How Do People Respond to Reputation: Ostracize, Price Discriminate or Punish?

Kay-Yut Chen Scott Golder Tad Hogg Cecilia Zenteno

Hewlett-Packard Labs, Palo Alto, CA

Abstract

We evaluated how people use reputation in a laboratory market and in Prosper, an online microfinance business. We found people use information on past behavior to ostracize previous poor performance in both cases. The laboratory market did not show significant price discrimination, but people used their ability to not fulfill contracts to punish poor performers. Price discrimination was significantly correlated with reputation in Prosper. Thus we find people apply multiple strategies to deal with reputation.

1 Introduction

Reputation in commerce helps ensure promised actions are taken without the expense of external enforcement [8]. eBay's feedback system is an example of reputation for e-commerce. Another example is in microfinance, where reputation helps people decide whom to loan money to. Establishing reputation has been studied in several contexts, such as the iterated Prisoner's Dilemma [1], both theoretically and empirically [4, 2, 7, 3]. These studies showed how information on past behavior helps reduce cheating and establish efficient markets, and mainly focus on market aggregate behaviors.

However, examining how individuals use this information is more relevant for designing web sites using reputation than aggregate behavior. In this paper we study how individuals use reputation information. We examine two contexts: 1) a set of laboratory experiments that mimic an eBay-like marketplace; and 2) field data from Prosper, an online peer-to-peer microfinance site.

We chose these two contexts for several reasons. First, reputation facilitates economic transactions in both contexts: people can choose not to fulfill their promised obligations and the system provides users with reputation information, i.e., summaries of other people's past behaviors. These environments allow ostracism and price-discrimination.

Second, the contexts have different underlying economics, providing generality for our observations. Specifically, the experimental market is two-sided while Prosper is only one-sided, with different available strategies. For example, direct punishment is not possible in Prosper but is available in the lab-based market. Furthermore, isolating the effect of price discrimination is more difficult in the two-sided case because it is necessary to control for the reputations of both the buyer and the seller. Moreover, people must build their reputation from scratch in the experiments while Prosper uses existing credit scores from the beginning. This difference is particularly relevant for developing e-commerce among people with no prior transaction history.

A third motivation for considering these two contexts is their differing strengths and limitations from a research perspective. Specifically, lab-based experiments have few unobserved variables, but involve only small groups over short periods of time. The field data from Prosper, on the other hand, involves many people over longer time periods, but is less controlled (e.g., we do not know the supply and demand profiles of the participants). Moreover, in Prosper new users continually join the system as lenders and borrowers while the participants remain the same over the course of a single experiment.

Unlike larger financial markets dominated by institutions, Prosper focuses on lending by individuals, corresponding to the e-commerce context of peer-to-peer transactions. To achieve high returns on Prosper, lenders must evaluate individual descriptions on the rationale for the loans, not just the overall credit score. This information helps lenders find situations for small-scale loans that the credit rating agencies have mispriced. Evaluating these details requires significant effort for each small loan decision. So identifying ways to automate evaluating this reputation information could benefit these microfinance decisions. By contrast, large, liquid financial markets have far larger trades, much more available information, and enough money to support many analysts. Thus studies of individual behavior on peer-to-peer microfinance sites has more scope for improving use of reputation than transactions involving large institutions.

In the remainder of this paper we first describe the laboratory experiments and Prosper microfinance system. We then describe how people used reputation information and discuss design implications.

2 The Reputation Contexts

2.1 Experimental Market

We use laboratory experiments [3] with a broad set of endogenous market choices to study how people use reputation. This market, created with standard experimental techniques [11], exchanged a single homogenous good through a double-sided auction in which both buyers and sellers could post new offers or accept an existing one. Participants can explicitly choose whom they wish to do business with in two ways. First, they can specify a filter with each offer, limiting who can accept it. Second, and more easily, they can simply refuse to accept offers posted by those with whom they do not wish to do business. These techniques also enable price discrimination, i.e., by only allowing high-reputation people to accept an offer made at a favorable price, or by only accepting an offer from a low-reputation person if the price is especially favorable.

At the conclusion of each period of the market, all participants decide which of their contracts, if any, to fulfill. Our experiments include noise: a preannounced probability (10%) of loss "in transit" for each delivery. This noise models, for example, the situation in a single transaction where the intention to pay late cannot be distinguished from an on-time payment delayed in transit. We conducted several experiments with groups of between 12 and 16 people, divided into buyers and sellers. Market prices converged reasonably well to equilibrium within 3 periods, as expected from prior studies [11]. Thus we could study reputation in the context of a rapidly equilibrating underlying market. The experiments compared three policies on historical information revealed to participants:

- Low information: People had historical information about only their own transactions.
- High information: People had historical information about all transactions between any buyer and any seller, summarized by the total value of contracts signed and value-weighted fraction delivered.
- Self-reported ratings: After contract fulfillment, people rated each of the other parties to their transactions (giving either a positive or negative rating). After everyone submitted their ratings, each person's average rating was made available to all participants, summarized by the total value of contracts and value-weighted fraction of contracts receiving a positive rating.

In the Self treatment, about 90% of the ratings corresponded to whether the good or payment was received, making the information available in the Self treatment close to that of the High treatment.

All experiments exhibited strong end-game effects: toward the end of an experiment, participants have little benefit from maintaining a high reputation and contract fulfillment rates declined significantly. We found about 10 periods in each experiment minimally affected by the end-game, providing an indication of the effects and dynamics of reputation likely to arise in the context of a long series of repeated transactions. We used this portion of the data for the analysis reported in this paper.

2.2 Prosper Microfinance System

Prosper.com is a peer-to-peer, social lending system. In Prosper, individuals seeking to borrow money initiate loan requests in the form of reverse price auctions; other individuals seeking an investment can bid on those auctions [12]. This system eliminates the bank as a middleman, and the borrower receives the best interest rate the market will provide. While many lenders have invested small amounts of money, some have invested hundreds of thousands of dollars [6].

Lenders and borrowers sign up with personal details, including social security number, that verify members' identities. Prosper auctions contain the potential borrower's credit history, including a list of delinquent accounts, bankruptcies, open credit lines, and credit score, given as a grade: AA, A, B, C, D, E, HR (high risk), and NC (no credit). Lenders use this information, along with a written description of the need for the loan, to assess the potential borrower's creditworthiness. The interest rate the bidder is willing to accept is one way to measure and compensate for the perceived risk associated with lending to a given borrower.

Prosper is a one-sided market. Lenders choose which borrowers to lend to, as well as the interest rate they will offer on each loan, by placing a bid on an individual loan request and specifying the dollar amount and interest rate. Borrowers have no control over whose money they accept. In practice, it does not matter; all bids are accepted so long as they are at or below the maximum interest rate the borrower will accept, and the lowest bids win to fund the loan. In general, a loan is funded by a combination of bids from different lenders, each contributing a portion of the full loan. Repayment terms are defined by the system, which acts as an intermediary, and there is no interaction between borrower and lender following the end of the auction.

Higher-risk borrowers paid higher interest rates than lower-risk ones [10]. A high credit grade, verified bank account, and lengthy profiles were positively associated with receiving more bids on one's loan auction, indicating Prosper functions like other credit markets, where interest rate reflects credit risk.

Prosper provides information[1] about listings (i.e., requests for loans), loans, bids, members and groups. In this analysis, we focus on listings, bids and members. The listings data set contains all the public information about the listings, including the potential borrower's ID, the amount of money requested, how many bids they received, their credit grade, debt-to-income ratio, and other information.

We examined activity between November 2005 and December 2006. 28,946 users placed 65,340 listings, of which 5,141 became loans, 31,150 expired without being fully funded, and 26,568 were withdrawn by the potential borrower before completion. The remaining 2481 listings were active and not yet funded at the end of the sample period. The bids data set contains the amount of the bid and the IDs of the bidder and the listing. If the bid was ultimately outbid, the minimum rate the bidder would have accepted is given. There were 514,952 bids made; 205,447 were outbid and 309,505 were winning.

3 Reputation Measures

For the laboratory experiments, we created a reputation measure based on information available to participants. This measure is a hypothetical function that may help explain decisions, not unlike utility functions in economics. The measure we used is the Bayesian beliefs of the probability of fulfillment given information observed about past fulfillment behavior. We assume participants' beliefs include the prior is the maximum entropy distribution (i.e. uniform) when there are no observations, fulfillment choices are independent and focus on the expected value of the resulting distribution. We analyzed buyers and sellers separately. This procedure gives the expected value of person j's probability of fulfillment to i as

$$P_{i,j} = \frac{1}{1-\nu} \frac{\beta(1-\nu, 2+s, 1+n-s)}{\beta(1-\nu, 1+s, 1+n-s)} \qquad (1)$$

where $\beta(z,a,b) \equiv \int_0^z t^{a-1}(1-t)^{b-1} dt$ is the incomplete beta function, and the observations n (number of contracts signed) and s (number fulfilled and not lost to noise) depend on the pair of people involved and $\nu = 0.1$ is the noise level.

We computed $P_{i,j}$ for each pair using the cumulative n and s data appropriate for each treatment. In all treatments users knew their personal histories of transactions with other players. Using the n and s values for these transactions gives the *Personal information* measure. The High and Self treatments also provided aggregate information, i.e., a summary of each person's fulfillment in all their contracts. Thus for these treatments we also computed an *Aggregate information* measure, where n and s include all contracts or reports for that person (in the High and Self treatments, respectively).

For Prosper, we use the borrower's credit rating as a measure of their reputation, and the loan's interest rate as the price of the loan.

4 Results

This section reports our observations on how people respond to reputation. We focus on three strategies. First, people ostracize those with low reputation by refusing to do business with them. Second, people respond to low reputation by offering worse terms for deals (e.g., higher interest rates for loans). Third, people reciprocate poor behavior as a form of punishment, and is only available in the two-sided market.

4.1 Ostracism

We found ostracism in both Prosper and the experimental market.

In the experimental market, we say a person *filters out* another one when he makes no offer to and does not accept any offer from that person. In the Low treatment, 15 out of 31 users filtered out someone at least once. For High and Self treatments these were 39 of 48 and 24 of 32, respectively. Thus about 70% of people filtered someone out at least once. The reputation of those filtered in was significantly higher than those filtered out for High and Self treatments with personal information (Table 1). Fig. 1 shows how this ostracism varies with reputation for the Self treatment. We found less filtering using aggregate information, which is only available for the High and

[1] http://prosper.com/tools/DataExport.aspx

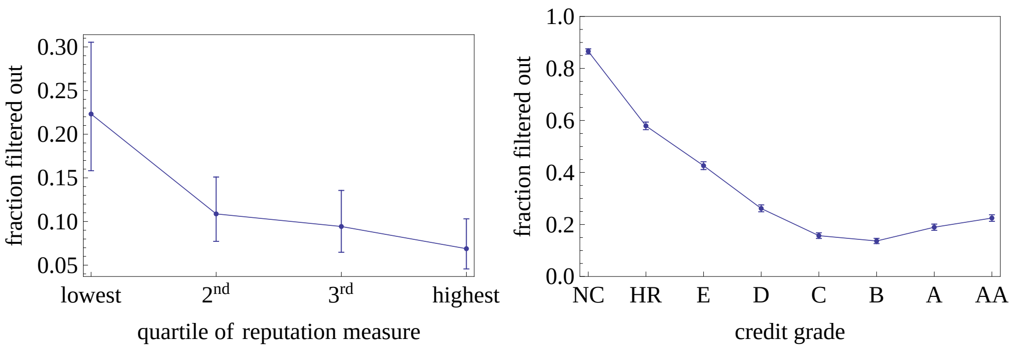

Figure 1: Filtering: proportions of offers filtered vs. reputation measure. Left: experiments based on quartile of Personal information reputation measure for the Self treatment. Right: Prosper based on credit grade. Error bars indicate the 95% confidence intervals for the estimates assuming independent samples.

Table 1: **Filtering.** Comparison of Personal information reputation measures for those filtered in and out. The p-value (Wilcoxon test) indicates whether the median value for those filtered in is larger than for those filtered out.

	median $P_{i,j}$ value		
treatment	filtered in	filtered out	p-value
Low	0.67	0.65	0.25
High	0.61	0.50	$< 10^{-3}$
Self	0.67	0.54	10^{-3}

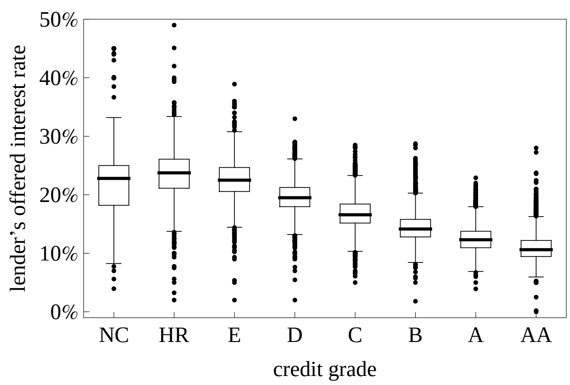

Figure 2: Price discrimination in Prosper. Average interest rates offered to each credit grade, for loan requests with at least one bid. The box extremes are the first ($Q1$) and third quartiles ($Q3$); the line in the box is the median. The points are outliers: values more than $1.5 \times (Q3 - Q1)$ away from $Q1$ and $Q3$, respectively. The horizontal tic marks, connected to each box with vertical lines, show the minimum and maximum values that are not outliers.

Self treatments. The strong reaction to personal information may arise from social preferences such as fairness and reciprocity [9].

To measure filtering in Prosper, we counted the number of lenders that made no bids for each credit grade. We compared these proportions, shown in Fig. 1, with pairwise two-sided binomial tests. All differences were significant: the largest p-value is 0.015, corresponding to comparing filtering for credit grades "B" and "C". Also, "B" was the least filtered grade, followed closely by grade "C". The active filtering of low grades matches the filtering behavior seen in the experimental market results. In both cases people significantly discriminate against those with the lowest reputation measures and make less distinction among those with moderate or high reputations. Lenders typically participate in multiple loans (a median of 23 bids) and loan to multiple credit grades: only 1.6% of lenders restrict loans to a single grade and only 6.1% to just two grades.

4.2 Price Discrimination

We found clear evidence for price discrimination in Prosper, but not in the experimental market.

For the experimental market, we regressed the price of each offer at each round on the offeree's[2] reputation, as well as the offerer's reputation as a control. We found no noticeable price discrimination in any treatment.

Prosper has clear evidence of price discrimination (Fig. 2). We regressed the interest rate on the credit grades, finding consistent and significant results (with p-value less than 2×10^{-16} for all grades): the better the credit score, the lower the interest rate.

With both filtering and price discrimination in Prosper, do people maximize expected return, e.g., by giving more loans to medium credit scores instead of

[2]If the offer had more than one offeree, we used the corresponding number of data points.

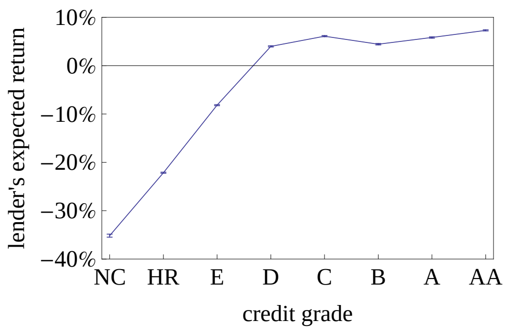

Figure 3: Expected return, as an annual percentage rate, in Prosper for loans to each credit grade. The small error bars indicate the standard error of the estimate based on variation of interest rates within each credit grade.

the highest scores because the higher return compensates for the increased risk? To address this question, Fig. 3 shows the expected return for each credit grade. Expected return is the average value of the repaid loan (e.g. 108% on an 8% loan) times the payment likelihood[3], minus the initial investment (100%).

The expected return decreases with the credit score, even though the interest rate charged increases (see Fig. 2). Thus many users make loans that, in aggregate, will not be repaid. Why this happens is an open question. One possibility is lenders overestimate their ability to select loans mispriced by the credit score, e.g., based on "sob stories" in the listings. Another possibility is new lenders are enticed by high interest rates for low credit grade borrowers and discount the default risk until they personally experience defaults. Moreover, lenders may lack the data they need to make sound financial decisions (and therefore better reputation mechanisms are necessary). To succeed, peer-to-peer lending systems must eventually have profitable lending all categories, either by trending toward higher interest rates for (or excluding) unprofitable borrowers.

4.3 Punishment

Social norms for cooperative behavior can be enforced through punishment, even if the punisher incurs some cost [5]. With the two-sided market, people can punishing poor behavior by not fulfilling contracts, although at a cost to themselves in terms of a lower fulfillment score, i.e., lower reputation. We found they do so: people tend not to fulfill contracts with those who didn't fulfill prior contracts [3]. This behavior is consistent with the social preference of reciprocity where a negative action is met with another negative action. A game theoretic interpretation indicates punishment strategies are used in repeated games to ensure cooperation via a "tit-for-tat" behavior [1]. However, punishment appears to be mainly a short-term response to others' behavior. Specifically, based on cumulative fulfillment history, Table 2 shows some difference in mean values of reputation between those whose contracts were and were not fulfilled, but the difference is only significant in the Self treatment.

Table 2: **Punishment.** Difference between the mean value of reputation of people whose contracts were fulfilled vs. those whose contracts were not fulfilled, using reputation based on Personal information. The p-values are from the Wilcoxon test of the difference.

	mean value		
treatment	fulfilled	not fulfilled	p-value
Low	0.58	0.55	0.33
High	0.58	0.56	0.67
Self	0.64	0.60	0.01

5 Discussion

We examined how people use reputation information in laboratory experiments and in the Prosper microfinance business. As a caveat on our results for the experiments, the reported p-values assume independent choices. Correlations within an experiment would reduce the effective number of independent points, leading to somewhat larger p values than we report. We address this issue to some extent by using several separate experiments.

Ostracizing is a response to low reputation in both scenarios. We found evidence of price discrimination in Prosper but not in the experiments. This difference may be because Prosper is a one-sided market while the experiments were two-sided. Specifically, in the one-sided market (Prosper), prices can be separated by borrower's reputation monotonically (i.e., higher reputation gives lower interest rate). However, in a two-sided market, the price depends on the reputation of both the buyer and the seller, leading to complicated decisions for users.

Our results suggest people use multiple reputation

[3]Determining the default rate is challenging: a loan that has not yet defaulted may default in the future, leading to an underestimate of eventual default rate. To reduce this error, the values in Fig. 3 were calculated using the default rates for the loans in our dataset as of Jan. 2008. Moreover, a loan may be repaid for a while before defaulting so the lender receives partial repayment and, possibly, some further return from the collection process (increasing the actual return).

strategies. Thus models that only consider single strategies (e.g. tit-for-tat) do not fully describe reputation use in general e-commerce settings.

The results have implications for web-based reputation tools. First, to support the ostracizing strategy, the tools should allow users to easily choose whom they want to do business with, such as automatically filtering out offers from people with reputations below a user-specified threshold. Such filtering tools are provided, for instance, by Prosper. Since people distinguish more among low than moderate or high reputations (e.g., Fig. 1), these tools could highlight those with fulfillment history significantly below most of the population. Second, to facilitate price discrimination, the web site should enable setting prices based on reputation and facilitate price discovery (i.e., how prices vary with users' reputations). Third, tools could help reduce the cost of punishment strategies, which somewhat counteract ostracizing since a person must accept a contract with a low-reputation trader in order to punish. With only aggregate fulfillment information available, other traders cannot easily distinguish a punishment strategy (which they may support) from not fulfilling contracts in general. Thus reputation mechanisms should facilitate users making their intentions clearer, e.g., by providing information on the past behavior of those whose contracts are not fulfilled. Reputation information should also match the incentive structure. One example is addressing the temptation to fulfill less on higher value contracts by showing fulfilled contract *values*, not just *number* fulfilled.

We found people responded more to personal experience than aggregate information in the lab experiments. Relying mainly on personal experience would not apply to e-commerce involving many small players, including Prosper, where repeated business opportunities are rare. The emphasis on personal experience suggests reporting information in a more personalized way, e.g., with comments, may be more effective than simple aggregates.

There are many fertile areas to extend our work. One issue is the kind of information people use. For example, Prosper encourages reputation building and self-policing by facilitating formation of social groups. Group leaders may have personal contact with members and vouch for borrowers. Borrowers can include a text justification and images with their loan request. An interesting question for future work is the extent to which this additional information affects the fulfillment expectations of lenders.

Another area for further study is the difference between one-sided and two-sided markets. We hypothesize the different results with respect to price discrimination were driven by the differences in market organization, despite other significant differences between the two scenarios. New experiments would allow us to address this issue by limiting fulfillment decisions to only the buyers or only the sellers.

These additional studies could aid the development of more effective web-based tools for using reputation in e-commerce, particularly situations involving small-scale transactions among many people.

Acknowledgements

We benefited from discussions with Scott Clearwater.

References

[1] R. Axelrod and W. Hamilton. The evolution of cooperation. *Science*, 211:1390–1396, March 1981.

[2] Gary E. Bolton, Elena Katok, and Axel Ockenfels. How effective are online reputation mechanisms? Technical Report 2002-25, Max Planck Institute of Economics, 2002.

[3] Kay-Yut Chen and Tad Hogg. Experimental evaluation of an eBay-style self-reporting reputation mechanism. In X. Deng and Y. Ye, editors, *Proc. of the Workshop on Internet and Network Economics (WINE2005)*, pages 434–443. Springer, 2005.

[4] Douglas V. Dejong, Robert Forsythe, and Russell J. Lundholm. Ripoffs, lemons and reputation formation in agency relationships: A laboratory market study. *J. of Finance*, XL:809–820, 1985.

[5] Ozgur Gurerk et al. The competitive advantage of sanctioning institutions. *Science*, 312:108–111, 2006.

[6] Anne Kadet. The banker next door. *SmartMoney.com*, 2007.

[7] Claudia Keser. Experimental games for the design of reputation management systems. *IBM Systems Journal*, 42(3):498–506, 2003.

[8] Daniel B. Klein, editor. *Reputation: Studies in the Voluntary Elicitation of Good Conduct*. Univ. of Michigan Press, Ann Arbor, 1997.

[9] Serge-Christophe Kolm and Jean Mercier Ythier, editors. *Handbook of the Economics of Giving, Altruism and Reciprocity*, volume 1 of *Handbooks in Economics*. Elsevier, Amsterdam, 2006.

[10] Sanjeev Kumar. Bank of one: An empirical analysis of peer-to-peer financial marketplaces. In *Proc. of the 13th Americas Conf. on Information Systems (AMCIS)*, 2007.

[11] Vernon L. Smith. *Bargaining and Market Behavior: Essays in Experimental Economics*. Cambridge Univ. Press, 2000.

[12] Aaron Steelman. Bypassing banks. *Region Focus*, pages 37–40, Summer 2006. Federal Reserve Bank of Richmond.