Can Online Trading Survive Bad-Mouthing? An Experimental Investigation

Ninghua Du   Hailiang Huang   Lingfang (Ivy) Li *

June 2012

Abstract:

Consumer ratings are crucial in creating and sustaining trust and trustworthiness in e-commerce markets. Thus, it is important to know whether online trading can survive bad mouthing among participants. We use controlled lab experiments to test whether market efficiency (measured by the percentage of successful trades) is affected by unfair negative ratings, and whether announcing the percentage of unfair ratings in the market makes any difference. We find that market efficiency is higher when rating information is provided than when no rating information is provided, even when unfair and ambiguous ratings are present. We also find that buyers behave differently when unfair rating information exists; however, no matter whether the percentage of unfair ratings is known, market efficiency is not significantly different from that in the market without unfair ratings.

Keywords: trust and trustworthiness, unfair rating, reputation systems, ambiguity, experiment

JEL Codes: C92, D03, L81

---

* Du is at the School of Economics, Key Laboratory of Mathematical Economics, Shanghai University of Finance and Economics(SHFUE), Shanghai 200433, P.R. China, email: ninghua.du@mail.shufe.edu.cn; Huang is at School of Information Management & Engineering, SHUFE, e-mail: hlhuang@shufe.edu.cn; Li is at the School of Economics, Key Laboratory of Mathematical Economics, SHUFE, email: lingfangivy.li@gmail.com. We thank Chun-lei Yang for valuable comments. Du and Li would like to thank the Economics Laboratory at SHUFE, Shanghai Pujiang Program, the 211 Project for Shanghai University of Finance and Economics, Shanghai Leading Academic Discipline Project (Project Number: B801) for support; Huang would like to thank the Natural Science Foundation of China (grant# 70901049), Shanghai Social Science Foundation (grant# 2011BGL012) for support and Peipei Guo for excellent research assistance. The three authors make equal contribution to this paper and are in alphabetical order.

1. Introduction

Global e-commerce has increased dramatically during the past decade. One important feature of online markets is that traders are usually anonymous and geographically dispersed. This increases the difficulty of legal enforcement of agreements in online markets. Thus, reputation systems have become essential mechanisms to establish and sustain trust among traders [1, 3, 5, 9, 14, 17, 18, 21, 24].

Given this importance of reputation systems, online markets constantly strive to improve their reputation systems to increase market efficiency, measured by the percentage of successful trades.[1] Nevertheless, unfair rating problems such as bad-mouthing and ballot stuffing still exist [1, 2, 11]. Sellers may provide good products but receive negative ratings due to factors beyond their control. For example, shipping companies may mishandle the item, buyers may misunderstand the seller's description of the items or be picky about the packaging, or competitors may pretend to be buyers and leave malicious negative ratings to weaken the seller's reputation. Recent empirical studies show that negative ratings have significant impacts on the probability of trade, selling price and profits [2, 7, 10, 19, 22, 25]. Therefore, it is important to empirically investigate the impact of negative distortions of reputation systems (i.e., bad-mouthing) on market efficiency. To accomplish this task, it is critical to consider the impact of unfair negative ratings on trust and trustworthiness (i.e., buying and shipping behavior in our experimental design) in the market.

[1] For example, in addition to the binary rating system where traders can leave positive, neutral, or negative feedback, eBay introduced a five-star rating systems to give traders more detailed feedback information on their trading partners in 2007.
(http://www.auctionbytes.com/cab/abn/y07/m05/i02/s02)

We use lab experiments to test whether announcing the percentage of unfair ratings makes any difference in market efficiency. We design four treatments in the experiment: no rating market (NRM), fair rating market (FRM), unfair rating market (URM) and ambiguous rating market (ARM). In the URM treatment, the participants are told the exact percentage of unfair ratings. In the ARM treatment, the participants are informed about the existence of unfair ratings but not the exact percentage of unfair ratings. Thus, the ARM treatment is designed to be closer to reality on eBay or other online markets, and the URM treatment is designed to examine whether telling traders the percentage of unfair ratings would improve market efficiency. NRM and FRM are the control treatments used to identify the impact of rating systems and the impact of unfair ratings, respectively.

Our experiments address a series of research questions. First, we inquire whether a contaminated reputation system that includes unfair negative ratings would still improve market efficiency more than the system with no rating information. We find that market efficiency improves when rating information is provided, even when unfair and ambiguous ratings are present. Second, we consider whether unfair and ambiguous ratings decrease market efficiency more than in the fair rating case. We show that, given the same rate of positive feedback for the seller, the percentage of buying in unfair and ambiguous markets is higher than in the fair market; however, these differences are not statistically significant. A third important research question is whether providing buyers and sellers the percentage of unfair ratings has any effect on their behavior. We examine this question by comparing the unfair market with the ambiguous market. We show that the previous buying experience has more impact on buyers in the ambiguous market than in the unfair market, especially when the buyer was cheated in the previous round.

The seller's behavior in the ambiguous market is not statistically different from that in the unfair market.

This paper contributes to the literature of both reputation systems and experimental economics. First, it is important for researchers and online market providers to know the impact of unfair negative ratings on market efficiency. If the unfair ratings decrease market efficiency dramatically, then we need to design mechanisms to solve for the problem. If traders can self-adjust their beliefs about shipping in the market and efficiency is not affected much, then there is less need to worry about unfair ratings in reputation systems. Second, since it is difficult to acquire field data on unfair ratings, we use controlled lab experiments to address the questions. From an empirical perspective, this paper provides experimental evidence that highlights the effect of knowing the percentage of unfair ratings on market efficiency. Our experiment data show that knowing the percentage of unfair ratings has an impact on buyers' behavior, but, more interestingly, it does not make any difference in market efficiency. This suggests that traders can always adjust their expectations well to the markets, and the reputation systems still work.

The remainder of the paper is organized as follows. Section 2 introduces the related literature and presents behavioral predictions. Section 3 reports the experimental design. Section 4 analyzes the results, and Section 5 concludes.

2. Related Literature and Behavioral Predictions

Reputation systems have been used to establish and ensure trust and trustworthiness in markets since the Middle Ages [17, 20]. Through the years, as markets have changed, so has the ability for traders to build trust. This became more complicated, as buyers and sellers no longer needed to meet face-to-face to do business. The Internet makes it easier for people who are separated by long distances and have never met before to trade. Legal enforcement is difficult, therefore online markets have developed reputation systems in which buyers and sellers leave feedback. These systems play an essential role in building trust and trustworthiness in the online market, and thus are crucial to sustaining market efficiency.

Much literature shows that a seller's reputation has an effect on her probability of sale and price, especially negative ratings [2, 7, 10, 21]. As reported by Cabral and Hortascu [7], a 1% increase in negative ratings causes a 7.5% decrease in prices; after an online seller receives her first negative rating, her weekly sales rate drops from a positive 5% to a negative 8%; an online seller's next negative rating arrives 25% more rapidly than the first one. Other than not sending the promised products, sellers may also get negative ratings due to factors beyond her control, such as problems created by the shipping companies, unreasonable buyer expectations, or malicious negative ratings from competitors. The negative ratings due to these factors are considered unfair negative ratings. Dellarocas [13] points out that the incidence of unfair negative ratings hurts market efficiency because sellers may be induced to be less trustworthy when unfair negative ratings are present.

Researchers have designed various mechanisms to solve the unfair rating problem. Conte and Paolucci [9] examine the social cognitive factors of unfair ratings. Whitby et al. [24] use a statistical filtering technique to exclude unfair ratings in Bayesian reputation systems. Dellarocas [11] proposes using controlled anonymity to avoid unfair negative ratings, and Miller et al. [21] design truth-eliciting mechanisms to promote truthful reports. Researchers also use lab or field experiments to explore various mechanisms to improve the current reputation systems [5, 6, 15].

However, there is limited empirical evidence on the impact of unfair negative ratings on market efficiency and whether traders can adjust.[2]

Using lab experiments, Du and Huang [14] show that market efficiency is not significantly different between a fair rating market and an unfair rating market where traders are informed about the percentage of unfair negative ratings. However, in real online markets, it is almost impossible for traders and market managers to know the exact percentage of unfair ratings in the market. Therefore, to make the study closer to reality, we extend the Du and Huang [14] study by considering ambiguous unfair ratings in the market. In this market, buyers are informed that unfair negative ratings exist, but not told the exact percentage of the unfair ratings.

Our behavioral predictions in this paper are based on the theoretical literature of reputation system design and behavioral literature in risk and ambiguity. Dellarocas [12] examines reputation system design in the pure moral hazard model using game theory. In the model, the probability of getting negative feedback for a high quality product is α. As the game repeats, sellers' long-term payoff will theoretically decrease with α, and, what is more, sellers are discouraged from cooperating. As Dellarocas [13] summarizes, "The incidence of 'unfair' negative ratings, thus, hurts market efficiency for two reasons: first, 'unfair' punishment directly reduces average seller payoffs; second, since sellers understand that, even if they cooperate, there is still a possibility that they might be punished, the difference in expected future profits that sellers obtain by cooperating vs. by cheating declines. This, in turn, reduces their incentives to cooperate." Based on the theory, we make the following behavioral predictions:

[2] Rice [27] also uses lab experiments to study reputation and uncertainty in online markets, but she manipulates the payoffs. In our paper, we directly manipulate the reputation scores. Rice [27] uses a subjective reputation system while we use an objective reputation system for better control.

H1: Sellers' shipping rates are lower when unfair ratings exist than when no unfair rating exists.

H2: Market efficiency (i.e., the percentage of successful trades) is lower when unfair ratings exist than when no unfair rating exists.

In the literature of risk and ambiguity, ambiguity refers to the probability that outcomes remain partially or completely unknown [16]. It is well-observed that many individuals are *ambiguity averse*, that is, they prefer to choose alternatives that imply low degrees of ambiguity [4, 8]. As shown in the literature, ambiguity aversion leads to conservative behavior. Hence we make the following behavioral prediction:

H3: Buyers in ambiguous unfair rating markets are less likely to buy than those in unambiguous unfair rating markets.

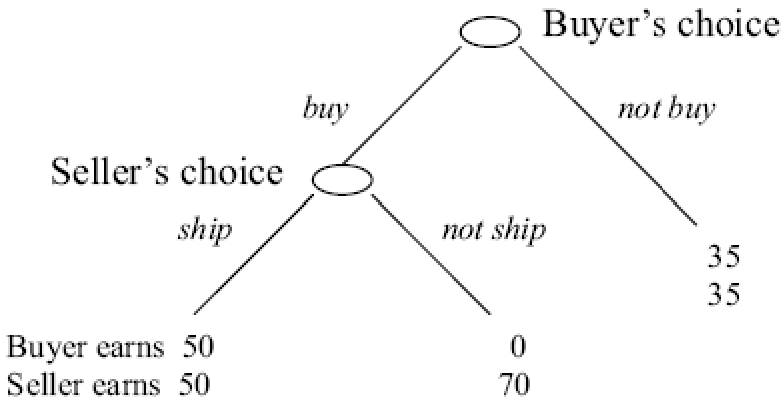
3. Experimental Design

We conducted tests at the Experimental Economics Laboratory, Shanghai University of Finance and Economics (SHUFE). The participants were recruited from a campus-wide list of undergraduate students who had previously responded to advertisements in public courses or on the web. There were 15 sessions (240 participants) total, and no participant was permitted to participate in more than one session.

All laboratory sessions were computerized using Visual Basic 6.0. Both the instructions and the information shown on the computer screen were in Chinese. There were 16 participants in each session, who played games for 30 rounds. In each round, each person had a 50% probability of being a buyer; we arranged the draws such that each person was a buyer 15 times and a seller 15 times. The buyers and the sellers were paired in each round under the commonly known restriction that nobody was matched with the same person in the same role more than once. For each participant, the role (the buyer or the seller) alternated at least once in every four rounds. The participants were anonymous to one another. Sample instructions are provided in Appendix A.

We adopt the trust game in Bolton, Katok, and Ockenfels [5] in our experiment design. The game structure is shown in Figure 1.

Figure 1. The Trust Game



In the game above, both the buyer and the seller are endowed with 35. The seller lists a product for sale at the price of 35. The product costs 20 for the seller and values 50 for the buyer. If the buyer chooses not to buy, then both the buyer and the seller keep their initial endowment. If the buyer buys, then the seller receives the price 35. Conditional on the buyer's buying decision, the seller has two choices: to ship or not to ship. If the seller chooses not to ship, then the seller receives the payment plus his endowment for a total of 70, and the buyer's payoff is 0. If the seller ships, then the seller receives the price minus the costs plus his endowment for a total of 50, and the buyer's payoff is 50 minus 35 for a total of 15. In the game, the gains from trade are realized only if the buyer buys and the seller ships. Therefore the corresponding market outcome is most efficient. However, with selfish preferences, the only Subgame Perfect Equilibrium in the game is the buyer chooses not to buy and the game ends.

It is well-established in the previous experimental literature that the framing of an experiment can affect subjects' choices. The work of Samuelson [26] suggests that the subjects confront decision problems by looking for an analogous situation in the real world and by applying the most suitable real-world behavior to the experiment. For example, if we tell a subject she is taking the role of a buyer, then she may think that as a buyer she should buy and so makes the buying choice even though she feels that buying is not optimal for her. To control for the framing effect as a potential confound, we used abstract terms in the instructions, as shown in Table 1.

Table 1: Abstract Terms Used in Experiment

Term Used in Experiment	**Meaning**
The first mover	The buyer
The second mover	The seller
Choice A	Not to buy
Choice B	To buy
Choice C	Not to ship
Choice D	To ship
Label X	Negative rating
Label Y	Positive rating
The probability of mislabeling Y with X	Negative unfair rating

In this section, we use these abstract terms to introduce treatments in our experiments.

There were three sessions for the no rating market treatment (NRM), where no rating information was given to the first mover before his decision. The first mover either clicked the button A or the button B on his computer screen. If he clicked A, the game was finished. Otherwise the game continued and the second mover clicked either C or D. The three sessions were conducted on the same day and took about 45 minutes (including the time for reading instructions). The average payment was 43.5 yuan in RMB (the exchange rate was $1 = 6.85 yuan), including a 10 yuan show-up fee. Since the average hourly wage in Shanghai for a college graduate is about 20 yuan, 43.5 yuan is a considerable amount for undergraduate students.

There were four sessions for each of the fair rating market (FRM), the unfair rating market (URM) and the ambiguous rating market (ARM) treatments. In the FRM treatment, the second mover earns a label X if she chooses C and a label Y if she chooses D. In the URM treatment, the second mover's choice C gives her a label X for sure while the choice D gives her an 80% chance of receiving a label Y, and 20% of chance of receiving a label X. These probabilities are controlled by computer and are common knowledge to the participants. The ARM treatment is mostly the same as the URM treatment, with the only difference that after choosing D the probabilities of receiving Y or X were not given in the instructions. In each of these three treatments, the first mover could (beginning in the second round) click the button "Summary Information" and/or the button "Detailed Information" before choosing A or B. Summary information provided the aggregate number of label X and the total number of label Y earned by the second mover and detailed information provided round-by-round history of labels, i.e., X or Y earned by the second mover given the first mover's choice B. Under our design, the buyer observes the seller's true shipping history in the FRM treatment, but observes a contaminated history in the URM and ARM treatments where a seller received unfair negative ratings 20% of the time. In each treatment, all four sessions were conducted on the same day, and each session took about 55 minutes. The average payments in FRM, URM, and ARM treatments were 47.8 yuan in RMB, 47.3 yuan in RMB, and 47.7 yuan in RMB, respectively, each including a 10 yuan show-up fee.

4. Analysis

In this section, we first present summaries of our data and statistical tests, then proceed to formal regressions to investigate the determinants of the observed behavior. As we shall see, both the level of trust and the level of trustworthiness are significantly and substantially higher than the no rating case when the seller's shipping history is provided, whether or not the history is contaminated by bad-mouthing, and whether or not the percentage of bad mouthing is announced in the market.

4.1. Data summary and statistical tests

Table 2 shows the average rates of trust (i.e., buying in our experiment) and trustworthiness (i.e., shipping in our experiment) in each treatment. Providing the seller's past history of shipping increases the average buying rate by 20 percentage points and increases the shipping rate by 40 percentage points compared to the NRM treatment. The shipping rate in the FRM treatment is the highest among all four treatments.

Table 2: Buying and Shipping Rates, by Treatment

Treatment	Sessions	Buying rate	Shipping rate	Market Efficiency
No Rating Market	3	.402 [.491] (720)	.333 [.484] (289)	.149 [.357] (720)
Fair Rating Market	4	.622 [.485] (960)	.729 [.432] (597)	.467 [.500] (960)
Unfair Rating Market	4	.628 [.484] (960)	.645 [.497] (603)	.430[.497] (960)

Ambiguous Rating Market	4	.623 [.485] (960)	.644 [.466] (598)	.425 [.495] (960)

Standard errors are in brackets. The number of observations in each cell is in parentheses.

The most conservative statistical tests treat each session as only one independent observation. We use the Wilcoxon-Mann-Whitney rank sum test [23] to compare the behavior across treatments (results not shown in table). Buying rates are higher in all FRM sessions than in any NRM session; this leads to a test statistic of $Z = 2.12$, which indicates that the difference is significant at $p = 0.034$.[3] Buying rates are also higher in every URM session and ARM session than in any NRM session; again, we have a test statistic of $Z = 2.12$, indicating that the difference is significant at $p = 0.034$. The pairwise comparisons across FRM, URM, and ARM show that there is no significant difference across these three treatments (with $Z \leqslant 0.15$, $p \geqslant 0.885$) in buying rates. This contradicts our theoretical prediction H3.

We find very similar patterns when we compare the shipping rates across all four treatments and the market efficiency across all four treatments. The shipping rates are all higher in FRM, URM, and ARM, when comparing with the NRM treatment ($Z \geqslant 2.12$, $p \leqslant 0.034$), while the pairwise comparisons across FRM, URM, and ARM indicate no significant difference (with $Z \leqslant 0.87$, $p \geqslant 0.387$). This contradicts our theoretical prediction H1. Since the gain from trade is realized only if the buyer chooses buying and the seller chooses shipping, we use the percentage of trades where the buyer buys and the seller ships as a measure of market efficiency. Efficiency is not significantly different across FRM, URM, and ARM treatment (with $Z \leqslant 0.58$, $p \geqslant 0.564$), but these three treatments are more efficient than in NRM treatment ($Z \geqslant 2.12$, $p \leqslant 0.034$). This observation contradicts our theoretical prediction H2.

[3] All statistical tests are two-tailed, except where otherwise indicated. All probabilities are rounded to three decimal places.

We next consider the behavioral patterns over time. Figures 2 and 3 show the patterns for buying and shipping rates, respectively, while Figure 4 shows the pattern for market efficiency.

Figure 2: Buying Rates Over Time

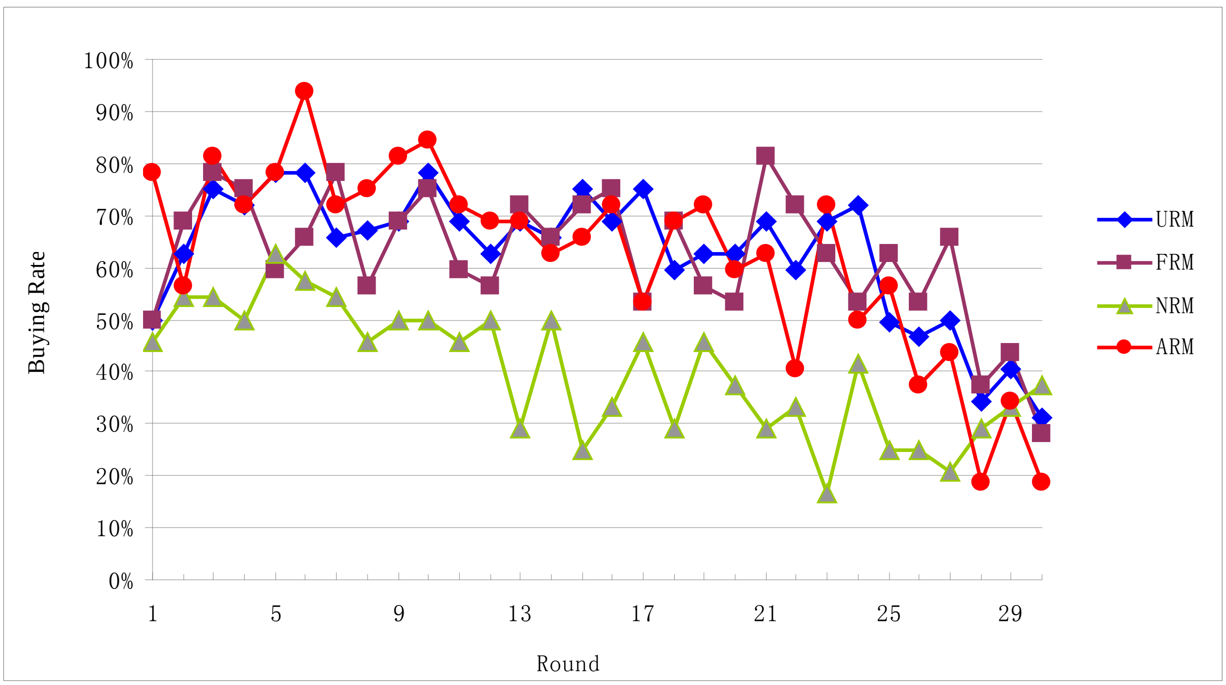


Figure 3: Shipping Rate Over Time

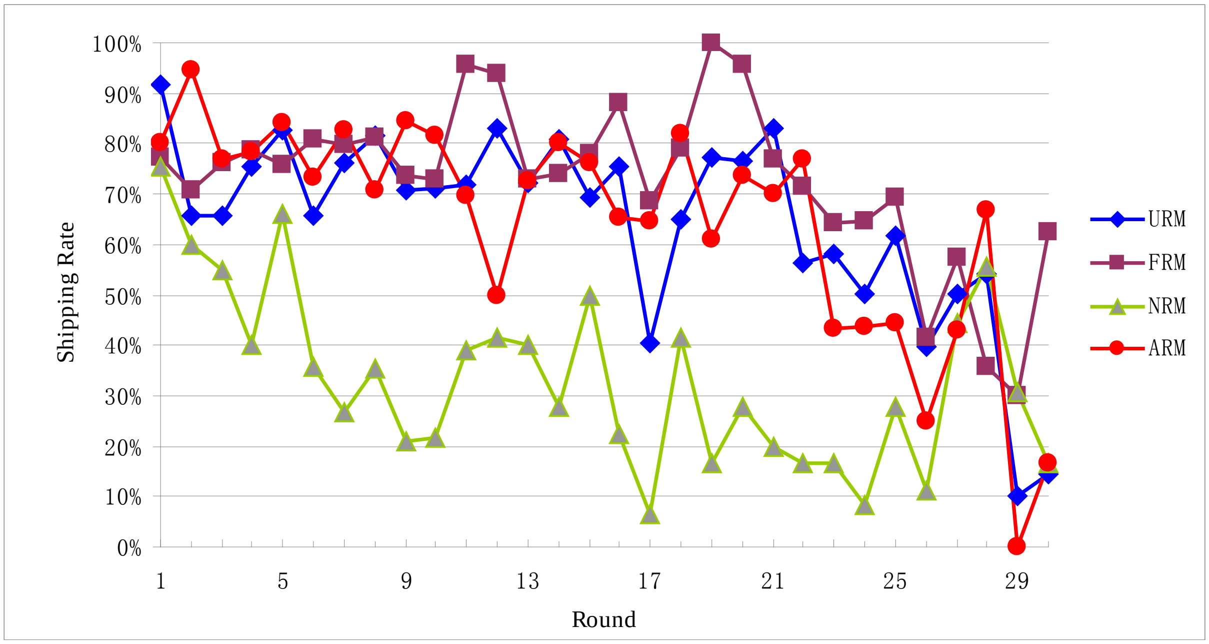
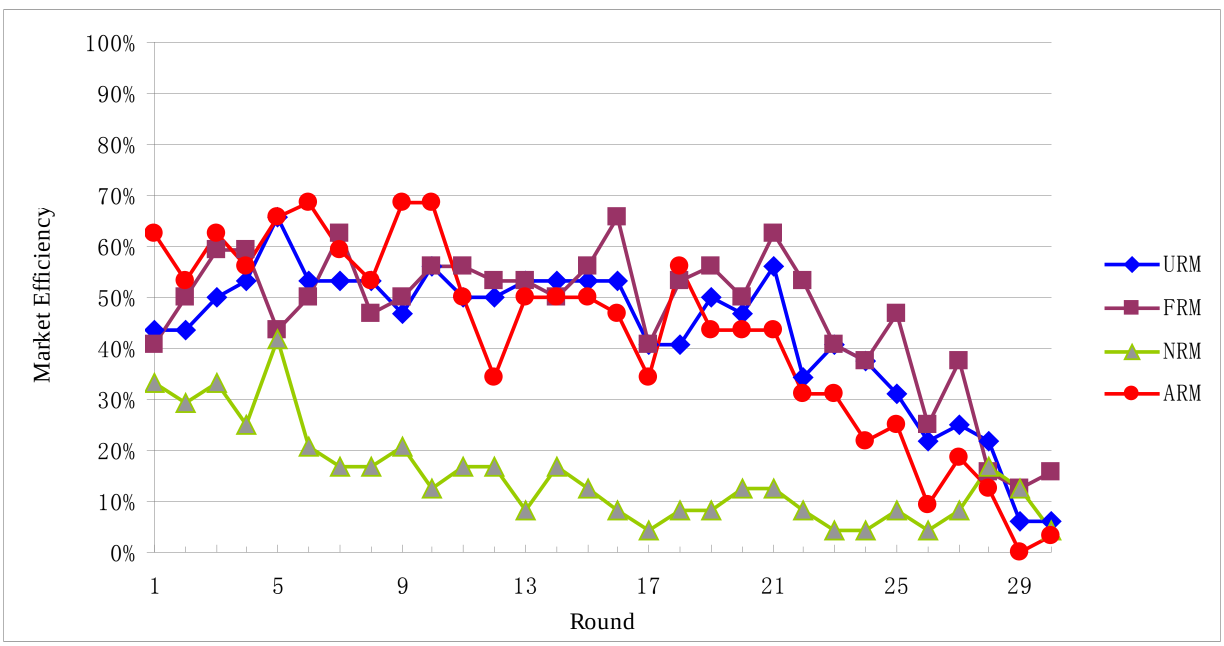


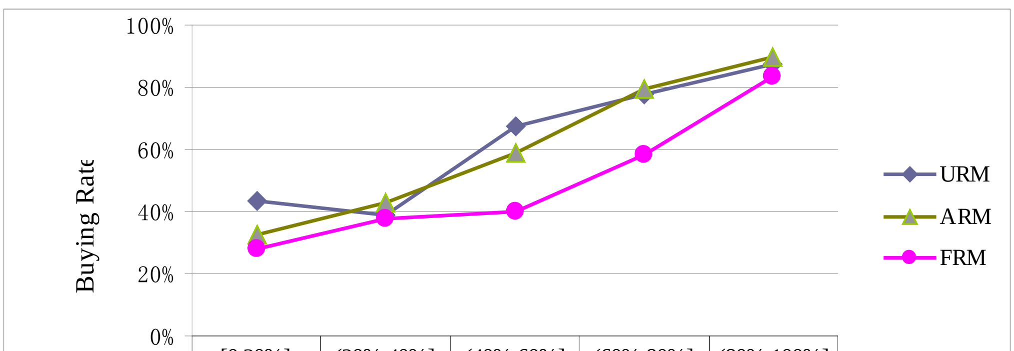
Figure 4: Market Efficiency Over Time (Measured by the Percentage of Trades Where the Buyer Buys and the Seller Ships)



We see a substantial decline over time in both buying rates and shipping rates in all treatments, and market efficiency also drops substantially. The pattern in buying rates over time is nearly identical for the FRM, URM, and ARM treatments, while the decline in buying rates is a bit more moderate in the NRM treatment. With respect to shipping rates, there is a drop in the NRM treatment in the early periods. The shipping rates in the FRM, URM, and ARM treatments decrease primarily in the ending periods. Finally, the market efficiency drops steadily in the FRM, URM, and ARM treatments, while it is always low in the NRM, particularly after the first few periods.

How responsive are buyers to the seller's shipping history? Buyers elected to check this history by clicking the button for either summary information or detailed information about 72.5% of the time in FRM, 63.8% in URM, and 73.3% in ARM, with no trend over time. Figure 5 shows the buyer's buying rates conditioned on the seller's historical rating information in the FRM, URM, and ARM treatments.

Figure 5: The Buyer's Buying Rate Conditioned on the Seller's Percentage of Positive Ratings

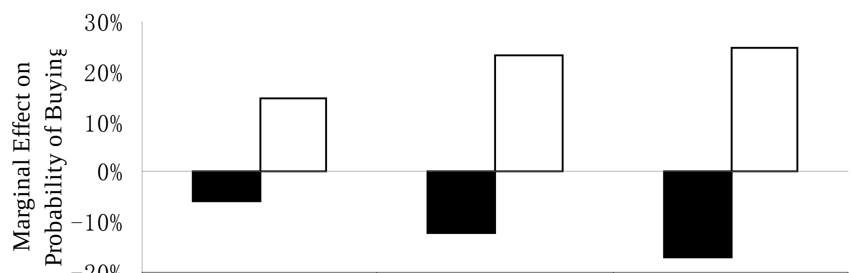


	[0,20%]	(20%,40%]	(40%,60%]	(60%,80%]	(80%,100%]
URM	43.3%	39.1%	67.3%	77.9%	87.3%
ARM	32.6%	42.9%	59.1%	79.3%	89.9%
FRM	28.0%	37.8%	39.8%	58.1%	83.5%

There is a clear positive relationship between the buyer's buying rate and the paired seller's historical rating information. Given the same percentage of positive feedback, the buyers in URM and ARM are more likely to buy from the sellers, compared to those in FRM. This reflects the fact that, as the historical rating information in URM and ARM are contaminated by unfair negative ratings, buyers in URM and ARM try to incorporate this into their choices. When a buyer in URM or ARM observes the paired seller's percentage of positive ratings, he knows that the true percentage of positive ratings should be higher. This explains the findings reported in Figure 5.

Figure 6 shows the marginal effects on the probability of buying in FRM, URM, and ARM, conditioned on the buyer's last experience of shipping decision from her counterpart.[4]

Figure 6: Marginal Buying Conditioned on the Last Experience of Shipping across Different Rating Markets



	FRM	URM	ARM
■ No Shipping	-5.9%	-12.3%	-17.0%
□ Shipping	14.8%	23.3%	24.9%

The marginal effects in FRM are relatively low compared to URM and ARM. In ARM, the marginal effect is even stronger than in URM. These observations are reasonable since in URM and ARM the ratings are contaminated by exogenous bad-mouthing, and the buyers rely more on their own recent experience. In ARM, the buyers don't even know the real percentage of bad-mouthing so they rely on their own experience even more.

[4] The marginal effect on X conditioned on event Y compares X right after event Y happens and X right before event Y happens.

According to the statistical tests, our theoretical predictions H1, H2, and H3 are not supported by the experimental data. To better understand this, we study the determinants of trust and trustworthiness in the following subsections. Since the current research does not cover why our predictions H1, H2 and H3 are not supported, we articulate these determinants in a heuristic way.

4.2. The determinants of trust: the impact of reputation information

We perform random effects probit regressions (with clustering on the individual level) regarding the determinants of buying and shipping, taking into account variables that could be expected to affect such behavior. The regressions for the determinants of buying are shown in Table 3. In these regressions, we first include all variables and then eliminate those explanatory variables that are not significant.

Note that the rate of buying decreases slightly (but significantly) over time in all FRM, URM, and ARM treatments.

Table 3: Determinants of Trust

Independent variables	FRM (1)	URM (2)	ARM (3)	Combined (4)
Constant	-0.046 (0.266)	-0.042 (0.269)	0.180 (0.315)	0.119 (0.191)
URM	--	--	--	0.220 (0.176)
ARM	--	--	--	0.221 (0.186)
Period	-0.046*** (0.007)	-0.065*** (0.008)	-0.066*** (0.007)	-0.059*** (0.004)
Detail	-0.876*** (0.191)	-0.916*** (0.176)	-0.578*** (0.157)	-0.793*** (0.100)
Exp_Shipping_Percent	0.589* (0.334)	0.989*** (0.337)	0.896** (0.363)	0.852*** (0.196)
Exp_Shipping_Last	0.195 (0.151)	0.243* (0.140)	0.178 (0.127)	0.206** (0.080)
Last_Feedback	0.364** (0.177)	0.603*** (0.185)	0.416** (0.184)	0.434*** (0.103)
FRM_PositiveRating	1.712*** (0.171)	--	--	1.740*** (0.156)
URM_PositiveRating	--	2.432*** (0.239)	--	2.300*** (0.201)
ARM_PositiveRating	--	--	1.028*** (0.204)	1.149*** (0.191)
URM_BM_Rate	--	0.175 (0.280)	--	0.096 (0.261)
ARM_BM_Rate	--	--	1.059*** (0.307)	1.012*** (0.322)
N	832	843	856	2531
LL (log likelihood)	-441.44	-403.19	-467.10	-1318.74
Wald $\chi 2$ p-value	137.47 [0.000]	166.36 [0.000]	145.55 [0.000]	443.48 [0.000]

Standard errors are in parentheses; ***, **, and * indicate significance at p = 0.01, 0.05, and 0.10, respectively (two-tailed tests). Dummy Detail is 1 when the detailed rating information is viewed and 0 otherwise. Exp_Shipping_Percent indicates the percentage of shipping that the buyer has experienced. Dummy Exp_Shipping_Last indicates the buyer's last experience of shipping decision from his counterpart (1 for shipping and 0 for not shipping). Last_Feedback indicates the paired seller's last feedback (1 for positive and 0 for negative feedback). PositiveRating indicates the percentage of positive ratings of the paired seller. BM_Rate is the percentage of bad-mouthing that the buyer suffered when he was a seller in previous rounds. The prefix FRM indicates the variable is 0 if not FRM treatment (similar for URM and ARM).

One category of determinants is related to the information variable. As already seen in Figure 5, the revealed rating information of the seller in the overall data significantly affects the buyer's buying decision. When comparing URM and FRM in empirical model specification in column (4) of table 3, the coefficient of the seller's rating history in URM is significantly higher than that in FRM (Wald χ^2=5.60, p=0.018). This shows that in URM the buyers incorporate the fact that the seller's rating history is contaminated by 20% bad-mouthing when making buying decisions. The coefficient of rating history in URM is also significantly higher than that in ARM (Wald χ^2=19.16, p=0.000). This is because the buyers do not observe the rate of bad-mouthing, thus are unable to evaluate the seller's rating in ARM. Another interesting finding is that the buyers in ARM use their own experience of bad-mouthing as an indicator of the real bad-mouthing in the market. As the percentage of bad-mouthing experienced increases, buyers become more generous in evaluating the seller's shipping history and tend to buy more. This observation does not appear in URM. The corresponding Wald test in column (4) of table 3 gives us χ^2=4.92, p=0.027.

A second category of determinants deals with the personal experiences of the buyer. The buyer's experience of shipping has a significant impact on buying in all of the regressions. Given control for experience in all of the previous rounds, the impact of the last experience of shipping is not significant in FRM and ARM, but is still significant in URM.

4.3. The determinants of shipping

The regressions for the determinants of shipping are shown in Table 4. Once again, we first include all variables and then eliminate those explanatory variables that are not significant.

Again, we observe that there is a steady decline in shipping over time in URM and ARM treatments, which reflects the trends displayed in Figure 3.

Table 4: Determinants of Shipping

Independent variables	FRM (1)	URM (2)	ARM (3)
Constant	1.030*** (0.185)	0.660** (0.162)	0.875*** (0.169)
Period	-0.004 (0.018)	-0.051*** (0.017)	-0.033** (0.017)
Exp_NegRating_Amount	-0.410*** (0.072)	-0.225*** (0.062)	-0.279*** (0.060)
Exp_Shipping_Last	0.442*** (0.138)	0.708*** (0.122)	0.531*** (0.127)
Exp_Buying_Amount	-0.045 (0.042)	0.096* (0.039)	0.017 (0.040)
Exp_Bad_Mouthing	--	0.156* (0.086)	0.208** (0.089)
N	545	549	543
LL (log likelihood)	-255.09	-295.67	-292.47
Wald χ^2 p-value	68.02 [0.000]	83.39 [0.000]	85.72 [0.000]

Standard errors are in parentheses; ***, **, and * indicate significance at $p = 0.01$, 0.05, and 0.10, respectively (two-tailed tests). Exp_NegRating_Amount is the sum of negative ratings received by the seller. Dummy Exp_Shipping_Last is the seller's last experience of shipping decision from her counterpart when she was a buyer (1 for shipping and 0 for not shipping) . Exp_Buying_Amount indicates the sum of buying that the seller has experienced. Exp_Bad_Mouthing is the sum of bad-mouthing received by the seller.

As there is no buyer's rating information available for the seller, the only relevant category of a determinant for shipping is her experience. As shown in the regressions, the sellers in all of the rating information markets behave in a similar pattern. On average, as the negative ratings a seller receives increases, the seller behavior becomes less trustworthy in all treatments. Meanwhile, in URM and ARM, when the experience of being bad-mouthed increases, sellers become more trustworthy to try to recover their own rating scores. This contradicts our theoretical prediction H1 based on Dellarocas [13]. Another interesting finding is that a seller's experience as a buyer in the last round of meeting a trustworthy seller has a positive impact on her own shipping in the current round. This reflects indirect reciprocity in all three markets with rating information: If someone cooperates with me but I have no chance to meet this person again, I will cooperate with a different person when possible.

5. Concluding Remarks

In this study, we use a controlled lab experiment to study the effect of unfair ratings on market efficiency and test whether informing buyers of the exact percentage of unfair ratings makes any difference to market efficiency and traders' behavior. Our study shows that providing the sellers' rating information of trustworthiness (i.e., shipping in our experiment) improves market efficiency. This echoes findings in the previous literature. We further show that the existence of bad-mouthing doesn't affect market efficiency in the experimental markets, even when the buyers do not know the percentage of bad-mouthing in the market and the overall rate of negative unfair ratings is constant.

A previous article by Du and Huang [14] shows that when buyers know the percentage of bad mouthing, they will fully incorporate this piece of information so that the contaminated rating mechanism still works. In this study, we further find that even when buyers do not know the bad-mouthing rate in the market, the buyers adjust their beliefs on the percentage of bad-mouthing by their own trading experience and, again, the bad-mouthing does not affect buyers' trust. This finding is limited to the case where the percentage of unfair negative ratings is constant. Contrary to our predictions, buyers' buying rates are not lowered when the percentage of bad-mouthing is not common knowledge. Since the buyers' self-adjustment incorporates the distortion of bad-mouthing, the contaminated reputation systems still work. The analysis also shows that, after receiving unfair negative ratings, the sellers behave more cooperatively to try to recover their rating scores. Overall, the sellers' propensity to cooperate in the bad-mouthing market is no different from that in the fair rating markets. Based on these results, we conclude that markets do not need to worry about the presence of unfair ratings in reputation systems since traders will self-adjust, particularly when the percentage of unfair negative ratings is constant.

References

[1] R. Bhattacharjee, A. Goel, Avoiding Ballot Stuffing in eBay-like Reputation Systems, Third Workshop on the Economics of Peer-to-Peer Systems, Philadelphia, PA, (2005).

[2] S. Ba, P. Pavlou, Evidence of the effect of trust building technology in electronic markets: Price premiums and buyer behavior, MIS Quarterly, (2002) 243-268.

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

[4] S.W.Becker, F.O. Browson, What Price Ambiguity? Or the Role of Ambiguity in Decision-Making, Journal of Political Economy, 72, (1964) 2-73.

[5] G. Bolton, E. Katok, A. Ockenfels, How Effective Are Electronic Reputation Mechanisms? An Experimental Investigation, Management Science, (2004) 1587-1602.

[6] G. Bolton, A. Ockenfels, The limits of trust, in: K. Cook, C. Snijders, V. Buskens (Eds.) eTrust, Russell Sage Inc., New York, 2009, pp.15-36.

[7] L. Cabral, A. Hortaçsu, The Dynamics of Seller Reputation: Evidence from Ebay, Journal of Industrial Economics, 58(1), (2010) 54-78.

[8] C.F. Camerer, M. Weber, Recent Developments in Modelling Preferences: Uncertainty and Ambiguity, Journal of Risk and Uncertainty, 5, (1992) 325-370.

[9] R. Conte, M.Paolucci, Social cognitive factors of unfair ratings in reputation reporting systems, Web Intelligence, 2003. WI 2003. Proceedings. IEEE/WIC International Conference. Issue Date: 13-17 Oct. (2003) 316-322.

[10] C. Dellarocas, The digitization of word-of-mouth: Promise and challenges of online feedback mechanisms, Management Science, 49 (10), (2003), 1407-1424.

[11] C. Dellarocas, Building trust on-line: The design of robust reputation mechanisms for online trading communities. In M. N. Doukidis, G. and N. Pouloudi, (Eed.) Information Society or Information Economy? A combined perspective on the digital era, IdeaBook Publishing, 2004, pp.95-113.

[12] C. Dellarocas, Reputation Mechanism Design in Online Trading Environments with Pure Moral Hazard, Information System Research, (2005) 209-230.

[13] C. Dellarocas, How Often Should Reputation Mechanisms Update a Trader's Reputation Profile? Information Systems Research, 17, (2006) 271-285.

[14] N. Du, H. Huang, The Effects and Mechanism of Unfair Ratings on Online Trading: An Experimental Investigation, Journal of Management Sciences in China, forthcoming.

[15] M. Fan, Y. Tan, A. B. Whinston, Evaluation and Design of Online Cooperative Feedback Mechanisms for Reputation Management, IEEE Transactions on Knowledge and Data Engineering 17 (2), (2005) 244-254.

[16] D. Frisch, J. Baron, Ambiguity and Rationality, Journal of Behavioral Decision Making, 1, (1988) 149-157.

[17] A. Greif, Reputation and coalitions in medieval trade: Evidence on the maghribi traders." Journal of Economic History , 49(4), (1989)857–882.

[18] L.I. Li, Reputation, Trust, and Rebates: How Online Auc-tion Markets Can Improve Their Feedback Mechanisms, Journal of Economics and Management Strategy, 19 (2), (2010a) 303-331.

[19] L.I. Li, What is the Cost of Venting? Evidence from eBay, Economics Letters, 108, (2010b) 215-218.

[20] P.R. Milgrom, D.C. North, B.R.Weingast, The role of institutions in the revival of trade: The law merchant, private judges, and the champagne fairs. Economics & Politics, 2(1), (1990) 1–23.

[21] N. Miller, P. Resnick, R. Zeckhauser. Eliciting honest feedback: The peer prediction method, Management Science, September, (2005)1359-1373.

[22] P. Resnick, R. Zeckhauser, J. Swanson, K. Lockwood, The value of reputation on eBay: A controlled experiment, Experimental Economics 9(2), (2006) 79–101

[23] S. Siegel, N. J.Castellan, Nonparametric Statistics for the Social Sciences, Boston: McGraw-Hill, 1988.

[24] A. Whitby, A. Jøsang, J. Indulska, Filtering Out Unfair Ratings in Bayesian Reputation Systems, Proceedings of the Workshop on Trust in Agent Societies, at the Autonomous Agents and Multi Agent Systems Conference (AAMAS2004), New York, July 2004.

[25] Q. Ye, R. Law, B. Gu. The impact of online user reviews on hotel room sales. International Journal of Hospitality Management, 28(1), (2009) 180-182.

[26] L. Samuelson. Analogies, Adaptation and Anomalies. Journal of Economic Theory, 97, (2001) 320–366.

[27] Rice, S. Reputation and Uncertainty in On-line Markets: An Experimental Study. Information Systems Research (forthcoming).

Appendix A: Sample instructions

Instruction (FRM)

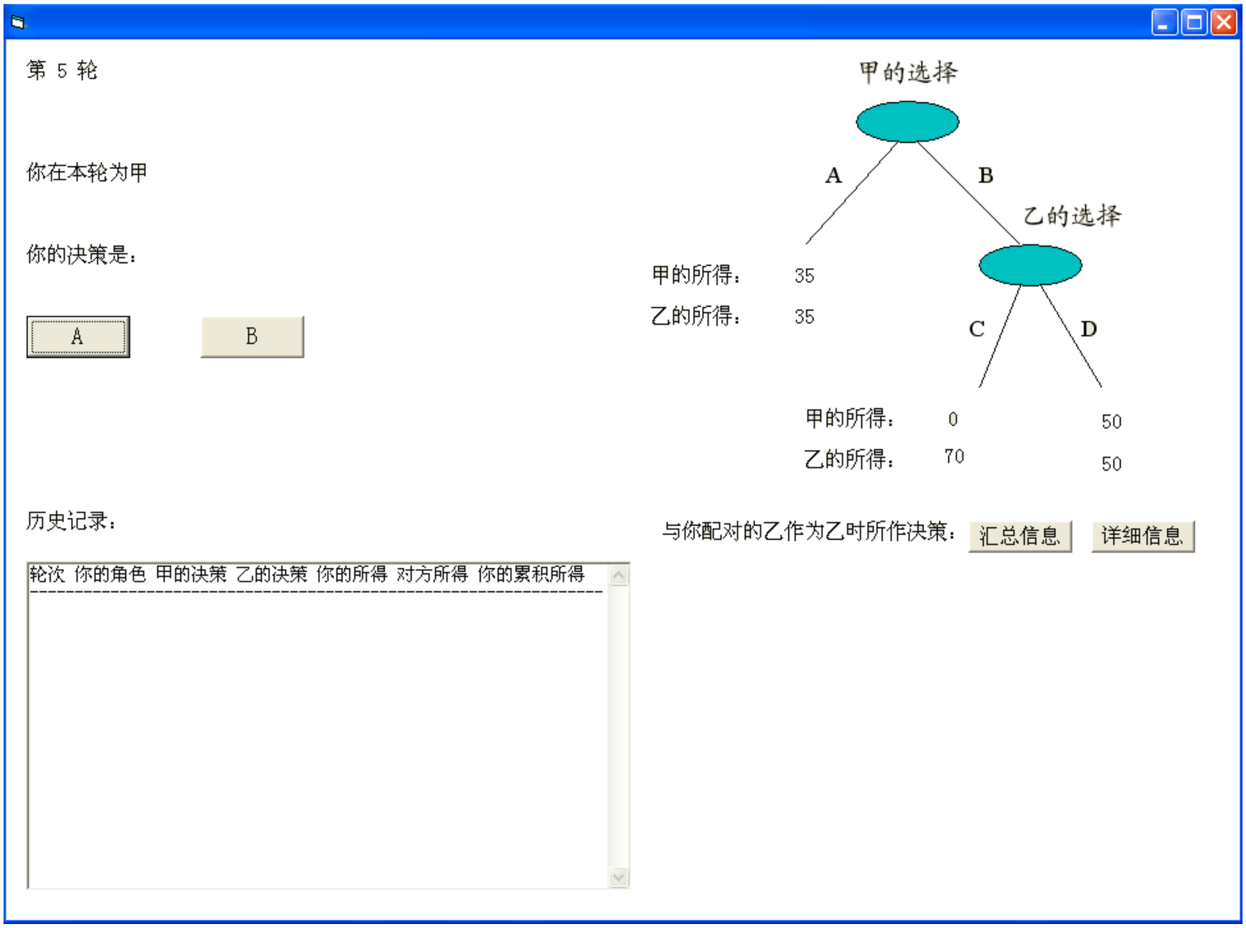
Welcome to our experiment. You will earn lab money depending on how you and others decide. At the end of experiment, you can have your lab money exchanged for RMB. You will get 3 RMB for 100 points.

There are 16 participants in this room. The 16 participants will be randomly re-matched in pairs among themselves round by round for a total of 30 rounds. At the beginning of each round, you will be randomly selected to be either the first mover or the second mover, as displayed on your screen. The matching is anonymous. However, you will take the role of the first mover and second mover at least once each in every 4 rounds.

In each round, the first mover is to first choose between A and B. If A is chosen, by clicking the button "A" under "Your decision is" as shown on the screenshot, this round ends. Both the first mover and the second mover get 35 pts. If B is chosen, then it is the second mover's turn to choose between C and D. And she will be able to click the buttons "C" or "D" under "Your decision is". If C is chosen, the first mover gets 0 pt while the second mover gets 70 pts and the second mover earns a label X for the round. If D is chosen, both the first mover and the second mover get 50 pts, and the second mover earns a label Y for the round.

Beginning in the second round, the first mover will be able to view some information about the matched second mover's past labels earned *as second mover*. As shown in the screenshot below, he or she could click on either "Summary" or "Detail." With Summary, he would see the numbers of label X and label Y his partner earned in all past rounds as second mover herself. With Detail, he will get the exact sequence of those labels with precise round information.

At the end of each round, the first-mover and second-mover decisions including the associated payoffs in the pairing will be recorded under "History" on the screen, as an add-on entry round by round.


第 5 轮
你在本轮为甲
你的决策是：
A
B
历史记录：
轮次 你的角色 甲的决策 乙的决策 你的所得 对方所得 你的累积所得
甲的选择
A
B
乙的选择
甲的所得： 35
乙的所得： 35
C
D
甲的所得： 0 50
乙的所得： 70 50
与你配对的乙作为乙时所作决策：
汇总信息
详细信息