How Effective are Online Reputation Mechanisms?

An Experimental Investigation

Gary E Bolton*
Smeal College of Business, Penn State University and Harvard Business School, USA

Elena Katok
Smeal College of Business, Penn State University and Harvard Business School, USA

Axel Ockenfels
Max Planck Institute for Research into Economic Systems, Jena, Germany

May 2002

Online reputation – "feedback" – mechanisms aim to mitigate the moral hazard problems associated with spatially distant exchange among strangers by providing traders with the type of information available in small groups, where members are frequently involved in one another's dealings. We compare trading in a market with feedback to a market without, as well as to a market in which the same people interact with one another repeatedly (partners market). We find that, while the feedback mechanism induces quite a substantial improvement in transaction efficiency, it also exhibits a kind of public goods problem in that, unlike the partners market, the benefits of trust and trustworthy behavior go to the whole community and are not completely internalized. We discuss the implications of this perspective for improving these systems.

---

* Correspondence to Gary Bolton, 310 Beam, Smeal College of Business, Penn State University, State College, PA 16870; gbolton@psu.edu. Ockenfels gratefully acknowledges the support of the Deutsche Forschungsgemeinschaft through the Emmy Noether-program.

1. Introduction

The scope of transactions on electronic market platforms makes it difficult to resolve trading disputes through standard legal channels. Trades commonly span multiple, uneven jurisdictions. Due to the scope of the necessary investigations, the costs of law enforcement often exceed the monetary stakes involved. As a consequence, many electronic trading platforms, including Amazon.com, Cnet.com, eBay.com, Half.com and Yahoo.com, have developed electronic reputation mechanisms to substitute for legal enforcement. (Others platforms get help from online services that collect and provide information about business partners.)

While online reputation mechanisms vary considerably in detail, all are based on the common understanding that providing information about a trader's past behavior − information from earlier transactions with other parties − allows the market to distinguish between the trustworthy and the fraudulent.[1] The sale of used books (along with other used merchandise) on Amazon.com's site provides a simple but illustrative example. Amazon does not actually sell used books; rather Amazon acts as a middleman for independent dealers, some brick-and-mortar bookstores but also individuals selling from private collections. Sellers state the price and describe the book's condition, and Amazon posts this information alongside its own new book offerings. Buyers pay through Amazon, who takes a percentage, but sellers ship directly to buyers. The moral hazard problem inherent in the seller's side of the deal – stipulating the book's condition and the shipping – is mitigated through a system in which buyers are invited to post "feedback" on the transaction. Favorable comments are posted in blue and unfavorable comments in red. Buyers see a seller's feedback when deciding whether to make a purchase.

[1] The electronic production of trust and trustworthiness is fundamentally different from the mechanisms that evolved in human societies over long periods of time in traditional non-electronic interactions (see Brynjolfsson and Smith, 2000, Dellarocas, 2001, and Resnick and Zeckhauser, 2001, for a more detailed comparison). There is, however, nothing new in relying on reputation based on feedback of third parties to curtail moral hazard in the marketplace. For instance, certain groups of eleventh century Mediterranean merchants screened their market partners who were often at the other side of what was then a difficult to navigate sea, through the exchange of letters reporting on who had been reliable and who had not (Grief, 1989; see also Milgrom et al., 1990).

Recent studies of online auction platforms confirm that reputation mechanisms have at least some of the desired economic effect. Reputable sellers are more likely to sell their items (Resnick and Zeckhauser, 2001), and can expect price premiums (e.g., Lucking-Reiley et al., 1999).[2] But alongside the positive evidence, there is also evidence of room for improvement. The U.S. Department of Justice reports that

> "According to the Federal Trade Commission and Internet Fraud Watch, fraudulent schemes appearing on online auction sites [most of these have reputation mechanisms] are the most frequently reported form of Internet fraud. These schemes, and similar schemes for online retail goods, typically purport to offer high-value items - ranging from Cartier® watches to computers to collectibles such as Beanie Babies® - that are likely to attract many consumers. These schemes induce their victims to send money for the promised items, but then deliver nothing or only an item far less valuable than what was promised (e.g., counterfeit or altered goods)." [Brackets added.][3]

In order to improve reputation mechanisms, we would like to better understand how they work, where they are successful and where they fall down. Towards this end, we conducted a laboratory investigation to measure the effectiveness of online reputation mechanisms and to provide some insight into the behavior that they induce. (We will return to discuss the complementary relationship between our experiment and previous field studies in a moment.) The moral hazard problem in the trading environment we study concerns shipping. A buyer chooses whether to purchase an item (at a fix price). If a purchase is made, the seller must then decide whether to ship or simply keep the buyer's money. In one market, we introduce a reputation system that tracks the seller's history of shipping decisions and provides this information to prospective buyers (reputation market). To determine the value added by the reputation mechanism we benchmark against the amount of commerce observed when no reputation mechanism is present (strangers market). Comparing strangers and reputation markets

[2] Analogous results come from Ba and Pavlou (forthcoming), Houser and Wooders (2001), Melnik and Alm (2001), and Ockenfels (2002). Brynjolfsson and Smith (2000) compared pricing behavior at 41 Internet and conventional retail outlets and also identify internet sellers' trustworthiness as one important factor that affects market outcomes.
[3] See http://www.internetfraud.usdoj.gov/#What Are the Major Types of Internet. Other evidence comes from the U.S. Internet Fraud Complaint Center (IFCC), a partnership of the U.S. White Collar Crime Center and the Federal Bureau of Investigation, that recently reported that 63 percent of the 49,711 formal complaints received in 2001 involved either non-delivery of merchandise, non-payment, or auctions, and that these numbers are rising rapidly.

provides a measure of the improvement in trust (decision to buy) and trustworthiness (decision to ship) the reputation mechanism brings about. We find that the reputation system improves trading efficiency quite a bit, although the improvement falls substantially short of 100 percent.[4]

Can we hope to do better? More specifically, can we hope to do better with a better-designed reputation mechanism – and what might such a mechanism look like? Online feedback mechanisms aim to emulate small group reputation mechanisms. Members of small groups know one another's history in substantial detail if only because they are frequently involved in one another's dealings. So the natural thing to do is to benchmark the large group reputation mechanism (reputation market) with cooperation in small communities, the smallest being situations where the same two people interact together all the time (partners market). Besides giving us a sense of the upper bound of the performance for these kinds of systems, an analysis of the differences in partners and reputation trading patterns sheds light on what the reputation mechanism does well and where it falls short.

Whereas field studies maximize external validity – they look directly at the field reputations mechanisms in question, with the least filtering possible – the comparative advantage of a laboratory study is, in a word, control. The distinction between partners, strangers and reputation markets tightly controls for the influence of repeated interaction, and also provides a measure of the performance of the reputation mechanism relative to worst case (strangers) and best case (partners) benchmarks.[5] This sort of control requires us, almost by definition, to abstract away from many of the complexities found in the field. As such, the comparison between the strangers and the reputation market will show that features common to all electronic

[4] How effective reputation mechanisms are in circumstances where there is endogenous price formation, two-way rating (both buyer and seller moral hazard issues), and self-reporting of the transaction will not be addressed here. The experimental paradigm we describe can be modified to address these issues. Also see the summary section.

[5] In addition, in the reputation market we control for the noise in feedback production (always truthfully provided in our experiment), we control the distribution of individual valuations and knowledge of these valuations in all markets, and we focus on the effect of reputation on the probability of trade keeping the price fixed across markets. While many of these factors, and there influence, are important field phenomena, interesting in their own right, and deserving of broader investigation, experiments allow us to separate the effect of each factor, and in a systematic way. Doing so in the field can be difficult; see, for example, Dellarocas (2001) on the difficulty of accounting for the online reputation mechanisms create to manipulate feedbacks and to secretly switch identities.

reputation mechanisms are *sufficient* to explain many of the important economic effects observed in the field. At the same time the comparison between the reputation and the partners market – a comparison that would be difficult, if not impossible, to conduct in the field – suggests that reputation mechanisms suffer a kind of public goods problem, in that the benefits of trust and trustworthiness behavior induced by the reputation mechanism go to the whole community and are not completely internalized. This observation suggests new directions for economic theory and new things to look for in further field studies. We will come back to this point in the conclusion section. First we describe the experiment and the data.[6]

2. Three market experimental design and some theoretical considerations

2.1 The basic market platform

The same set of buyer-seller encounters is common to all three markets of the experiment. Each time there are 16 traders interacting over a period of 30 rounds. At the beginning of each round, each trader is matched into a pair, with one person assigned the role of buyer and the other the role of seller. The matching process varies with market, as does the information the buyer has about the seller (variations are taken up below). The role assignment in all markets is random under the commonly known restriction that each trader is a buyer half the time and a seller half the time.

[6] For more discussion on the relationship between experimental and field studies see Roth (forthcoming) and Ariely et al. (2002).

Figure 1. The buyer-seller encounter

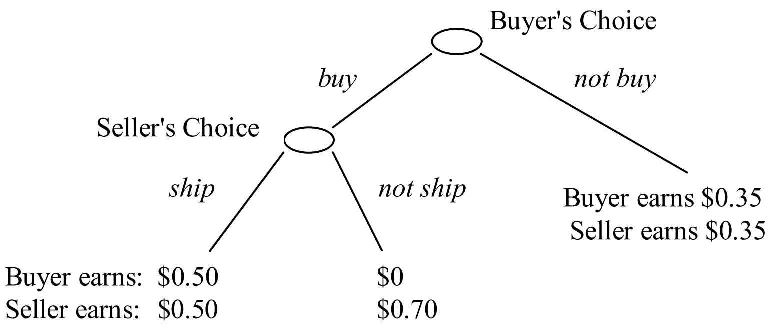


Figure 1 illustrates the buyer-seller encounter. Game moves and payoffs can be interpreted in the following way. Both the seller and the buyer are endowed with $0.35, which is the payoff when no trade takes place. The seller offers an item for sale at a price $0.35 which has a value of $0.50 to the buyer. The seller's cost of providing the buyer with the item, costs associated with executing the trade, shipping, handling etc., as well as production costs,[7] is $0.20. So each successfully completed trade increases efficiency by creating a consumer surplus of $0.15 and a $0.15 profit for the seller. If the buyer chooses to *buy* the item, he sends his endowment of $0.35 to the seller, who then has to decide whether to ship the item, or whether to keep both the money and the item. If the seller does not ship the item he receives the price plus his endowment of $0.35 for a total of $0.70. If he ships, he receives the price minus the costs plus his endowment for a total of $0.50, whereas the buyer receives his value of the item. If the buyer chooses not to buy the item, both players keep their endowment.

Under the assumption that the buyer-seller encounter is one-shot, the seller, once he receives the money from the buyer, has no pecuniary incentive to be trustworthy and to ship the item. Anticipating this, the buyer may not trust the seller, so that trade does not take place, even though it would make everybody better off. This is the basic dilemma online reputation mechanisms are designed to solve.

[7] Production costs where either the seller only produces the item once he knows the demand, or the product is produced before the buyer's decision is known but costs are not sunk (e.g., when the item can be resold at a price equal to production costs).

2.2 How the three markets differ and the experimental protocol

In the strangers market, buyers and sellers are randomly paired in each round. Buyer and seller are anonymous to one another, and no information about one another's history of action is made available. The random matching scheme is public knowledge.

Matching is done the same way in the reputation market (we used the same random pairing rotation used in the strangers market), but here buyers are given the feedback on the seller's shipping decisions prior to choosing whether to buy. The feedback includes a summary of the number of times the seller shipped in the past, as well as a round-by-round history of their shipping decisions. Appendix A includes a typical buyer's computer screen.

In the partners market, matching is fixed; that is, the same two subjects are always paired together, and this is public knowledge. The same kind of feedback information available to reputation market buyers is available to partner market buyers. In this way, the information structure is parallel across the two markets, albeit the information is redundant in the sense that it is telling the buyers things the buyer has himself experienced.

Each of the three markets consists of three sessions. There are 16 subjects per session (48 per market) for a total of 144 participants in the experiment. All sessions were conducted in March and April of 2002[8] at the Laboratory for Economic Management and Auctions in the Smeal College of Business, Penn State University. Subjects were Penn State University students, mostly undergraduates, from various fields of study who volunteered through an on-line recruitment system. Cash was the only incentive to participate. Upon arrival at the laboratory participants were seated at the computers, separated by partitions. They were asked to read the instructions. (See Appendix B for the written instructions given to subjects, and illustrative examples.) To create public knowledge, the monitor read instructions to subjects out loud, after which consent forms were signed and collected. Subjects then played several practice

[8] The partners sessions were held on 3/27/2002 1 PM, 3/27/2002 2 PM and 3/39/2002 10 AM; the reputation sessions were held on 3/15/2002 1 PM, 3/15/2002 2 PM, and 4/08/2002 4 PM; the strangers sessions were held on 3/21/2002 9 AM, 3/21/2002 10 AM, and 3/29/2002 11 AM.

games in a sequence of roles that was chosen at random, with the computer as partner making its moves at random.[9] Once familiar with the game interface, subjects played the 30 actual rounds. Upon completion of the session, each subject was privately paid his or her earnings in cash plus a $5 show-up fee. They then exited the lab.

2.3 Theoretical considerations

Overall, theory suggests reasons to believe that we will see fewer trades in the strangers market than in either the reputation or partners market, but little reason to suspect a difference in trades between reputation and partner markets. First, however, consider the simplest Nash equilibrium analysis, one that assumes complete information, and implies no trades – in any of the three markets: Consider again the game in Figure 1 and suppose that we are in the 30th and final round of play. In terms of pecuniary rewards, the seller's optimal action, regardless of the game's history, is *not ship*, and so the buyer's optimal action is to exercise the outside option. This remains true regardless of what seller feedback the buyer has, or of whether the buyer and seller are randomly matched or partnered. Backing through the game then shows that, in Nash equilibrium, there can be no trading in any of the 30 rounds: Since there is no trading in the 30th round, there is no incentive to ship in the 29th round, again independent of feedback or matching considerations, and so again the optimal action for the buyer is the outside option, and so on back to the 1st round.

There are, however, theoretical, as well as behavioral, reasons to believe that this simple analysis overstates the difficulty of trading.[10] For one, the no-trade equilibrium is not robust to minor perturbations with respect to plausible player beliefs about others' behavior and payoffs – at least for the markets were information about seller past histories is available. In a famous paper, Kreps et al. (1982) demonstrate that if each player assesses a (small) positive probability that his partner is 'cooperative' (i.e., he prefers to cooperate (defect) if the other cooperates

[9] In order to encourage subjects to explore the features of the game interface (the point of practice), practice game payoffs were displayed as the Marx brothers: Chico, Groucho, Harpo and Zeppo.

[10] In fact, as we will see in the next sections, there is a lot of trading in all markets that systematically responds to the different available information channels across markets.

(defects)), then sequential equilibria exist wherein perfectly rational players cooperate until the last few stages. Intuitively, if rational players believes there are even a small number of cooperative-types, then they have an incentive to cooperate themselves in order to encourage cooperative-types to cooperate in future rounds.

The same intuition can be applied to our game: Suppose buyers believe there are a small proportion of sellers who will ship, say, out of a sense of social obligation.[11] They buy only if the seller has shipped for orders received in the past; giving sellers incentives to ship in order to continue to receive buy orders in the future. Note that this argument requires the buyer to be able to observe what the seller has done in the past – a condition that holds in the reputation and partner market, but not strangers (the same is true of Kreps et al.'s model). This (admittedly informal argument) suggests, then, more trading in reputation and partners markets than in strangers, but offers little reason to think that trading levels will differ across reputation and partners.[12]

A second reason to believe there would be more trading then suggested by the complete information Nash argument is that people have been shown to be more myopic than assumed by the backwards induction this argument relies upon to show there is no incentive to cooperate (e.g., Selten and Stöcker, 1986). Rather than go to the end of a long game and reason back, most people appear to have trouble reasoning more than one or two steps ahead. Myopia has less bite with the strangers market because in this case all a seller need realize is that the decision made

[11] Bolton and Ockenfels (2000) show that a model in which people are assumed to care about relative as well as pecuniary payoffs can explain data patterns across a wide variety of experimental games, including bargaining, social dilemma, and market games. This model suggests that some (but not all) sellers will ship when they are trusted, even in the 30th round. Explicit connection to the Kreps et al. model is made in their Section VI.D. Güth and Ockenfels (forthcoming and 2001) review the economic theory literature on the evolution of preferences in trust games not unlike the one we study. They show that depending on the details of the institutional environment trustworthy behavior may survive evolutionary competition. These two complementary lines of research suggest that trustworthiness and trust can also have non-strategic causes and can be exhibited even in one-shot encounters or in the last round of repeated interaction. Our experiments measure the extent to which trust and trustworthiness is intrinsically or strategically motivated.

[12] As with Kreps et al.'s model, the formal incomplete information model would be technically demanding, and we will not attempt to work such out here. We note, however, that our data suggests that new models, different from the standards models discussed here, will be necessary to explain how traders respond to reputation mechanisms, and our data contains some clues about what such model would look like.

now will not be observable to future partners, making each encounter, in essence, a separate game. In the reputation and partners markets, however, shipping behavior is reported to future partners, information the seller can dismiss only if he does the complete backwards induction – and is confident that future buyers will do so as well. Consideration of myopic behavior, then, suggests the same pattern of trading as consideration of incomplete information.[13]

In sum, the unique Nash equilibrium, assuming complete information, has no buying and no shipping along the equilibrium path for the games in any of the three markets. This argument, however, is not robust to incomplete information about player payoffs or to player myopia with respect to backwards induction. Both of these considerations suggest that there will be more trading in reputation and partners market than in strangers. They do not, however, suggest any difference in trading in reputation and partners owing to the base assumption – common to all of these arguments – that information about past behavior drives future cooperation.

One final theoretical consideration: That this game has a finite number of rounds makes for a stringent test of reputation systems to induce transactions, since the monetary value of having a good reputation will diminish as the last round approaches, even if all players are quite myopic (consider the 29th round). We could have designed a finite game to simulate an infinitely repeated game by inserting some probability of stopping after each round of play.[14] In these games, there are equilibria in which rational actors – assuming none are too risk averse – cooperate every round (although not cooperating in *any* round is always an equilibrium as well). But now the market structure overstates things in the opposite direction: In reality, at any given time, there are market actors who are in the market short term, perhaps for the last time. In our set-up, all traders exit the market at a publicly stated time, and in this sense it is the tougher test.

[13] To give a more precise example, assume that myopic subjects play the game as if it is infinitely repeated. It is easy to see that in this case simple trigger strategies that call from buyers to punish untrustworthy sellers by never trusting them again support cooperative equilibria in the partners and reputation markets if the discount rate is sufficiently high (see e.g. Ockenfels, 2002, for the theory in a related scenario), but not in a strangers market when no information about the opponent's history is available (regardless of the players' discount rates).

[14] While this would eliminate the stopping issue, it would also introduce risk-over-stopping as a factor, the influence of which is not directly observable and so difficult to measure. Alternatively, we might not reveal the ending round to subjects, but this would have the same drawbacks.

3. Experimental Results

We first describe the basic treatment effects we observed. Theory suggests that the key to understanding differences in transaction behavior across markets (we find substantial differences) is how buyers react to information about the sellers they are paired with. If buyers discriminate between sellers who have been trustworthy in the past and those who have not been, then sellers will have an incentive to be trustworthy. For this reason, we look at buyer behavior in some depth.

3.1 Treatment effects

The major treatment effects have to do with trading patterns. These can be measured three ways: *efficiency* or the percentage of potential transaction completed (Figure 2), *trust* or the percentage of buy orders given (Figure 3), and *trustworthiness* or the percentage of shipped items, conditioned on buy orders (Figure 4). In all three figures, the treatment data has been aggregated across sessions.

Figure 2. Efficiency measured as how often the gain from trade is realized, by round

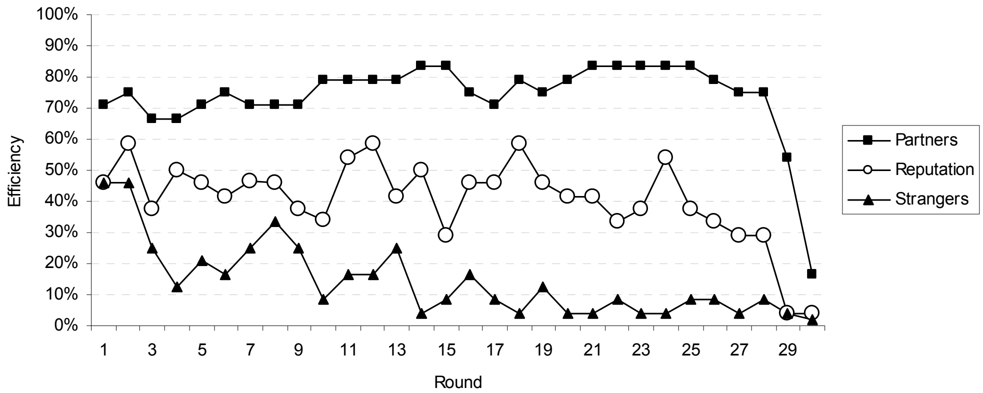


Figure 3. Trust measured as the percentage of buying per round

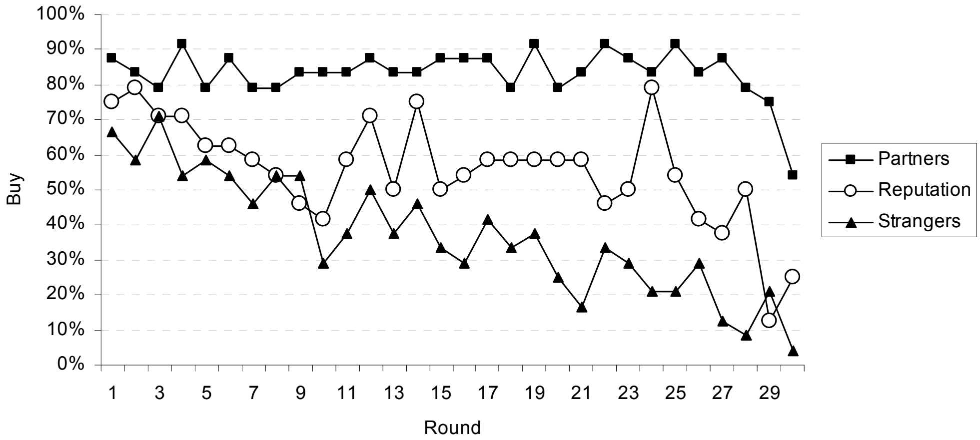
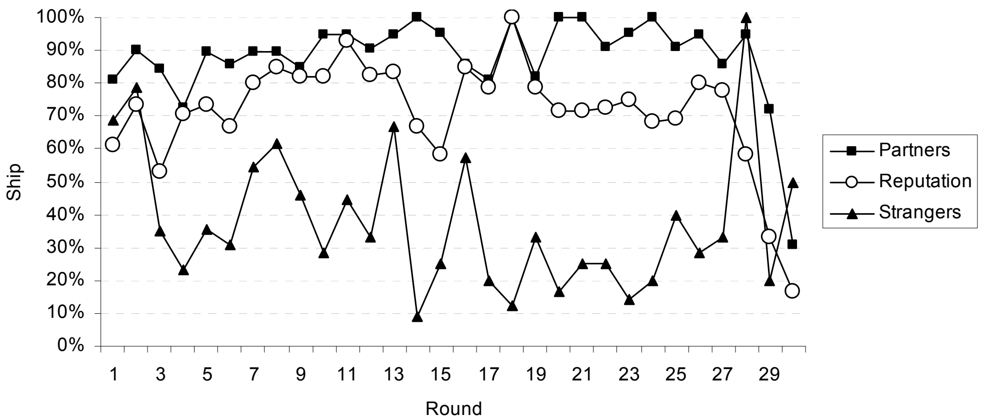


Figure 4. Trustworthiness measured as percentage of shipping per round



The same pattern is evident in all three figures: There is the least efficiency, trust and trustworthiness in the strangers market, more of all three in the reputation market, and more still in the partners market. For instance, averaged over all rounds, reputation yields 2.8 times the efficiency of strangers, and partners yields 1.8 times the efficiency of reputation. Pair-wise *t*-tests, in which sessions are taken as the individual observations, verify these observations ($p < 0.025$ in all cases save for comparing average buying between strangers and reputation, $p = 0.08$, one-tail tests, assuming equal variances).[15]

[15] A non-parametric rank test on session observations yields similar results, with $p = 0.05$ for all but the strangers-reputation comparison where $p = 0.10$ (one-tail tests).

Table 1 breaks the data out by sessions and reveals how the markets' differences in matching schemes and information flows affect the variability of trading outcomes.

Table 1. Percentage of efficiency, buying, and shipping, by session

Treatment	Session	Efficiency	Buy	Ship*
Strangers	I	0.158	0.391	0.411
	II	0.046	0.204	0.224
	III	0.225	0.517	0.435
	Mean	0.143	0.371	0.357
	coef. var.	0.633	0.424	0.323
Reputation	IV	0.517	0.671	0.767
	V	0.367	0.538	0.682
	VI	0.338	0.458	0.736
	mean	0.407	0.556	0.728
	coef. var.	0.236	0.193	0.059
Partners	VII	0.738	0.833	0.885
	VIII	0.671	0.792	0.847
	IX	0.808	0.875	0.924
	mean	0.739	0.833	0.885
	coef. var.	0.093	0.050	0.043

* conditioned on buying

For all gauges (efficiency, buy, ship), the coefficient of variation is highest for strangers, less than half as large for reputation, and then smaller still for partners. Thus, on average, trading patterns are less reliable in strangers markets, more so when a reputation mechanism is available, and more so still in partners relations.

There are also marked differences in the trading dynamics across treatments. Observe from Figures 2, 3 and 4 that while there is a clear downward trend in efficiency, trust and trustworthiness over all rounds in the strangers treatment, trading volumes appear to be rather stable in reputation and partners, save for the very last two rounds when trading collapses.

Model 1 in Table 2 supports this observation for the buyer behavior with the help of a random effect probit regression.[16] There is not only much more trust *per se* in partners than in

[16] The more sophisticated Models 2 and 3 examine *individual* trading patterns and will be discussed below. Analogous probit models for the seller behavior are omitted but yield the same qualitative conclusions (see also Table 3 below).

Table 2. Random effects probit models, buyers[a]

Maximum likelihood estimates (and two-sided *p*-values) for buyer behavior
Dependent variable = "1" for *buy*

Independent variable	Model 1	Model 2	Model 3
CONSTANT	0.533	0.347	0.524
	(.0040)	(.0185)	(.0001)
REPUTATION	-0.020	0.200	
= 1 if buyer is from reputation treatment, and 0 else.	(.9473)	(.4347)	
PARTNERS	0.963	1.48	0.852
= 1 if buyer is from partners treatment, and 0 else.	(.0001)	(.0000)	(.0011)
TOTALSHIPreputation			0.0616
= number of seller ships prior to last order.			(.0014)
TOTALNOSHIPreputation			-0.124
= number of seller no ships prior to last order.			(.0144)
SHIPLASTreputation			0.212
= 1 if reputation seller shipped last order, and 0 else.			(.2111)
NSHIPLASTreputation			-0.646
= 1 if reputation seller did not ship last order, and 0 else.			(.0005)
SHIPLASTpartners			1.330
= 1 if seller in partners shipped last order, and 0 else.			(.0000)
NSHIPLASTpartners			-.697
= 1 if seller in partners did not ship last order, and 0 else.			(.0100)
CBSH		0.045	-0.005[b]
= number of past times item was shipped to buyer.		(.0180)	(.8386)
CBNH		-0.412	-0.386[b]
= number of past times buyer bought but not shipped.		(.0000)	(.0000)
ROUNDstrangers	-0.062		
= round in strangers treatment, and 0 else.	(.0000)		
ROUNDreputation	-0.019		
= round in reputation treatment, and 0 else.	(.0006)		
ROUNDpartners	0.006		
= round in partners treatment, and 0 else.	(.4806)		
LAST2ROUNDstrangers	-0.151	-0.390	-0.404
= 1 if round 29 or 30 in strangers treatment, and 0 else.	(.6414)	(.1649)	(.1671)
LAST2ROUNDreputation	-0.903	-0.944	-0.974
= 1 if round 29 or 30 in reputation treatment, and 0 else.	(.0000)	(.0000)	(.0000)
LAST2ROUNDpartners	-1.15	-1.200	-1.322
= 1 if round 29 or 30 in partners treatment, and 0 else.	(.0000)	(.0000)	(.0000)
RHO	0.399	0.456	.444
(random effects)	(.0000)	(.0000)	(.0000)
Number of observations	2160	2160	2160
Log-likelihood	-1087.77	-1056.57	-988.67
χ^2 *p*-value	.0000	.0000	.0000

[a]Analogous estimates for fixed effects linear models are given in Appendix B.
[b]History for Partner's buyers does not include last transaction.

reputation and strangers, as indicated by the treatment dummy PARTNERS, but controlling for end-game effects the trust shown by partners is remarkably stable over time: The ROUNDpartners coefficient is small and not significant. ROUNDreputation is also small, but significantly negative, indicating a slight downward trend of trust in reputation. However, it is still significantly larger than ROUNDstrangers (two-tail p = 0.0106, Wald test). Since the coefficient for the dummy REPUTATION is close to zero and not significant, Model I explains the larger percentage of trust in reputation by the fact that trust in strangers declines more rapidly over time. At the same time, there are large and significant end-game effects in both reputation and partners but not in strangers, indicated by the LAST2ROUND variables.[17] In the next sections we examine the underlying causes for these strong treatment effects by looking at how individual behavior varies across markets.

3.2 Comparing strangers and reputation: The strategic benefits of trust and trustworthiness

Theory (Section 2.3) suggests a reputation mechanism can help to realize efficiency enhancing trade on large, anonymous market platforms *if* buyers condition their behavior on the shipping history of their sellers, so that sellers have strategic incentives to avoid spoiling their reputation.[18] The data supports this view.

Figure 5 shows the marginal effects on the probability of trust in all markets depending on whether the seller shipped the *last* order ('trustworthy') or not ('untrustworthy').[19] Strangers cannot distinguish between whether the seller shipped the last order or not, so that the marginal effect is close to zero regardless of the seller's history. (That the effects are negative in both cases is due to the facts that the bars do not include "newbies" − sellers who have not been trusted yet and who are therefore typically encountered in early rounds − and that buy rates in strangers decline rapidly.) Buyers in reputation, on the other hand, strongly condition their

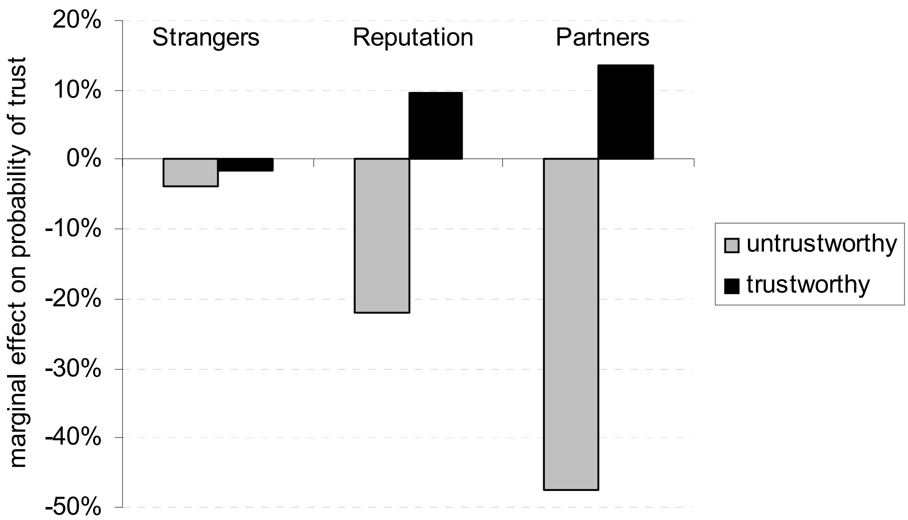
[17] For endgame effects, we focus on the last *two* rounds because, while player roles were switched randomly, they were told they would be a seller (buyer) for half the rounds, so that in round 29 a seller may be in his last round as a seller and thus has no strategic reason to be trustworthy.

[18] As eBay puts it: "A user's feedback is a key factor people use to determine whether or not they want to trade with that user. What feedback you give or receive is an important part of your trading reputation at eBay. […] If you're a buyer, checking a seller's Feedback Profile before you make a bid is one of the smartest and safest moves you can make. This Feedback Profile answers many questions about how a seller does business." (http://pages.ebay.com/services/forum/feedback.html).

[19] Figure 5 and the next Figure 6 are illustrations of the main effects. A more thorough statistical analysis of buyer behavior across treatments that takes into account the whole feedback history and that controls for other important factors will be discussed in Section 3.4.

behavior on the seller's last feedback. In absolute terms, they trust with probability 33 percent if the seller did not ship the last order, and with about twice this probability, 65 percent, if he shipped the last order. The simple descriptive statistics make it clear that buyers do indeed condition their buying decision on shipping history. (Models in Table 2, to be discussed in a moment, show this more formally.)

Figure 5. Marginal trust conditioned on last feedback across treatments*



* The base rate (the zero line) is the average buy over all encounters for each treatment separately (37.08 percent in strangers, 55.56 percent in reputation and 83.22 percent in partners).

This kind of conditional buying is rational since the seller's history has predictive power for his future performance. Table 3 presents a random effect probit for sellers. We can see that shipping the last time both a reputation and partner market seller received a buy order is a significant predictor of whether the seller will do so this time. (The coefficient for LASTSHIPstrangers is significant as well but with a negative sign.) Further, a last decision to ship is more highly predictive of shipping this time in partners than in reputation markets (two-tailed $p = 0.0121$, Wald test).

Table 3. Random effects probit model, sellers

Maximum likelihood estimates (and two-sided p-values) for seller behavior

Dependent variable = "1" for *ship*

Independent variable		Independent variable	
CONSTANT	0.190 (.2818)	SHIPLASTstrangers	-0.366 (.0713)
REPUTATION	0.278 (.2666)	SHIPLASTreputation	0.350 (.0168)
PARTNERS	0.557 (.0267)	SHIPLASTpartners	0.898 (.0000)
ROUNDstrangers	-0.037 (.0008)	LAST2ROUNDstrangers	-0.112 (.8711)
ROUNDreputation	-0.005 (.6154)	LAST2ROUNDreputation	-1.757 (.0066)
ROUNDpartners	-0.007 (.4589)	LAST2ROUNDpartners	-1.833 (.0000)
RHO	0.204 (.0000)	Number of observations Log-likelihood χ^2 p-value	1267 -561.63 .0000

[a]Variable interpretations are analogous to those given in Table 2.

Finally, the strong end-game effect exhibited by both buyers and sellers in reputation (Tables 1 and 3, respectively) additionally supports the view that it is the strategic incentive to ship created by the 'shadow of the future,' as opposed to, say, an intrinsic preference for being trustworthy, that largely drives the efficiency-enhancing effect of the reputation mechanism.[20]

Returning to Figure 5, buyers in the partners market respond even more strongly to the sellers' histories than do buyers in partners. Why is it that buyers in reputation markets, who have access to similar history information, are less affected? The next section provides an answer.

3.3 Comparing reputation and partners: The public benefits of trust and trustworthiness

The central thrust of our arguments in this section will be that, unlike in a partner relationship, feedback and past experience do not perfectly overlap in reputation markets. This creates effects of trusting and being trustworthy that are not internalized by the feedback mechanism in the reputation market as they are in the partners market. In this section, we

[20] Of course, the fact that sellers in strangers ship some 36 percent of the time indicates that not all trustworthy behavior can be explained by strategic response to the pecuniary incentives.

explain this phenomenon with some simple descriptive statistics to illustrate the points. In the next section, we discuss the analogous inferential evidence.

First, observe that a trusting buyer in a reputation market generates valuable feedback information for *other* buyers who meet the same seller in the future. A trusting buyer in a partner relationship generates the same valuable feedback information − but entirely for himself. Thus, the informational benefits from trusting (as opposed to the pecuniary gains from trade) are internalized in trades among partners but not in a reputation market. As a consequence, all other things equal, the overall benefits from trusting are smaller in reputation.[21]

This informational dilemma is particularly apparent if a buyer is matched with a newby, a seller with no feedback history yet, because then there is no evidence of whether the opponent can be trusted or not. The average trust in newbies is about 65 percent in strangers, 77 percent in reputation and 93 percent in partners.[22] While the difference between the strangers market and the other markets can be explained by the fact that only newbies in the strangers market have no strategic reason to be trustworthy, the difference between reputation and partners may be explained by the informational dilemma. In particular, observe that buying from a newby in our reputation treatment yielded an expected payoff of 31 cents, less than the 35 cents from not buying. Buying from a seller who shipped the last order, on the other hand, yielded an average payoff of 40 cents, and buying from a seller who did not ship the last order yielded 17 cents.

Thus, a buyer in the reputation market is better off trusting somebody if he or she has already been shown to be trustworthy. But a buyer does not have an incentive to generate the information that is needed by the market to discriminate between trustworthy and untrustworthy sellers, because the production of feedback is costly but not beneficial to the buyer. In contrast,

[21] There is a related but distinct public goods problem of feedback provision observed Resnick and Zeckhauser (2001), among others: Once the transaction is concluded, buyers have no incentives to provide others with feedback about their experience. (Resnick and Zeckhauser report that on eBay nevertheless feedback was provided in about half the time.) Since in our experiments feedback was produced automatically, this public good problem was not an issue in the lab (and in this sense our results overestimate the merits of a reputation mechanism that is based on *voluntary* feedback production). The public good problem identified above has to our knowledge not been addressed before.

[22] Four buyers in reputation never faced a newby and so are not included in these statistics.

trusting buyers in partners markets only produce feedback information for their private use, say, to execute a simple conditional buying strategy. Both costs and benefits are fully internalized.

There are not only public benefits in reputation markets from trust but also from trustworthiness. The critical evidence for this comes from examining buyers' reactions to their *own* history. Extending the scope of Figure 5, Figure 6 additionally provides some sense of the relative importance of own experience, defined either as 'rewarded' if the buyer's *last* order was shipped, or 'exploited' if not (later, our statistical models will provide more formal evidence). The figure shows that, regardless of the market environment, if a buyer was treated well in the past he is more likely to trust in the future. While strangers cannot distinguish between trustworthy and untrustworthy sellers, they do condition their behavior on whether they were rewarded or exploited in the past: Rewarded buyers trust substantially more often than exploited buyers. In reputation and strangers least trust is shown if the buyer was exploited and faces an untrustworthy seller and most trust is shown if he was rewarded and faces a trustworthy seller.

There is, however, an important difference between reputation and partners markets: In partners markets, histories are perfectly aligned. That is, a buyer's trust is rewarded if and only if his seller is trustworthy, so that the two history effects always cumulate.[23] In reputation, on the other hand, the feedback effect is diluted by the effect from the buyer's own history. A trustworthy seller is less trusted when the buyer's own experience was bad and an untrustworthy seller is more trusted when the buyer's own experience was good. Thus, there is a wedge driven between buyers' and sellers' histories in reputation markets that is (at least partly) responsible for why the marginal response to both the seller's positive and negative feedback is on average weaker in reputation compared to partners (as shown in Figure 5). The wedge does not explain, however, why overall average trust is smaller in reputation. But note that because both the seller's and the buyer's history affect the evolution of trust, trustworthiness creates a public benefit in the reputation market. A seller who ships in the reputation market benefits only through the improved own reputation, but *other* sellers will profit from the induced good history

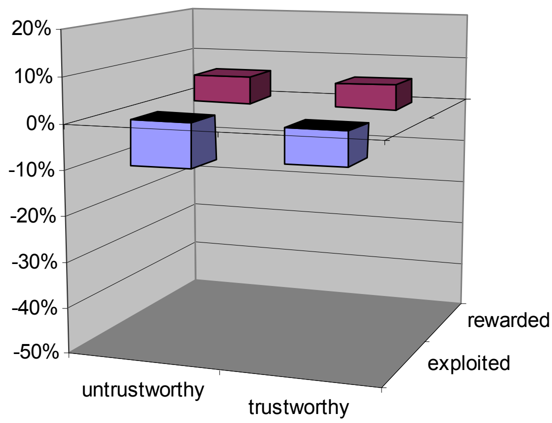
[23] Consequently, the bars concerning the partners market in Figures 5 and 6 have the same lengths.

of his buyer. On the other hand, a seller who ships in partners not only contributes to his own good reputation but also contributes to a good history of *his own* future buyer. In other words, the trust-inducing effects of trustworthiness are fully internalized in partners but not in reputation. As a consequence the overall private benefit from trustworthiness is weaker in reputation markets, which reduces the incentive to trade of both sellers and buyers.

Figure 6. Marginal effects of the buyers' most recent history and the sellers' most recent feedback on the buying probability*

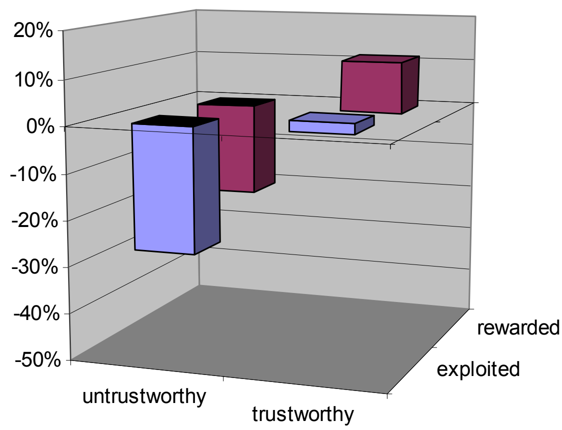
Strangers

average buy = 37 percent



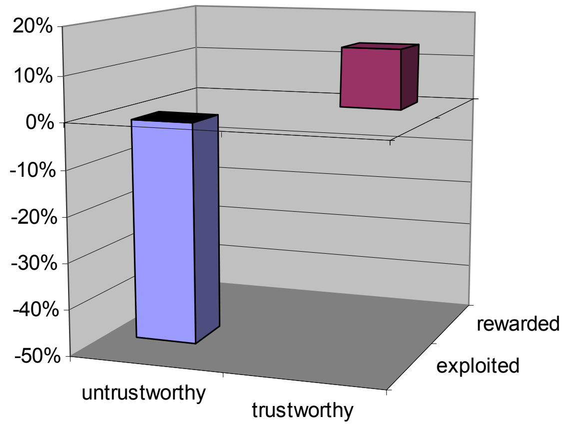
Reputation

average buy = 56 percent



Partners

average buy = 83 percent



* As in Figure 5, in each graph, the zero level is normalized with respect to the respective "average buy" that includes all encounters.

3.4 A probit model of buyer behavior

Models 2 and 3 in Table 2 provide formal inferential support for the findings in the last section. They also provide some further insight into how buyers evaluate seller histories.

Model 2 replaces the round effect variables with variables reflecting the buyer's history with shipping: the cumulative number of times he has bought and had the item shipped (CBSH) and the cumulative number of buys that were not shipped (CBNH). Figure 6 suggests that buyer decisions are influenced by their experience with seller shipping. The alternative hypothesis is that the differing learning effects we see in Figure 3 are due to the differing rules of the three markets; so, for instance, there is less information for buyers to process in strangers and this is plausibly why the experience effect is greater in strangers than in the other treatments (see Section 3.1). If differing rules, not buyer experience, were responsible for the dynamic across rounds, then we would expect both C··H coefficients in Model 2 to be about the same, and certainly Model 2 should fit no better than Model 1. In fact, the C··H coefficients differ sharply and the likelihood of Model 2 is higher than that of Model 1, indicating that buyer experience is quite a good explanation for the differing dynamic across markets.[24] What we see in Model 2 is

[24] Models 1 and 2 are not nested and so not conducive to a formal inference test of the buyer experience hypothesis. If we re-estimate model 1 with three CNBH*x* variables, one for each treatment *x*, all three of the new variables are

that negative experiences – making a purchase and having a seller fail to ship – erodes buyer trust in the market, while positive experiences have substantially less effect.

The influence of buyer history persists in Model 3, where we add variables reflecting the information buyers have about sellers at the time they decide whether to purchase. For the reputation market, we add variables to reflect the cumulative shipping history of the seller (TOTALSHIPS and TOTALNOSHIPS) as well as variables to reflect the most recent shipping history (SHIPLAST and NSHIPLAST). In the partners market, the seller's cumulative history is already reflected in the buyers C··H variables. So for partners, we need add only most recent history variables. We also drop the REPUTATION variable since this has not been significant in the other two models.[25]

All of the information coefficients in Model 3 have the expected signs, save that for CBSH, but this estimate is very small and not significant. There are two main observations. First, both reputation and partner buyers weight recent observations more heavily than older ones (p = .0000 for both markets, Wald test). Second, in reputation markets, NSHIPLAST has a more reliable effect on buyer decision than SHIPLAST, whereas partner buyers weight recent positive and negative history about the same; in fact, SHIPLAST and NSHIPLAST are, in absolute terms, virtually identical.

We also see the information wedge effects described in the last section: First, buyers are more likely to trust newbies in the partners market than in the reputation market as indicated by the significant coefficient for PARTNERS (while REPUTATION is insignificant). Second, a partner seller who has been trustworthy in the recent past is granted higher trust (SHIPLASTpartners is greater than SHIPLASTreputation, two-tail p = 0.0007, Wald test). Third, we can see the greater incentive sellers have to be trustworthy due to the internalization of buyer history. Note that a decision *not to ship* has very similar immediate total negative effect

negative and significant (two-tail $p < 0.01$ in all cases). A full model, with history variables broken out by treatment, is given in Appendix B.

[25] Including it would make little difference to the estimates of Model 3 and the coefficient of REPUTATION would still not be significant. There are, however, some indications that REPUTATION is collinear with SHIPLASTreputation and NSHIPLASTreputation.

on buying in both Reputation and Partner markets: compare NSHIPLASTrep + CBNH to NSHIPLASTpar + CBNH. The difference is that the seller in the systems incurs only about 65 to 70 percent of this cost since the CBNH becomes part of the history of a buyer he is unlikely to meet again in the near future.

3.5 Payoffs

One would correctly conclude from the analysis of the efficiency reached by the different markets in Section 3.1 that average payoffs are smallest in strangers, larger in reputation, and larger still in partners. In fact, the strategic incentives created by reputation systems also translate into positive *correlations* between trust(worthiness) and payoffs. The Spearman rank correlation coefficients between subjects' total payoffs and the total number of ship and buy in the strangers markets is negative (−0.307, p = .034, for the frequency of buy and −0.038, p = .795, for ship), making all trade activities unprofitable. Sixteen out 48 subjects receive total payoffs that are smaller than in the Nash equilibrium in which nobody ever buys or ships. A reputation mechanism, on the other hand, makes trust and trustworthiness lucrative behavior. The corresponding coefficients in the reputation market are significantly positive (.288, p = .047, for buy and .504, p = .000, for ship) and not a single subject made payoffs smaller than equilibrium payoffs. Finally, the pecuniary incentives to trade are strongest in partners markets (.706, p = .000, for buy and .728, p = .000 for ship). While two subjects in this market received less than they could expect in equilibrium (both subjects never shipped to their partners but vainly tried to buy from them), 34 buyers and 31 sellers did their part of the exchange 14 or 15 times, numbers that are reached only once in strangers and reputation together.

4. Summary

Our experiment shows that market participants systematically respond to the strategic incentives created by a reputation mechanism in large, anonymous markets. Reputation markets perform better than strangers markets with respect to straightforward measures of efficiency, trust, trustworthiness, and stability. The strong end-game effect also supports the view that behavior is mainly driven by the strategic considerations. But at the same time our experiments indicate that the strategic incentives to trade in reputation markets are weaker than the incentives in partner relations, the kind of relations the reputation mechanisms try to imitate. Reputation markets, unlike partners markets, drive a wedge between buyer and seller histories. Consequently, both trust and trustworthiness create public benefits not internalized by online reputation mechanisms.

At first glance, our finding that buyers condition their behavior on their own experience (and thus create public benefits of trustworthiness) is a puzzle. Standard strategic arguments do not easily capture this observation. However, there are two plausible non-strategic approaches that are in line with this pattern. First, adaptation theories predict that people tend to choose strategies that performed well in the past (such as Roth and Erev's, 1995, reinforcement learning theory or Selten's, 1988, learning direction theory). Hence, if trust was rewarded it has a higher probability of being chosen again. Second, the own experience effect may reflect a non-strategic (backward-looking) reciprocal motive. Market participants may not be willing to cooperate in a market in which they were exploited. This argument appears to have more bite in the partners market where *not buy* can be straightforwardly interpreted as a reciprocal punishment for unfair behavior against oneself. The fact that the own experience effect also occurs in the reputation market suggests, however, that such kind of unfairness aversion is also relevant among strangers (similar observations have been made by Blount, 1995, and Bolton et al., 2001, among others).

There is strong reason to believe that one's own experience is important to the decisions field traders make. Online reputation mechanisms typically create all kinds of incentives to manipulate feedbacks making them far less reliable than the feedbacks truthfully generated in our experiment.[26] Ones own history is informative of the severity of these problems. In this

[26] See e.g., Dellarocas (2001). eBay distinguishes 4 forms of fraudulent feedbacks: defensive and offensive shill feedback (using secondary eBay User IDs or other eBay members to artificially raise the level of your own feedback or to leave negative comments for another user), feedback extortion (demanding any action of a fellow user that he or she is not required to do, at the threat of leaving negative feedback), and feedback solicitation (offering to sell

sense, we believe that our findings, if anything, underestimate the influence of buyer histories. Consequently, as we argued in Section 3.3, trustworthiness in the field is likely to create public benefits. If this effect has something to do with adaptation (as opposed to reciprocal responses), a reputation mechanism may be able to multiply these public benefits by informing all market participants about the shipping probability in the whole market, and not only about the trustworthiness of individual traders. If buyers also take the experiences made with other sellers into account (a hypothesis that can be tested within our framework), then a market with high levels of trustworthiness may be able to push trust to higher levels, while of course a market with already low shipping probabilities would deter even more buyers from trusting.

The same set of extenuating circumstances suggests that the influence of the information dilemma may be even larger in the field than in our experiment. Recall that the informational dilemma occurs when, for instance, buyers try to avoid trusting newbies (sellers with no feedback information yet) but are willing to trust sellers with good feedback. That is, when hardly anybody wants to bear the costs associated with verifying trustworthiness. However, in the field newbies can mitigate the associated market failure by initially charging lower prices in order to get their reputation started. They should be willing to do so, because a good reputation is a private good for the seller. Conversely, in many online market platforms participants can change their identity at no costs. This creates a stronger incentive for newbies not to ship, because buyers typically cannot distinguish a 'real' newby, who trades the first time, from a 'deceptive' newby, an experienced seller who got rid from his bad reputation. So, all other things equal, the probability that a newby is not trustworthy is likely to be higher on such platforms, which aggravates the informational dilemma.[27] Thus, any platform should try to avoid identity changes (see Friedman and Resnick, 2001, for more theoretical reasons why a platform should gain control over the agents' identities and how this can be realized).

---

feedback, trade feedback undeservedly, or buy feedback); see http://pages.ebay.com/help/community/investigates.html.

[27] Of course, the conflicting impacts of endogenous pricing and free identity changes can be examined with the help of new experiments within our framework.

Other, more immediate design implications from our study come from the analysis of how buyers respond to feedback patterns. In particular, buyers in our experiment put more weight on negative than positive feedback, and more on recent than old feedback. The emphasis on recent negative feedback has also been reported in field studies (Lucking-Reiley et al., 1999, and Resnick and Zeckhauser, 2001, among others[28]). Given this trader predilection, a cumulative measure may not be appropriate because it hides information critical to the buyers' decision to trust.

In closing, we emphasize that we do not claim to have identified *all* possible reasons for why partners markets perform better than reputation markets. Besides the public good characteristics of trust and trustworthiness, one might also think that the direct reciprocal contact in partners *per se* yields more cooperation. Actually, the fact that there is more trading activity in the closing rounds in partners than in reputation seems to indicate that even absence from any strategic incentives to be trusty or trustworthy, partner traders managed to create a partnership that is more profitable than what is possible in reputation markets. There is no hope that such non strategic effects can be imitated by a reputation mechanism regardless of how clever it is designed.

[28] Lucking-Reiley et al., 1999, note that this finding is in line with various studies in risk management and marketing and provide some references

References

Ariely, D., A. Ockenfels, and A.E. Roth (2002). “An Experimental Analysis of Ending Rules in Internet Auctions.” Working Paper, Harvard Business School.

Blount, S. (1995). “When Social Outcomes Aren’t Fair: The Effect of Causal Attributions on Preferences.” *Organizational Behavior and Human Decision Processes*, 63, 131-44.

Bolton, G.E., A. E. Katok, and A. Ockenfels (2001). “What's in a Reputation? Indirect Reciprocity in an Image Scoring Game.” Working Paper, Smeal College of Business Administration, Penn State University.

Bolton, G.E and A. Ockenfels (2000). "ERC: A Theory of Equity, Reciprocity and Competition." *American Economic Review*, 90, 166-193.

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

Brynjolfsson, E., and M.D. Smith (2000). “Frictionless Commerce? A Comparison of Internet and Conventional Retailers.” *Management Science*, 46, 563-585.

Dellarocas, C. (2001). “The Design of Reliable Trust Management Systems for Online Trading Communities.” Working paper, MIT Sloan School of Management.

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

Greif, A. (1989), “Reputation and Coalitions in Medieval Trade: Evidence on the Maghribi Traders,” *Journal of Economic History*, 49, 857-882.

Güth, W. and A. Ockenfels (2000). “Evolutionary Norm Enforcement.” *Journal of Institutional and Theoretical Economics*, 156, 335-347.

Güth, W. and A. Ockenfels (forthcoming). “The Coevolution of Trust and Institutions in Anonymous and Non-anonymous Communities.” in: M.J. Holler, H. Kliemt, D. Schmidtchen and M. Streit (eds.), *Jahrbuch für Neue Politische Ökonomie*, 20, Tübingen: Mohr Siebeck.

Houser, D. and J. Wooders (2001). “Reputation in Auctions: Theory and Evidence from eBay.” Working Paper, University of Arizona.

Kreps, David M.; Milgrom, Paul; Roberts, John and Wilson, Robert (1982), "Rational Cooperation in the Finitely Repeated Prisoners' Dilemma," *Journal of Economic Theory*, *27*, 245 252.

Lucking-Reiley, D., D. Bryan, N. Prasad, and D. Reeves (1999). “Pennies from eBay: the Determinants of Price in Online Auctions.” Working paper, Vanderbilt University.

Melnik, M.I., and J. Alm (2001). "Does a Seller's e-commerce Reputation Matter?" Working paper, Georgia State University.

Milgrom, P., D. North, and B. Weingast (1990). "The Role of Institutions in the Revival of Trade: The Law Merchant, Private Judges, and the Champagne Fairs." *Economics and Politics*, 2, 1-23.

Ockenfels (forthcoming). "New Institutional Structures on the Internet: The Economic Design of Online Auctions." In: M.J. Holler, H. Kliemt, D. Schmidtchen and M. Streit (eds.), *Jahrbuch für Neue Politische Ökonomie*, 20, Tübingen: Mohr Siebeck.

Ockenfels, A. (2002), "Reputationsmechanismen auf Internet-Marktplattformen." Working paper, Max Planck Institute for Research into Economic Systems, Jena.

Resnick, P. and Zeckhauser, R. (2001). "Trust Among Strangers in Internet Transactions: Empirical Analysis of eBay's Reputation System." Draft prepared for NBER workshop.

Resnick, P., R. Zeckhauser, E. Friedman, and K. Kuwabara (2001). "Reputation Systems: Facilitating Trust in Internet Interactions. Working Paper, mimeo.

Roth, A. E. (forthcoming). "The Economist as Engineer: Game Theory, Experimental Economics and Computation as Tools of Design Economics." Fisher Schultz lecture, *Econometrica*.

Roth, A.E., and I. Erev (1995). "Learning in Extensive-Form Games: Experimental Data and Simple Dynamic Models in the Intermediate Term." *Games and Economic Behavior*, 8, 164-212.

Roth, A.E., and A. Ockenfels (forthcoming). "Last-Minute Bidding and the Rules for Ending Second-Price Auctions: Evidence from eBay and Amazon Auctions on the Internet." *American Economic Review*.

Selten, R. (1998). "Features of Experimentally Observed Bounded Rationality." Presidential Address, *European Economic Review*, 42, 413- 36.

Selten, Reinhard, und Rolf Stöcker (1986). "End Behavior in Sequences of Finite Prisoner's Dilemma Supergames." *Journal of Economic Behavior and Organization*, 7, 47-70.

U.S. White Collar Crime Center and the Federal Bureau of Investigation, "IFCC 2001 Fraud Report," http://www1.ifccfbi.gov/strategy/statistics.asp.

Appendix A. Subject instructions and buyer screen

BELOW ARE THE WRITTEN INSTRUCTIONS THAT WERE GIVEN TO SUBJECTS IN THE REPUTATION TREATMENT. INSTRUCTIONS FOR THE OTHER TREATMENTS WERE PARALLEL, THE ONLY DIFFERENCES BEING THE DESCRIPTION OF THE REPUTATION SYSTEM (REMOVED FOR STRANGERS), OR THE DESCRIPTION OF PARTNER ROTATION (PARTNERS).

General. The purpose of this session is to study how people make decisions. If at any time you have questions, raise your hand and a monitor will happily assist you. From now until the end of the session, unauthorized communication of any nature with other participants is prohibited.

During the session you will play a game that gives you an opportunity to earn cash. At the end of the session, you will be paid your earnings plus a $5 show-up fee. Decisions and payments are confidential: No one will be told your actions or the amount of money you make.

[The figure that appeared here is the same as Figure 1 in the text of the paper.]

Description of the game. You and the other participants in the room (but not the monitors) are the players in the game. The game proceeds in a series of rounds. Each round, each player is randomly matched with another player to trade a (fictional) commodity. First, one of the players, the "Buyer," chooses to either *buy* or *not buy*. If the Buyer chooses *not buy*, then the game ends and both players receive $.35. If the Buyer chooses to *buy*, then the game continues and the other player, the "Seller," makes a decision to *ship* or *not ship*. *Ship* pays each player $0.50 while *not ship* pays the Buyer nothing and the Seller $0.70.

The game will last for 30 rounds. You will be a Buyer for half of the rounds, and a Seller for the other half. When you switch between roles is a matter of random chance, so you may be in one role for more than one round in a row before switching to the other role, and the pattern of switching may be different for you than for other players in the game.

Seller's feedback history. For each game played, the computer will record whether the Seller chose *ship* or *not ship* (if the Seller did not get to move, the computer records nothing). This feedback will then be made available to all future Buyers that are matched with this Seller. The feedback will include a summary of the number of times the Seller shipped in the past, as well as a round-by-round history of their shipping decisions, beginning with the most recent decision. Buyers will see this feedback history prior to making their buy decision.

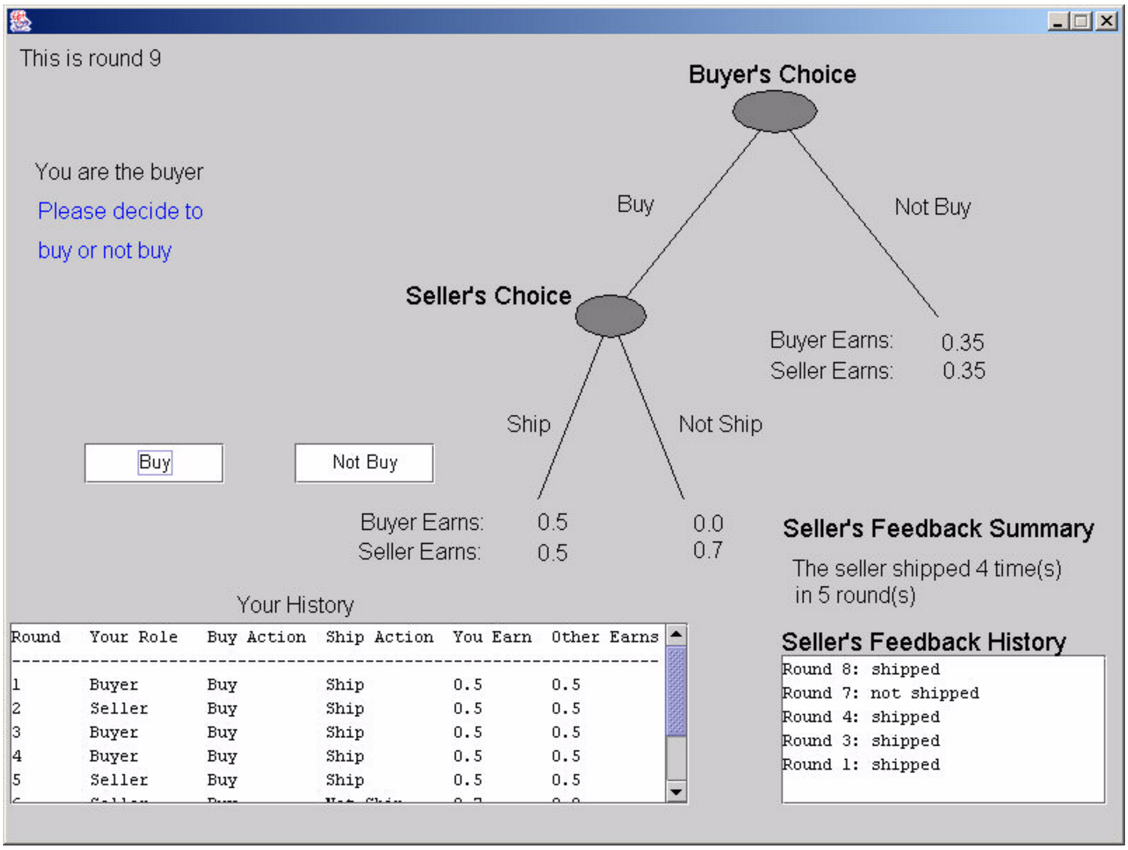
Pairings. All partner pairings are anonymous: Your identity will not be revealed to the person you are playing with either before, during or after the game. You will never be matched with the same player in the same role more than once.

Money earnings. You will be paid your earnings from all of the rounds of the game (plus a $5 show-up fee) in cash.

Practice games. When the monitor gives the OK, play some practice games. Your partner for the practice games will be the computer. It has been programmed to choose its moves at random. The practice games will allow you to experience the game from both the Buyer and Seller's perspective. Practice until you feel comfortable with the game and its rules.

Consent Forms. If you wish to participate in this study, please read and sign the accompanying consent form. The consent form explains your rights as a subject as well as the rules of confidentiality that will be adhered to regarding your participation.

Figure A1. Buyer screen



Appendix B.

Table A2. Fixed effects linear models, buyers[a]

OLS estimates (and two-sided p-values) for buyer behavior

Dependent variable = "1" for *buy*

Independent variable	Model 1	Model 2	Model 3
CONSTANT	---	---	---
REPUTATION	---	---	---
= 1 if buyer is from reputation treatment, and 0 else.			
PARTNERS	---	---	---
= 1 if buyer is from partners treatment, and 0 else.			
TOTALSHIPSreputation			0.021
= number of seller ships prior to last order.			(.0000)
TOTALNOSHIPreputation			-0.035
= number of seller no ships prior to last order.			(.0092)
SHIPLASTreputation			0.009
= 1 if reputation seller shipped last order, and 0 else.			(.8697)
NSHIPLASTreputation			-0.261
= 1 if reputation seller did not ship last order, and 0 else.			(.0000)
SHIPLASTpartners			0.167
= 1 if seller in partners shipped last order, and 0 else.			(.0023)
NSHIPLASTpartners			-0.224
= 1 if seller in partners did not ship last order, and 0 else.			(.0005)
CBSH		0.000	-0.006[b]
= number of past times item was shipped to buyer.		(.9103)	(.1428)
CBNH		-0.133	-0.129[b]
= number of past times buyer bought but not shipped.		(.0000)	(.0000)
ROUNDstrangers	-0.184		
= round in strangers treatment, and 0 else.	(.0000)		
ROUNDreputation	-0.007		
= round in reputation treatment, and 0 else.	(.0003)		
ROUNDpartners	0.001		
= round in partners treatment, and 0 else.	(.6651)		
LAST2ROUNDstrangers	-0.005	-0.069	-0.070
= 1 if round 29 or 30 in strangers treatment, and 0 else.	(.9434)	(.2393)	(.2227)
LAST2ROUNDreputation	-0.301	-0.263	-0.263
= 1 if round 29 or 30 in reputation treatment, and 0 else.	(.0000)	(.0000)	(.0000)
LAST2ROUNDpartners	-0.213	-0.136	-0.111
= 1 if round is 29 or 30 in partners treatment, and 0 else.	(.0010)	(.0225)	(.0610)
Number of observations	2160	2160	2160
Adjusted R-squared	0.368	0.401	0.442
F-test p-value	.0000	.0000	.0000

[a]These are analogous estimates for Table 2 in the text.

[b] History for Partner's buyers does not include last transaction.

Table A2. Random effects probit models, buyers[a]

Maximum likelihood estimates and two-sided *p*-values for buyer behavior

Dependent variable = "1" for *buy*

Independent variable	Coefficient	P-value
CONSTANT	0.512	0.0008
PARTNERS	0.835	0.0016
SHIPTOTALreputation	0.077	0.0009
NOSHIPTOTALreputation	-0.089	0.1194
LASTSHIPreputation	0.270	0.1520
LASTNOSHIPreputation	-0.566	0.0073
LASTSHIPpartners	1.184	0.0007
LASTNOSHIPpartners	-0.776	0.0127
CBSHstrangers	0.003	0.9691
CBSHreputation	-0.067	0.1171
CBSHpartners	0.055	0.2151
CBNHstrangers	-0.382	0.0000
CBNHreputation	-0.334	0.0000
CBNHpartners	-0.474	0.0000
LAST2ROUNDSstrangers	-0.420	0.1896
LAST2ROUNDSreputation	-0.926	0.0001
LAST2ROUNDSpartners	-1.649	0.0000
RHO	0.442	0.0000
Number of observations	2160	
Log-likelihood	-985.78	
F-test *p*-value	.0000	