John F. Kennedy School of Government
Harvard University
Faculty Research Working Papers Series

The Value of Reputation on eBay: A Controlled Experiment

Paul Resnick, Richard Zeckhauser, John Swanson, and Kate Lockwood

July 2002

RWP03-007

The Value of Reputation on eBay: A Controlled Experiment

Paul Resnick, Richard Zeckhauser, John Swanson, and Kate Lockwood[1]
The latest version of this working paper can be found at http://www.si.umich.edu/~presnick/papers/postcards/
Initial Version: June 21, 2002
This Version: July 1: 2002

Abstract

Many empirical studies assess the effectiveness of reputation mechanisms, such as eBay's Feedback Forum. These investigations involve products ranging from pennies to collector guitars; they vary widely in their conclusions on how well reputation systems perform. Part of the explanation for the disparity among prior studies is that they merely collect samples from the eBay population. Such observational studies significantly increase the number of other variables that are left uncontrolled. This makes it difficult to isolate the effects of reputation on auction outcome.

In our main experiment, we worked with an established eBay auctioneer to sell matched pairs of items -- batches of vintage postcards -- under his extremely high reputation identity, and under newcomer identities with little reputation. Our second experiment followed the same format, but compared sales under newcomer identities with and without negative feedback. Having controlled the content of the auctions, and the presentation of item information, we were able to minimize the effects of variables other than reputation. As expected, the established identity fared better. The price difference was 7.6% of the selling price. Back-of-the-envelope calculations indicate that this amount is reasonable, given the level of risk that buyers incur. Surprisingly, one or two negative feedbacks for our new IDs had no price effects, even though these sellers had few positives.

Introduction

As the Internet grows as a means of executing transactions, each buyer's array of possible sellers is mushrooming. On auction sites, like eBay, users already buy and sell things from others across the nation and around the world. Despite opening many new venues, this electronic bazaar stresses some of the foundations of the traditional market place. Any economic transaction involves some amount of information asymmetry. In traditional markets, a buyer can usually "squeeze the orange", e.g., inspect the vintage plate, before buying. Beyond this, the hostage of a seller's reputation, possibly built over many years, including the cost of her physical facility and her standing in the community, together with repeat transactions, keeps her honest and diligent.

Sales over the Internet lack these tools of reputation. No seller has long been in the electronic market. A seller's physical facility may be her kitchen, and virtually no buyer knows a seller's standing in the community. To be sure, some sellers, such as Dell and L.L. Bean, borrow reputations from elsewhere. However, for tens of thousands of sellers,

[1] Resnick and Lockwood, School of Information, University of Michigan; Zeckhauser, Kennedy School, Harvard University; Swanson, Johnninaswanson, eBay, and Mission Viejo, California.

there is no outside instrument of reputation, and there are few repeat transactions with individual sellers. In such circumstances, the temptation to sellers to misrepresent products, e.g., exaggerate their quality or misrepresent their provenance, are great. So too is the temptation to sloth, to ship slowly or sloppily after receiving payment.[2] This necessarily lowers the price that the buyer is willing to pay, since they are forced to assume some risk for the quality and utility of the good being traded. Unless sellers can provide sufficient information about product quality and their own quality in transaction fulfillment, low-quality products and sellers will drive out those of high quality and the market will shrivel (Akerlof, 1970).

The two primary ways this problem is overcome in the "real world" are standardization and reputation. Standardized products reduce the difference between the buyer's and seller's knowledge of product quality: a Big Mac is a Big Mac, and a SONY digital camera is a SONY digital camera whether you purchase one in Atlanta or in Anchorage. The advantages of standardization, which are communicated through brand reputation, are so powerful, that chain restaurants have almost driven mom-and-pop eateries off of the major highways crossing America. Reputation makes it possible to trust that a seller has revealed important information about the product and that the seller will perform well in the fulfillment phase.

Online, standardization is also useful, but only works for products that are standardized and known by buyers to be so. Seller reputation, too, can be quite useful, but it does not work in its accustomed way in the physical world. First, with many many times more possible transaction partners, repeat transactions are less likely and reputation can't wait to build from first-hand experience. Your corner store has a strong interest in keeping you coming back. Internet sellers, by contrast, could flourish for a period without repeat business, given the vast scale of the marketplace. They could even build a reputation quickly, with the intent of cashing it out. This gives Internet sellers less of an incentive to be trustworthy in the present transaction. Second, word of mouth spread of reputation is less effective on the Internet. Unlike your local mechanic who has been recommended by your neighbor, it is much harder to "ask around" to determine the reliability of an Internet seller.

Internet players have struggled to find a substitute for traditional seller reputations. Important systems have been introduced to enable the systematic elicitation and distribution of reputational information. These systems collect information on the past behavior of a seller, or for that matter of a buyer, and then make that information available to potential future transaction partners. Because people know that their behavior now will affect their ability to transact in the future, not only with their current partner but with unknown others as well, they are less likely to engage in opportunistic behavior. Moreover, less reliable players will be discouraged from joining the marketplace. The purpose of reputation systems is to inform buyers about whether

[2] At eBay and most other on-line auction sites, the norm is for the buyer to send payment first, then for the seller to send the good. Escrow services are available to withhold payment until after acceptance of the shipment, but they are rarely used.

potential trading partners are trustworthy, and thereby to make chiseling and cheating rare and losing propositions.

Though disadvantaged in the respects described above, Internet markets have some significant advantages in establishing reputations. First, any information that is gleaned can be near-costlessly tallied on a continuing basis, and written assessments can readily be assembled. Second, that information can be costlessly transmitted to millions of potential customers. (By contrast, word of mouth distribution loses vasts amounts of information, with different buyers having significantly different assessments of the same seller.) It also entails a per-telling cost. Third, the Internet has the potential, though at present not the reality, for sophisticated processing of information, e.g., using Bayesian calculations, and for using micropayments to induce careful and honest assessments from transactors (Avery, Resnick and Zeckhauser, 1999).

The eBay Reputation System

There are many sites with reputation systems of some sort. The eBay system is undoubtedly the biggest and best known. On eBay, there are millions of items available to bid on at any time. The eBay reputation system enables users to leave feedback about interactions with each other. The system is transaction based, so that in order to leave feedback for each other, two users must actually have completed an auction. After the auction ends, the buyer and seller each have the opportunity to rate each other's performance with either a 1 (positive), a 0 (neutral), or a –1 (negative). Users also have the opportunity to leave a text comment and rated individuals can respond to comments that they feel were unfair. Users' net reputation scores are calculated as the count of distinct users who gave positive feedback minus the count of those who gave negative feedback. The seller's net reputation score -- positives less negatives -- is automatically displayed on the auction page for each item he lists. Hence, potential buyers see this rating before bidding. A buyer can choose to click on the net score in order to see a more detailed break down into positive, negative, and neutral over a series of time periods, as shown in Figure 1. The buyer can then scroll down to see individual comments, with the

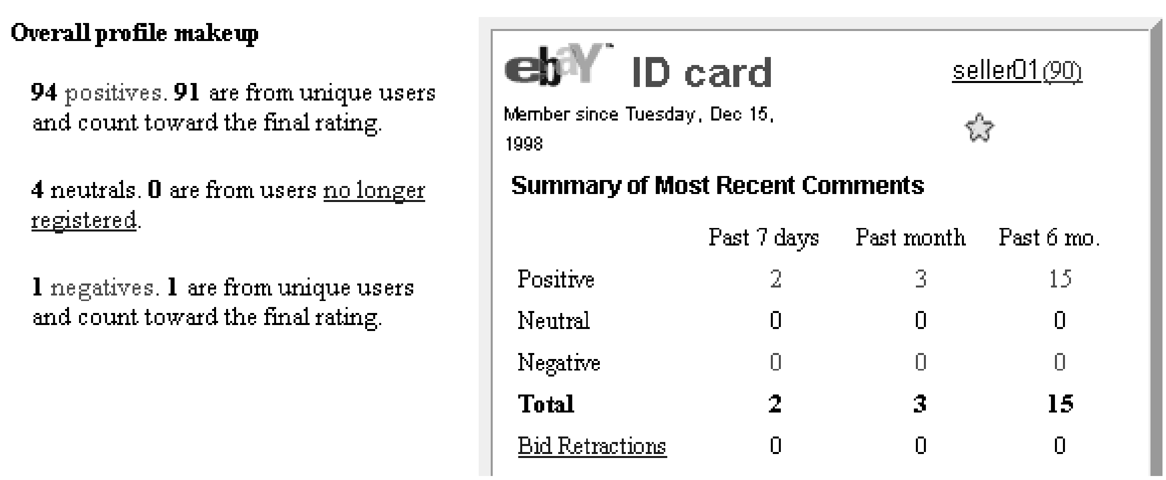

Figure 1: A sample summary of ebay feedback for a seller (displayed after click on reputation score)

most recent ones shown first. There is no search program to find negatives, however. A user who is new to the system starts with a net feedback score of zero and has a sunglasses icon displayed next to his or her screen name for the first 30 days of membership. Users may change their eBay identities by registering again, but must then start all over as new users with a zero reputation score.

Expectations About Feedback and Reputations on eBay

Game theory and the economics of information would have clear implications about the nature of feedback behavior and the returns to reputations on eBay. Our focus in this analysis is on sellers, since buyer reputations matter little. The seller can simply wait to get paid. If buyers are uncertain about seller trustworthiness, they will reward better seller reputations by raising their offers, even though each buyer is only concerned about his own welfare. Indeed, if it is costly to maintain a reputation for high quality, then a good reputation needs to be rewarded by at least the cost of building one. A bad reputation or a decline in reputation should incur a loss that exceeds the benefit from opportunistic behavior (Shapiro, 1983). Thus, in equilibrium, a good reputation must command a price premium.[3] Since sellers who get negative feedback can start over relatively easily, posing as genuine newcomers, to provide incentives for veteran sellers to protect their good reputations, buyers need to impose some penalty on sellers with no feedback at all (Friedman and Resnick, 2001).[4] Finally, no buyer will provide information to help determine seller reputations, since to do so incurs a cost, and free riding is hard to punish.[5]

Evidence to date indicates that eBay reputations do not illustrate pure rational game-theoretic processes in action. Resnick and Zeckhauser (2001), hereafter RZ, found that even though the incentive to free-ride is clear, half of the buyers on eBay, and three fifths of the sellers, provided feedback. This suggests that a high level of courtesy pertains on eBay. After a satisfactory transaction, you provide a relatively low cost positive feedback just the way you provide a thank you in everyday discourse.

A striking feature about eBay feedback is that it is so positive. Sellers received negative feedback only 1% of the time, and buyers 2% (RZ). (It seems unlikely that transactions could be quite this favorable, which presents a puzzle to be explained.) Given their rarity, negatives should be much more consequential than positives in affecting a seller's overall reputation.

---

[3] Brick and mortar retailers may be rewarded in part through esteem in the community, but such rewards seem unlikely on eBay. Conceivably such rewards such as self esteem or adherence to internalized ethics could lead to good behavior despite insufficient direct economic rewards.
[4] The penalty could also be imposed by the system designer; e.g., eBay could charge sellers who develop bad reputations. At present, apart from fraud situations, eBay imposes no penalties.
[5] If buyers think that sellers can develop reputations as reciprocators, then they could provide positive feedback to get a positive feedback of their own, to bolster their own reputations. However, there is no need for a buyer to have a reputation, unless they too are sellers. EBay merely adds together feedback secured as a buyer and as a seller.

For a seller, what constitutes a good reputation in the eBay feedback system? The answer depends on how buyers behave. There are many possibilities. At one extreme, buyers may in effect be Bayesians, effectively incorporating information not only from reputation scores but from a seller's product, geographic location, written comments, whether they have an expensive website, etc. Such buyers would be to statistics as Moliere's Monsieur Jourdain was to prose, unknowing users.

At the opposite extreme, buyers may employ simple heuristics that are far from optimal, as human decision makers are known to do (Tversky and Kahneman, 1974). Thus, a fraction of buyers may merely trust newcomers, assuming that new sellers have the same reliability as old, until proven otherwise. Some buyers may just worry about net reputation scores and rarely click through to even get their breakdown. Buyers who do click through may employ crude and biased algorithms for aggregating positives and negatives to judge seller reliability.

We conducted a controlled field experiment, detailed below, to determine how eBay seller reputations work in practice. In judging the results, we shall be concerned about understanding the nature of buyer behavior. Are they rational game theorists, do they merely follow heuristics in the spirit of behavioral decision, or do they accord with social norms? We have already tipped our results: as a group, buyers engage in some element of each.

Prior Empirical Studies

A large number of empirical studies of the effects of eBay's reputation system on sales have been undertaken in the last few years. We are aware of twelve, as summarized in Table 1. All follow a similar logic, though the details vary in important ways. Apart from one laboratory experiment, each is an observational study of a particular category of items.[6] Mean prices for the items studied ranged from $32.73 to $1620.93. These are extremely high-priced items for eBay. Informal study shows that the median item selling in 1999 went for less than $15.

The studies either identify or create a set of items whose sellers had varying reputations and correlates the reputations with auction outcomes, while controlling for possible confounds. The results do not produce a clear picture. Some suggest that negative feedback is more important, others that positive feedback plays the salient role. Some suggest that the effects on sale price are non-existent or tiny. At the larger end of an effect size for positive evaluations, HW finds that for their sample of Pentium chips a move from from 0 to 15 positive evaluations raises the price by about 5%, or $12. LBPD, looking at collectible coins, finds that a move from 2 to 3 negatives cuts the price by 11%, about $19 from a mean price of $173.20.

---

[6] There are other controlled studies of eBay auctions, but none that we know of that focus on reputations. Thus Katkar and Lucking-Reiley (2000) is a laboratory experiment, but it focuses on the effect of reserve prices on final prices.

Initials or shorthand	Citation	Items Sold	Mean price	Remarks
HW	(Houser and Wooders, 2000)	Pentium chips	$244.40	Positive feedback increases price; negative feedback reduces it
LBPD	(Lucking-Reiley et al, 2000)	Coins	$173.20	No effect from positive feedback; negative feedback reduces price
Eaton	(Eaton, 2002)	Electric guitars	$1620.93	Negative feedback reduces probability of sale, but not price of sold items
LIL	(Lee, Im, and Lee, 2000)	Computer monitors and printers	Not given	Negative feedback reduces price, but only for used items.
BP	(Ba and Pavlou, forthcoming)	Music, software, electronics	$232.30	Online laboratory experiment in the field: subjects responded with trust level and estimated price premiums for auction listings with different feedback profiles spliced in. Positive feedback increased estimated price, but negative feedback did not have an effect.
KW	(Kauffman and Wood, 2000)	Coins	Not given	No significant effects, but negative feedback seems to increase price (!) in univariate analysis
BH	(Bajari and Hortascu, 2000)	Coins	$47	Both positive and negative feedback affect probability of modeled buyer entry into the auction, but only positive feedback had a significant effect on final price.
KM	(Kalyanam and McIntyre, 2001)	Palm Pilot PDAs	$237.93	Positive feedback increases price; negative feedback reduces price
RZ	(Resnick and Zeckhauser, 2001)	MP3 players, Beanie Babies	Not given	Both forms of feedback affect probability of sale but not price contingent on sale.
MS	(McDonald and Slawson, 2000)	Dolls	$208.36	Higher net score (positives – negatives) increases price

MA	(Melnik and Alm, forthcoming)	Gold coins	$32.73	Positive feedback increases price; negative feedback decreases price
DH	(Dewan and Hsu, 2001)	Collectible stamps	$36.56	Higher net score increases price

Table 1. Summary of related empirical studies

Almost all these studies are in working paper form. Their results are highly inconsistent, producing vastly disparate findings on the main question, the effects of reputations. Moreover, a few of the studies have serious flaws in either methods or analysis. Thus, it would be impossible at this time to conduct a highly informative meta-analysis, or to try to develop a single theory that accounts for all the results. Given that the results clash, some are likely to prove spurious. It is useful, however, to explore the design space of such experiments, to understand the power and limitations of work to date, to frame our own experiment, and to help serve as a guide for future work.

While most of the studies explored the impacts of reputation on selling prices, some looked instead or in addition to the impacts on other relevant factors. For example, reputations affect the probability of sale (Eaton, RZ), the probability of individual bidders entering the auction (BH), and the number of bids (BH, MS). Seller reputation also affects potential buyers' subjective assessments of the trustworthiness of the seller (BP). In theory, such subjective trustworthiness assessments should help determine whether various buyers enter the auction and how high they are willing to bid. Hence, such assessments influence the probability of sale and price if sold.

These studies all control in some way for variability in underlying product values. Without such controls, as LBPD demonstrate, missing variable bias can produce misleading results, including positive reputation scores that appear to lower prices while negative reputations raise them. Two ways have been used to control for value of the product, so as to correct this missing variable bias. One is to study auctions of identical products, such as computer processors (HW), specific collectible items (MS, RZ), or coins whose value comes from their gold content rather than their collectible value (MA). Extending that idea, hedonic regression seeks to control for the effect of a well-defined feature set among non-identical items, an approach taken by (LIL). An alternative approach is to include book value or market price in regressions to control for the differences in item values (LR, BH, DH, BP, KM).

Reputations may matter more for some types of products than others. Generally, transactions that involve expensive items, less standardized items, and used items are riskier than transactions of inexpensive, new, and highly standardized items. However, for expensive items, eBay payment insurance and credit card insurance reduce the risks to buyers. It's been expected that reputation matters more for riskier items and this was mostly supported by the results. BP showed worse effects of negative feedback on the price for high-valued items; LIL showed negatives only had a significant effect on used/refurbished items. On the other hand, KM showed that positives increased price significantly only for new items.

Previous studies specify reputation in their regression models in one of three ways. Some count only negatives (LIL, KW), or only net score (MS, DH), the count of positives minus the count of negatives. The latter heavily weights the effects of positives, since negative feedback is less than 1% of the total at eBay (RZ). Including only negative feedback in a regression model creates the danger of attributing positive effects on price to negative ratings, since positive and negative feedbacks tend to increase with the number of sales. Including only positives produces less bias, though it overlooks the independent effects of negative feedback.

Most studies include positives and negatives separately, as continuous variables. KM specifies the percentage of negatives. It seems unlikely that one additional positive evaluation has the same effect for someone with 400 prior positive feedbacks as for someone with no prior feedback. It seems more reasonable to expect that doubling one's feedback count would have the same effect anywhere in the scale, so it is common to specify the logarithms of the positive and negative counts (BP, BH, HW, LBPD, RZ, MS).

An alternative to the log transformation is to partition the observations into reputation groups according to the overall reputations and consider the effects within each group. Eaton divides observations into low, medium, and high reputation groups by the absolute number of seller's feedback ratings (<40, 40-100, and >100). LIL partitioned observations into two groups, with a cutoff of 55. The results show a significant effect of negative feedback for used/refurbished items in the large evaluation group but not in the smaller groups.

The results of these studies differ about whether positives or negatives, or neither, or both matter. BP finds that only positives have a significant effect on prices paid while the negatives do not matter significantly. BH shows that both positives and negatives matter for auction entry, with the negatives having a stronger effect, and that only overall reputations, essentially the positives, affect the high bid. KM find that not only the overall reputations but also the percentage of negatives significantly affect the final price. LBPD finds that negatives, but not the overall reputations, have a significant effect on price. RZ found that neither mattered for price of sold items but both mattered for the probability of sale.

Many auctions do not result in transactions, and frequently items do not even receive bids. We expect that reputation will affect probability of sale as well as price. If we truncate the sample by selecting only sold items or censor the outcome variable by treating the minimum bid as selling price, but utilize an Ordinary Least Squares regression, we will underestimate the effect of a positive factor, such as a strong reputation. The intuition is straightforward. The weaker is this positive factor, the more exceptional the positive error will need to be to get included in the sample. Examining only sold items will thus disproportionately include strong upside outliers when reputation is weak, leading to a biased underestimate of reputation benefits.

The papers reviewed have employed a variety of techniques to overcome this bias. BH, LBPD, and MA model with Tobit or censored normal regression to correct for it. However, traditional censored regression models are somewhat problematic as correctors, because the items that sell for their minimum bid effectively pile up at that value. By contrast, traditional Tobit and censored normal analyses assume that the uncensored values are normally distributed. We take up the censoring problem again in the methods section.

In addition to price, there are many other possible confounds that should be controlled. Many studies control for whether a picture of the item is displayed as part of the auction and whether credit cards are accepted. Shipping cost may differ depending on the shipping policies and item value. Since in most cases the buyer pays for shipping, the shipping costs may affect bidding, especially for inexpensive items where shipping costs account for a large proportion of the total expense. Date of auction closing, time of day that the auction closes, and length of the auction also get controlled for (LBPD, MA, LIL, HW, DH).[7] Lastly, the completeness and quality of the description page may affect bidding. This is especially troubling, since it may be correlated with positive feedback: as a seller gains more experience, he tends to acquire both positive feedback and skill at listing items in a way that will appeal to buyers. Unfortunately, there is no clear way to measure and control for the quality of the description.

Many papers include the number of bids as a control variable in regression models. However, LBPD argues, and we agree, that the number of bids is an endogenous indicator of the impact of reputation on price and should not be an independent variable in a simple regression model. Number of bids or bidders can be treated as an outcome variable (BH), or may be modeled simultaneously as both an effect of reputation and independent contributor to price (MS).

The observational studies surveyed here rely on being natural quasi-experiments. There is always a danger that unknown or otherwise unmeasured covariates of reputation are the real cause of outcome differences that will be mistakenly attributed to the reputation score. For example, the responsiveness of sellers to e-mails may be the real reason for buyers' willingness to bid high, but it is beyond our measurement. In a controlled experiment, the potential covariates are under the control of the experimenter, so that the effect of reputation can be clearly attributed to reputation alone. BP secured the benefits of a controlled trial by presenting subjects with actual eBay listings spliced together with fake feedback profiles. They then asked subjects about their trust in the seller and the price premium they would pay. This allows for comparison across feedback profiles while holding the item presentation constant. This gain must be balanced against some sacrifice in external validity, since the controlled laboratory condition differs from the real world. The actual behavior of real bidders who are interested in acquiring the items is not observed.

[7] In some sense, such controls argue that the seller may not be maximizing, or that the goods sold have other intrinsic differences, e.g., should have different length auctions.

A controlled field experiment entails considerably more pains to secure data, but combines the best of both worlds in terms of interpretation of results. The experimenter can manipulate what is presented to subjects, carefully varying features on some dimensions while holding others constant, yet getting reactions from subjects in the natural setting of interest.[8]

Methods

Procedures

We worked with an established eBay seller with a high reputation (net score above 2000 as of the beginning of the study, with just one negative). In real life, that reputation belongs to coauthor John Swanson, who runs a business with Nina Swanson dealing primarily in vintage postcards. They typically lists dozens of items for sale each week on eBay under the ID johnninaswanson, and also sells items live at postcard shows and other events.

In addition to selling items using his established ID, our established seller -- hereafter often ES -- created seven new eBay IDs, each starting with no feedback. ES created matched pairs of items, to be sold under different IDs. He provided the same great service (communication, packaging, shipping) when listing items under any of the IDs, but buyers looking only at the item listings and seller information seen on eBay would not have known that the same dealer was behind all the IDs. At the completion of each auction, detailed information was collected about the bids placed, the selling price of the items, and the feedback of both the buyer and the seller at the time of the auction.

Our primary experiment, lasting twelve weeks, was the first stage. It compared sales of 200 matched pairs under ES's established ID, johnninaswanson, to sales under the new IDs. Information on results was gathered directly from the eBay webpage, using a spider to collect data. This data was then double checked against the records kept by our seller.

Since the postcards were not standard items, the matching, which was done on both subject and value, required the dealer's judgment. This led to a further precaution against possible bias: a random device determined which item in each pair was sold under the ES ID and which under a new ID. By matching items rather than trying to control for variation in item value, the controlled experiment allows us to examine the effects of reputation on sales of used, unique items, in our case vintage postcards. For such items, there is a great deal of information asymmetry between seller and buyer about item condition, and no established book values to guide buyers.

The reputations of the new IDs throughout the first stage, though uniformly positive, were brief. The first stage did not test the effects of negative feedback. In the second stage, we were concerned whether and how much negative feedback – which by being scarce had much more information content -- in a brief reputation would hurt sellers. To do this, we purchased items from three of the new IDs in order to give them each one or two negative comments. The second stage of the experiment, lasting three weeks,

[8] Such manipulations, of course, are subject to the oversight of human subjects committees.

involved sales of 35 matched pairs solely under new IDs, to compare those with and without negative comments.

The typical item was a "lot" of vintage postcards, titled something like "Vintage Valentine Postcards (36)", where the number in parentheses indicated the number of cards in the lot. The item description indicated the general condition of the cards and provided photos of one or a few cards in the lot. The dealer followed his usual practices to determine a minimum starting bid for each item. The range was $4.99-$49.99 with a mean starting bid of $13.13 and a median of $9.99. Informal analysis suggests that listings of this value are quite typical on eBay, not only in the vintage-postacard category, but overall.

Over the course of the first stage of the experiment five of the new identities presented 20 items each for sale, and two new IDs presented 50 items each. This was done to allow for accumulation of different amounts of feedback. The items were divided into five sets of pairs, grouped by listing price. Each set contained 20 pairs (40 items total) and was sold in two different weeks, which in turn were separated by a week in the middle to ensure that there would be no overlap in their availability. For half of the pairs in any set, again determined at random, the ES (established seller) ID sold its item in the first week and for the other half, the new ID sold its item first. Two weeks later, the other ID sold the other item of the pair. In any set, 10 items were sold by each high volume seller, and four by each low volume seller. To illustrate how this counterbalancing worked: the first set of 40 items had starting (minimum) prices in the $9.99-14.99 range. Half of each pair sold in week one and the other half sold in week 3 of stage 1. For half of the pairs, the new ID sold its item in week 1 and the ES ID sold its matching item in week 3. For the other half, the order was reversed. The starting bids for the items were balanced among the seven new identities as closely as possible.

Several steps were taken to make it difficult for buyers to notice that an experiment was underway, and we received no communications suggesting that any bidders noticed item pairings or other elements of an experiment. Items were listed in a category that typically has thousands of items for sale. Each of the new identities used a slightly different format for listing items (e.g., "36 Valentine Postcards, Vintage"). Each of the new identities had its own email address from which correspondence emanated. Finally, the two halves of each matched pair were listed for sale in different weeks rather than at the same time. Care was taken to assure that each seller ID listed the item using the same information (selling price, tax, shipping cost, payment methods and description), while maintaining a unique look and feel to its listings. The listings were created using AuctionHelper, the same program used by our experienced seller to list his items under the ES ID. By giving each seller a unique look, we were able to avoid making it apparent that all sellers were being operated by the same person. However, by including all the same information, we tried to keep the items matched as closely as possible.

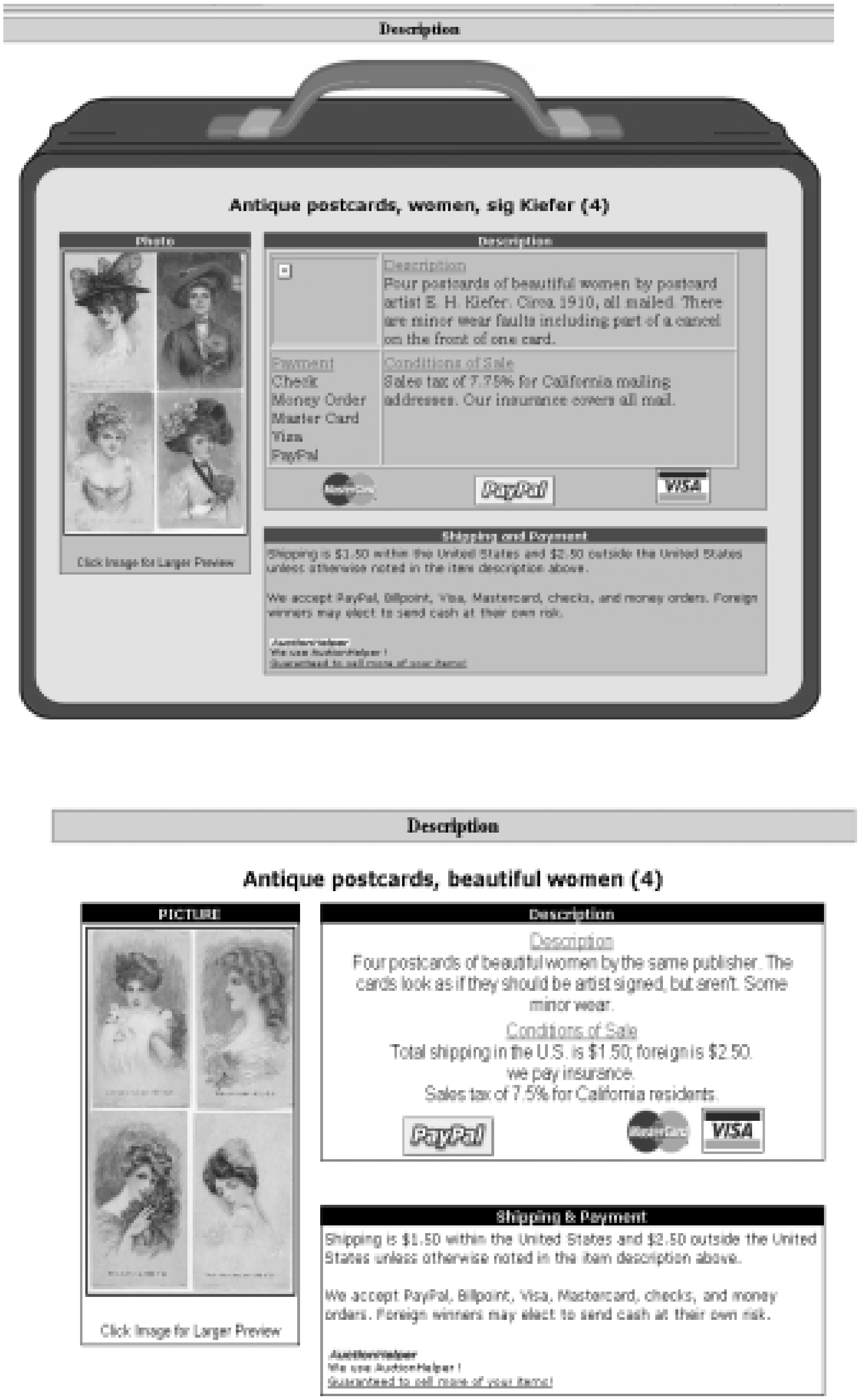

Figure 1. Sample item descriptions from two of our new IDs. Note that both have the same information available to the buyer, but have different layouts to maintain the impression that they are being sold by different sellers.

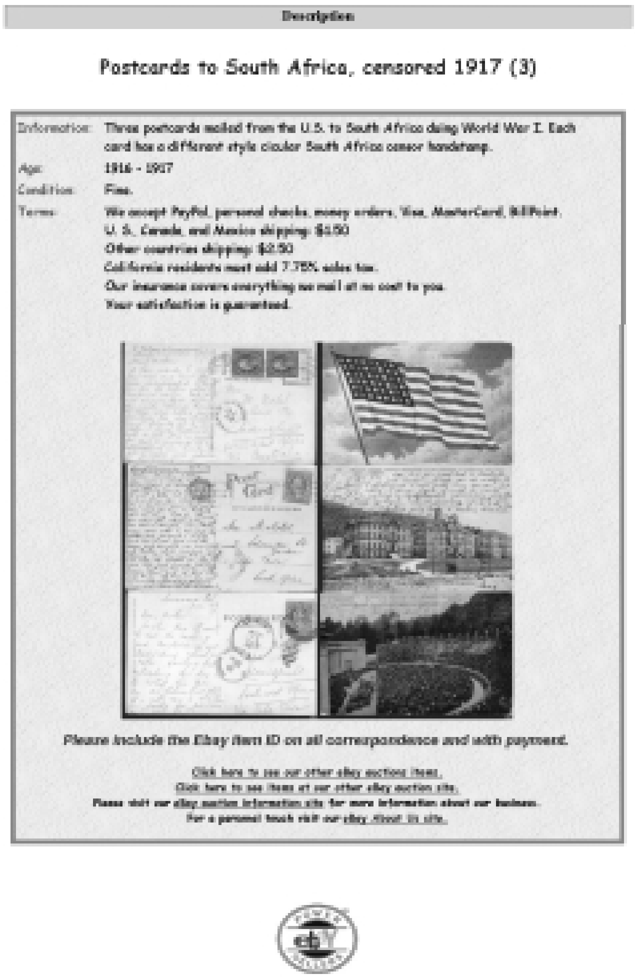

Figure 2. A sample item listing for our established ID seller created using the same software package used for our new ids.

Two item pairs were discarded from the analysis because the dealer accidentally listed them with different starting prices under the two IDs. This left a total of 198 pairs for analysis from stage 1.

Each of the new IDs began with no prior feedback. During the course of the experiment, feedback from buyers in previous time periods became visible. As mentioned earlier, at the end of the first stage, none of our sellers had any negative feedback. Our high volume sellers had 17 and 12 positive feedbacks and our low volume sellers had between 5 and 14 positive feedback points, with an average of 9.2 positive points.

After stage 1 and before stage 2, we purchased items from three of the new IDs and provided one or two negative feedback comments to them. In a previous study, we analyzed reasons for negative feedbacks at eBay and found everything from slow shipping to sellers who cashed checks but never sent items (RZ). All of the negative feedback comments we left indicated either that the item received was not as described or was in worse condition than in the auction listing text, though we purposefully did not provide any details. In all cases the user that gave the feedback was itself a fictitious entity with zero feedback. The “item not as described” negative highlights one of the big problems with Internet auctions: the inability of potential buyers to decide whether to trust the seller's description of an item. Figure 3 illustrates the feedback profile for one of our IDs at the beginning of stage 2.

Figure 3. The feedback comments page for one of our new seller IDs showing the negative we left for him for stage 2 of the experiment. User IDs and item IDs have been removed to protect the privacy of buyers.

Stage 2 of the experiment began after we had given the negative feedbacks. New IDs with similar amount of positive feedback were paired, as shown in Table 4. Each dealer then sold 35 more matched pairs of items, with half of each pair listed under a new ID with no negative feedback and the other listed by a new ID with negative feedback. As in the first stage, items were listed in separated weeks, with order of listing counter-balanced between the two IDs in each pair. Randomization was used as before. The two high-volume IDs from stage 1 were paired together. The next pair was formed from the two new IDs with the next most feedback, and the last pair was the two IDs with the next highest scores after that. The new ID that ended stage 1 with the lowest feedback score (4) was not used in stage 2.

Volume in Stage 1	ID with Positives Only	ID with Negatives
High	17+, 0-	12+, 1-
Low	11+, 0-	14+, 2-
Low	9+, 0-	7+, 1-

Table 4: Feedback profiles of new IDs before the start of stage 2 of the experiment.

Hypotheses

The ES ID has a much better reputation than any of the new IDs. This will translate into greater revenues, with revenue from future sale of unsold items estimated at a fraction of the initial minimum bid. Revenues come both from probability of sale and prices given a sale. We first consider a subsidiary hypothesis to hypothesis 1.

H1A: The ES ID seller will have a higher probability of sale.

With probability of sale determined, we address our main hypothesis about the returns to reputations.

H1: The ES ID seller will reap greater revenues than the new ID sellers in stage 1.

After some amount of positive feedback, a seller will be proven trustworthy, and additional positive feedback may have little or no impact on buyers' assessment of trustworthiness. The gradient of willingness to bid against reputation will differ among buyers. Thus, as the new IDs accumulate positive comments, perhaps even just a few, they should be trusted more.

H2: New ID prices will improve as new buyers accumulate some positive feedback in stage 1.

We expect that buyers will give little weight to the one negative feedback of the ES ID, given the positive feedback from more than 2000 distinct buyers. However, in the second stage, when seller IDs with relatively few positive feedbacks also have negative feedback, we expect buyers to distrust them.

H3: In stage 2, the IDs with negative feedback will reap lower revenues than those without negative feedback.

We have a subsidiary hypothesis to **H3.**

H3A: In stage 2, the ID with two negative feedbacks will reap lower revenues then the IDs with just one negative feedback.

Analysis

We first analyze the effects on probability of sale. We conducted sign tests to determine whether listings under the ES ID were more likely to sell than under the new IDs: if they are equally likely, then the median of the difference between them should be 0. Similarly, in the second stage, we check whether listings under IDs without negative feedback were more likely to sell. In addition, we examined whether the probability of sale was correlated between the two halves of a pair. Independence would suggest poor pairing by our experienced seller. Strong correlation, by contrast, would suggest the pairings were well done. Assuming that pairings were well done, this experiment does not enable us to distinguish whether some pairs were overpriced and others underpriced, or whether there is just substantial randomness in eBay auctions.

To analyze the impact of reputation, we compare prices paid to sellers whose IDs have quite disparate reputations. In stage 1 of the experiment, IDs 1 and 2 are the ES ID and one of the new IDs, and we expect that the ES ID will get higher prices on average than the new IDs. In the second stage of the experiment, the two IDs are both new, one with negative feedback and the other without, and we expect the ones without negative feedback to get higher prices.

To analyze price impacts, we focus on the log of the ratio of selling price to starting price for the different IDs. Our experienced seller tries to set starting prices as a consistent fraction of the price the item is likely to command. Hence, we expect actual selling prices to be lognormally distributed, leaving aside the truncation below the starting price, and a pile up at that price. Given the selection procedure for starting price, and our expectation of lognormal prices, we focus on the ratio of sale price to starting price, rather than the absolute difference. Following our lognormal expectations, our basic unit of data is thus: $\ln\left(\frac{id1\_sale\_price}{start\_price}\right)$. This log transformation avoids heteroskedasticity problems, and reduces the impact of outliers on our analysis. At stage 1, our primary interest is to compare $\ln\left(\frac{id1\_sale\_price}{start\_price}\right)$ to $\ln\left(\frac{id2\_sale\_price}{start\_price}\right)$. Since we are comparing the two logs, the start_price term drops out. The logarithm of the ratio of the ES ID sale price to the new ID sale price is given by: $\ln(id1\_sale\_price)$ - $\ln(id2\_sale\_price)$.

We could merely compute this difference, and have a good estimate of the (log of the) return to reputation but for one complication. There are three distinct types of outcomes from an eBay auction. The item may not sell, indicating that the highest willingness to pay was less than the minimum bid. There may be only one bidder (so the item sells for exactly its starting bid), indicating the *second highest* willingness to pay was at most the minimum bid. Or an item may have more than one bidder, in which case it sells for the second highest value plus the minimum bid increment. (Note, eBay does not allow bidders to raise prices by more than the minimum increment, though they can leave advance instructions to continue to bid up to some price.) We might expect the second highest valuations to be normally distributed, but because of the minimum bid we do not observe the second highest valuations when there is 0 or 1 bidder.[9] Over the whole of Stage 1, 40.7% of items failed to attract a starting price bid.

Unsold items complicate the analysis. One possibility would be to consider only pairs that sold under both identities. This would bring two unfortunate developments: reduce the sample size quite a bit, and introduce a truncation bias. Statistical methods such as Tobit and other censored-normal regression methods are available, but they do not fit our data set in a natural way because many sold items pile up at the minimum bid level, violating the normality assumption.

[9] We could have conducted the experiment with $.01 minimum bids in order to avoid or at least minimize this problem in analysis, but for ecological validity of the experiment, we preferred to have the dealer follow his normal sales practices as closely as possible.

We resolved this problem by estimating the dealer's reservation value for the unsold items and treating that as the current sale price.[10] In effect, we were computing the dealer's effective revenues from all of the items, those sold and those not sold.[11] Our dealer's usual practice is not to immediately relist an unsold item, even though eBay permits this without charging an additional fee. Instead, our dealer rotates his stock, waiting about six months before relisting the item, with a lower starting bid.[12] Alternatively, he sells the item live at a convention or a show. In either case, he incurs handling and inventory costs and has a lower expected sales price. Although he has not carefully analyzed his records, he estimates that when an item does not sell initially, he eventually gets, on average, about 50% of the initial listing price, after accounting for other costs incurred.[13] In the remainder of the paper, whenever we refer to the sales price, we refer to either the actual sales price or, if the listing did not result in a sale, 50% of the initial listing price.

HI: Our first analysis of the paired differences between $\ln(id1\_sale\_price)$ and $\ln(id2\_sale\_price)$ is non-parametric. We performed both a sign test and a Wilcoxon signed rank test. By looking only at the ranks of the prices, we avoided all censoring problems, but still lost observations when both items in a pair went unsold. The setup of the auction, with identical starting prices and increments for a pair, tended to produce ties, which worked against any hypotheses of difference.

We then turned to use magnitude of price differences, not just their ranks. That required parametric tests. Our primary parametric test is a paired t-test determining how likely it is that the ES ID and the new IDs had the same mean price, i.e., whether the difference in logarithms above is 0. Effectively, reputation is our only independent variable. We have already controlled for the starting price by relying on ratios of sale price to starting price.

As part of our results, we considered two possible confounds. Confound 1. The order in which the two items from a pair were listed might have mattered. If only a few buyers are interested in a particular type of item, the first sale may remove one of those few buyers from the market. In addition, since buyers can search for closed items containing particular words, it's possible that when the second half of a pair was listed, some buyers may have found the first listing and used its closing price to inform their bidding. As mentioned previously, the listings were counter-balanced so that half the items were

[10] An unsold item loses value in three ways: the hassle of relisting, the downgrading in value because it was learned it had not sold at the minimum price in the prior auction, and the stale goods problem. The second is the biggest factor.

[11] It might be argued that the minimum bid should be treated as the seller's reservation value, so that the profits from no sale and sale at the minimum bid are equal. But the seller's rational strategy will be to set a minimum bid higher than the reservation value in the spirit of monopoly pricing. By setting a higher minimum bid, the seller extracts additional revenue from those buyers who turn out to be the only bidders, while forgoing only transactions that would have yielded 0 profit. This analysis may be complicated by strategies for encouraging people to pay enough attention to one's listing. Bajari and Hortascu model such endogenous bidder entry explicitly (BH).

[12] Relisting at a lower bid confirms that his reservation value is lower than the starting bid that he sets.

[13] It is, of course, negative information that an item has not sold. As a check, we reran our analyses with an estimate of 80% of the initial starting bid. This yielded similar results, though smaller in magnitude.

listed by ES ID first, and half second. Two tests were performed to check for this confound. First, instead of comparing the ES ID price to the new ID price, we compared the first sale to the second sale, regardless of which ID listed the item, using the analogous sign and t-tests described above. Second, the difference in log of sale price between the two IDs was regressed against whether the item was listed first or second.

Confound 2. Though we attempted to make them similar, there may have been unforeseen differences among the new seller IDs. Perhaps some had more attractive names or formats for listing items. To check for that, a regression against dummy variables for the new seller IDs was run.

H2: To test the hypothesis that new ID prices will improve as they get even a few feedbacks, we regressed the difference in log prices against the number of positive feedbacks the new seller had as of the closing time of the auction. This ranged from 0 in the initial week to as high as 12 for some IDs in the last week.

Results

Stage 1: ES ID versus new IDs

We first consider probability of sale, and do some background calculations before turning to **H1**. Table 2 shows the results on probability of sale in stage 1, comparing the ES ID to the new IDs. A chi-square test concludes that the probability of sale was not independent between the two selling IDs (P<.001). The new IDs were much more likely to sell when the ES ID sold and vice versa; this correlation suggests that ES did a good job of pairing items. That the correlation is far below 1 indicates that other factors, including chance, sometimes gave the new IDs an advantage in a matched pair, and sometimes the ES ID. That some items sold and some did not indicates that ES could not perfectly guess the appropriate starting bid.[14]

		New ID		
		Not sold	Sold	Total
ES ID	Not sold	47	27	74
	Sold	40	84	124
	Total	87	111	198

Table 2. Number of sales in stage 1, analysis by pairs

H1A looks first at the likelihood of sale under the two IDs. Overall, the ES ID listings sold 63% of the time, the new ID listings 56% of the time. A one-sided sign test on the difference between ES_id_sold and new_id_sold approaches significance, Pr(#positive >= 40 | mean of differences=0) = Binomial(n = 67, x >= 40, p = 0.5) = 0.0710.

[14] Is ES had perfect knowledge, implying perfect price discrimination, all items would sell at the starting price.

Our main hypothesis, **H1**, related to revenues. We tested it first with a nonparametric comparison looking within pairs, and subsequently parametrically to assess the magnitude of revenue differences.

1. *Nonparametric tests*. We looked at the sign of $\ln(ES\_id\_price) - \ln(new\_id\_price)$. An item not selling was assumed to get some constant fraction less than 1 of the starting price, which incorporates differentials in the probability of sale. A one-sided non-parametric sign test was significant: Pr(#positive >= 81 | median of the differences = 0) = Binomial(n = 139, x >= 81, p = 0.5) = 0.0308. The Wilcoxon signed rank test also concludes that the difference is significant (P=0.0285). See Table 3.

Sign	Observed	Expected
Positive	81	69.5
Negative	58	69.5
Zero	59	59
All	198	198

Table 3. Observed versus expected values of $sign[\ln(ES\_id\_price) - \ln(new\_id\_price)]$.

2. *Parametric test of magnitude difference*. We had a primary concern with the magnitude of the mean price difference, not merely whether one existed. To compute this, we assigned a reservation value of 0.5 times the starting bid for unsold items. The mean difference of the logs was 0.0734, which corresponds to the ES ID, reaping 7.6% higher revenue on average. A one-sided t-test was significant (P=0.0346). We shall discuss the magnitude of this 7.6% later, looking from the buyer's side. But we should note that to compute the returns to reputation, we must take account of the value of the goods sold, including eBay's listing fees.[15] For example, if the sellers on average had costs equal to 50% of the sales prices, then a 7.6% greater revenue on average would mean a 15.2% higher profit.

Tests for the two possible confounds turned up negative results. Whether the ES ID sold an item first or second had no significant effect, whether looking at ratios to starting price, or the log of that ratio. The means were nearly identical, and sign, rank, and t-tests all failed to rule out that the mean of the difference was 0.

The hypothesis that the first few positive feedbacks for the new IDs would make a noticeable difference in the selling is not supported. In the third wave, weeks 5 and 7, the new IDs received higher prices overall than the ES ID (the only wave in which they did so). This might suggest that the new IDs improved in performance over time. However, comparing waves one and two (weeks 1 through 4) to waves four and five (weeks 9-12) tells a different story. The new IDs actually fared slightly better, relative to the ES ID, in

[15] The fee structure at eBay is somewhat complicated. For our items, there is an insertion fee ranging from $.30 for items with minimum bids less than $10 to $1.10 for products with minimum bids between $25-50. In addition, when a product sells, there is a final value fee, 5% of the first $25 and 2.5% of the remaining value up to $1000. (Our highest-priced item sold for $154, which incurred a final value fee of $4.48). We ignored listing fees in our analyses of revenue from sales.

weeks 1-4 than in weeks 9-12, though the difference did not approach significance. Regression on the number of positive feedbacks for the new IDs also failed to yield a significant coefficient.

There were not significant differences among the new IDs. The F-test comparing the models with and without fixed-effects for the sellers was not significant, and the coefficients on the individual dummies were not significant either. Even regressing against only a single dummy for the one ID that had higher average log price than the ES ID yields an insigificant coefficient .

As a final check, we ran one regression including all of the variables tested in the models above. None of the coefficients proved significant.

Stage 2: New IDs without negative feedback versus new IDs with negative feedback

Table 4 shows the results on probability of sale in stage 2, comparing the new IDs with and without negative feedback. Overall, the IDs without negatives sold 46% of the time, and those with negatives sold 40% of the time, but the difference is far from significant given the small sample size. The small sample size also makes it impossible to rule out the hypothesis that the probability of sale within a pair was independent under the two IDs (chi-squ(1) = 1.2281; P = .268) although the trend, together with the results from first stage, suggest that they might well be correlated.

		With Negatives		
No Negatives		Not sold	Sold	Total
	Not sold	13	6	19
	Sold	8	8	16
	Total	21	14	35

Table 4. Probability of sale in stage 2

Surprisingly, **H3**, which posits that negatives in a brief reputation will hurt revenues, was not confirmed. To test for the hypothesized effect, following our approach in testing **H1**, we compared the log of the price ratios, substituting half the minimum bid when there was no sale. As before, the starting price drops out. Table 2 summarizes the sign of $\ln(no\_negs\_id\_price) - \ln(with\_negs\_id\_price)$. Even though the IDs without negatives were slightly more likely to sell, the items they sold went for lower prices more often than they went for higher prices. The difference in prices was actually negative, though the Wilcoxon signed rank test shows that it is far from significantly so. Using a parametric test, and looking at magnitudes, the overall mean of log price was slightly higher for the IDs with negative feedback, though a paired t-test does not reject the hypothesis of no-difference. The check for heterodskedasticity shows no correlation between the difference in log prices and the initial starting bid.

Sign	Observed	Expected
Positive	9	10
Negative	11	10
Zero	15	15
All	35	35

Table 5. Observed versus expected values of $sign[\ln(no\_negs\_id\_price) - \ln(with\_negs\_id\_price)]$.

Finally we turned to **H3A**, the conjecture that the seller with two negatives would do worse than the sellers with one. A regression with fixed effects for the different pairs of IDs showed no significant differences. A simple comparison of means shows that in two of the three pairs the ID with negatives actually had a higher average, though not significantly so. One of those pairs included the ID with two negatives.

Discussion

While there was a difference in prices between the established seller ID and the new IDs in stage 1, the magnitude of that difference, estimated at 7.6%, was fairly small. There was no apparent difference between the new sellers with a few positives and zero versus one or two negative comments. There are many possible explanations for why the returns to reputation were not bigger.

One possibility for why the ES ID did not have a larger advantage is that it had one negative comment, albeit more than 2,000 positive comments, while the new IDs did not have any in stage 1. Perhaps some buyers paid more attention to the one negative than the 2000 positives. This explanation is quite implausible given the results from the second stage, where negatives had no effect, even for selling IDs that had less than 20 positives.

A second possibility is that there may be other markers of product or seller quality that are sufficient to inform buyers. Some of these markers include accepting credit cards, posting high quality images, having clearly delineated policies about insurance and shipping, item descriptions that convey expertise about the product, and using software that helps high-volume dealers post and manage their transactions. Both the ES ID and all the new ID postings had all these markers, even though their feedback histories varied.

A third possibility is that reputation may matter more for higher-priced items. While we rejected that hypothesis within the range of selling prices of this experiment, all bu tthree of the items in the experiment had starting prices of $25 or less, and the overall median selling price for sold items was $14.99. It may be that for items in a higher price bracket the effects of feedback may be greater, and there is some evidence from other empirical studies for this hypothesis, as discussed in the related work section. On the other hand, there are lots of items listing for $25 or less on eBay, indeed the majority, so it would still be surprising if the widely touted feedback system had only minor impacts on those transactions. Moreover, we deliberately selected items for which the information asymmetry was large, Typically there were more items in a lot than could be shown in the on-screen images, items are all unique or rare so that there is no established book value, there is no official grading scheme that a buyer could use to dispute that a seller had failed to adequately describe the condition of the old postcards, as would be the case with collectible coins, for example.

The lack of an effect from negative feedback could merely be a result of the small sample size, though the trends did not even point toward an effect. It's possible that buyers discounted the negative feedback because it was very terse and did not explain exactly what was wrong with the item, or because the feedbacks were given by buyers who had no feedback themselves. For the seller ID with two negatives, buyers may have suspected something fishy because the two negatives, both from IDs with no feedback themselves, were both posted within a minute of each other (in retrospect, an experimenter's error).

Our primary conjecture, however, is that most buyers simply didn't bother to click through to look at the detailed feedback, and instead merely relied on the overall score (number of positives minus number of negatives) that eBay displays as part of the item listing. This suggests that eBay might do well to make negative feedback a little more salient for buyers. The simplest idea would be to display the number of positives and negatives separately rather than just providing a composite score, although many more sophisticated statistics could be computed and displayed.

Perhaps the most interesting explanation for the limited price premiums from reputation is that, since most transactions turn out well, buyers rationally accept the risks of an occasional transaction being less than fully satisfactory. At the very worst, the buyer will lose the full purchase price and get nothing in return. A simple model of buyer's expected profit illustrates why the premium for the ES ID should be so small.

Let $r =$ the buyer's reservation value for the item, $v_i =$ price paid to seller i, and $p_i =$ the buyer's belief about the probability that the transaction will seller I will result in a bad transaction. For simplicity, assume the worst case, that when a transaction with seller i goes bad the buyer loses v_i. Then the expected profit from the purchase from v_i is $(r - v_i)(1 - p_i) - v_i p_i = -v_i + r(1 - p_i)$.

Assume that buyers pay prices such that their expected consumer's surplus is the same (for identical items), regardless of their beliefs about the seller's trustworthiness, paying lower prices when they think there is more risk.[16] Then, for purchases of identical items, we have
$-v_1 + r(1 - p_1) = -v_2 + r(1 - p_2)$. Rearranging and simplifying we get

$$\frac{v_1}{v_2} = 1 + \frac{r}{v_2}(p_2 - p_1)$$

Consider a pair of sales, one by the ES ID in our experiment, the other by a new ID. Since the ES ID has a long history of feedback, taking feedback at face value, the buyer

[16] A more appropriate but complex model would have buyers set their maximum bids so as to assure expected 0 profit. But because winning bids are the second highest willingess to pay (plus a bid increment), buyers realized profit is typically greater than 0. Hence, analysis of the model would have to take into account expected effects on selling prices of changes in buyers' maximum bids. We stick to the simpler model here for illustrative purposes.

estimates $p_1 = \frac{1}{2000} = .0005$. We have observed in our experiment that $\frac{v_1}{v_2} = 1.076$. Suppose further that $\frac{r}{v_2} = 1.25$: the buyer pays on average only 80% of her reservation value for the item, even to the seller that she trusts less. Solving for p_2, we find that in order for the price ratio to be 1.076, the buyer must believe that $p_2 = .06$. To get to a more reasonable estimate[17] of, say $p_2 = .02$, we must have $\frac{r}{v_2} = 3.83$. That is, buyers would have to be paying only about a quarter of their reservation values on average in order to justify a price premium as high as 7.6% for the ES ID seller.

Thus, given the actual population of sellers on eBay, where problems seem to be rare (or at least are rarely reported in the form of negative feedback), we should expect a low price premium, and 7.6% does not seem unreasonable. Note also, this 7.6% applies to revenues. The profit differential, which subtracts out costs, would be greater.

We recognize, of course, that the participants in the eBay Feedback Forum are not sophisticated Bayesians. There are many equivalents of the noise traders from financial markets. The market outcome may be over- (or under-) rewarding sellers with positive reputations. From the marketplace perspective, over-rewarding would be beneficial. It would be maintaining the incentives for seller quality. If all sellers were strategic, and we assume that the market is in equilibrium, then we should expect that the cost of being a high quality seller (including the lost profits from not cheating people) should be just enough to balance out the 7.6% price premium.

If sellers were under-rewarded, fraud would be a concern. However, whereas it might be worthwhile to sell imitation watches, it is hardly economic to fake postcards. Moreover, if a seller merely took the money and never sent the goods, both eBay and the legal authorities would come after him. On items like postcards, where batches sell for less than $25, fraud should not be a major concern.

We suspect that the vast majority of sellers are not strategic in the sense of misrepresenting their products or lowering their efforts at delivery just because they can get away with it. Normal hype, e.g., "beautiful item" or "must be seen," is to be expected. Fraud and severe shirking may just not be worthwhile, except for a few bad outliers, who Feedback Forum will ultimately identify.

Of course, we also should not necessarily assume that the market is in equilibrium, particularly since it is difficult for even very sophisticated buyers to estimate the true probability of problems in transactions, or for sellers to estimate the return to reputations. Even securing the available information on a seller's reputation may be costly. Our stage

---

[17] From a large data set of eBay transaction, using a logistic regression model predicting the probability of negative or neutral feedback on the next transaction, based on seller's prior feedback, the estimated probability for a seller with no prior feedback was .0191 (Resnick and Zeckhauser, 2001).

2 results suggest that many buyers do not bother to find out about negatives, at least when reputations are brief and negatives are presumably rare.[18] Extensive reputations bring their own problems. There are many classes of negatives, some important, some not. Scrolling through to find the nature of negative comments is costly, and hardly worthwhile with long reputations and low-priced items.

Conclusion

The eBay Feedback Forum illustrates Yhprum's Law. (Yhprum is Murphy spelled backward.) Systems that shouldn't work sometimes do. Though buyers have no incentive to leave feedback, half of them do. That fraction raises the spectre of selection bias, and suspicion is raised further because only 1% of feedback is negative. Yet each week hundreds of thousands of items – very few of them assuredly new or standardized -- get sold on eBay. This implies that their sellers are trusted.[19] Anecdotal accounts and virtually all studies find an effect of seller's reputation on the price he receives. However, there is little agreement on what elements of reputation are most important.

Working with an established eBay seller, with an extremely favorable reputation, we conducted the first controlled experiment to assess the returns to reputation. Looking at matched pairs of items – batches of vintage postcards -- our established seller received 7.6% more on average than new sellers. Surprisingly, a second experiment found that negative feedback in the brief reputations of new sellers did not affect their prices. The 7.6% premium for an established seller of a low-priced but used and hardly standardized item is shown to be reasonable.

References

Akerlof, G. A. (1970). "The Market for "Lemons": Quality Uncertainty and the Market Mechanism," The Quarterly Journal of Economics 84(3): 488-500.

Avery, C., P. Resnick and R. Zeckhauser (1999), "The Market for Evaluations," The American Economic Review 89(3): 564-584.

Ba, S. and P. A. Pavlou (forthcoming). "Evidence of the Effect of Trust Building Technology in Electronic Markets: Price Premiums and Buyer Behavior." MIS Quarterly.

Bajari, P. and Ali Hortacsu (2000). Winner's Curse, Reserve prices and Endogenuos Entry: Empirical Insights from eBay Auctions. Working paper, Dept. of economics, Stanford University. Available on-line at http://siepr.stanford.edu/papers/pdf/99-23.html.

[18] There are countervailing considerations in deciding whether a longer reputation makes it worthwhile to click through to get numbers of negatives, positives and neutrals. A long reputation probably suggests a better reputation, since bad sellers get eliminated from the system. This makes looking for negatives less worthwhile. On the other hand, with a long reputation the information received is more likely to be statistically informative.

[19] If sellers were distrusted, they still could sell at a heavily discounted price. However, if they were selling honestly, they would be better off in regular markets, suggesting that absent a substantial element of trust eBay would be like a used-car market that prohibited tire kicking and test drives. The lemons problem would be severe.

Dewan, H. (2001). Trust in Electronic Markets: Price Discovery in Generalist Versus Specialty Online Auctions. Working paper, available on-line at http://databases.si.umich.edu/reputations/bib/papers/Dewan&Hsu.doc.

Eaton, D. H. (2002). Valuing Information: Evidence from Guitar Auctions on eBay. Working paper, Murray State University. Available on-line at http://databases.si.umich.edu/reputations/bib/papers/eatonpaper.pdf.

Friedman, E., and P. Resnick (2001). "The Social Cost of Cheap Pseudonyms." Journal of Economics and Management Strategy, 10(2): 173-199.

Houser, D., and J. Wooders (2000). "Reputation in Auctions: Theory, and Evidence from eBay", Working Paper, Dept. of Economics, University of Arizona. Available on-line at .

Kalyanam, K.. and S. McIntyre (2001). Returns to Reputation in Online Auction markets. Working paper W-RW01-02, Santa Clara, CA, Santa Clara University.

Kauffman, R. J., and C. Wood (2000). Running up the Bid: Modeling Seller Opportunism in Internet Auctions, working paper 00-18, University of Minnesota, available on-line at http://misrc.umn.edu/wpaper/WorkingPapers/MISRC_RunningUpBid.pdf

Lee, Z., I. Im, et al. (2000). The Effect of Negative Buyer Feedback on Prices in Internet Auction Markets. In Proceedings of the 21st International Conference on Information Systems. Available on-line at http://papers.weburb.net/showpaper.php3?ID=104&CategoryID=3.

Lucking-Reiley, D., D. Bryan, et al. (2000). "Pennies from eBay: the Determinants of Price in Online Auctions", Working Paper, Vanderbilt. Available on-line at http://www.vanderbilt.edu/econ/reiley/papers/PenniesFromEBay.pdf.

McDonald, C. G., and V. C. Slawson Jr. (2001). Reputation in an Internet Auction Market. Working paper, available on-line at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=207448.

Melnik, M. I. and J. Alm (2002). "Does a Seller's Reputation Matter? Evidence from eBay Auctions." Journal of Industrial Economics forthcoming.

Resnick, P. and R. Zeckhauser (2001). Trust Among Strangers in Internet Transactions: Empirical Analysis of eBay's Reputation System. Working paper available on-line at http://www.si.umich.edu/~presnick/papers/ebayNBER/index.html.

Shapiro, C. (1983). "Premiums for High Quality Products as Returns to Reputations." Quarterly Journal of Economics, 98(4): 659-680.

Tversky, A., and D. Kahneman (1974), "Judgment Under Uncertainty: Heuristics and Biases," Science, 185, 1124-31.