this issue; in fact, this question can only be addressed in a setting such as ours. We do not find a significant effect on efficiency as compared to the first reputation system. However, our results offer suggestive evidence that the additional information provided by this new system is used advantageously by buyers and sellers.

2. Material and Methods

2.1. Background

Previous work has emphasized the necessity of trust and reputation in the context of e-commerce applications in business-to-business [1, 46], business-to-consumer [6, 23, 38, 45], and consumer-to-consumer interactions [43, 50, 52]. This research is relevant for organizations that facilitate trade among agents in consumer-to-consumer e-commerce, such as eBay, Amazon Marketplace, and others. This research is also applicable to business-to-consumer e-commerce where users purchase products and rate quality (see [35] and [17] for an overview of the role that feedback plays). Reputation systems have been recognized as a way to increase trust [7, 37, 42, 51] signal information [41], reduce information asymmetry [14, 48, 50, 54], and act as a collaborative sanctioning system [35, 52].

Although previous experiments have studied the role of seller reputations with heterogeneous goods [11], all previous market experiments that have addressed reputation mechanisms explicitly employ a setting of homogeneous goods [7, 18, 19, 47, 48]. The general finding of related work is that the reputation system increases the number of successful transactions that take place, thereby unilaterally improving efficiency of the market. Many of these experiments analyze the value of various reputation mechanisms; however, this understanding may be limited in practical applicability because buyers and sellers may behave differently in the framework with heterogeneously valued goods.

Empirical evidence from eBay suggests that buyers and sellers behave differently when heterogeneous goods are available. In particular, sellers may engage in reputation building by selling many low value goods to increase reputation, and then selling high value goods and defaulting [4, 8]. eBay's recent changes to the cost structure made the practice of arranging for false transactions in order to gain reputation even cheaper and more feasible [20]. Anecdotal evidence suggests that some users of eBay know that this type of behavior is occurring. For example, a participant in a focus group commented, “For example I’ve seen people selling laptops with pretty high ratings and I go and read their feedback and the only thing they’ve sold is like clothes or very cheap goods, and they go for very high feedback” [30].

Diverse empirical findings have been reported for different types of goods, suggesting that different feedback systems must be designed to take such differences into account [39]. Empirical work has found that reputation matters in eBay auctions for several different types of homogeneous goods, such as computer processors and guitars [22, 33]. Other work has found that reputation matters for heterogeneous goods, such as rare coins of varying qualities [41]. Our contribution is an experiment that allows us to control many factors that cannot be controlled in empirical work, such as quality aspects of the good and communication between the buyer and seller that can occur online. In our experiment, we systematically change the types of goods available on the value/price dimension only. While empirical work can compare quality in “collector” auctions, such as for rare coins, even very experienced buyers may not always be fully aware of the quality of the good, or the quality may be perceived differently by different buyers. The experiment is the best way to abstract from differences in knowledge about quality for heterogeneous goods, but simply using goods that have a different experimental dollar value that is commonly known. The unique benefit of approaching this problem using an experimental methodology allows us to directly compare the effect of different reputation systems on decision-making. Moreover, the careful comparison of different reputation systems has been advanced as an important area of interest in DSS research [40].

The reputation mechanisms used in most e-commerce sites do not account for the value of previous transactions, which are only available in the detailed reputation. For example, on eBay, buyers must click through several pages to read details about transactions to find their value, and this is time consuming. It has been suggested that buyers may not look at the detailed reputation on eBay [48]. Empirical work has shown that buyers pay more attention to feedback when buying expensive products and less attention when buying cheaper products on eBay [48]. In addition, buyers pay more attention to feedback incurred as a seller than as a buyer [53]. Therefore, it is important to investigate whether detailed feedback is a necessary aspect of an effective reputation system.

2.2 Experimental Environment

The objective of this project is to investigate various reputation systems in a heterogeneous good setting. In the experiment, we consider a market with goods that are different on the value/price dimension. We also consider a case of buyer and seller transactions in which all individuals participate in both markets for high and low value goods.[4] The payoffs are chosen to assure that a sufficient volume of trade occurs in each type of good. The setup does not allow for price-setting for consistency with previous literature [7]. The decision tree is displayed in Figure 1.

In each period of our experiment, sellers are first asked whether they would like to offer for sale a high value or a low value item. Each seller is only allowed to offer one item. Then, buyers are shown a screen with each seller's offer and reputation score, and enter the market one by one to buy items (see Figure 2 for a screenshot of a buying screen). The buyers enter in a random order as determined by the computer in each period. When buyers enter the market, they are given the choice to buy one available item from any human seller, or to buy an item from the computer (the computer item is lowest value and acts as an outside option in case the buyer does not want to incur any risk). Once an item is purchased, the item is removed from the buying screen and no future buyer can purchase it. However, there are unlimited items offered by the computer. After all buyers make their decisions, sellers are given the choice of whether or not to send the good to the buyer. The computer always sends the item to the buyer. If no buyer bought his good, the seller skips and waits until the next period. At the end of the period, the actions of each participant and the participant's individual earnings are displayed on the outcome screen.

[FIGURE 1 ABOUT HERE]

[4] Note that this special case would be likely to occur in the following example. If a seller was transacting in the same *type* of items but sold different quantities of each of them- for example, selling a set of 10 or 1,000 miscellaneous postage stamps. Postage stamp collectors may choose to purchase the smaller set or the larger set, but all potential buyers would view both types of listings.

[FIGURE 2 ABOUT HERE]

Six sequences of nine periods each were conducted in every session. Using multiple sequences per session gives subjects time to learn and gain experience. The use of sequences was previously suggested in Camerer and Weigelt's [10] work. The computer automatically records a change in the seller's reputation, and automatically displays the seller's reputation scores on the buying screen in treatments with reputation. The reputation scores are cumulative for the duration of the sequence, but revert to the starting values at the start of each new sequence. Buyers do not have a reputation.

There are seven human sellers and four human buyers in each session. We chose this design for two reasons. First, we wanted to maximize the chance that there would be an excess supply of each type of item. In this way, buyers were not constrained to buying one item over another for reasons of lack of supply most of the time. Second, sellers essentially "drop out" of the market over time by reneging, as buyers no longer want to purchase items from these sellers. The most efficient outcome (which maximizes the buyer and seller's joint payoff) occurs when the seller offers the high value item, the buyer buys the item, and the seller sends the item. For example, total surplus for four successful high value transactions is 4x110=440 (this is calculated using the experiment parameters provided in Figure 1 – 4 transactions * (seller payoff of 70+ buyer payoff of 40). If instead, four transactions occur for the low value item, efficiency falls to 4x95=380 (using experiment parameters provided in Figure 1 – 4 transactions * (seller payoff of 60 + buyer payoff of 35), or 380/440=86% of efficiency.

2.3 Treatments

We consider three treatments, as summarized in Table 1: *No Reputation*, *Simple Reputation* and *Separate Reputation*. The *No Reputation* treatment serves as a baseline. In the case when there is no reputation system, the buyer does not have information about any seller's past actions. To measure the benefit of reputation, this treatment is compared with the *Simple Reputation* treatment, which uses a reputation system similar to one previously studied in the homogeneous good setting and generally used in e-commerce settings. The reputation mechanism carries information similar to the information each buyer would have in a partner matching setting. In the *Simple Reputation* treatment, reputation is updated in the same way regardless of whether the transaction was for a high or low value item. Specifically, the buyer sees two numbers – the first number indicates how many times the seller has sent an item and the second number indicates the percentage of times the seller has sent an item (number of times seller has sent an item, divided by the number of times an item has been bought from the seller, whether the seller sent it or not). eBay has a similar, but somewhat more complex, reputation system, called the "feedback market", where two reputation numbers are displayed based on buyers' voluntary rating of transactions as "positive," "negative," and "neutral." eBay's reputation system then updates a number indicating the seller's reputation (number of positives minus the number of negatives) and another number indicating the percentage of times the seller has received positive feedback (see [3]).

In the *Separate Reputation* treatment, reputation is updated separately for the high value good and low value good transactions. Buyers see separate numbers indicating outcomes of transactions in the high value market and the low market. Specifically, the buyer sees four numbers – the first two numbers indicate how many times and the percentage of times the seller has sent a high value item, and the second two numbers indicate how many times and the percentage of times the seller has sent a low value item. This treatment is compared to the *Simple Reputation* treatment to examine the effect of restoring information to a level comparable with information provided in experiments with homogeneous goods.

[TABLE 1 ABOUT HERE]

2.4 Experimental Procedures

The experiment was conducted at the Vernon Smith Experimental Economics Laboratory. Participants were recruited from a pool of undergraduate students from Purdue University. A total of 132 subjects participated in 12 sessions, with 44 subjects participating in each treatment and 11 subjects participating in each session. Subjects were randomly divided into groups of sellers or buyers, and remained in that designation for the entire session. Each subject participated in only one session of this study, although some subjects had participated in other economics experiments that were unrelated to this research. Upon arrival, subjects drew a number out of a bingo cage that determined which computer station they would be sitting at. Computers were designated as buyer/seller stations in advance.

The computerized experimental sessions were run using z-Tree [24]. Subjects were given the instructions at the beginning of each part and the experimenter read the instructions aloud. In the first part, subjects' risk attitudes were elicited using a multiple price list of 15 simple lotteries, see Appendix I [32]. The second part was the market experiment (instructions in Appendix II). Subjects also completed a quiz on understanding.

At the end of the experiment, 3 periods out of each sequence were randomly selected for payment using a bingo cage draw. Thus, subjects were paid for 18 total periods, with experimental dollars converted to U.S. dollars at the rate of 50 experimental dollars =$1. Subjects earned about $13 for their participation in the market, as well as a $5 show-up fee, $1 per correct answer on a quiz (up to $4 total) and risk elicitation task earnings ($0-$3). The duration of each session was approximately 1.5 hours.

3. Theory

3.1 Stage Game Equilibrium Predictions

The goal of this paper is to document buyer and seller behavior when trade in differently valued goods is possible. We do not aim to test alternative models; instead, we proceed by providing some basic predictions. Multiple equilibria exist in simpler settings than ours, including mixed strategy [10] and the pure strategy "full reputation" equilibria [31].

We first define the predictions in the stage game, and then proceed to document the effects of reputation. The stage game solution is similar to any round of the *No Reputation* treatment, because with random re-matching, buyers cannot update beliefs about the seller's probability to send. The theory assumes exogenous matching of sellers and buyers. Assume a buyer believes that the seller of a particular item will choose to send it (rather than renege) with probability p_H for the high value item and p_L for the low value item. The risk-neutral buyer will then purchase an item if and only if:[5]

$$p_H > p_H^* = [\ u(c)-u(b_H)\ ] / [\ u(a_H)-u(b_H)\ ] \tag{1}$$

$$p_L > p_L^* = [\ u(c)-u(b_L)\ ] / [\ u(a_L)-u(b_L)\ ] \tag{2}$$

Where a is the payoff for the buyer if he receives the item, b is the payoff for the buyer if he does not receive the item, c is the payoff for the buyer from the outside option, and $u()$ is the utility function of the buyer, $u(x)=x$. p_H^* and p_L^* are called the 'threshold' probabilities, and are known to both sides.

The seller, in turn, chooses a complete strategy including whether to offer the high or low value item and whether to send this item. When both threshold conditions are met ($p_H > p_H^*$ and $p_L > p_L^*$), the probability that a buyer buys either item is 1 and the seller maximizes his utility by offering high (low) when his utility from the high (low) payoff is greater than the low (high). When only one condition is met ($p_H > p_H^*$ or $p_L > p_L^*$), the seller maximizes his utility by offering the corresponding item (since items in a market with $p < p^*$ will result in no earnings for the seller). When neither condition is met, the seller is indifferent between offering high or low since the probability that the buyer will buy is 0 (resulting in no earnings for the seller in either case).

Literature distinguishes types of sellers between those who prefer to send and those who prefer not to send, separate from reputational concerns [10]. Types are the innate behavioral characteristics of each subject and range from strictly egoistic to strong welfare preferences [13].[6] We propose that sellers maximize the utility function $u_S(e_s,e_b)=e_s+ge_b$, where e_s is the seller's payoff, e_b is the buyer's payoff, and g is the weight that the seller places on the buyer's payoff. This utility formulation is similar to that proposed in related work (see [9], [12], [16]). We assume g to be uniformly distributed on the interval [0,1].[7]

---

[5] Note that the predictions of the model depend on the assumptions made, including the assumption on the form of the utility function of the buyer and seller.

[6] Note that while some related work induces cooperative sellers through changes in their payoff function, we rely on heterogeneity of subjects' social preferences to create enough trust to facilitate trade.

[7] The assumption on the uniform distribution of g is by design restrictive but allows us to provide explicit predictions for trade in high versus low value goods. Other distributions exist for which the results would hold, for

The range of preferences we describe could be classified as Standard Preference (SP), Medium Preference (MP), and High Preference (HP) types – these we defined based on the particular setup of our game and each seller's social preference parameter *g*, but can be generalized to other settings as well. HP sellers prefer to send both the high and low value good: $g>(y-x)/(a-b)$, where x is the payoff to the seller from sending the good (high or low) and y is the payoff from not sending. MP sellers prefer to send the low value good but renege on the high value good; $(y_H-x_H)/(a_H-b_H)>g>(y_L-x_L)/(a_L-b_L)$, and SP sellers prefer to renege on both the high and low value good; $g<(y-x)/(a-b)$. Intuitively, think of the HP types placing the greatest weight on their partner's payoff, and the SP types placing the lowest (or no) weight on their partner's payoff. Sellers know their own type, but buyers only know the distribution of types.

We now proceed by calculating the equilibrium solution using the experiment parameters. In the stage game, assuming that the buyer is risk neutral and that his payoff is exactly equal to earnings from the game, we can calculate the probability threshold requirements for each market as specified in (1) and (2), $p_H \geq 0.93$ and to be $p_L \geq 0.80$.[8]

Further, the assumption on the distribution of *g*, as well as the assumption on seller preferences for SP, MP and HP types, allow us to calculate the proportion of each type of seller in our market – we arrive at 20% SP, 8% MP, and 72% HP.[9] Figure 3 displays a number line indicating values of *g* and corresponding seller types for our experiment parameters.

[FIGURE 3 ABOUT HERE]

We can calculate the probability that a seller will send a high value item, $p_H = 0.72$ (the proportion of HP types) and the probability that a seller will send a low value item, $p_L = 0.80$ (the proportion of HP and MP types). Given that $p_H < p_H^*$ and $p_L > p_L^*$, the seller's dominant strategy in the stage game is {Offer Low, Send} for the HP and MP types, and {Offer Low, Don't Send} for SP types.

---

example Normal (0.25, 0.05) – $P_H = 0.27$ and $P_L=0.84$, Beta (2, 2) – $P_H=0.81$ and $P_L=0.90$, Beta (2,3) – $P_H=0.66$ and $P_L=0.82$.

[8] This simple calculation is done by assuming that the buyer's utility is $u(x)=x$, and solving $p_H > (c-b_H) / (a_H-b_H)$ and $p_L > (c-b_L) / (a_L-b_L)$ using the earnings parameters of the experiment that are provided in Figure 1. The buyer will only buy if he believes the probability the seller will send is greater than this calculated threshold.

[9] To do this calculation, we substitute the payoffs for the seller and the buyer in Figure 1 into (2) and calculate the threshold values for *g* for each preference type.

And the buyer's dominant strategy is: {Buy if Offered Low, Don't Buy if Offered High}. Thus, the pure strategy equilibrium solution using the experiment parameters is:

HP/MP types: {Offer Low, Buy, Send}

SP types: {Offer Low, Buy, Don't Send}

3.2 Long-run Equilibrium and the "Reputation" System

A reputation mechanism acts as a signal about the type of seller one is dealing with (one who prefers to send, or one who prefers to renege). Buyers can "punish" sellers who renege by refusing to purchase items from these sellers. A key characteristic of the framework with reputation is that sellers of the type who prefer not to send are aware that buyers may "punish," and act as if they prefer to send during the sequence in order to attract buyers in future periods. Once a seller reneges once, he reveals himself as a type who prefers not to send and buyers do not buy from this seller in future periods.

Under fixed matching, there exists a 'full reputation equilibrium' (as in [31]) in which the buyer always buys (as long as the seller has never reneged) and the seller always sends in every period but the last (regardless of seller type). In this equilibrium, the buyer's beliefs in any period are p_H^1 and p_L^1 (the belief in period 1) if the seller has always sent. However, if the seller has reneged on the low value good, he must be SP type and the buyer updates his belief $p_H^t = p_L^t = 0$. If the seller has reneged on the high value good, he must be SP or MP type and the buyer updates his belief $p_H^t=0$ and $p_L^t=0.29$. Assuming no time discounting, at any time *t,* buyers buy if:

$$p_H^t > p_H^{*t} = ([\, u(c)-u(b_H)\,] \,/\, [\, u(a_H)-u(b_H)\,])\text{\^{}}(t-1) \quad (3)$$

$$p_L^t > p_L^{*t} = ([\, u(c)-u(b_L)\,] \,/\, [\, u(a_L)-u(b_L)\,])\text{\^{}}(t-1) \quad (4)$$

Using our experiment parameters and $T = 9$ (T = number of periods), we arrive at the following threshold values for the high value good, 0.53 at t=1, 0.57 at t=2, 0.61 at t=3, 0.65 at t=4, 0.70 at t=5, 0.75 at t=6, 0.81 at t=7, 0.87 at t=8 and 0.93 at t=9 and low value good, 0.13 at t=1, 0.17 at t=2, 0.21 at t=3, 0.26 at t=4, 0.33 at t=5, 0.41 at t=6, 0.51 at t=7, 0.64 at t=8 and 0.80 at t=9. Since $p_H^{*1} = 0.53$ and

$p_L{*}^{1} = 0.13$, in all periods until $p_H^t > p_H{*}^t$ the seller chooses to send the item that generates the greatest utility for him, which is the high value item for all *g* as long as the probability of attracting a buyer is 1.[10] While there are numerous mixed-strategy equilibria (e.g., in periods after $p_H^t > p_H{*}^t$ the buyer begins mixing between *Buy, Don't Buy,* and the seller mixes between *Send, Don't Send* as in [10] or **Error! Reference source not found.**, we solve the pure strategy equilibrium and arrive at the following equilibrium strategies:

HP types: {Offer High, Buy, Send} in periods 1-5, {Offer Low, Buy, Send} in periods 6-9

MP types: {Offer High, Buy, Send} in periods 1-5, {Offer Low, Buy, Send} in periods 6-9

SP types: {Offer, High, Buy, Send} in periods 1-5, {Offer Low, Buy, Send} in periods 6-8, {Offer Low, Buy, Don't Send} in period 9

We do not see unraveling in this pure strategy equilibrium because buyer beliefs are not updated, buyers choose to buy the low value item with probability 1 in period *T*, and sellers have greater utility from sending in period *T-1* plus reneging in *T*, relative to reneging in period *T-1* and having no buyer in *T*.

3.3 Heterogeneous Goods and Information

In a setting with heterogeneous goods, the simple reputation system does not carry full information about the seller's past decisions. For example, sellers who renege in the high value market can still continue to attract buyers in the low value market if buyers believe that they are MP types. We provide some intuition for the differences between the two systems.

Under the simple reputation system, sellers signal their type with some uncertainty; that is, sellers with a good reputation score may either be MP or HP types. We introduce a new reputation system, which we suggest improves the value of the signal and reduces the potential for sellers to use successful low value transactions as a reputation-building tool for transactions in the high value market (as in [4, 8].

[10] Again, we can calculate this using the parameters from Figure 1, whereby 70+40g > 60+35g for all levels of *g* that apply to HP types, 150-250g > 60+35 g for all levels of *g* that apply to MP types, and 150-250g > 75-40g for all SP types.

Under the separate reputation system, a good reputation score in the high value market sends a clear signal to the buyer that the seller is an HP type, while a good reputation score in the low value market sends a signal to the buyer that the seller is at least an MP type. We propose that buyers will react to this signal by being more likely to buy a high value good from a seller with a good reputation in the high value market.

In a setting with heterogeneous goods, the seller who reneges in the high value market does not exit the experiment entirely as in settings with one good; rather, he may continue by offering low value goods and attracting buyers in the low value market if buyers believe he may be an MP type. The separate reputation system also assists buyers in identifying MP types after a seller renege. Under the simple reputation system, reneging in either market identifies the seller as either an SP or MP type. Under the separate reputation system, reneging in the high value market identifies the seller as either an SP or MP type, and reneging in the low value market identifies the seller as an SP type.

Suppose that under the simple reputation system a seller reneges in either market. This identifies the seller as an SP or MP type and no buyer will buy a high value good from this seller in the future. Since there is some probability that the seller is an MP type, buyers may still buy the low value good from this seller. Reneging in either market has the same result for the seller.

Suppose that under the separate reputation system a seller reneges in the high value market. This identifies the seller as an SP or MP type. Because the buyer knows with certainty that the renege occurred in the high value market, there is a greater likelihood that the buyer will buy the low value good after a renege in the high value market in separate reputations versus in simple reputations. Suppose that under the separate reputation system an SP type seller reneges in the low value market. This identifies the seller as an SP type, and no buyer will buy from this seller in the future. Under the separate reputation system, reneging in the low value market results in no future trade for the seller in either market.

4. Results and Discussion

Sub-sections 4.1-4.2 provide an analysis and discussion of the results from the *No Reputation* and *Simple Reputation* treatment. *Separate Reputation* treatment results are mostly relegated to the footnotes due to the similarities between these results and the *Simple Reputation* treatment results. Section 4.3 explores seller offer and buyer purchase decisions in more detail for all treatments. Section 4.4 provides a discussion of the suggestive differences between *Simple* and *Separate.* Finally, Section 4.5 introduces further discussion of the "end period effect."

4.1 Overview

The results address the question of whether transactions in the high versus low value good are affected in the same way by the introduction of reputation systems. We find that individuals respond to the reputation system differently for high as compared to low value goods. As predicted in Section 3, we find a greater number of successful transactions in the treatments with reputation relative to the treatment without reputation. Without a reputation system, most transactions occur in the low value market, while with a reputation system, most transactions occur in the high value market. This is due to both the increase in seller offers and buyer purchases of the high value good in treatments with reputation.

Figure 4 shows the number of high and low value goods offered by sellers in the *No Reputation* and *Simple Reputation* treatments. Under no reputation, the high value good is offered a total of 254 times (27% of all offers), while under *Simple Reputation*, the high value good is offered a total of 648 times (69% of all offers).[11]

[FIGURE 4 ABOUT HERE]

Figures 5 and 6 show the number of each good purchased and the proportion of goods purchased among those that were offered, respectively. A greater number of total goods are purchased from human sellers in the *Simple* reputation treatment relative to the *No Reputation* treatment (Wilcoxon Mann-

[11] Under *Separate* reputation, the high value good is offered 570 times (60% of all offers).

Whitney p-value < 0.01).[12] In aggregate, a greater proportion of available high value goods are bought in the *Simple* reputation treatment relative to the *No Reputation* treatment. However, a similar proportion of available low value goods are bought in both treatments. The high value good is bought 428 times (49% of all chances to buy) in the *Simple* reputation treatment, while it is bought only 26 times (9% of all chances to buy) in the *No Reputation* treatment. The low value good is bought 142 times (36% of all chances to buy) in the *Simple* reputation treatment, while it is bought 328 times (34% of all chances to buy) in the *No Reputation* treatment.[13] Despite the similar proportions, due to the level of offers, fewer low value goods are purchased in the *Simple* reputation treatment relative to the *No Reputation* treatment (Wilcoxon Mann-Whitney p-value < 0.01).[14]

[FIGURE 5 ABOUT HERE]

[FIGURE 6 ABOUT HERE]

Result 1: *Buyers are buying more goods from human sellers overall in the* Simple *treatment as compared to the* No Reputation *treatment.*

Result 2: *Sellers are offering and buyers are buying more of the high value good in the* Simple *treatment as compared to the* No Reputation *treatment.*

Figure 7 displays the proportion of reneges of all purchased goods. There is significantly more reneging in both goods in the *No Reputation* treatment (65 total reneges) than in the *Simple* treatment (26 total reneges) (Wilcoxon Mann-Whitney rank sum p-value = 0.00).[15] Note that the frequency of reneging

[12] We aggregated the number of times each buyer bought any good from a seller (not from computer seller) across sessions 2-6, and conducted a non-parametric Wilcoxon rank-sum test. When comparing *Simple* to *No Reputation* or *Separate* to *No Reputation,* we find statistically significant differences (p-values = 0.00 and 0.01, respectively).

[13] The high value good is bought 402 times (48%) and the low value good is bought 137 times (32%) in the *Separate* reputation treatment – these numbers are similar to the *Simple* reputation treatment.

[14] We aggregated the number of times each buyer bought the low value good across sessions 2-6, and conducted a non-parametric Wilcoxon rank-sum test. . When comparing *Simple* to *No Reputation* or *Separate* to *No Reputation,* we find statistically significant differences (both p-values = 0.01).

[15] We aggregated the number of reneges for each seller across sessions 2-6 separately for the high and low value good, and a Wilcoxon Mann-Whitney rank-sum test finds statistically significant differences in high value good reneges between *Separate* and *No Reputation* and *Simple* and *No Reputation* (p-values 0.02 and 0.00, respectively) and statistically significant differences in low value good reneges (both p-values 0.01). Aggregating by session, we find similar results – there are statistically significant differences between *Separate* and *No Reputation (p-values 0.00 and 0.05 for high and low, respectively) and between *Simple* and *No Reputation* (p-values 0.09 and 0.00, for high and low, respectively).

of high value goods looks highly volatile in the *No Reputation* treatment on the graph because very few high value goods are bought in this treatment.

Result 3: *There is greater reneging in the* No Reputation *treatment versus in the* Simple *treatment.*

[FIGURE 7 ABOUT HERE]

Figure 8 displays the efficiency of the *Simple* and *No Reputation* treatments. Efficiency is calculated as a proportion, taking the total earned in the market in any period and dividing this total by the maximum surplus that would have been achieved if all sellers offered the high value item, all buyers bought, and all sellers sent their items. Efficiency is significantly greater in the treatments with reputation as compared to the treatment without reputation (Wilcoxon rank-sum *p*-value 0.00 for both tests).[16]

Result 4: *There is significantly greater efficiency in the treatments with reputation as compared to the treatment without reputation.*

[FIGURE 8 ABOUT HERE]

We also find an "end-period effect" in seller offers, buyer purchases, and seller reneges in the *Simple* treatment. As evidenced by Figure 4, sellers substitute offers of high value goods for low value goods in later periods (while 73% of offers in Period 1 are for the high value good, this proportion drops to 58% in Period 9).[17] In the *No Reputation* treatment, this effect is not as great (31% of offers in Period 1 are for the high value good, and 25% of offers in Period 9 are for the high value good). As displayed by Figures 3 and 4, this "end period" effect exists for buyers as well. In *Simple,* 87% of buyers select an item offered by a human seller in period 1, compared to 8% of buyers in period 9. In *No Reputation*, 53% of buyers select an item offered by a human seller in period 1, compared to 42% of buyers in period 9.[18] Finally, there is a strong end-period effect of reneging in the last periods in the *Simple* treatment, which is

[16] We aggregated efficiency across each session for sessions 2-6 and conducted a Wilcoxon rank-sum test, efficiency is statistically significantly greater in *Separate* and *Simple* as compared to *No Reputation (*both *p*-values 0.00).
[17] In *Separate*, 75% of offers are high value in period 1, compared to 47% in period 9.
[18] In *Separate,* 57% of buyers select an item offered by a human seller in period 1, compared to 17% in period 9.

not observed in the *No Reputation* treatment (Wilcoxon Mann-Whitney rank-sum p-value = 0.01).[19] While the "end-period effect" has been observed in related work, the shift from offers of high value goods to offers of low value goods in final periods could only be investigated in our heterogeneous good design. Also, as evidenced by Figure 8, efficiency declines over time in the *Simple* treatment but does not have as great of a decline in the *No Reputation* treatment – a discussion of this result continues in section 4.4.

Result 4: *There is an "end-period effect" of fewer high value good offers, fewer overall purchases, and greater reneges in later periods (and specifically, in the last period) for the treatments with reputation.*

Result 4 suggests that "false reputation building" may be a good explanation for behavior. In the treatments with reputation, SP/MP type subjects build false reputations and then renege in the last period. Knowing this, buyers are less likely to purchase items from human sellers in the last period (note that this result is not supported by the equilibrium solution). In the treatment without reputation, subjects do not renege more in the last period than in any other period.

4.2 Detailed Analysis of Buyer and Seller Decisions

To investigate seller offer decisions further, we estimated a logit model with subject fixed effects for each treatment with seller offer decisions as the dependent variables, and the results are summarized in Table 2 (clustered on session). We use a dummy variable that is equal to 1 if the seller's reputation (percentage of items sent) is 100%. As predicted in Section 3, for the *Simple* and *Separate* treatments, the choice by sellers to offer a high value good decreases in periods 8 and 9.[20] Also, the choice to offer a high

[19] We aggregated reneging across sessions 2-6 for each subject both for the first 7 periods and for the "end periods", periods 8 and 9. Even considering the greater number of periods in the non-end-period group and the fewer goods bought in the end periods, there is statistically significantly more reneging in *Separate* and *Simple* in periods 8 and 9 as compared to all of periods 1-7 combined, a Wilcoxon Mann-Whitney rank-sum test finds statistically significant differences (p-values 0.06 and 0.01 for the high value good and p-values of 0.12 and 0.04 for the low value good, for respective treatments). No such statistically significant differences exist for the *No Reputation* treatment (p-values 0.48 and 0.56, for high and low value goods, respectively). Similar results are found when aggregating for each sequence rather than for each subject.

[20] We do not include reputation or period 8/9 dummies in specification 3. Including period 8 and 9 dummies in specification 3 results in qualitatively similar estimates for the remaining coefficients, while the period 8 and 9

value good in the last period is positively correlated with choosing to offer a high value good in the current period. Having a buyer in period *t*-1 increases the likelihood of offering a high value good in period *t* in the treatments with reputation, but actually has a negative effect on offering the high value good in the treatment without reputation. Also, having an "unblemished" reputation of 100% in the high value market results in the seller being more likely to offer the high value item, while having an "unblemished" reputation of 100% in the low value market results in the seller being less likely to offer the high value item in specification (1).

[TABLE 2 ABOUT HERE]

To investigate buyer purchase decisions, we estimated a multinomial logit model for each treatment with buyer's purchase choices as the dependent variables, and the results are summarized in Table 3. We include only instances where both a high and low value good were available to the buyer (the outside option was always available). End-period effects are observed in the treatments with reputation: buyers are less likely to purchase the high value good and more likely to purchase the outside option in periods 8 and 9.[21] The end period effect is discussed in more detail in section 4.4. We find that having a seller send an item in the previous period significantly increases the likelihood of buying a high or low value good in the current period. We also find that buyers are more likely to buy items from sellers with 100% send reputation.

[TABLE 3 ABOUT HERE]

4.3 Simple versus Separate Reputation

On aggregate, similar buyer behavior is observed in the *Separate* versus *Simple* treatments. No statistically significant differences in seller offer frequencies, buyer purchase decisions, or seller send

---

dummies are not statistically significant, which is in line with the prior prediction that there should not be an "end period effect."

[21] In fact, another probit model was run with the inclusion of Periods 8 and 9, and it was found that these are not statistically significant.

decisions are observed across the two treatments on aggregate.[22] The proportion of reneging was not statistically significantly different in either treatment.[23] Efficiency is not statistically significantly different across the two treatments (p-value 0.90).[24] Given that the simple reputation system was very effective at increasing successful transactions, and especially successful transactions in the high value market, it is possible that this is partially due to the lack of room for improvement (average efficiency across the 9 rounds was 42% under no reputation, and increased to 72% under simple reputation and 73% under separate reputation.)

Table 4 summarizes the reneging and subsequent ability to successfully transact for the treatments with reputation. Consistent with our predictions, attracting a future buyer following a renege is unlikely (overall, 82% of reneges resulted in no future buys). Buying in the low value market after a high value market renege is more likely than buying in the high value market after a renege. We also observe that attracting a future buyer following a renege is not very likely under either reputation system; however, following a renege in the low value market, several buyers continue to buy under the *Simple* reputation system but no buyers continue to buy under the *Separate* reputation system.

[TABLE 4 ABOUT HERE]

Figure 9 displays reneging percentages under each reputation system. Recall that sellers reneging on the low value good under the simple reputation system may be identified as an MP or SP type, while sellers reneging on the low value good under the separate reputation system will definitely be identified as an SP type. Sellers renege less on the low value good under the separate reputation system as compared to the simple reputation system, although these differences are minor (0% versus approximately 2% in

[22] We aggregated the number of times each seller offered the high value item across sessions 2-6, and conducted a non-parametric Wilcoxon rank-sum test. Comparing *Simple* to *Separate,* there is no statistically significant difference (p-value 0.34). In regards to buyer purchase decisions, comparing *Simple* to *Separate,* there is no statistically significant difference (p-value 0.70). No statistically significant differences in renege rates are found between *Simple* and *Separate* (p-values 0.39, high value; 0.79, low value). Similar results are found when aggregating for each sequence rather than for each subject.

[23] We aggregated the proportion of reneges in each treatment by sequence and found that there is not a statistically significant difference in renege proportions in the high value market (Wilcoxon rank-sum p-value 0.36) or in the low value market (p-value 0.99).

[24] A Wilcoxon rank-sum test using efficiency in each session finds that efficiency is not different between *Separate* and *Simple* (p-value 0.90).

periods 1-7). Our conjectures in section 3.3 suggest that under the separate reputation system it is never optimal to renege on the low value good in the penultimate period, and we find a 0% renege rate in period 8 in *Separate*. However, under the simple reputation system, the seller may renege on the low value good in period 8, believing that buyers may be uncertain about his type in period 9. Consistent with this, we find a renege rate of 16% in period 8 in *Simple.* Further analysis is not feasible due to the limited number of reneges in this study, and future work is needed to explore these issues further.

[FIGURE 9 ABOUT HERE]

4.4 Crowding out Trust

Since reputation no longer plays a role in the last period of the treatments with reputation, buyers can no longer utilize reputation to "punish" sellers who do not send the item. In this case, sellers who preferred to send continue to send, while sellers who preferred not to send under no reputation system renege. Thus, behavior in the last period should mimic the behavior under no reputation, where every period is like the last period of reputation treatments. However, we find that in the *Simple* and *Separate* treatments, the end period effect of reduced buying and increased reneging of high and low value goods results in lower efficiency in the later periods as compared to any period in the *No Reputation* treatment. This may suggest that the existence of the reputation system acts to "crowd out" inherent trust in sellers.

We conjecture that this finding is due to the change in perception that buyers have about sellers when a reputation system is present. For example, buyers in the treatment with a reputation system may believe that it is needed because most sellers defect, while buyers in the treatment without a reputation system may not hold this belief since the experiment instructions suggest that buyers can trade with sellers over the course of the experiment.

Changes in perception resulting from a change in institution have been observed in a field study in a daycare in Israel in which parents sometimes arrived late to collect their children, but imposing a fine on arriving late actually resulted in *more,* rather than *fewer,* late arrivals [27]. In this case, the imposition of the fine caused the perception to change – before the fine, parents felt guilty if they did not pick up their child on time, while after the fine, parents could pay a fine and the feeling of guilt was reduced. Similarly, the reputation system may have caused sellers to feel less guilt about not sending the item to the buyer, and may have caused the buyers to become less trusting of sellers.

This reaction to the reputation system can be partially explained by *trust responsiveness. Trust responsiveness* is a psychological mechanism explaining the tendency for trustees to fulfill trust because trustees do not want to “let down” the truster in the treatment without reputation [2, 21, 29]. Under a reputation system, *trust responsiveness* may be lower because trustees are more focused on maintaining reputation than on fulfilling beliefs of trust placed on them by sellers. Further, buyers may not place trust on sellers when a reputation system exists, instead using the reputation system as a contract of sorts.

This behavior has practical implications for understanding reputation systems. If the existence of the reputation system reduces inherent trust in sellers, it is even more important to assure that the reputation system is the most efficient possible. Further, removing a reputation system without instituting another may result in less trust than would have been observed initially without the reputation system. Future sessions in which individuals first participate in an experiment with a reputation system and then sequentially participate in an experiment without a reputation system could shed light on changes in perceptions of trustworthiness of sellers caused by changes in institution.

5. Conclusions

We study the functioning of reputation systems in heterogeneous good markets with anonymous agents who have the opportunity to cheat. Our finding is that when there is no reputation system, subjects do not trade sufficient quantities of the high value good. We also find that the reputation system affects transactions in high and low value goods differently. The reputation system is a valuable decision support system for both buyers and sellers, because efficiency is increased when a reputation system is present. Efficiency is increased for three reasons. First, a greater number of high value goods as compared to low value goods are offered under a reputation system as compared to under no reputation system. Second, a greater proportion of the high value goods are bought relative to low value goods under a reputation system. Third, less reneging occurs under a reputation system in both goods. Reneging results in lower efficiency in the current period. Reneging under a reputation system also leads to a lack of buyers in future periods for that seller, resulting in lower efficiency in subsequent periods.

These results suggest that reputation systems are especially important for organizations that derive profits from sales of high value or quality goods, as customers are more likely to buy high value goods as compared to low value goods when a reputation system is available. An important concern for e-commerce activity is customer retention [5, 49]. Because customers may exit the market after encountering a seller who defaulted, reputation systems that facilitate successful transactions are important for customer retention, especially for high valued goods.

The design aspects of reputation systems, such as amount and type of information to display, play a role in determining the success of organizations that conduct business in electronic exchanges [17, 35]. We introduced a new reputation system, which displays reputation separately for each type of good, thereby increasing information. We find that efficiency is unchanged as compared to the simple reputation system. However, there is suggestive evidence that the separate reputation system is a more valuable decision support than the simple reputation system. First, buyers do respond to the new reputation system, being more likely to purchase high value goods from sellers with 100% high value reputation, and being more likely to purchase low value goods from sellers with 100% low value reputation. Second, the separate reputation system improves the identification of seller types, and we find preliminary evidence indicating that buyers use the system.

An extension of this study is to set up new experiments in which reneging is more likely in both markets in order to further investigate differences between the simple and separate reputation systems. A further extension is to make feedback posting voluntary. It is expected that buyers, when offered the choice, may post feedback with different probability for a high value good as compared to a low value good. Further, buyers incur a small implicit cost (time and effort) when posting a feedback – future research can investigate the impact of this cost on voluntary feedback provision in a heterogeneous good setting. While the current study investigating heterogeneous goods in which goods differed on the value/price dimension, future work could also investigate differences in type of goods (e.g., books versus clothing). Future work could also investigate a context with different markets for buyers (e.g., some buyers who only purchase one type of good and other buyers who purchase the other type of good).

Acknowledgements:

I am grateful to Tim Cason for excellent guidance and support. I thank Jack Barron and Roman Sheremeta as well as seminar participants at Purdue University, participants at the June 2009 Economic Science Association meeting, and anonymous referees of this journal for helpful comments. I thank Andrew Simon and Amanda Chuan for valuable research assistance. I thank the International Foundation for Research in Experimental Economics and John Umbeck for generous funding of this project. Any remaining errors are mine.

References

[1] N. Archer, Y. Yuan, Managing business-to-business relationships throughout the e-commerce procurement life cycle, Internet Research: Electronic Networking Applications and Policy 10 (2000) 385-395.

[2] M. Bacharach, G. Guerra, D.J. Zizzo, The Self-Fulfilling Property of Trust: An Experimental study, Theory and Decision 63 (2007) 349-388.

[3] P. Bajari, A. Hortacsu, Economic insights from internet auctions, Journal of Economic Literature 42 (2004) 457-486.

[4] R. Bhattacharjee, A. Goel, Avoiding Ballot Stuffing in eBay-like Reputation Systems, SIGCOMM '05 Workshops, 2005.

[5] A. Bhattacherjee, An empirical analysis of the antecedents of electronic commerce service continuance, Decision Support Systems 32 (2001) 201-214.

[6] G. Bock, J. Lee, H. Kuan, J. Kim, The progression of online trust in multi-channel retailer context and the role of product uncertainty, Decision Support Systems 53 (2012) 97-107.

[7] G. Bolton, E. Katok, A. Ockefels, Cooperation among strangers with limited information about reputation, Journal of Public Economics 89 (2005) 1457-1468.

[8] J. Brown, J. Morgan, Reputation in online markets: The market for trust, California Management Review 49 (2006) 61-81.

[9] C. Camerer, Progress in Behavioral Game Theory, The Journal of Economic Perspectives 11 (1997) 167-188.

[10] C. Camerer, K. Weigelt, Experimental Tests of a Sequential Equilibrium Reputation Model, Econometrica 56 (1988) 1-36.

[11] T.N. Cason, L. Gangadharan, Environmental labeling and incomplete consumer information in laboratory markets, Journal of Environmental Economics and Management 43 (2002) 113-134.

[12] G. Charness, E. Haruvy, Altruism, equity, and reciprocity in a gift-exchange experiment: an encompassing approach, Games and Economic Behavior 40 (2002) 203-231.

[13] G. Charness, M. Rabin, Understanding social preferences with simple tests, Quarterly Journal of Economics 117 (2002) 817-869.

[14] C.H. Chua, J. Wareham, Self-regulation for online auctions: an analysis, Twenty-third international conference on information systems, 2002.

[15] B.J. Corbitt, T. Thanakankit, H. Yi, Trust and e-commerce: a study of consumer perceptions, Electronic Commerce Research and Applications 2 (2003) 203-215.
[16] M. Costa-Gomes, K.G. Zauner, Ultimatum Bargaining Behavior in Israel, Japan, Slovenia, and the United States: A Social Utility Analysis, Games and Economic Behavior 34 (2001) 238-269.
[17] C. Dellarocas, The digitization of word of mouth: Promise and challenges of online feedback mechanisms, Management Science 49 (2003) 1407-1424.
[18] C. Dellarocas, Reputation Mechanisms, in: T. Hendershott (Eds.), Handbook on Economics and Information Systems, Elsevier, Oxford, 2006, pp. 629-660.
[19] C. Dellarocas, M. Fan, C. Wood, Self-Interest, Reciprocity, and Participation in Online Reputation Systems, Center for eBusiness at MIT, Paper 205, 2004.
[20] F. Dini, G. Spagnolo, Buying reputation on eBay: Do recent changes help? International Journal of Electronic Business 7 (2009) 581-598.
[21] M. Dufwenberg, U. Gneezy, Measuring Beliefs in an Experimental Lost Wallet Game, Games and Economic Behavior 30 (2000) 163-182.
[22] D.H. Eaton, Valuing information: Evidence from guitar auctions on eBay, Journal of Applied Economics and Policy 23 (2005) 1-19.
[23] F.N. Egger, "Trust me, I'm an online vendor": Towards a model of trust for e-commerce system design, Conference on Human Factors in Computing Systems, 2000.
[24] U. Fischbacher, Z-tree: Zurich toolbox for ready-made economic experiments, Experimental Economics 10 (2007) 171-178.
[25] S. Ganeson, Determinants of long-term orientation in buyer-seller relationships, The Journal of Marketing 58 (1994) 1-19.
[26] C.L. Gardner, J.R. Marsden, D.E. Pingry, The design and use of laboratory experiments for DSS evaluation, Decision Support Systems 9 (1993) 369-379.
[27] U.R.I. Gneezy, A. Rustichini, A fine is a price, Journal of Legal Studies 29 (2000) 1-17.
[28] S. Grabner-Krauter, E.A. Kaluscha, Empirical research in on-line trust: a review and critical assessment, International Journal of Human-Computer Studies 58 (2003) 783-812.
[29] G. Guerra, J. Zizzo, Trust responsiveness and beliefs, Journal of Economic Behavior & Organization 55 (2004) 25-30.
[30] A.J. Haywood, Online auctions: User experience insights from eBay, Chimera Working Paper 2006-10, 2006.
[31] P.J. Healy, Group Reputations, Stereotypes, and Cooperation in a Repeated Labor Market, American Economic Review 97 (2007) 1751-1773.
[32] C.A. Holt, S.K. Laury, Risk Aversion and Incentive Effects, American Economic Review 92 (2002) 1644-1655.
[33] D. Houser, J. Wooders, Reputation in auctions: Theory, and evidence from eBay, Journal of Economics & Management Strategy 15 (2006) 353-369.
[34] S.L. Jarvenpaa, N. Tractinsky, M. Vitale, Consumer trust in an internet store, Information technology and management 1 (2000) 45-71.
[35] A. Josang, R. Ismail, C. Boyd, A survey of trust and reputation systems for online service provision, Decision Support Systems 43 (2007) 618-644.
[36] D.J. Kim, D.L. Ferrin, R. Rao, A trust-based consumer decision-making model in electronic commerce: The role of trust, perceived risk, and their antecedents, Decision Support Systems 44 (2008) 544-565.
[37] D. Kreps, R. Wilson, Reputation and Imperfect Information, Journal of Economic Theory 27 (1982) 253-279.
[38] M.K. Lee, E. Turban, A trust model for consumer internet shopping, International Journal of Electronic Commerce 6 (2001) 75-91.

[39] Z. Lee, I. Im, S.J. Lee, The effect of negative buyer feedback on prices in internet auction markets, Proceedings of the Twenty First International Conference on Information Systems, 2000.
[40] L. Liu, M. Munro, Systematic analysis of centralized online reputation systems, Decision Support Systems 52 (2012) 438-449.
[41] M.I. Melink, J. Alm, Seller Reputation, Information Signals, and Prices for Heterogeneous Coins on eBay, Southern Economic Journal 72 (2005) 305-328.
[42] P.A. Pavlou, S. Ba, Evidence of the Effect of Trust Building Technology in Electronic Markets: Price Premiums and Buyer Behavior, Management Information Systems Research Quarterly 26 (2000) 243-268.
[43] P.A. Pavlou, D. Gefen, Building effective online marketplaces with institution-based trust, Information Systems Research 15 (2004) 37-59.
[44] J.A. Quelch, L.R. Klein, The internet and international marketing, Sloan Management Review 37 (1996) 60-75.
[45] C. Ranganathan, S. Ganapathy, Key dimensions of business-to-consumer web sites, Information and Management 39 (2002) 457-465.
[46] P. Ratnasingham, K. Kumar, Trading partner trust in electronic commerce participation, Proceedings of the International Conference on Information Systems (ICIS), 2000.
[47] P. Resnick, R. Zeckhauser, Trust Among Strangers in Internet Transactions: Empirical Analysis of eBay's Reputation System, Advances in Applied Microeconomics 11 (2002) 127-157.
[48] P. Resnick, R. Zeckhauser, J. Swanson, K. Lockwood, The value of reputation on eBay: A controlled experiment, Experimental Economics 9 (2006) 79-101.
[49] S.S. Srinivasan, R. Anderson, K. Ponnavolu, Customer loyalty in e-commerce: an exploration of its antecedents and consequences, Journal of Retailing 78 (2002) 41-50.
[50] S.S. Standifird, Reputation and e-commerce: eBay auctions and the asymmetrical impact of positive and negative ratings, Journal of Management 27 (2001) 279-295.
[51] A. Ule, A. Schram, A. Riedl, T. Cason, Indirect Punishment and Generosity Toward Strangers, Science 326 (2009) 1701-4.
[52] G. Zacharia, P. Maes, Trust management through reputation mechanisms, Applied Artificial Intelligence 14 (2000) 881-907.
[53] J. Zhang, The roles of players and reputation: Evidence from eBay online auctions, Decision Support Systems 42 (2006) 1800-1818.
[54] M. Zhou, M. Dresner, R.J. Windle, Online reputation systems: Design and strategic practices, Decision Support Systems 44 (2008) 785-797.

Table 1 – Summary of Treatments

	Type of Reputation	Number of Sessions	Number of Subjects
No Reputation Treatment	None	4	44
Simple Reputation Treatment	Simple	4	44
Separate Reputation Treatment	Separate	4	44
Total:		12 sessions	132 subjects

Table 2: Logit – Seller's Offer Decision by Treatment

Dependent Variable: Seller's Offer [1 if high value decision]	**(1)** **Separate**	**(2)** **Simple**	**(3)** **No Reputation**
Period 8	-0.58*	-0.82**	
[1 if t=8]	(0.24)	(0.21)	
Period 9	-0.80**	-0.99**	
[1 if t=9]	(0.23)	(0.21)	
1/Sequence	0.43	-0.41	0.75**
[inverse of sequence order]	(0.28)	(0.24)	(0.25)
Decision Lag	1.53**	1.85**	1.86**
[1 if high value decision in t-1]	(0.18)	(0.15)	(0.17)
Had buyer lag	0.39*	0.27	-0.40*
[1 if had buyer in t-1]	(0.19)	(0.16)	(0.17)
Reputation 100 Dummy - High	0.94**	0.29	
[1 if reputation is 100% in high (all) goods]	(0.19)	(0.18)	
Reputation 100 Dummy - Low	-1.06**	(omitted)	
[1 if reputation is 100% in low goods]	(0.22)		
Observations	1120	1120	1140
# Subjects	28	28	28

Note: All results are from logit models with subject fixed effects clustered by session. Standard errors in parentheses. Asterisks indicate *** p<0.01, ** p<0.05, * p<0.10

Table 3: Multinomial Logit: Buyer Purchase Decisions by Treatment (Outside Option as Base Value)

Dependent Variable, Buyer's Purchase Decision	**(1)** **Separate Reputation**	**(2)** **Simple Reputation**	**(3)** **No Reputation**
[1 if high value]	**High Value**	**High Value**	**High Value**
Period 8	-4.92**	-5.27**	
[1 if t=8]	(1.06)	(1.30)	
Period 9	-20.83**	-18.93**	
[1 if t=9]	(1.23)	(0.44)	
1/Sequence	0.59	-0.41	2.10*
[inverse of seq. order]	(0.59)	(0.78)	(0.89)

Partner Coop. Lag	2.79**	0.04	0.32
[1 if received t-1]	(0.72)	(0.43)	(0.79)
Rep. 100 Dm. - High	36.82**	32.58**	
[1 if 100% high/all]	(1.79)	(0.55)	
Rep. 100 Dm. - Low	35.48**	(omitted)	
[1 if 100% low]	(1.34)		
Constant	-1.37	1.07	-3.75**
	(0.79)	(0.79)	(1.15)
[1 if low value]	**Low Value**	**Low Value**	**Low Value**
Period 8	-3.72**	-	
[1 if t=8]	(0.91)	(0.20)	
Period 9	-19.07**	-1.98**	
[1 if t=9]	(0.75)	(0.33)	
1/Sequence	0.41	-0.37	0.59
[inverse of seq. order]	(0.65)	(0.22)	(0.44)
Partner Coop. Lag	3.21**	0.55**	0.98*
[1 if received t-1]	(0.71)	(0.18)	(0.50)
Rep. 100 Dm. - High	34.66**	1.72**	
[1 if 100% high/all]	(1.84)	(0.16)	
Rep. 100 Dm. - Low	37.28**	(omitted)	
[1 if 100% low]	(0.63)		
Constant	-1.88*	-0.13	-0.99*
	(0.78)	(0.89)	(0.49)
Observations	568	611	443
# Subjects	16	16	16

Note: All results are from multinomial logit models with subject clustering. Standard errors in parentheses. Asterisks indicate ** p<0.01, * p<0.05. Abbreviations can be translated with the following: "Coop" – Cooperative, "Rep." – Reputation, "Dm" – dummy, "Avail." – available."

Table 4: Reneging Behavior and Frequency of Attracting a Future Buyer[25]

	Separate Reputation			***Simple Reputation***		
Find buyer for:	*high only*	*low only*	*neither*	*high only*	*low only*	*neither*
Ren. High Only	1	4	19	0	3	29
Ren. Low Only	0	0	0	1	2	3

[25] These frequencies are obtained by counting the number of cases where a renege occurred in a period before period 9, and then looking to see if a future buy occurred. In all cases, the future buy occurred only one other time in the sequence.

Figure 1 – Decision Tree for Each Period (Sellers first choose which to offer)

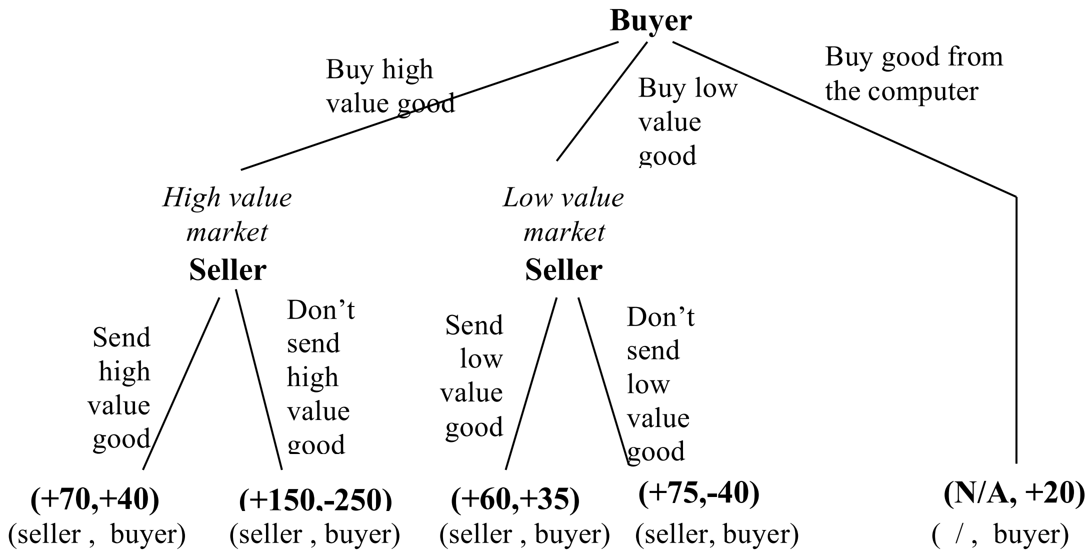


Figure 2: Buying Screen with 3 Purple "High Value" Items, 2 Orange "Low Value Items" and 2 Pink "Computer Seller" items on offer – 2 items already purchased

Figure 3: Values of Social Preference Parameter *g* and Seller Types

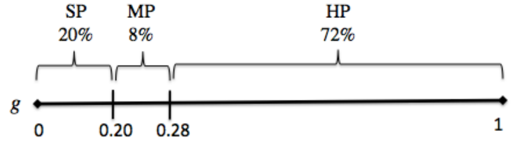


Figure 4: Seller's Choice of Good Offered over Periods 1-9, Sequences 2-6 Aggregated

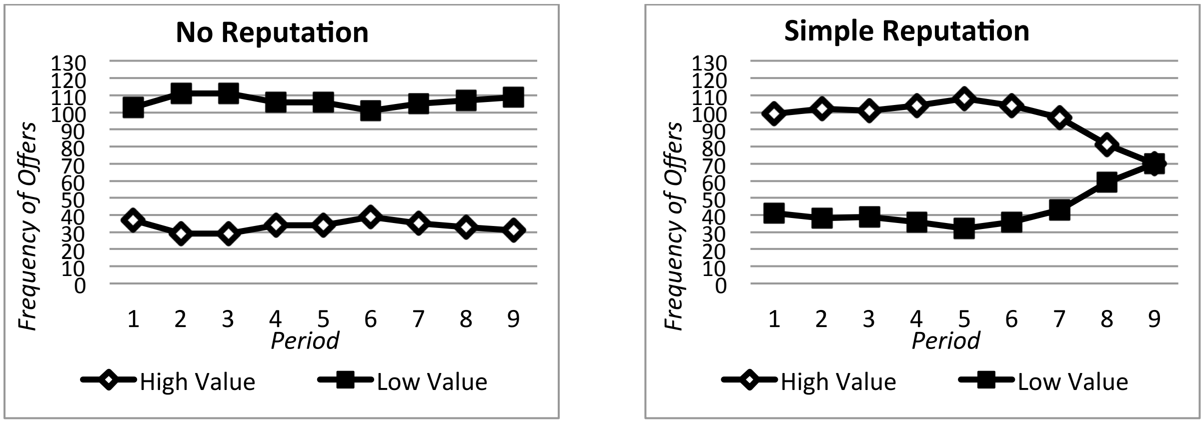


Figure 5: Buyer's Choice of Good over Periods 1-9, Sequence 2-6 Aggregated

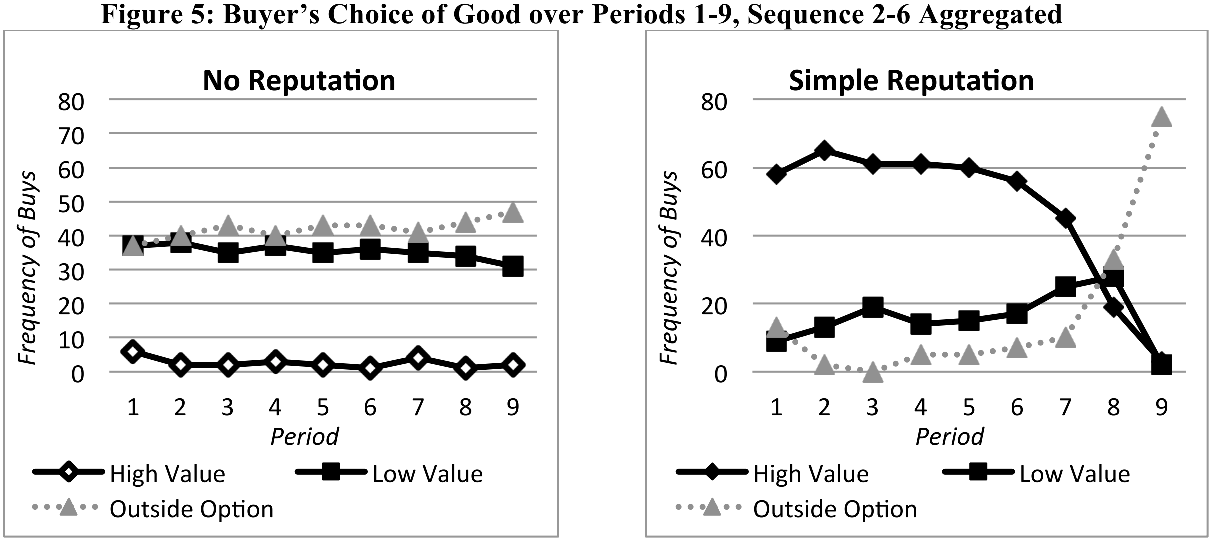


Figure 6: Percentage of Goods Bought of Offered, Periods 1-9, Sequences 2-6 Aggregated

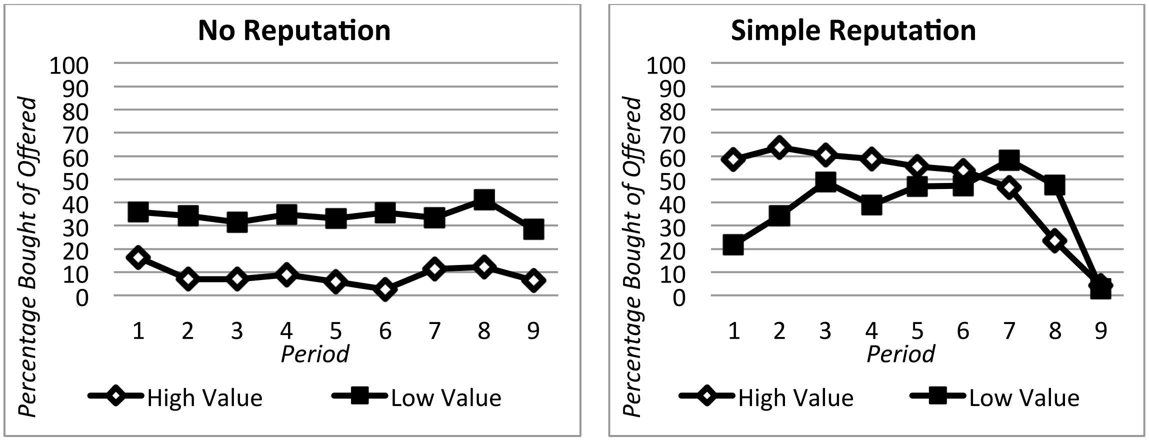


Figure 7: Reneging Percentages, Sequences 2-6 Aggregated

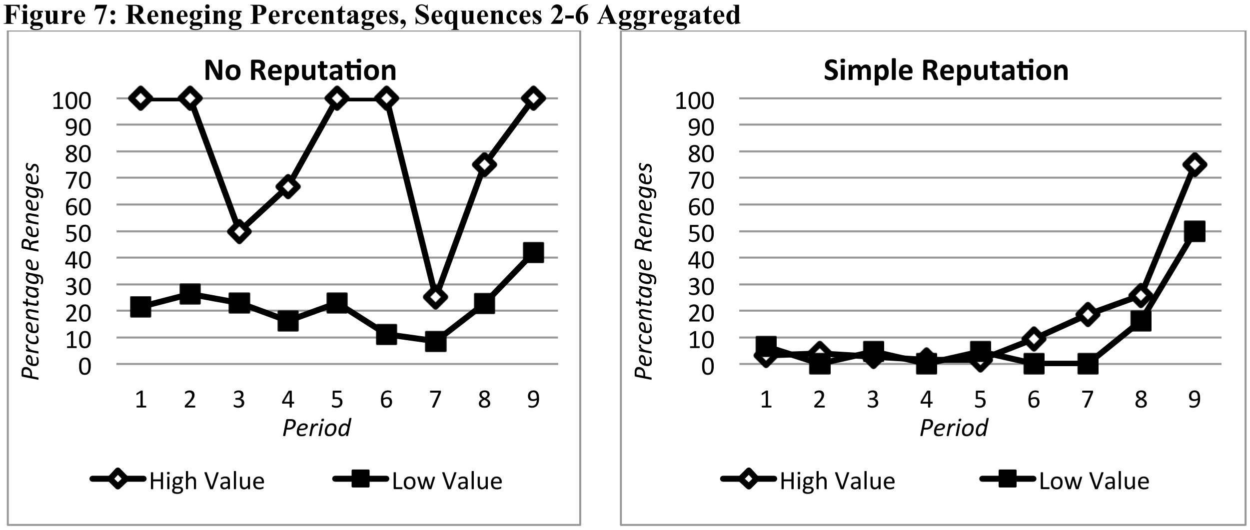
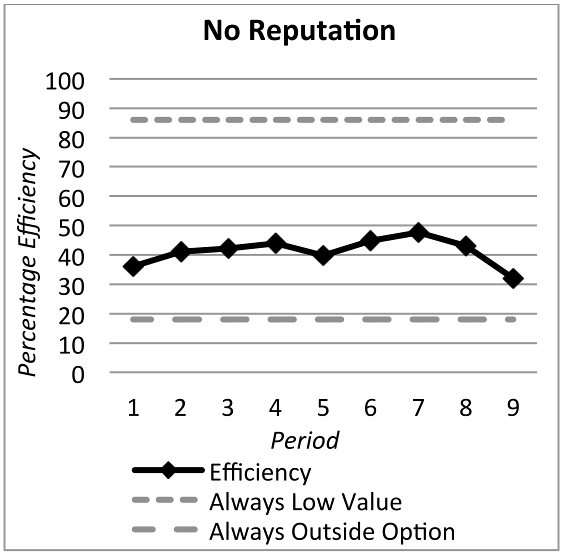


Figure 8: Efficiency, Periods 1-9, Sequences 2-6 Aggregate



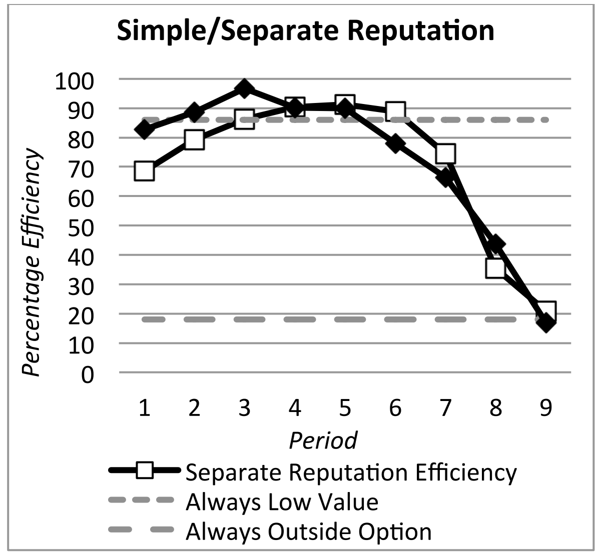


Figure 9: Reneging Percentages, Periods 1-9, Sequences 2-6 Aggregated