Self-Selection, Slipping, Salvaging, Slacking, and Stoning: the Impacts of Negative Feedback at eBay

Tapan Khopkar
University of Michigan
School of Information
Ann Arbor, MI, USA
tkhopkar@umich.edu

Xin Li
University of Michigan
Department of Economics
Ann Arbor, MI, USA
xinl@umich.edu

Paul Resnick
University of Michigan
School of Information
Ann Arbor, MI, USA
presnick@umich.edu

ABSTRACT
Analysis of usage history for a large panel of eBay sellers suggests that both seller and buyer behavior change in response to changes in a seller's feedback profile. Sellers are more likely to stop listing items right after receiving a negative feedback. Sellers who continue listing do not seem to improve their performance in order to salvage their reputations. Instead, sellers get more negative feedback after receiving a negative feedback. One reason is that observed negative feedback appears to be symptomatic of a temporary decline in the seller's quality, which is also reflected in other transactions around the same time. Receipt of negative feedback might also cause a decline in seller quality, but we find only weak evidence of that. Empirical evidence does support a second hypothesis, that buyers appear to be more willing to give negative feedback to sellers who have recently received other negative feedback.

Categories and Subject Descriptors
K.4.4 **[Computers and Society]**: Economics

General Terms
Economics.

Keywords
Reputation Systems, Electronic Commerce, Feedback, Trust, Stoning

1. INTRODUCTION
The Internet brings together buyers and sellers separated by physical boundaries, opening endless avenues for trade [3]. With the expanding reach of the Internet, electronic markets are emerging as an increasingly important factor of the economy. eBay, touted as 'The World's Online Marketplace' is among the most noteworthy examples of electronic markets.

In a trade environment like eBay, buyers have limited information about product quality and seller reliability, at the time of a transaction. Thus, electronic markets like eBay are "ripe with the possibility of large-scale fraud and deceit" [8]. The information asymmetry between buyers and sellers can reduce market efficiency. Most directly, risk averse buyers may miss beneficial transactions. In addition, adverse selection may lead high quality sellers to abandon the market, creating a 'Lemons Market" as modeled by Akerlof [1]. Buyers are unwilling to pay the full price of the high quality good because they are unable to distinguish those goods from lower quality goods. Thus, the sellers of high quality goods are unable to get the full value of their items, and choose to leave the market.

The Feedback Forum at eBay, where buyers and sellers leave feedback about each other, reduces the information asymmetry, by telling buyers whether previous customers were satisfied with the seller. Both buyers and sellers can leave post-transaction feedback about each other; each feedback contains a numerical rating (+1, 0, or –1), and a text comment.

In principle, reputation systems like the Feedback Forum can improve the efficiency of marketplaces in three ways:

- **Signals.** A seller's feedback history can serve as a signal to buyers of how risky it is to purchase from that seller. This allows each buyer to choose sellers, and how much to bid, based on the buyer's level of risk aversion.
- **Sanctions.** Sellers will strive to avoid negative feedback, in order to avoid adverse future impacts on their sales.
- **Selection effects.** Because buyers will be better able to distinguish high quality from low quality, high quality sellers will not leave the market. Indeed, the low quality and fraudulent sellers may be driven from the market, leaving a higher overall quality level and less risk even for those buyers who do not carefully monitor the signals about the trustworthiness of individual buyers.

Prior research has tried to document the extent of some of these effects. One study provided a model showing that past feedback is somewhat useful as a signal predicting future customer satisfaction [9]. Many studies have examined the sanctioning effect by quantifying the impact of a seller's feedback profile on the probability that an item will sell and the price it will receive. Generally, these employ a cross-sectional methodology, comparing sales between sellers and including variables in regression models to try to control for differences among sellers and listings. One study employs a field experiment that compares sales by a single seller using different personas with different feedback profiles. See [2, 10] for a survey of results. In general, the studies of the sanctioning effects suggest that a better reputation is of some value to a seller, though the studies differ somewhat in their assessments of the particular impacts positive and negative feedback in a profile.

This paper analyzes other aspects of how seller and buyer behavior change as seller feedback profiles change, other than buyer bidding behavior. In particular, we analyze self-selection by sellers, changes in seller quality levels, and changes in buyer feedback giving patterns. Our approach is based on analysis of a large panel of sellers. We track whether their sellers continue to list items, and when they do, whether they get positive feedback, negative feedback, or no feedback at all.

When a seller gets a negative feedback, it marks a significant change in a feedback profile. An earlier study found that neutral and negative feedback constituted only about 1% of all feedback [9]. Anecdotal evidence from observation of eBay discussion boards and attendance at a conference of eBay users suggests that getting a negative feedback is a traumatic event for many.

Thus, we contrast what happens before receiving a negative feedback with what happens afterward. Seller behavior, both before and after receiving negative feedback, is reflected in whether and how frequently they list items. Buyer behavior is reflected in whether listed items sell. Both buyer and seller behavior are reflected in the feedback that buyers give to sellers. For example, a higher probability of negative feedback may reflect worse performance by the sellers, or it may reflect a greater willingness of buyers to report dissatisfaction, or some combination of those.

Our analysis is most similar in spirit to that in [4]. We ask many of the same questions but there are major differences in our analytic approach and we reach quite different conclusions.

2. THE IMPACT OF NEGATIVE FEEDBACK

A negative feedback may result from random factors, or it may be an indicator that the quality of the seller's goods or services has declined. We will refer to such a quality decline as *slipping*.

Rather than just reflecting a change in the seller's quality, negative feedback can cause changes. The seller's own behavior may change and the community's attitude towards her may also change. Let us examine each of these in detail.

2.1 Seller Behavior Change

The seller may stop listing items entirely, what we will refer to as *self-selection*. This would be a sign that the system is working as intended to weed out sellers of lower quality, though it could be that some high-quality sellers who get a negative feedback as a result of a genuine misunderstanding are also being driven from the system.

A seller's quality also may change as a result of receiving a negative feedback. The seller may be more careful and conservative in describing her items, more responsive in her communication, and faster and more careful in her packaging and shipping, in an effort to *salvage* her reputation. Contrariwise, sellers may deliberately offer lower quality service after receiving a negative feedback, a practice that we will refer to as *slacking*. Like slipping, slacking will be observable in the form of less positive feedback on future transactions and more negative feedback. The difference is that, with slipping, the quality decline happens around the time of the transaction that eventually gets negative feedback, while with slacking it happens only after the seller receives the negative feedback.

Slacking could occur for psychological reasons, if sellers get angry or discouraged after receiving a negative feedback. It could also occur as part of equilibrium behavior that creates an incentive for sellers not to get negative feedback in the first place. Cabral and Hortacsu [4], following Diamond [6], construct an equilibrium for a model where getting negative feedback provides evidence that the seller is of an opportunistic type rather than a type that always gives high effort. Once the seller has been found out as not always giving high effort, buyers expect the seller not to give high effort in the future, and it is rational for the sellers to act in accordance with that buyer belief. Whether slacking occurs for psychological reasons or as a way to maintain an equilibrium of high effort prior to receiving a negative, if we find evidence of slacking, we would consider it a negative consequence of the design of the feedback system, as it would reflect sellers offering lower quality than they were capable of after receiving a negative feedback.

2.2 Buyer Behavior Change

Clearly, a negative feedback reflects unfavorably on the seller. A negative feedback may result from factors beyond the control of the seller, such as a package lost in the mail. Still, natural Bayesian updating of beliefs will lead buyers to be more suspicious of a seller after she has received a negative feedback.

As mentioned earlier, our analysis will focus on how buyers' increased suspicions affect their feedback giving behavior rather than their bidding behavior. Resnick and Zeckhauser [9] suggest that buyers would be more willing to cast another stone at an already disreputable seller, what they call a *stoning* effect. That is, receiving a negative feedback would make the user more likely to receive another. This could occur for one of two reasons. First, buyers may be willing to forgive a single bad behavior but want to punish sellers who exhibit a pattern of bad behavior. Second, a buyer may interpret what happened in his own transaction differently depending on the suspicions raised by the seller's previous feedback. For example, if an item appears to be damaged in shipment, a previous negative feedback suggests that the damage was more likely to have been the seller's fault.

Dellarocas [5] shows that a reputation system can create the same sanctioning effect no matter how much of the seller's history buyers see. When only feedback for the most recent n transactions are displayed (or buyers only pay attention to the recent feedback), buyers can only sanction sellers for up to n rounds after they receive a negative feedback, but if the per-round punishment in terms of lost revenue is sufficiently high, the sellers will be deterred from getting a negative feedback. If the per-round punishment is limited, however, then it will be necessary to extend sanctioning over a longer period of time. Stoning can be viewed as a strategy that probabilistically extends a seller's sanctioning period after receiving negative feedback—even if many buyers are paying attention only to a seller's most recent feedbacks, because of stoning there is an increased probability of getting another negative feedback, which keeps a negative feedback in the "recent history" for longer.

Messages on eBay's message boards suggest that sellers believe buyer stoning to be a common phenomenon. On August 22, 2004, one user wrote:

> "I have been buying and selling on eBay for almost four years and have over 1000 positive feedbacks with four negs and six neutrals. Something that I have noticed, is that the Negs/Neutrals seem to happen in spurts.
>
> I think that people - especially buyers - are reluctant to leave a Neg or Neutral for a seller who has an excellent feedback score. Likewise, they are more apt to leave the Neg or Neutral if they see that others may have recently left one as well..."

Another veteran eBay seller replied:

> "Goes without saying that period following a non-positive feedback is at high risk of getting another."

If stoning occurs, the probability of negative feedback should go up for transactions after the seller receives a negative feedback. When a transaction goes well, however, we hypothesize that the presence of a negative feedback should have no effect on whether the buyer provides a positive feedback or no feedback at all. Thus, unlike seller slipping and slacking, stoning should increase the probability that a seller receives negative feedback, but should not affect the probability that a seller receives positive feedback.

Table 1: summary of possible effects to be analyzed

Effect	Party	Description
Self-selection	Seller	Drop out after negative feedback
Slipping	Seller	Quality declines; that decline leads to negative feedback
Salvaging	Seller	Quality increases after negative feedback
Slacking	Seller	Quality declines after receiving negative feedback
Stoning	Buyer	More willing to give negative feedback

3. DATA ANALYSIS

For the analyses conducted in this paper, we selected panels of sellers and examined their transactions and feedback. The use of a panel dataset is important as it enables us to study the changes in individual users, and also account for user heterogeneity. The datasets we analyzed were derived from the following master datasets, provided by eBay in a form that stripped all personally identifiable details.

1) Items Dataset: contains transactional data for all the items listed for sale on eBay from February 1st 1999 to June 30th 1999.

2) Feedback Dataset: contains all the feedback data up to May 31st 1999.

3) Users Dataset: contains the id and registration dates for all the users who registered before June 30th 1999.

The participants' feedback profiles as of the times of transactions were not stored by eBay, but we were able to reconstruct measures of prior feedback similar to those eBay displayed to users. eBay calculates information in its feedback profiles in two ways. The first and perhaps more intuitive one treats the feedback as the unit of analysis. A feedback profile includes a count of the total number of positive feedbacks and the total number of negative feedbacks in a user's profile. This accounting method, however, makes it relatively easy to inflate a reputation score, by having a friend leave multiple positive feedbacks.

A second accounting metric treats the partnership as the unit of analysis. At most one positive feedback from each partner counts toward the seller's count of distinct "members who left positive feedback" and similarly for neutral and negative feedback. For example, if user A sold 5 items to user B and gave a positive feedback for all the items, it would be counted only once. The composite score that eBay displays next to a user's id at various places in the site, including on auction listings, is the difference between the number of transaction partners who left positive feedback and the number who left negative feedback. For our analyses that use measures of a seller's prior feedback history at the time of a transaction, we rely on the metrics that treat the partnership as the unit of analysis.

Analysis of the content of neutrals and negatives showed that both are used primarily to indicate problematic transactions [9]. Thus, we treat neutral feedback as negative for the purpose of our analyses.

3.1 Selection Effect

First let us examine whether sellers drop out when they get bad reputations, either not selling any more on eBay or switching to a new user id and starting over without any feedback. Of course, there will always be some attrition of users not continuing to sell on eBay, regardless of the state of their feedback profile. The question is whether attrition is higher among users with worse feedback profiles. To answer this question, we examined a panel of 76,956 users who joined eBay on or after February 1, 1999, received a feedback (as either buyer or seller) in the period April 11-30, 1999, and who had sold at least one item prior to receiving that feedback.

One indicator of a selection effect is whether the last feedback of a user was positive or negative. In the panel, 6.42% of users who listed no items after April 30 had a negative or neutral as their last feedback during April 11-30, while the overall percentage of negative (and neutral) feedback was only 1.37%[1]. This suggests some selection effect: the higher probability of negative feedback in last transactions indicates that users were more likely to stop selling after a negative feedback than after some other feedback.[2]

---

[1] The frequency of negative feedback was higher for this panel than in other datasets we consider in this paper because the sample is restricted to relatively new sellers, who tend to get somewhat more negative feedback.

[2] Cabral and Hortacsu [4] interpret similar data about higher percentages of negative feedback just before a seller drops out as evidence of seller profit-taking in advance of dropping out, rather than a decision after receiving negative feedback to drop out.

A more direct measure of the impact of feedback on whether users drop out comes from analysis of whether individual users sold again after receiving positive or negative feedback. To avoid statistical complications from repeated, overlapping measures, we randomly selected one feedback event for each user from the April 11-30 period: 2,091 were negative (or neutral), 74,865 positive.

When the randomly selected feedback was negative, 53.85% of the users listed another item before June 30^{th}. For those users whose randomly selected feedback was positive, 82.35% listed another item before June 30^{th}. This suggests quite a large selection effect based solely on receiving a single negative feedback.

There may be additional effects from a user's feedback profile beyond the impact of the most recent feedback. If we look into the user's history, are sellers who received a negative feedback recently more likely to drop out? Are sellers who have received more negative feedbacks more likely to drop out, regardless of the content of the most recent feedbacks? Are sellers who have received more positive feedbacks less likely to drop out, either because they value their accumulated reputations, or because the large amount of positive feedback is an indicator of sellers who are more committed to eBay?

To test for these other indicators of selection effects, and to see whether the impact of a negative in the most recent transaction is still strong when controlling for the cumulative effect of the seller's full reputation, we conducted a logistic regression. The dataset is the random selection of one feedback for each of the 76,956 sellers. The outcome variable is whether the user listed another item. The covariates are as follows.

- *fbscore* is the score of the current feedback, 1 if positive, else 0.
- *Posr* is the number of distinct partners who gave positive feedback prior to this transaction's closing time. As is customary in other empirical analyses, we employ a log transform on the number of positive feedbacks: we expect the marginal impact of another positive feedback to decline as the user accumulates more feedback.
- *Negr* is the number of distinct partners who gave negative or neutral feedback prior to this transaction's closing time.
- *Neg5* is 1 if there is at least one negative feedback among the seller's five most recent feedbacks as of the time this feedback. It is 0 otherwise.

We find that the seller's earlier positive feedback has a very small but positive impact on her probability of returning as a seller. There is a decrease in the probability of return with prior negative feedback, supporting the claim that there is a selection effect from accumulated negative feedback. A recent negative feedback has a larger effect. Controlling for long-term and recent history, the feedback on the current item is still a large and significant predictor of whether the seller will return for another transaction.

To understand the effect of the covariates, consider a hypothetical user with 50 positive feedbacks. If the user has no negative feedback, the probability of return is 84%. If the user has one negative feedback, which is not recent, the probability of return is 82%. If the user has one negative feedback which is relatively recent, but not for the last transaction, the probability of return is 73%. If the user has one negative feedback, and it is for the most recent transaction, the probability of return is 59%.

Table2: results of logistic regression predicting whether the user sells again

	listagain
ln(1+Posr)	0.036 (0.009)***
Negr	-0.120 (0.034)***
fbscore	1.294 (0.046)***
Neg5	-0.555 (0.066)***
constant	0.220 (0.045)***
Observations	76,956
Pseudo-R^2	0.013

Notes: Standard errors in parentheses
*** significant at 10%; ** significant at 5%; *** significant at 1%**

Thus, the strongest selection effect comes immediately after receiving a negative feedback, although there is a smaller ongoing, cumulative effect. We will return in the discussion section to the implications of the primacy of the most recent negative feedback and the comparatively smaller impact of accumulated positive feedback.

3.2 Salvaging, Slacking and Stoning Effects

In this section we look at the impact of negative feedback on the behavior of sellers who continue selling and the impact on the feedback giving practices of buyers. The analysis is again based on the transaction and feedback histories of a panel of users, in this case a panel of 9,655 users who sold at least one item meeting criteria to be described below.

One approach would be to consider only sellers who received negative feedback and to directly compare behavior before and after they received their first negative feedback. For example, Cabral and Hortacsu [4] show for a different dataset that sellers conducted more transactions before their first negative feedback than between receiving their first and second negative feedback.

We are concerned, however, that such an analysis would be biased, because either the dependent variable is censored or the sample is truncated. This is clearest in the analysis of number of transactions as a proxy for probability of getting negatives. If users with at least one feedback are the sample, the dependent variable for number of transactions after the negative is censored for those users who did not receive a second feedback. They would have conducted an unknown number of additional transactions before getting a negative. On the other hand, the sample of users with at least two negatives is a truncated sample--it leaves out those users who did not receive a second negative, and thus would have had, on average, more transactions after the first negative and before the second than those in the truncated sample.

Our approach is to conduct logistic regressions with the feedback on each transaction as the dependent measure. We compare outcomes for transactions that are in windows of time where the hypothesized slipping, slacking/salvaging, and stoning effects would occur. The logistic regressions also control for a seller's long-term feedback history.

The partnership is the unit of analysis. To avoid confounds from multiple transactions for the same partnership (e.g., partners may be less likely to leave second feedbacks since they know that eBay only counts one in its partner-based statistics) we analyze the outcome of only of the first transaction for each partnership. We exclude from the analysis partnerships where both parties joined eBay prior to the beginning of our transations dataset, as we are unable to determine if the first transaction in our dataset was truly the first one for the partnership. We consider only partnerships where the user in our sample was the seller in their first transaction, since the hypothesized effects such as slipping apply more clearly to sellers.

During the time period of our data, eBay did not require users to tie feedback to particular transactions, though they had the option to do so. Even when a buyer explicitly specified a transaction, if there were multiple transactions between the partners in a short time, we think that a feedback for any one transaction was often intended to cover them all. Thus, for each partnership's first transaction, we classified the outcome based on the first feedback recorded from buyer to seller within 42 days (six weeks) of transaction closing time.

We have feedback data only through May 1999. The transactions later in the dataset are less likely to have a feedback recorded within our dataset. To avoid any truncation effect, we consider only transactions that ended on or before April 10, 1999, i.e., 42 days before the end of the dataset.

To summarize, we analyze the feedback outcomes for first transactions within each partnership, where the user from the panel was the seller in the first transaction, where at least one of the partners joined eBay after February 1, and where the first transaction closed some time in the period Feb. 1- April 30. Of the 100,761 transactions meeting these criteria, 42.34% resulted in positive feedback, 0.42% in negative (or neutral) feedback, and the remaining 57.23% received no feedback.

Slipping, stoning, and slacking or salvaging behavior all would be reflected in the outcomes of transactions near those transactions that are known to have received negative feedback. To differentiate among the effects, we define more carefully the windows around those transactions where we would expect to see effects.

Figure 1 illustrates the definition of these windows for a single user's history. The transactions are shown from left to right in order of their closing times. The feedbacks received by the user are displayed along the same timeline, but arranged according to the time the feedback was received. Dotted lines connect feedbacks with the transactions they comment on. One negative feedback is shown (feedback a, associated with transaction A); the remainder of the feedbacks are positive.

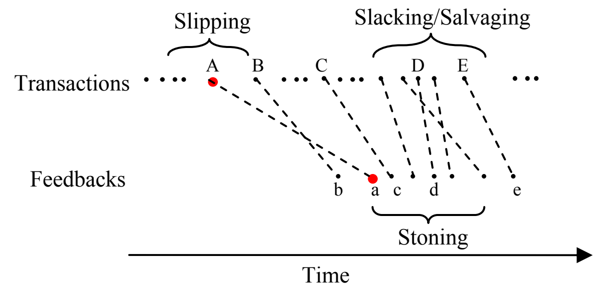

Figure 1. Windows generated from negative feedback "a".

The slipping window is defined for transactions that closed shortly before or after the known bad transaction (A in the figure). The idea is that a transaction known to get a negative may be indicative of a temporary decline in the seller's quality, perhaps due to a vacation or illness or family crisis. We (somewhat arbitrarily) assume that the window of decline lasts for a week, centered on the known bad transaction. A transaction that receives a negative is defined not to be in its own slipping window, since we will be analyzing whether being in a slipping window has an impact on feedback outcomes for transactions. A transaction is considered to be in a slipping window only if it is in the window defined by some other transaction. In the figure, transaction B is in A's slipping window, but A is not in a slipping window.

The slacking/salvaging window is defined for transactions that close during a recovery period. Intuitively, the recovery period lasts until the negative feedback becomes less salient in the user's profile. The basic format of a user profile in 1999 was similar to it's current format, although percentages of positive feedback were not calculated and displayed then as they are now. Additional positive feedbacks received after the negative would push the negative comment down on the screen, eventually requiring a user to scroll or even click to another page to see it. We (somewhat arbitrarily) defined the recovery period to last until five positive feedbacks were received. Thus, any transaction that closed in the time between receipt of the first negative and the first subsequent positive feedback (including transactions D and E in the figure) was classified as being in the slacking/salvaging window for that negative feedback.

The stoning window is also based on the same five feedback recovery period. A transaction is in the stoning window, however, based on the time when feedback for the transaction is received, not the time when the transaction closes. Thus, in the figure, C and D, but not E, are in the stoning window defined by feedback a. Transaction A is defined to be outside of its own stoning window.

For transactions that did not receive feedback, we can only estimate whether a recovery window was active at the time when the buyer might have contemplated giving feedback. For transactions that did receive feedback, the median time before feedback was about 21 days after transaction close. Thus, if a transaction did not receive any feedback, we classify it as being in a stoning window if a recovery period was active 21 days after the transaction.

The windows are overlapping, but sufficiently distinct to enable analysis. Tables 3-5 show the limited overlap. The vast majority of transactions are not in any of the windows defined by negative feedbacks, as we would expect since negatives are rare.

Table 3 Cross-tabulation of slipping and slacking windows

	slack window		
Slip window	0	1	Total
0	89.50%	2.54%	92.04%
1	7.19%	0.78%	7.96%
Total	96.68%	3.32%	100,761

Table 4 Cross-tabulation of slacking and stoning windows

	stone window		
slack window	0	1	Total
0	93.48%	3.20%	96.68%
1	2.44%	0.87%	3.32%
Total	95.93%	4.07%	100,761

Table 5 Cross-tabulation of stoning and slipping windows

	slip window		
stone window	0	1	Total
0	89.54%	6.39%	95.93%
1	2.50%	1.57%	4.07%
Total	92.04%	4.07%	100,761

As hypothesized, the three windows are all correlated with transaction outcomes, as shown in Table 6. The probability of negative feedback is higher for transactions in any of the windows than it is for transactions generally. The probability of positive feedback is lower, and the probability of no feedback is higher.

Table 6: Fraction of negative, positive and missing feedback received for transactions in the windows.

	All transactions	Slip window	Slack window	Stone window
Neg	0.42%	1.02%	1.14%	1.46%
Pos	42.34%	33.59%	31.91%	28.00%
None	57.23%	65.39%	66.96%	70.53%
Total	100,761	8,024	3,341	4,103

While the three windows identify possible effects that are temporally close to any particular negative, it could be that the observed correlations really reflect a long-term rather than short-term effect. In particular, receiving a negative feedback may be correlated with other negative feedback because it suggests that the seller's overall quality level is lower than that of other sellers. Under this hypothesis, it is the seller's overall percentage of negative feedback that is correlated with a higher percentage of additional negative feedback.

Figure 2 is consistent with this hypothesis. Transactions are grouped into bins based on the percentage of negative feedback

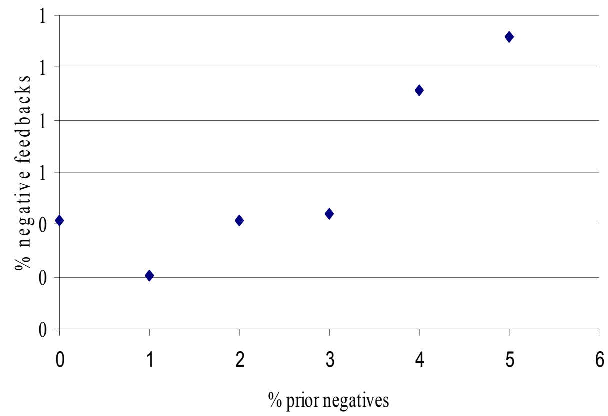

Figure 2: Percentage of prior negatives vs. fraction of receiving a negative feedback

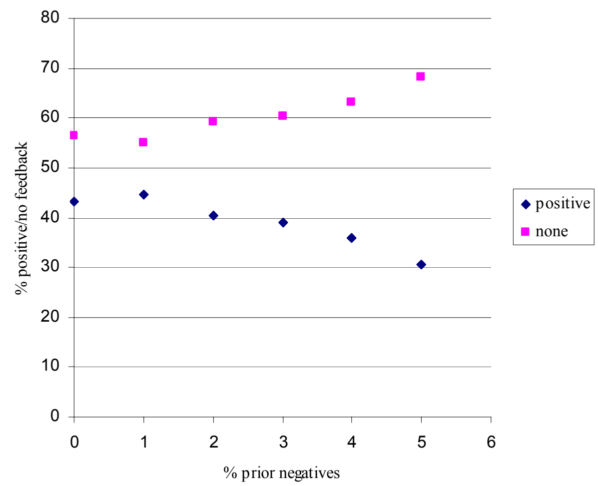

Figure 3: Percentage of prior negatives vs. fraction of receiving a positive/missing feedback

the seller had received for prior transactions. Transactions with 5% or more prior negative feedback are included in the 5% bin. The y-axis indicates the percentage of the transactions in each bin that received negative feedback. Figure 3 shows, using the same bins, the impact of prior negative feedback on the probability of positive and no feedback for a transaction. Thus, we include the percentage of negative feedback as a control variable in our logistic regressions.

We also include a proxy for the seller's experience level as a control variable. Figures 4 and 5 illustrate the correlation between the number of prior positive feedbacks and transaction outcomes. It appears that sellers tend to improve in quality over time, but that once sellers have accumulated a very large number of feedbacks, they are less likely to get any kind of feedback, perhaps because neither they nor their customers think that additional feedback is as important. To account for this

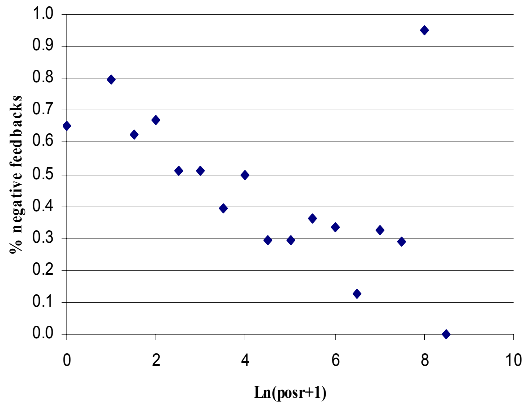

Figure 4. Log of prior positives vs. fraction of transactions receiving a negative feedback

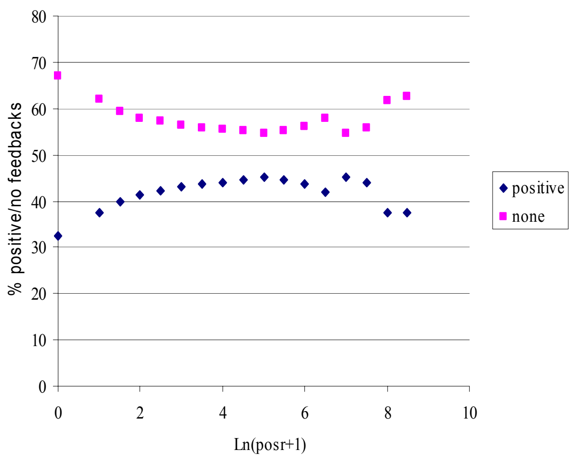

Figure 5. Log of prior positive vs. fraction of transactions receiving a positive/missing feedback

curvilinear effect, we include in the regressions a squared version of the experience variable as well as the variable itself.

To summarize, the covariates in the regressions are:

- *Slip*—1 if the transaction is in the slipping window defined by some other transaction's negative feedback, otherwise 0.
- *SlackSalvage*—1 if the transaction is in a slacking/salvaging window, else 0.
- *Stone*—1 if the transaction is in a stoning window defined by some other transaction's negative feedback, else 0.
- *pctneg*—the percentage of the seller's feedback prior the current transaction that was negative or neutral, capped at a maximum of 5%.
- *Logposr*—The log of the seller's number of prior positive feedbacks (1 is added to the count before taking logs, to ensure that the quantity is defined).
- *Logposr squared*—Logposr*logposr.

Two logistic regressions are reported in Table 7. The outcome variables are the probability of receiving a positive feedback and receiving a negative feedback. The regressions include a seller random effect, a per-user effect for the users in our panel data. Each user is assumed to have a particular innate quality level, composed of several immeasurable qualities such as honesty, proficiency in the mechanics of fulfilling the order, communication skills, and diligence. Quality is assumed to be normally distributed across the population of users.[3]

Table 7: Random effects logistic regression predicting the probability of positive feedback and negative feedback

	posfb	negfb
Logposr	0.198 (0.017)***	-0.179 (0.081)**
Logposr squared	-0.016 (0.003)***	0.001 (0.012)
% nfbr	-0.076 (0.010)***	0.157 (0.040)***
slip window	-0.122 (0.032)***	0.643 (0.151)***
slack window	-0.075 (0.047)	0.134 (0.224)
Stone window	-0.314 (0.040)***	0.631 (0.171)***
Constant	-0.717 (0.025)***	-5.316 (0.144)***
Observations	10,0761	10,0761
Number of userid	9,655	9,655

Standard errors in parentheses
*** significant at 10%; ** significant at 5%; *** significant at 1%**

Most of the univariate effects noted in Table 6 hold up in the multivariate, random effects regression. In particular, transactions in slipping or stoning windows were more likely to get negative feedback, and less likely to get positive feedback, even controlling for the other windows and the variables that capture features of the long-term feedback profile. The impact of being in a slacking window, however, is relatively small and not

[3] The regressions are conducted using stata's xtlogit command, which allows for the specification of a random effects model. A fixed effects model, while more attractive in some ways, would be inappropriate, especially for predicting negative outcomes, as it would effectively ignore any users who had no variability in their outcomes, including users who did not get any negative feedback.

statistically significant, suggesting that much of the apparent effect of being in a slacking window was a spurious attribution that more properly should be attributed to the other variables.

Again to illustrate the model's predictions, consider a hypothetical user who has 100 prior feedbacks: 99 positive and 1 negative. The predicted probability of receiving a positive feedback on the next transaction is 44.6% if the transaction is in none of the windows. This probability drops to 42.8% if the transaction is in a slacking window, 41.6% if in a slacking window, or. 37.1% if the transaction is in a stoning window. For this same hypothetical user, the predicted probability of getting a negative feedback on the next transaction is 0.26% if the transactions is in none of the windows. The predicted probability increases slightly to 0.29% if it falls into a slacking window but almost doubles (to 0.48%) if the transaction is in a slipping or stoning window.

4. DISCUSSION

The data strongly support the existence of self-selection among sellers. Immediately after receiving a negative feedback, the chance of dropping out increases significantly, and fades once the seller has received even one subsequent positive feedback, though there is still a small lingering effect. Sellers with more positive feedback are slightly less likely to drop out, but the effect is quite small.

It is not clear whether sellers who drop out are responding primarily to the psychological impact of receiving a negative or to an expected economic impact on future sales. Given the relatively small impact that a single negative seems to have on profits in the studies that have estimated this, it seems doubtful that the economic impact would be sufficient to drive sellers out of the market.

In addition, if economic impacts were critical to the decision, we would expect some of the sellers who apparently drop out to simply re-register, in order to start fresh with no feedback. This should happen only when the net value of the accumulated profile is worse than the value of a newcomer's profile. Thus, we should see a marked decline in the probability of dropping out as sellers accumulate more positive feedback. The actual effect of positive feedback on the probability of dropping out, however, was very modest, lending additional support to the idea that the psychological impact of a negative is more important than its economic impact.

While seller self-selection may be valuable to preserve the overall trustworthiness of the marketplace, from eBay's perspective it may be that the selection process is convincing too many sellers to refrain from participating. That is, good sellers may be dropping out of the system because negative feedback makes them feel unappreciated. This may explain, in part, why eBay encourages buyers to try to resolve problems with sellers directly before posting negative feedback on the system—fewer negative feedbacks may keep sellers with thin skins continuing to list items.[4]

[4] Of course, eBay may also discourage negative feedback in order to preserve an appearance that most buyers are satisfied most of the time.

For sellers who remain in the marketplace, negative feedback leads to more negative feedback. Part of this change seems to be due to slipping, a temporary underlying change in the seller's quality that leads both to the observed negative and a higher probability of negatives on other nearby transactions. Another part of the change seems to be due to stoning, a change in how buyers respond to sellers.

Our initial hypothesis was that stoning would only increase buyers' willingness to provide negative feedback; they would speak up in situations where they otherwise might have submitted no feedback. The analysis, however, shows that transactions in the stoning window were also far less likely to receive positive feedback. The most plausible explanation is that seeing a negative in a seller's profile may color a buyer's interpretation of everything that happens: not only are they likely to interpret ambiguous singals as indicators of bad behavior, but they may be more prone to noticing small problems that do not merit negative feedback but dissuade them from giving positive feedback.

Once slipping and stoning and the seller's prior reputation are accounted for, transactions in a slacking window do not have a significant impact on transaction outcomes. This suggests either that sellers do not change their quality level after receiving negative feedback, or that some users slack while others try to salvage their reputations, leaving no apparent net effect.

One caveat in interpreting our results is that the windows we defined based on the hypotheses of slipping, stoning, and slacking may really be picking up some other effects that differ between the windows. The results are consistent with our hypotheses of slipping and stoning, but do not rule out other possible explanations.

Another caveat is that, even if the effects exist, we may not have correctly classified transactions with respect to the true windows for those effects. For example, the seven day length for slipping windows was chosen arbitrarily. Perhaps seller quality actually declines during a longer window, which would overlap more with the other windows, especially the stoning windows. Thus, we might incorrectly attribute to stoning what is really a slipping effect. To check this, we recalculated longer slipping windows of 14 and 20 days and re-ran the regressions. The coefficients and standard errors barely changed, even with the 20-day slip windows, which included 63% of the transactions in stoning windows.

Another measurement error may result from the stoning window classification of transactions that did not receive any feedback. As noted above, we classified such transactions based on whether a recovery period was in effect 21 days after the transaction's close. But a buyer might have considered and decided against giving feedback somewhat earlier or later than that, and been influenced by the seller's then-current profile. If, for example, too many no-feedback transactions were classified into stoning windows, it would artificially inflate the magnitude of the stoning window's effect on positive feedback. But in that case it would tend to deflate the stoning window's effect on negative feedback. Thus, it is very unlikely that both of the effects of the stoning window are spurious results of our classification of no-feedback transactions.

5. CONCLUSION

The economic theory underlying reputation systems is well-understood, at least if participants are assumed to be rational actors. A seller's feedback profile acts as a signal about the quality of their future transactions, through some combination of indicating their underlying type and the strategic actions they will take. Buyers' responses to those profiles create a sanction for sellers who do get negative feedback. This deters bad behavior and causes those sellers who cannot meet the expectations of buyers to self-select out of the system.

In practice, both buyers and sellers may be acting more on emotion and less on calculation than the economic theories account for. Some motivations not contemplated in conventional utility models, such as a desire to punish wrongdoers even if one does not gain personally from the punishment [7], can lead to stoning, which helps to make a reputation system more robust. Our evidence suggests that stoning is a real phenomenon at eBay.

Theoretical models provide upper bounds on the amount of cooperation that is possible among purely self-interested actors. It may be that designs that take into account predictable actions that are not self-interested can lead to even more efficient market outcomes. For example, if buyers will punish sellers who receive a negative feedback even if the expected future behavior of those sellers is no different from that of other sellers, then in equilibrium sellers can continue to provide the same quality rather than slacking off after receiving a negative feedback.

On the other hand, psychological factors can also make a reputation have unintended and undesirable negative consequences. For example, if generally high quality sellers provide worse performance after a negative feedback because they are discouraged, or drop out entirely, efficiency declines. Such consequences may be a necessary evil in order to deter the entry of bad actors, and to encourage effort from good sellers. Our data, however, suggest that any slacking effect at eBay during the period of our study was small relative to the slipping and stoning effects.

In order to understand and improve the design of reputation systems in the field, it is important to understand how they are actually functioning. Empirical analysis of the kind provided in this paper is an important first step.

6. ACKNOWLEDGMENTS

This material is based upon work supported by the National Science Foundation under Grant No. IIS 0308006. We thank eBay for providing the data used in this analysis.

7. REFERENCES

[1] Akerlof, G.A., The Market for 'Lemons': Quality Uncertainty and the Market Mechanism, Quarterly Journal of Economics, 1970, pp. 488-500.

[2] Bajari Pat and Hortacsu Ali, Economic Insight from Internet Auctions, Journal of Economic Literature , v.42, June 2004, p. 457-486

[3] Bakos, Y. Reducing Buyer Search Costs: Implications for Electronic Marketplaces, Management Science, Volume 43, Number 12, Dec.1997.

[4] Cabral, Luís M B and Hortaçsu Ali. The Dynamics of Seller Reputation: Theory and Evidence from eBay. CEPR Discussion Paper No. 4345. April 2004.

[5] Dellarocas Chrysanthos . Analyzing the economic efficiency of eBay-like online reputation reporting mechanisms. Proceedings of the 3rd ACM Conference on Electronic Commerce. 2001.

[6] Diamond, Douglas W, Reputation Acquisition in Debt Markets, Journal of Political Economy, University of Chicago Press, vol. 97(4), pages 828-62. 1989

[7] Fehr Ernst, Gachter Simon ,Altruistic Punishment in Humans, Nature, 10 January 2002, 137-40.

[8] Kollock, P., The Production of Trust in Online Markets, in Advances in Group Processes, Vol. 16, edited by E.J. Lawler et al, JAI Press, Greenwich, CT, 1999.

[9] Resnick, Paul and Richard Zeckhauser. Trust Among Strangers in Internet Transactions: Empirical Analysis of eBay's Reputation System. The Economics of the Internet and E-Commerce. Michael R. Baye, editor. Volume 11 of Advances in Applied Microeconomics. Amsterdam, Elsevier Science. 2002.

[10] Resnick Paul, Zeckhauser Richard, Swanson John, Lockwood Kate. The Value of Reputation on eBay: A Controlled Experiment. Working Paper , available online at www.si.umich.edu/~presnick/papers/postcards/. 2004.