Bridging the Trust Gap in Electronic Markets

A Strategic Framework for Empirical Study

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

Elena Katok
Smeal College of Business, Penn State University, USA

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

July 2002

Trust that suppliers and buyers will keep their word is a necessary ingredient to a well functioning marketplace. Nowhere is the issue trickier than for electronic markets, where transactions tend to be geographically diffuse and anonymous, putting them out of the reach of the legal safeguards and the long-term relationships that build trust in the brick-and-mortar world. Many online platforms have turned to automated reputation systems as a way of giving traders a heads-up on who they are dealing with. Here we describe a strategic framework for thinking about these systems. We also describe some lab data that provides an initial sense of effectiveness. We find that reputation has substantial positive effect, but not enough to be a close substitute for personal relationships; this is so even though our laboratory test abstracts away from many of the problems reputation systems must confront in the field. The evidence suggests directions for improving automated reputation system performance.

* Correspondence to Elena Katok, 509H BAB, Smeal College of Business, Penn State University, State College, PA 16870; ekatok@psu.edu. Ockenfels gratefully acknowledges the support of the Deutsche Forschungsgemeinschaft through the Emmy Noether-program. Bolton and Katok gratefully acknowledge the hospitality of the Max Planck Institute for Research into Economic Systems during the summer of 2002. Katok gratefully acknowledges the support of the National Science Foundation award # DMI-0128588.

1 Introduction: The problem of trust in electronic markets

According to a recent report by the Aberdeen Group, the market for strategic e-sourcing tools is growing at a rate of 98% per year, significantly higher than the growth rate of the IT industry as a whole (Morphy, 2001). Most of the emphasis has been on creating ways for improving access to suppliers, facilitating competition, and driving down prices. This is the promise of the Internet: expanded, more efficient markets. But the promise is not without peril. Realizing gains from greater competition means moving away from the long-term relationships commonly used to safeguard *trust* in brick-and-mortar markets. Lack of trust could undo all the good of increased competition and then some (Arrow, 1974).

The potential and the pitfalls are evident in an illustrative case a group of insurance and B2B executives took up during a recent conference on trust in global B2B e-commerce: Sarah Jones works for the American clothing retailer Blazer Barn. She is searching the Internet for a new blazer manufacturer when she comes upon Hong Kong Blazer, posted on the B2B exchange Buyeverything.com. Hong Kong Blazer manufactures the style of blazer Sarah is looking for, at a good price, and offers delivery in six weeks, in time for the back-to-school season. The fact that Sarah has never heard of Hong Kong Blazer, let alone dealt with them, makes her nervous, but she decides to go ahead anyway. She completes the necessary contractual documents, some of which are in Chinese, a language she does not understand, and then waits with fingers crossed… Unfortunately things do not go well. The wrong style of blazer shows up several weeks late, the shipment short a thousand units. Hong Kong Blazer blames the shipping company, and anyway Sarah did not specify the order properly. Given the differences in American/China jurisdictions, Sarah has little legal recourse.[1]

The legal problems surrounding e-commerce are not limited to international trade; within the U.S., e-commerce transactions often span multiple jurisdictions with conflicting laws, making legal recourse difficult if not impossible (Federal Bureau of Investigation, 2002). Some

[1] Dennehy (2000), reports on this conference, and participant's reaction to this case.

of these problems are transient; over time we expect the law to become more effective. But even then, the law is unlikely to be a complete solution because, as in the brick-and-mortar economy, legal recourse is expensive and time consuming.

Brick-and-mortar markets heavily supplement the law with informal trust building mechanisms, particularly long-term relationships, but also word-of-mouth reputation. Both are usefully thought of as forms of reciprocity. Long-term trading relationships are a form of *direct reciprocity*; I continue to deal with you so long as you were reliable with me in the past (Axelrod, 1984). Word-of-mouth reputation is a form of *indirect reciprocity*; I continue to deal with you so long as you were reliable with others in the past (Alexander, 1987). As noted, moving away from long-term relationships is the flip side of exploiting the scope of Internet markets. But one need not leave word-of-mouth behind: Several electronic trading platforms have introduced online "feedback" mechanisms intended to institutionalize word-of-mouth reputation by storing and distributing information about the transactions of those they mediate.

The question we explore in this chapter is whether online reputation systems can fill the trust gap, or at least that part that long-term relationships filled. It turns out that in theory, the answer is yes: direct and indirect reciprocity can be equally effective. The latter may require more information to function properly, but information dissemination is precisely what the Internet excels at. That said, we also present laboratory evidence that the reality is more complicated: We find that even under conditions that are arguably ideal, reputation systems fail to match the performance of long-term relations, although importantly, reputation systems do a lot of good relative to not having anything at all. In addition, our experiments pinpoint some problems and point in directions to improve the systems.

Theory and experiment are compliments to field evidence, and it is important to interpret our results in the context of what is known from the field. In fact, formal reputation systems and markets have a long and successful history together. One particularly well-documented example dates to the beginning of the first millennium, involving a group of merchants known as the Maghribis, who operated and traded throughout the Mediterranean region. At the time, ship

navigation was a hazardous business. A merchant generally required a business associate to sell (or buy) the goods and handle profits (or expense money) at the other end of the transaction. The Maghribis developed a very explicit system for disseminating information about which associates could be trusted and which could not. When a Maghribi trader discovered an associate was cheating him, he was under obligation to write a letter to all the other Maghribis warning them to stay away from this associate. This word-of-mouth system provided a powerful incentive for associates to mind their reputation less they lose business (Greif, 1989). It is essentially a one-way reputation system of the kind we investigate in Section 2.

The development of reputation systems was also an important factor in the expansion of trade that took place during the European Renaissance (Milgrom, North, and Weingast, 1990). At a deeper (evolutionary) level, research suggests that reputation mechanisms play a central role in all human social behavior, as a regulator of moral systems (Alexander, 1987). This idea has recently been analyzed in the form of an 'image scoring game,' a game we investigate in Section 3 to illustrate some of the issues in two-way reputation systems.

The new electronic feedback mechanisms essentially institutionalize the old word-of-mouth methods. eBay, for example, hosts a 'feedback forum,' enabling buyers and sellers to rate each other after the transaction. This information is then made available to all potential future transaction partners. Bizrate.com asks consumers to complete surveys on registered retailers, and then converts this information into store ratings. In the B2B arena, Reputation.com provides software that allows businesses to track and internally disseminate information about vendor performance across the organization. Other reputation mechanisms on the Internet include 'expert' evaluation of past performance such as expertcentral.com, product review such as epinions.com. While diverse in form, all of these methods are based on the same fundamental

principle of indirect reciprocity: the dissemination of information on individual past performance with others can be used to build trust and trustworthiness.[2]

The reliability and limits of these modern systems are only beginning to be investigated. Field studies of online auction platforms find that reputation mechanisms have at least some of the desired economic effect, though these studies do not present a completely coherent picture. Some find that reputable sellers are more likely to sell their items but with no price effect (e.g., Resnick and Zeckhauser, 2001), other studies find that negative feedback reduces the attainable price (e.g., Lucking-Reiley, Bryan, Prasad, and Reeves, 1999), and a few studies find that positive feedback induces price premiums (e.g., Houser and Wooders, 2000).[3] Further field work should clear up some of the discrepancies. But there are also obstacles. The complexity of the natural environment makes it difficult to control for the many extraneous factors that come into play. For example, present technology is unable to wholly prevent people from changing their online identity and thus escaping their online reputation. So is a finding on a system's effectiveness due to the reputation system *per se* or because technology has not yet figured out how to prevent people from changing their online identity? The answer says a lot about how to go about improving things. But because the amount of ID fraud is presently anybody's guess, it's hard to say.

In the rest of this chapter, we describe a strategic framework for studying electronic reputation systems, one that was mapped out by game theorists in years prior to the Internet, but is nevertheless very useful for thinking about reputation and e-commerce. Our experiments test whether these systems work as theory says they should. Our laboratory experiments permit us to

[2] See http://pages.ebay.com/services/forum/feedback.html. for eBay's feedback forum, similar to mechanisms used by Amazon and Yahoo, among others. For Bizrate's system, see http://www.bizrate.com/content/about.xpml. Traditional providers of reputation information, such as accounting firms or business credit rating companies are also active on the Internet.

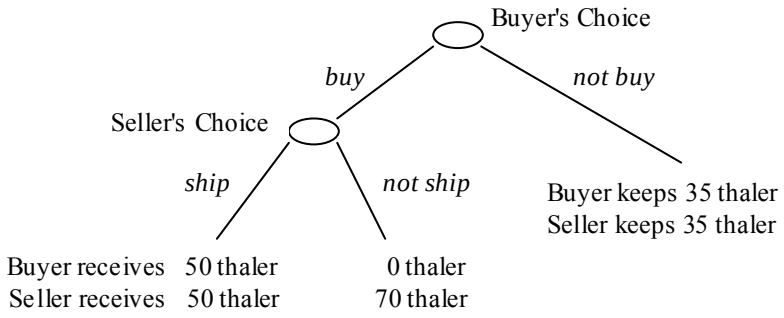
[3] Other field studies include Ba and Pavlou (forthcoming), Dewan and Hsu (2001), Melnik and Alm (forthcoming), and Ockenfels (2002). Resnick, Zeckhauser, Swanson, and Lockwood (2002) conduct a controlled field experiment using eBay's online reputation mechanism and survey the earlier literature. Brynjolfsson and Smith (2000) compared pricing behavior at 41 Internet and conventional retail outlets. They identify Internet sellers' trustworthiness as one important factor.

control extraneous factors, making it easier to puzzle out cause and effect. (We can add back the extraneous factors, in a controlled way, studying them as well). The lab also permits us to create parallel markets to compare reciprocity in long-term relationships to reciprocity under reputation systems—difficult to do under field conditions. Of course there is no free lunch; lab control sacrifices immediacy to the natural environment. But the insights gained in the lab can help us understand what we see in field, and may lead to ideas for improvements that would be difficult to come about in any other way.[4]

2 One-way reputation mechanisms: When the suppliers' trustworthiness is an issue

At base, reputation systems are intended to facilitate cooperation in situations where it would otherwise pay people not to cooperate. In a market setting, cooperation typically involves avoiding so-called 'moral hazards' like the false representation of the quality of goods or late payment (or maybe not paying at all).

Figure 1. The buyer-seller transaction problem



To illustrate how reputation systems work, consider a simple electronic market with a typical moral hazard. There are many buyers and sellers in this market. For convenience, we think of the transactions as taking place over a series of rounds. Every round, each buyer and

[4] For detailed discussion on the complimentary relationship between experimental and field studies see Ariely, Ockenfels, and Roth (2002), Bolton, Katok and Ockenfels (2002), and Roth (forthcoming).

each seller goes to the market (actually, to their computers) with a budget of 35 thalers (a fictional currency). This is the total amount available for purchases or expenses for that round. Each seller offers to sell a single unit of some homogenous good. We will suppose this is a competitive market, so sellers are price takers; the good trades for 35 thalers. The good has a use value of 50 thalers for each buyer, and the seller's cost of providing a buyer with the item is 20 thalers. The decisions buyer and seller face each round are then illustrated in Figure 1. If the buyer chooses not to buy, both players keep their budgets. If he chooses to buy, he sends his 35 thalers to the seller, who then has to decide whether to ship the good, or whether to keep both money and good. If the seller does not ship, he receives the price plus his budget for a total of 70 thalers. If he ships, he receives the price minus the cost plus his budget for a total of 50 thalers, while the buyer receives his value of the good.

The moral hazard problem is shipping. If, as is common in electronic markets, the buyer-seller encounter is one-time or anonymous – buyer and seller are effectively strangers – the seller, once he receives the money from the buyer, has no monetary incentive to be trustworthy and to ship. Of course, some sellers may be principled and ship anyway. But those less attuned to principle (or those whose principle is their self-interest) may not ship. Anticipating this, and absent a way of distinguishing the trustworthy sellers, the buyer may decline to buy, so that trade does not take place, even though this would make everybody better off.

Of course, our electronic market is highly stylized (simplified). Still, it captures the essence of the moral hazard problem in several real-world electronic markets. Amazon.com, to give a visible example, now offers used goods at its site right along side the new goods. Amazon actually acts as the middleman for the used transactions; the actual sellers range from brick-and-mortar stores to individuals selling personal items they no longer want. The site posting includes the purchase price the seller wishes to receive plus a description of the item and the shape it is in. A willing buyer sends the money to Amazon, which takes a cut, and passes the rest along to the seller. The seller is then supposed to ship the item within a pre-specified amount of time. While this market differs in several important respects from our stylized illustration (something we

return to discuss in Section 4), the moral hazard problem is essentially the same (although in the Amazon case there are two seller moral hazards, the second having to do with describing the quality of the good).

Returning now to the stylized market, suppose rather than strangers, buyers and sellers are effectively partnered together every round; that is, exclusively the same buyer has the opportunity to purchase from exclusively the same seller. Intuitively, this changes everything; so too in formal theory. Partnered players are effectively playing a repeated game, permitting the buyer to condition his purchase decision on the past behavior of the seller. A simple discrimination strategy can provide the seller with an incentive to ship: The buyer buys if and only if the seller shipped in the last round. So long as the seller does not discount the value of future sales too strongly, the seller then has a monetary incentive to ship. In sum, there should be trading. We are implicitly assuming that both buyer and seller are in the market endlessly, so that there is always future trading to consider. This is a simplifying assumption we will deal with in a moment. The sort of equilibrium in which players are cooperative now in order to elicit cooperation later is common to many repeated games and is known in the literature as a 'Folk theorem equilibrium' (to denote that no one knows who originally came up with the argument).

But electronic markets tend to involve trading among people and businesses that are more like strangers than they are partners – the reach of electronic markets is one of their main advantages. The question then is whether the kind of reputation strategy that can support trade in the partners market can also support trade in the strangers markets. In theory, the answer is that it can, and in these circumstances, rather easily at that. At first, this may seem surprising; we tend to think there is something special about direct reciprocity that is about dealing with the same person over-and-over: I scratch your back so long as you scratch mine. But the message inside the Folk theorem is that the key to cooperation is not the players' *interaction* per se but rather the *information* that is available to the players. One quick way to see this is to suppose the partnered buyer has no short-term memory (much like the lead character in the recent movie *Memento*); so the moment a round of the market ends, the buyer forgets what happened.

Obviously the incentive for the seller to be honest disappears with the information in the buyer's head.

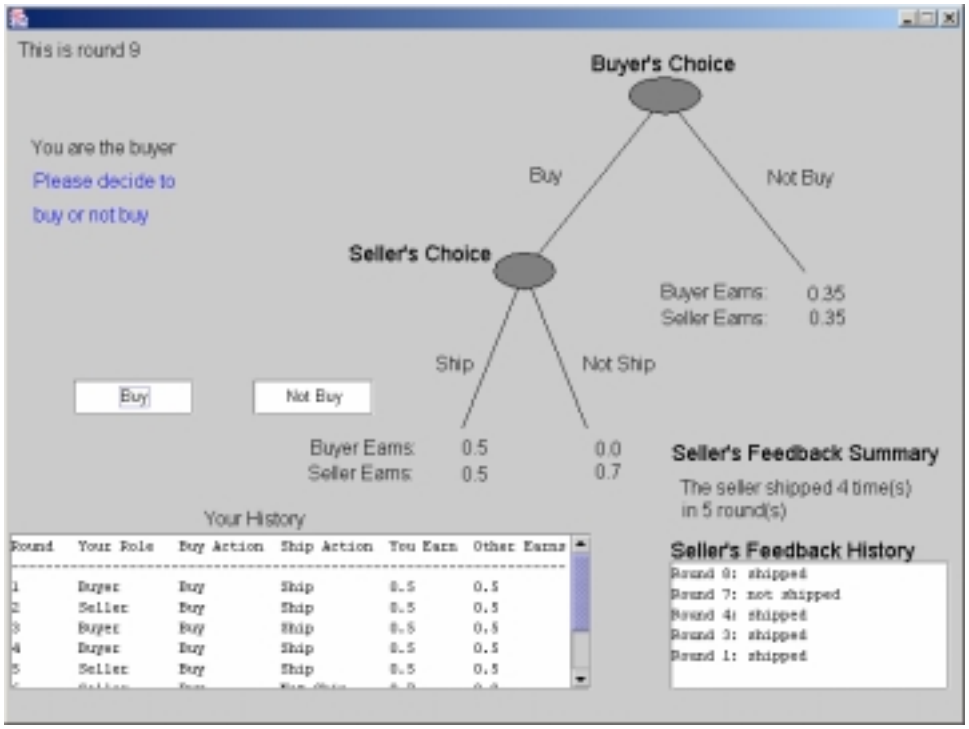
The information that the Folk theorem equilibrium effectively leverages is reputation. So long as a good reputation (in this case, a reputation for reliable shipping) is rewarded and a bad reputation is punished, the seller has an incentive to maintain a good reputation and avoid a bad one, independent of who's doing the rewarding or punishing. With incentives in place, the buyer can then be confident that a seller – even one unmoved by moral principle – will avoid falling into moral hazard. So in our stylized market, with strangers trading, providing buyers with a history of a seller's past shipping record should be enough to support trading.

In fact, this is precisely the insight that market platforms such as Amazon.com, eBay.com and Bizrate.com, among others, attempt to exploit. In theory, the system is a close, if not perfect, substitute, in the trust sense, for a one-on-one long-term buyer-seller relationship. (The field systems raise some additional issues we will ignore for the moment, concerning incentives to fill out buyer feedback forms and truthful reporting of feedback. We'll take them up in Section 4.)

But is the theory correct? That is, can a large group of strangers armed with information about reputation really trade as effectively as partnered traders? While plausible enough, it is a proposition that is virtually impossible to test in the field. For while there are sites like Amazon where strangers with reputation systems trade, there are few, if any, parallel markets involving just partners. We can, however, set up parallel electronic markets in the laboratory, and compare the performance of partners with that of strangers (with as well as without a reputation system). We did precisely this in Bolton, Katok and Ockenfels (2002).

In each of our experimental markets there were eight buyers and eight sellers. Each trader interacted with the market via a computer interface, meaning traders were effectively anonymous to one another (all were at computers in the same room, but these are separated by partitions so that no one can see what others are doing). The traders were Penn State University students, mostly undergraduates and studying in various fields. The market involved actual cash incentives: in fact, payoffs were as in Figure 1 save measured in U.S. cents.

Figure 2. Buyer computer interface in the reputation markets



The experiment consisted of three kinds of markets. In the *strangers markets*, buyers and sellers were randomly matched each round. No information about past histories (that is, about reputation) was made available. Traders in the *reputation markets* were matched in the same way but now, a buyer was shown feedback consisting of the shipping history of the seller he is matched with, prior to making a purchase decision. The computer interface of a buyer in this market is reproduced in Figure 2. Traders in the *partners markets* were paired with the same partner in every round and so had the same feedback available to them as in the reputation treatment. Each kind of market was simulated three times. We never used the same traders in more than one simulation, so in total the experiment involved some 144 participants.

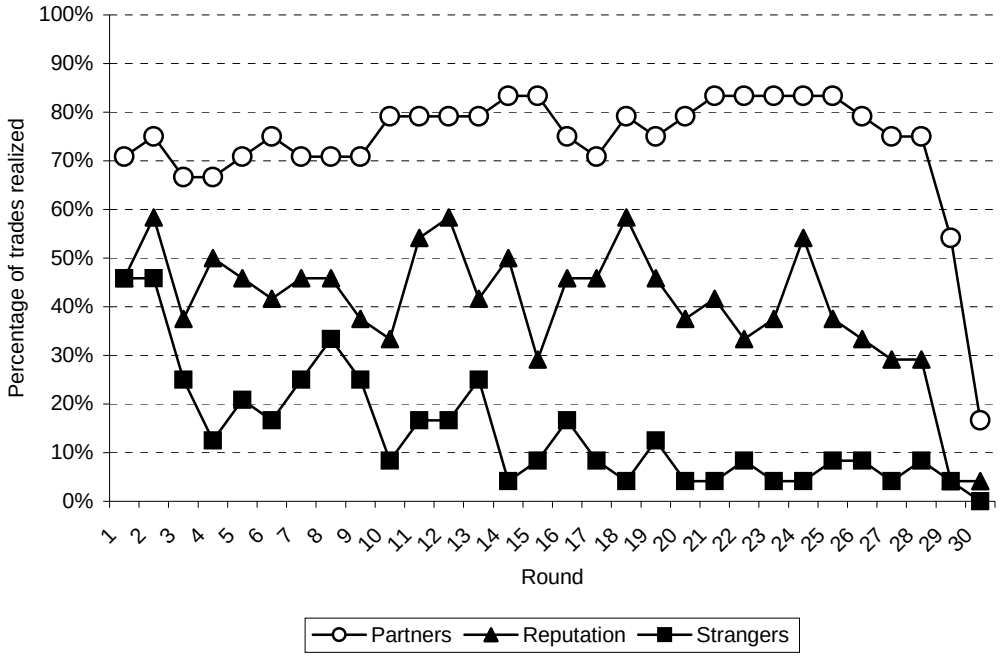
Each market simulation ran for 30 rounds, a fact that all traders were told up front. By theory, we would expect to see few trades in the final round or two – in all of the markets. To see this, consider again the game in Figure 1 and suppose that we are in the 30th round of play. In terms of monetary rewards, the seller's optimal action, regardless of the game's history, is *not*

ship, because since there are no future encounters there are no reputational benefits after round 30. A buyer who notices this should therefore be wary of buying. This remains true regardless of what seller feedback the buyer has, or of whether the buyer and seller are randomly matched or partnered. Good reputation has market value only if it can be leveraged in the future.

Note that if we push this reasoning to its ultimate end, *all* trade unravels: If the seller thinks there will be 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 in the 29th round the buyer should be wary of trade, ditto the 28th round, all the way back to the 1st round. There are several reasons, however, to think that things will not turn out this badly, some are theoretical reasons, others are behavioral. For one, it is unlikely that all buyers and sellers will see the unraveling argument we just gave; some will just not reason it through while others will continue to trade out of principle. The thing is that even if you are able to reason it through, but some of your partners do not, then you may be better off *acting* as if you too do not. That is, if your partners think reputation is valuable for future rounds, then you are better off acting like it is too; at least for some time, at least until all players can see that reputation will not pay any more (Kreps, Milgrom, Roberts, Wilson, 1982). Studies suggest that many people have trouble looking ahead more than one or two steps, so from this we would expect less trading in the last two rounds (e.g., Selten and Stöcker, 1986).

Of course, if as in the strangers markets, there is no information about reputation available, one need not reason far ahead at all to see the moral hazard problem: Since trade is always anonymous, one can see immediately that the seller has no monetary incentive to be trustworthy. So, in sum, theory leads us to believe that there will be more trading in the reputation and partners markets than in the strangers markets. Since, in theory, information is the only important thing to supporting trade, we expect to see about the same amount of trade in both reputation and partners, although in both cases we would not be surprised if the trade fell off in the last couple of rounds of the market.

Figure 3. Trading in shipping markets, by round, averaged over all simulations



The main results of the experiment, the trading rates in each kind of market, are displayed in Figure 3. First observe that trading rates in the strangers markets start fairly strong (about 45% on average), but quickly tail off to less than 10%. In these markets, but a few sellers are trustworthy, something buyers quickly pick up on. While trading in the reputation markets starts at about the same level as in the strangers markets, the trading levels remain high until the last two rounds, where, as suggested by theory, they crash. Overall trading levels are substantially higher in the reputation markets than in the strangers markets. But they are substantially higher still in the partners markets, hovering around 80 percent for most rounds until they too crash at the end. (All of the noted differences are significant at the 5% level, two-tail tests.)

Overall, the experiment indicates that there is some truth in the theory – but also some important problems. On the one hand, information about reputation does succeed in elevating trade. The strategic nature of reputation is evident from the low trading rates in both the strangers markets and in the last two rounds of all the markets. There are some people who are

trustworthy on the basis of principle, but there are many who need an economic incentive. On the other hand, partnered traders manage to maintain substantially higher levels of trade. This might indicate there is something special about partnered relationships; we cannot entirely rule that out. However, further analysis (available in our paper) suggests that much of the difference is due to the differences in the way information flows through the two markets, providing the traders with different incentives. In both of these markets, a buyer choosing to trust can be thought of as investment; I trust to see if you are the kind of seller who is trustworthy. We have already discussed how a seller choosing to be trustworthy can be seen as an investment. It turns out that the return on these investments is higher in the partner treatments (trusting yields information about your partner; being trustworthy fills your partner with future trust) than in the reputation markets (trusting yields information useful to other buyers; some of the value of trustworthiness goes to a buyer you may never deal with again). These public good aspects of trust and trustworthiness in the reputation treatments prevent the electronic reputation system from being as effective as lasting relationships.

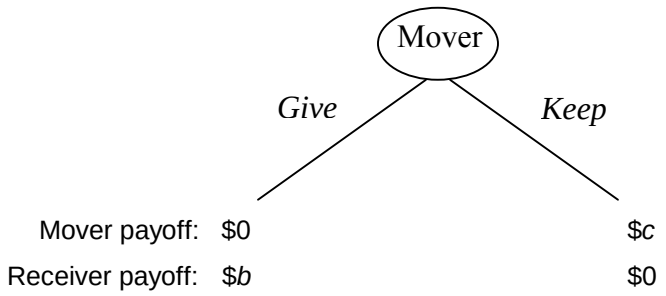
3 Two-way reputation systems: Markets where both sides' reputation is an issue

In many markets, reputation is a two-sided affair, with the trustworthiness of both buyer and seller at issue. It turns out that, in theory, these kinds of reputation systems, which track the reputation of both sides of the market, require a great deal more information to facilitate cooperation than do systems for one sided markets. Showing this in a full blown market setting, like the one we used in the last section, turns out to be rather complicated since other, superfluous issues get in the way. The base issues, however, can be captured by the extraordinarily simple 'image scoring game' (Nowak and Sigmund, 1998).

As simple as it is, the image scoring game is also ambitious. It aims to illustrate how reputation can be used to facilitate cooperation in just about any social situation, market or otherwise, where voluntary cooperation, and so the reputation of *all* participants is an issue.

As with the market we studied in the last section, the image scoring game conceives of the group interacting over a series of rounds. Again, in each round, people are paired off at random. Whether it is because the group is interacting anonymously (say, through computer interface) or because the group is large, we may assume that the people so partnered are strangers to one another. One person in the pair, designated the title of 'mover,' is given the opportunity to give a favor to the other, designated the title of 'receiver.' These designations are assigned randomly, so over many rounds, each player is a mover about half the time and a receiver the other half. Giving a favor would cost the mover c and benefit the receiver $b > c > 0$. Figure 4 illustrates the situation.

Figure 4. When Mover meets Receiver in the image scoring game



But why would the mover want to help the receiver? The strictly self-interested answer is he wouldn't – unless, of course, giving now induced a future mover to give when he is the receiver. In fact, notice that the efficient outcome in this game, the outcome that maximizes the total social benefits, is for everyone to give when they are the mover. The rub is that if everyone else gives, then I make more money by keeping when in the role of the mover (while graciously accepting others' beneficence when in the role of receiver). This is where reputation can help, by providing the information necessary to reward those who give with giving and punishing those who do not with keeping.

At first, this game might seem miles away from the buyer-seller auction context we mentioned at the beginning of the section. But in terms of the reputation issue at stake, the two

are actually quite similar. Both involve evaluating everybody's reputation, and cooperating, whether that means 'giving' or 'trading,' only if their reputation warrants it. The main difference is that in the market situation there is simultaneity to the evaluation – both buyer and seller evaluate the other's reputation at the same time. In the image scoring game, the evaluation is one-at-a-time. One-at-a-time makes for a more lucid discussion, but does not do violence to the basic reputation issue.

And the basic reputation issue is more complicated than before. To see why, consider the kind of discriminating strategy that worked so well for the one-sided case. It required a relatively small amount of information about the image score (reputation): the mover gives if he knows the receiver played give the last time as a mover, and keeps if the receiver last played keep. Before this was a sound strategy for all, effectively curbing moral hazard and supporting cooperation. To see where it runs into trouble here, suppose you are the mover matched with someone who last player *keep* as a mover. Do you really want to play *keep* as the discriminating strategy stipulates? If you do, then the next time you are the receiver, you can expect the mover to play *keep* on you (if others too play the discriminating strategy). Consequently, you make more money playing *give* (lose c now, pays b later) than playing *keep* (gain c now, pays 0 later). The problem is that if enough people decide to *give* to keepers then it pays to be a keeper. And if it pays to have a bad reputation, then why have a good one? Why cooperate and give?

Of course, if enough people are willing to punish keepers, say out of a sense of social obligation, or perhaps because they do not think far enough ahead about the incentives, then cooperation may subsist on discriminating strategies in spite of the flaw. It would be nice if this were so because the amount of information about a person's reputation necessary for the system to function – what they did last time as a mover, something we will call *first order information* – is minimal. Another reason to suppose that it might work, is that the Maghribi trader system had a similar flaw – what incentive did merchants have to expend time, paper, ink and postage to expose a dishonest associate? – but the system nevertheless appears to have functioned well for many years.

But for the moment, suppose first order information is not enough. What information would be necessary to fix things? We could add *second order information* to the image score. Now the receiver's image score would include not only what he did last time as a mover, but also what the receiver he faced did last time as a mover. For example, the image score might state that the receiver "last played keep with a player who last played give." We call this second order information because of its recursive nature. This amount of information pushes the unraveling problem back by a step. To see this, consider a mover who, for the first time, encounters a receiver who played keep on a giver. To support his punishment, keeping on a keeper would have to be rewarded, meaning that there needs to be giving to someone who gives to a keeper – which is not consistent with self-interest since keeping on a keeper pays more. So now players would have to think *two steps* ahead, and be confident others do so as well, before cooperation would unravel. Is that enough?

In theory, the only way to be really confident that cooperation will not unravel is if either a complete history of the game is available to all players, or if some sort of mechanism or institution is available to process and to provide the necessary information honestly (Milgrom, North and Weingast, 1990, Kandori, 1992, and Okuno-Fujiwara and Postlewaite, 1995). For this reason, some theorists have cautioned that indirect reciprocal systems might not be stable outside of very small groups where the information demands are relatively modest.

There are, however, reasons to believe that these systems are more stable than strict theory would suggest. As mentioned earlier, experimental research, much of it on prisoner's and other dilemma games, finds that people's ability to do backwards induction is rather limited, and that they tend to be myopic in their ability to look ahead. We might therefore conjecture that second order information is enough to support a discriminating strategy, or at least that second order information would yield more giving than first order information.

To find out, we conducted an experiment, running the image scoring game under three different information conditions: no information, first order information and second order information (Bolton, Katok and Ockenfels, 2001). Subjects were Penn State University students,

mostly undergraduates from various fields of study, and recruited by fliers posted around campus. In total, there were 192 participants. We ran two image scoring games for each information condition, each game with 16 subjects playing for 14 rounds. Each round, subjects were anonymously paired, interfacing with one another via computers. The value of a gift, *b*, was \$1.25 and the cost of giving, *c*, was \$0.75. Subjects knew that they would be in each role, mover or receiver, for half the trials (7 times) and roles would generally rotate between rounds.

Figure 5. Cooperation in the image scoring game, by round, averaged over sessions

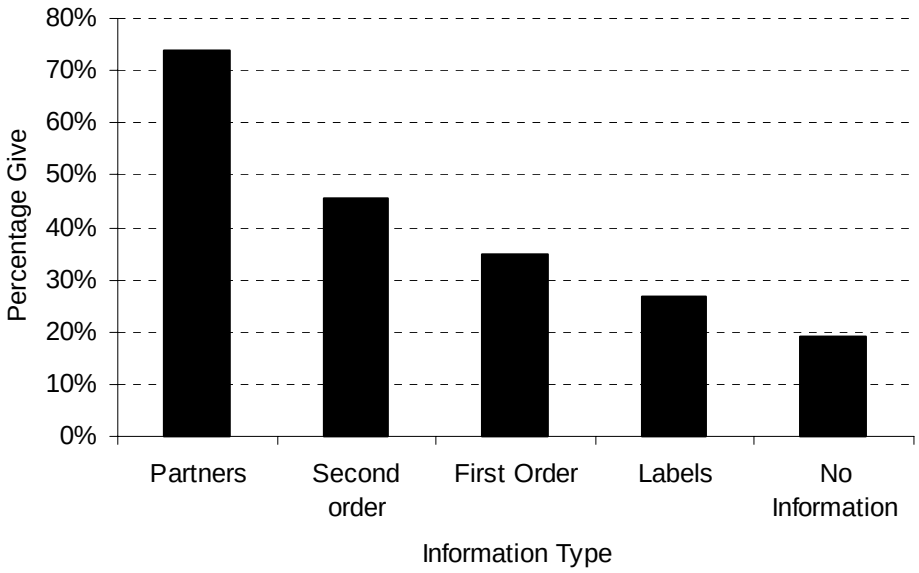
The main results of the experiment appear in Figure 5. When there is no information about reputation available, giving quickly tails off to rather low levels (19% averaged over all rounds). First order information greatly increases giving (35%) and second order information increases it further still (46%). The hypothesis that more information increases giving yields significance at the 5% level. Note in all cases, and as in the buyer-seller experiment of the previous section, giving tails off in the last two rounds pretty much independent of the information condition, again evidence of the strategic nature of reputation building.

But while information does improve cooperation, there is still considerable room for improvement. One thing we might try is to process *all* the information available for our players so that they might then apply a simple discrimination strategy in a way that cannot be cheated on (Kandori, 1992). For instance, in each round we might label each player as a member of either the 'matcher' club or the 'non-matcher' club according to the following rule:

- In the first round, everybody is a matcher.
- In every round after that, a player's label is updated as follows:
    - If the player gave to a matcher the last time he was mover, he is a matcher
    - If the player kept on a non-matcher, he is a matcher
    - If the player did anything else, he is a non-matcher

Now consider a discrimination strategy that stipulates giving to a matcher and keeping on a non-matcher. If everyone follows this rule, then everyone will stay a matcher and there will be 100 percent giving. Moreover, you cannot benefit by cheating. If you keep on a matcher, you become a non-matcher which lines you up to be punish since the next time you are matched with a mover, he will keep on you. And punishment is now with impunity: keeping on a non-matcher allows a mover to maintain matching status – he won't be punished for doing the right thing. So when *all* information is processed in this way, the discriminating strategy looks pretty air tight. At least that's the theory.

Figure 6. Giving levels in image scoring games under different information type



It turns out, though, that this particular theory does not work so well in practice, or at least that is what further tests we ran seem to show. A summary of these tests is given in Figure 6, together with the previous results for the sake of comparison. We also ran a 'partners' image scoring game, quite similar in nature to the partners buyer-seller markets. As with the latter, partners induces far more cooperation than in any other version of the game. But the labeling scheme does but a little better than the no information games and somewhat less well than even first order information. That's a surprise, and one for which we have not yet been able to come up with a detailed explanation. There are other versions of the labeling scheme that are, in theory, just as effective as the one we described. We have tried a couple of these with no better luck than what you see in Figure 6. The most we can say is that it appears that people respond more favorably to reputational reports about recent past *actions* than they do to reputational reports that *filter* actions.

4 Discussion and summary

In most supply chain transactions, successful buying and selling has traditionally been based on lasting relationships or secured by effective laws. The anonymous nature of online B2B environments and the opportunity of global and flexible exchange patterns in the Internet, however, tend to weaken the roles of repeated interaction and legal institutions. This poses an elementary problem to the effectiveness of e-supply chains unless mechanisms are developed that promote trust and trustworthiness and that function well even in adverse online environments.

Economic theory, field evidence and laboratory evidence all point in the same direction: The online trust gap can to some extent be filled with the help of online reputation mechanisms. Such mechanisms institutionalize word of mouth and thus make trust and trustworthiness profitable traits − even absent enforceable laws and personal relationships. Mutually beneficial trade among distant strangers in anonymous online markets is in principle feasible.

At the same time, however, our empirical research indicates that electronic reputation mechanisms do not easily completely substitute for partner relationships. The risk of trading with a cheater is in all our studies higher when reputation information is disseminated by an automated system than when it flows directly between permanent trading partners.

Our studies are all done in a highly controlled laboratory environment – but it's an environment that arguably favors reputation systems; that is, there is even more reason to worry about these systems outside the lab. These systems aim to curb moral hazard by collecting and disseminating the kind of information that is available to trading partners in personal, repeated relationships. But both collecting and distributing this information is more difficult in the field than it was for us in the lab. In particular, online reputation mechanisms typically rely on the *voluntary* provision of feedback information of experienced market participants. But when providing this information is costly, providing feedback is already a cooperative effort in itself,

because the benefits of feedback information go to others.[5] In this sense, reputation mechanisms appear to shift the dilemma to another level rather than solving it − agents are supposed to cooperate with the reputation mechanism rather than with their trading partners. On eBay, for instance, sellers are only rated 50 percent of the time (Resnick and Zeckhauser, 2001). Furthermore, traders may have incentives to manipulate feedback information to, say, artificially raise a confederate's reputation, or to impugn a competitor. Finally, as noted in Section 3, the amount of information needed in a two-way system may be enormous and easily go beyond the scope of the traders' information processing capabilities.

We hasten to add that these problems are not the end of the story but rather the beginning. Ours and others' research offers promising ways to improve the effectiveness of automated reputation systems:

- The public good aspect of trust and trustworthiness identified in Section 3 can be weakened by inhibiting identity changes, and by informing all market participants about trustworthiness indicators of the whole market, and not only about the trustworthiness of individual traders.[6]
- Our findings point to the kind of information statistics that are sufficient, and those that are not needed to make the agents trust in each other. In particular, a cumulative measure of reputation, as applied by most online reputation services, does not seem to be appropriate because it hides information critical to the buyers' decision to trust.
- Incentives to provide information can be created by 'micro-payments.' If, in addition, these payments depend on the predictive value of the feedback for the future performance of the seller, then incentives to create honest feedback are obtainable (Miller, Resnick and Zeckhauser, 2002).

[5] Recall that because the informational benefit of a feedback goes to others, we also observe less trust in Bolton et al.'s (2002) reputation market than in the partner market.

[6] See Bolton et al. (2002) for the details, and see Friedman and Resnick (2001) for theoretical reasons why one should gain control over the agents' identities and how this could be technically realized.

The research on automated reputation systems is just beginning, and the existing reputation systems on the Internet are still in their infancy. The emerging data suggest that there is plenty of room for improvement. But they also provide clues of how a clever architecture of electronic reputation mechanisms based on theory and empirical research might just be able to finally fill the trust gap.

References

Alexander, R.D. (1987). *The Biology of Moral Systems*. New York: Aldine De Gruyter.

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

Arrow, Kenneth (1974). "The Limits of Organization." New York: Norton, York.

Axelrod, Robert (1984). "The Evolution of Cooperation." New York: Basic Books.

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*.

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., A. E. Katok, and A. Ockenfels (2002). "How Effective are Online Reputation Mechanisms? An Experimental Investigation." Working Paper, Smeal College of Business Administration, Penn State University.

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

Dennehy, Michelle (2000), "Making B2B Trustworthy," special report posted on www.Auction.com.

Dewan, S., and V. Hsu (2001). "Trust in Electronic Markets: Price Discovery in Generalist versus Specialty Online Auctions." Working paper, University of Washington.

Federal Bureau of Investigation (2002). "New Report Shows what Internet Scams Cost Americans Most." Press Release on April 9, 2002.

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.

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

Kandori, M. (1992). "Social Norms and Community Enforcement." *Review of Economic Studies*, 59, 63-80.

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 (forthcoming). “Does a Seller’s e-commerce Reputation Matter?” *Journal of Industrial Economics*.

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.

Miller, N., P. Resnick, and R. Zeckhauser (2002). „Eliciting Honest Feedback in Electronic Markets.“ Working paper, Harvard University.

Morphy, E (2001) “Putting the Strategy Back in Your Sourcing.” *iSource* http://www.isourceonline.com/article.asp?article_id=2049.

Nowak, M.A. and K. Sigmund (1998). “Evolution of Indirect Reciprocity by Image Scoring.” *Nature*, 393, 573-577.

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

Okuno-Fujiwara, M and A. Postlewaite (1995). "Social Norms and Random Matching Games." *Games and Economic Behavior*, 9, 79-109.

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, J. Swanson, and K. Lockwood (2002). “The Value of Reputation on eBay: A Controlled Experiment.” Working Paper, University of Michigan.

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

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.