Analyse & Kritik 26/2004 (© Lucius & Lucius, Stuttgart) p. 185–202

Gary E. Bolton/Elena Katok/Axel Ockenfels

Trust among Internet Traders

A Behavioral Economics Approach[*]

Abstract: Standard economic theory does not capture trust among anonymous Internet traders. But when traders are allowed to have social preferences, uncertainty about a seller's morals opens the door for trust, reward, exploitation and reputation building. We report experiments suggesting that sellers' intrinsic motivations to be trustworthy are not sufficient to sustain trade when not complemented by a feedback system. We demonstrate that it is the interaction of social preferences and cleverly designed reputation mechanisms that solves to a large extent the trust problem on Internet market platforms. However, economic theory and social preference models tend to underestimate the difficulties of promoting trust in anonymous online trading communities.

1. Introduction: Trust on Internet Market Platforms

In a traditional market place, trust plays little role in the exchange of a good of recognizable quality. Being physically together in time and space allows buyers to inspect the item before buying, and permits a 'simultaneous' exchange of money and object. The situation is quite different in computer-mediated online markets such as eBay's auction house and Amazon's market platform for used books. Transactions on these platforms are characterized by asynchronous (sequential) actions of anonymous traders, operating at spatially disperse locations. In such a setting, trust becomes a critical issue.

It is not just the scattering of trade in space and time that pose a challenge to Internet exchange, it is also the medium of communication *per se*. Computer-mediated communication makes it more difficult to signal trustworthiness and to promote cooperation than 'richer' communication media such as face-to-face communication (Frank 1988; Brosig et al. 2003). Trading via computer networks also allows buyers and sellers to freely choose a trader identity, different from one's true 'offline identity', that can be changed, often with ease. In addition, lasting personal relationships on Internet market platforms are infrequent; unlike in traditional market environments, one-shot interaction is the rule rather than the exception (Resnick/Zeckhauser 2002). Thus, cyberspace makes it particularly difficult to develop social and economic bonding that supports the emergence of trust and trustworthiness in more traditional markets.[1]

[*] We thank Bernd Lahno for very helpful comments. Ockenfels gratefully acknowledges the support of the Deutsche Forschungsgemeinschaft through the Emmy Noether-program. Bolton and Katok both gratefully acknowledge the financial support of the National Science Foundation.

[1] There are more difficulties. For example, unlike in traditional auction markets, such as

To deal with the problem, many online market platforms implement electronic reputation mechanisms that collect, process, and distribute large amounts of information about past trading activities of the market participants. These reputation mechanisms attempt to emulate traditional word-of-mouth networks and thus promote trust and trustworthiness among strangers. On eBay's 'feedback forum', for instance, market participants continuously assess and reassess their transaction partners. Figure 1 provides a sample of the type of feedback that is left on the site. Each user's assessment is made public and so forms his individual reputation that others can condition their behavior on. As eBay's founder Pierre Omidyar writes in a letter to all users:

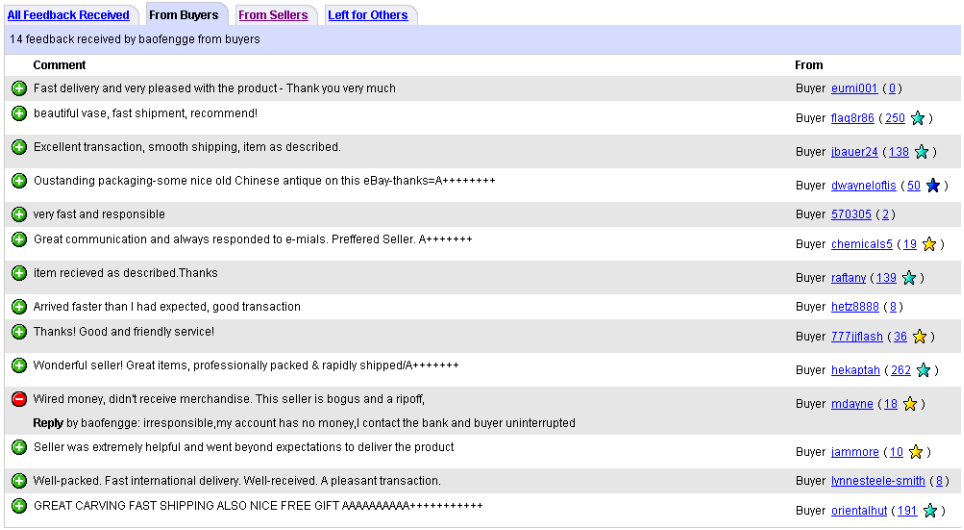
> "By creating an open market that encourages honest dealings, I hope to make it easier to conduct business with strangers over the net. Most people are honest. And they mean well. Some people go out of their way to make things right. I've heard great stories about the honesty of people here. But some people are dishonest. Or deceptive. ... It's a fact of life. But here, those people can't hide. We'll drive them away. Protect others from them. This grand hope depends on your active participation. Become a registered user. Use our Feedback Forum. Give praise where it is due; make complaints where appropriate."
> http://pages.ebay.com/community/news/founders-letter.html

In this paper, we describe a simple experimental game that captures the key problems of trust and trustworthiness on Internet market platforms from an economics point of view. We present new experiments and survey other evidence indicating why and when people do or do not trust. We also compare the empirical evidence with economic theories of decision making and institutional design. The two critical findings from this work that we will emphasize are, first, that understanding trust and trustworthy behavior requires us to consider a mix of human motives, no one motive being sufficient, and second, that the expression of these motives is sensitive to the institutional and strategic environment. The important implication is that trust and trustworthy behavior is neither wholly a matter of social norms and morality, as it is sometimes popularly assumed, nor entirely a matter of institutional design, as economists sometimes assume. Understanding how trust and trustworthiness is, and can be, sustained in an environment like the Internet will require a coming to grips with the interplay between the two.

The paper is organized as follows: In section 2, we motivate and describe the game. In section 3, we analyze one-shot encounters of the game from the point

---

in Christie's and Sotheby's auction houses, transactions among private individuals conducted on online market platforms such as eBay and Amazon are typically not secured; that is, eBay and Amazon neither attest to the credibility of the private sellers nor do they guarantee item quality. Also, selling on the Internet is typically cheap, so that signaling trustworthiness through 'fancy store fronts' is not possible (although fancy homepages or elaborately designed offers may well signal professional commitment and competence). Finally, formal contracting and legal enforcement is more costly (at least relative to the item's value) than in traditional markets.



All Feedback Received | From Buyers | From Sellers | Left for Others

14 feedback received by baofengge from buyers

Comment	From
Fast delivery and very pleased with the product - Thank you very much	Buyer eumi001 (0)
beautiful vase, fast shipment, recommend!	Buyer flaq8r86 (250)
Excellent transaction, smooth shipping, item as described.	Buyer jbauer24 (138)
Oustanding packaging-some nice old Chinese antique on this eBay-thanks=A++++++++	Buyer dwayneloftis (50)
very fast and responsible	Buyer 570305 (2)
Great communication and always responded to e-mials. Preffered Seller. A+++++++	Buyer chemicals5 (19)
item recieved as described.Thanks	Buyer raftany (139)
Arrived faster than I had expected, good transaction	Buyer hetz8888 (8)
Thanks! Good and friendly service!	Buyer 777jiffash (36)
Wonderful seller! Great items, professionally packed & rapidly shipped/A+++++++	Buyer hekaptah (262)
Wired money, didn't receive merchandise. This seller is bogus and a ripoff, **Reply** by baofengge: irresponsible,my account has no money,I contact the bank and buyer uninterrupted	Buyer mdayne (18)
Seller was extremely helpful and went beyond expectations to deliver the product	Buyer jammore (10)
Well-packed. Fast international delivery. Well-received. A pleasant transaction.	Buyer lynnesteele-smith (8)
GREAT CARVING FAST SHIPPING ALSO NICE FREE GIFT AAAAAAAAA++++++++++	Buyer orientalhut (191)

Figure 1: An example of buyer feedback at eBay (left for a seller of Chinese antiques)

of view of recent behavioral theories that seek to explain when trust and trustworthy behavior will arise. In section 4, we focus on institutional explanations of how trust and trustworthiness can be created even if traders are ultimately guided by their material self-interest; these explanations all pivot on repeated encounters. Section 5 summarizes the findings.

2. The Trust Dilemma

Typically, in online trading, a seller posts an offer on the market platform that includes a description of the item and its condition, and a price at which he is willing to sell. Then, a willing buyer sends the money to the seller. Finally, upon receiving the money, the seller is supposed to ship the item to the buyer.[2]

With no outside control, having received the money, the seller can profit from not sending the item or sending poorer quality than promised. Anticipating this moral hazard, buyers may not be willing to buy. As a consequence, trading that would make everybody better off would not take place. This is the essential trust dilemma that online markets need navigate. The game-tree in figure 2 captures this dilemma (for the moment, ignore the number in the brackets).

Both the seller and the buyer are endowed with 35, which is the payoff when no trade takes place. The seller offers an item for sale at a price of 35 which has a value of 50 to the buyer. The seller's cost of providing the buyer with the

[2] Here we assume that the seller fixes the price, such as on Amazon's used book market. However, all arguments in this paper hold equally if the price is endogenously determined, such as in eBay's auctions (in this case the auction winner is the buyer).

item – costs associated with executing the trade, shipping, handling, as well as production costs – is 20. So each successfully completed trade creates a consumer surplus of 15 and a net profit of 15 for the seller. If the buyer chooses to *buy* the item, he sends 35 to the seller, who then has to decide whether to ship the item. If the seller does not ship, he receives the price plus his endowment of 35 for a total of 70. If he ships, he receives the price minus the costs plus his endowment for a total of 50. If the buyer chooses not to buy the item, no trade occurs.

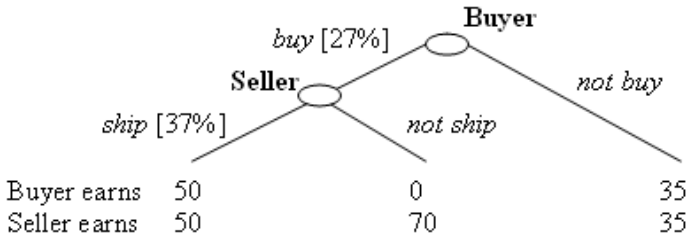


Figure 2: The basic trust game

The game captures the critical features of online trade. First, traders move sequentially because they are dispersed in space and time. Second, the seller's pecuniary motive dictates to keep the money along with his endowment and to not ship the item. In this case, the buyer would lose his endowment and end up with nothing. If the buyer anticipates that the seller does not ship, he will not buy and thus no transaction will take place. This is the prediction of economic theory assuming that traders are only guided by pecuniary concerns. Third, interaction is anonymous. In particular, the name attached to the seller does not reveal information about his trustworthiness.[3] Fourth, the trust-game is one-shot, meaning that there is no common history or common future among traders that could give them the opportunity to reward or punish each other for past behavior or to develop other kinds of social or economic bonding with each other. (As we will see later, however, even one-shot interactions can be strategically linked via information flows.) And fifth, since legal enforcement is typically too expensive to be a credible threat, prosecution when the item is not shipped is not an option in the game.[4] In the next sections, we discuss

[3] Trustworthiness is no issue for known reputable sellers such as barnesandnoble.com. But we will concentrate here on trading among *anonymous individuals* on market platforms with large numbers of buyers and sellers.

[4] The game abstracts away a couple of features that may be relevant to online markets. For instance, we assume that there is no uncertainty about the item's quality (and so there is no reason to have a first game stage in which the seller describes the item's condition). However, the seller's strategy 'not ship' can easily be reinterpreted as 'lying about the item's true quality'. Also, in principle one can easily introduce legal enforcement and courts (see, e.g., Güth/Ockenfels forthcoming, for an example), for, say, high-stake transactions.

experimental evidence using the basic trust game as a vehicle for studying trust and trustworthiness.

3. Trust in One-Shot Games: Behavioral Explanations

3.1 The Nature of Trust

(a) Distributional aspects

From the point of view of economic analysis, one can hardly think about a more trust-unfriendly environment than online market platforms; yet, many of these platforms flourish. So what makes these platforms work? One possible answer is that our basic trust game does accurately describe the market institution but that people *behave* in a different way than the economic analysis suggests. In particular, it is sometimes argued that people are not as selfish as economists typically assume, and that traders care about morals. In fact, in trust games and related anonymous one-shot games (like the prisoner's dilemma game and the ultimatum game), psychologists, sociologists, experimental economists and others have identified a couple of non-pecuniary motives that may drive behavior. Most prominently and actively discussed in the recent economics literature are concerns for fairness (Fehr/Schmidt 1999; Bolton/Ockenfels 2000) and reciprocity (Rabin 1993; Dufwenberg/Kirchsteiger forthcoming). Some authors also put forward an individual concern for efficiency (Charness/Rabin 2002).

All these social preference models assume that traders care about their own monetary payoff but that some traders may additionally be concerned with the social impact of their behavior. Reciprocity models conjecture that people tend to be kind in response to kindness and unkind in response to unkindness, while fairness models posit that some individuals may have a preference for equitably sharing the efficiency gains from trade. That is, in our trust game, reciprocity models suggest that a seller ships because the buyer was so *kind* to buy, whereas fairness models suggest that he ships because otherwise the payoff *distribution* would be unfair. Models of efficiency striving behavior, finally, suggest that traders care about the sum of payoffs. That is, a seller in our trust game may ship because this increases the pie to be distributed—independent of reciprocal and distributive fairness.

One important implication of all these models is that it is not the buyers' social concerns that drive the trading activities but the sellers' social motives. As long as the sellers are strictly selfish and rational, there will be no shipping and thus no buying by rational buyers. Even a perfectly fair-minded buyer, always striving for equitable outcomes, would prefer 35 for both players over 70 for the seller and zero for himself (this holds, e.g., regardless of whether we apply Fehr and Schmidt's or Bolton and Ockenfels' model here). Similarly, in reciprocity models, a buyer is only willing to buy if the seller reciprocates by shipping (again, this holds regardless of the specific reciprocity model). Finally, a buyer concerned about efficiency cannot increase efficiency by sending money to a seller who is unwilling to materialize the efficiency gain by completing the

trade. Thus, the only plausible reason to buy in our basic trust game is that the buyer sufficiently strongly believes that the seller will ship. By this view, the preferences of the buyer are largely irrelevant: shipping in our basic trust game induces a selfish, a reciprocal, a fair-minded or an efficiency striving buyer to trust.

At first glance, the statement that trust is the belief that the seller ships (with a sufficiently high probability) may sound like a trivial statement, but in fact it is highly sensitive to the details of the trust game. To see why, let us consider a very simplified 2-trader version of "ERC", the fairness model developed by Bolton and Ockenfels (2000). In this model, each trader i is motivated by his pecuniary income y_i (equivalent to the payoffs given at the end nodes of the trust game tree) but also prefers fair over unfair final payoff distributions. A payoff distribution is perceived as unfair if a trader either earns more or less than the trading partner, that is, in the simplest form, if $(y_i - y_j)^2 > 0$.[5] Traders motivated this way can be characterized by a 'motivation function' or utility u_i:

$$u_i(y_i, y_j) = y_i - r_i(y_i - y_j)^2$$

Utility is increasing in pecuniary income but decreasing in the inequality of income distribution. Not all traders suffer equally from unfairness; the parameter $r_i \geq 0$ measures the trade-off of trader i between the (sometimes) competing motives. Setting $r_i = 0$ yields the standard economic model of complete selfishness. The distribution of r_i in the population is supposed to follow some cumulative distribution function $F(r_i)$.

In our basic trust game, a buyer i will only trust if his expected payoff from trusting, $p(50 - 0r_i) + (1 - p)(0 - r_i(-70)^2)$, is larger than his certain payoff from not trusting, 35, where p is the probability that a trusted seller will ship.[6] Equivalently, the condition for trust is:

$$r_i < (50p - 35)/(4900(1 - p)).$$

That is, buyer i is only willing to buy if shipping is likely (p is large) and if he is sufficiently *selfish*, that is, if r_i is sufficiently *small*. Why? The buyer can avoid being exploited by the seller by not buying the item. Thus, if fairness (or exploitation aversion) is the main motive, he should not buy. He will only run the risk of being exploited and buy the item when the expected *material* gain of trade is sufficiently large, and if he is sufficiently motivated by the material gain relative to his inequity aversion. In other words, in our basic trust game, *pecuniary motives drive trust*. Buying is motivated by trust in shipping, not by non-pecuniary incentives.

While the simple model we present here oversimplifies certain complexities (and the general ERC model addresses some of these complexities), it sharply

---

[5] The general ERC model assumes that a trader may care about his own relative standing in a reference group, but the model formulation we use here works equally well in this particular 2-traders setting. Also, the general ERC model does not restrict the functional form of inequality aversion in any way. In particular, it is not restricted to be linear, nor quadratic, nor symmetric around equality.

[6] A seller j will ship if $50 > 70 - r_j(70)^2$, or equivalently $r_j > .004$; thus, $p = 1 - F(.004)$.

explains why the act of buying can be identified as trust in our basic trust game and is not confounded by other motives. By the same token, what has been called trust in other standard trust games studied in the literature may have little to do with trust in a reciprocal response. In sequential prisoner's dilemma games or in the well-known investment game by Berg et al. (1995) cooperating, respectively investing, increases efficiency, so that trusting may theoretically reflect a concern for efficiency. The following example demonstrates the difficulties of identifying trust when the basic game is even only slightly changed:

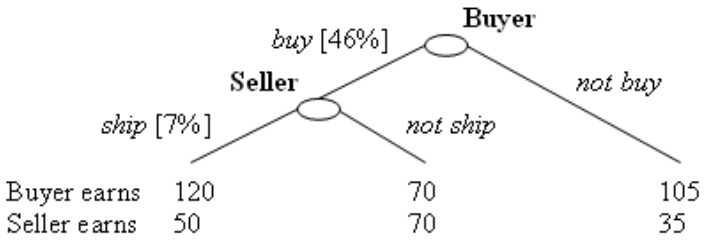


Figure 3: The asymmetric trust game

The asymmetric trust game in Figure 3 is the same game as our basic trust game in Figure 2 with the exception that the buyer gets an additional payment of 70 regardless of the game outcome. While this small change has no effect on the standard economic analysis based on (marginal) pecuniary incentives, it may have a big effect on behavior in a world with fairness. Assume that some traders care about equitable outcomes as described in the model above. This model predicts that *no* seller in the asymmetric trust game, regardless of his r_j, will ever ship, because not shipping maximizes pecuniary outcomes *and* is equitable.[7] That is, this model predicts that there is no shipping, regardless of the extent to which the seller population is selfish or fair. The interesting point is that even if he anticipates that there is no shipping, a fair-minded *buyer* may still want to buy. By buying he sacrifices pecuniary payoff in exchange for a fair distribution.[8] In particular, a buyer i prefers to buy if

$$70 > 105 - r_i(70)^2, \text{ or equivalently if } r_i > .011.$$

However, this buying cannot be interpreted as trust in shipping: no seller will ship and all buyers know that. Buying, in the asymmetric trust game of Figure 3, might, however, be motivated by fairness. Confounding motives for trust like this are largely avoided in the basic trust game.

(b) Procedural aspects

[7] Shipping yields for the seller $50 - r_j(50 - 120)^2$; not shipping yields 70 which is greater for all $r_j \geq 0$.

[8] These effects have also been studied in the so called dictator game (see, e.g., Forsythe et al. 1991).

Trust necessarily involves the risk of being exploited. An interesting question is whether the risk of a bad outcome is treated differently when it is due to uncertainty of the seller's morals or whether it is just determined by a chance move (e.g., there is always a risk that the item is accidentally lost or damaged in the mail). Standard economic theory does not distinguish between different sources of risk when the probabilities for the different states of the world are equivalent. Bohnet and Zeckhauser (forthcoming) demonstrate in an experiment, however, that the risk is perceived differently in a social context than in a non-social context, something that is not captured by any of the social preference models mentioned so far. The game they study is similar to our basic trust game, but the shipping decision is made, not by a seller, but a chance move. The following figure illustrates their game within our simple framework:

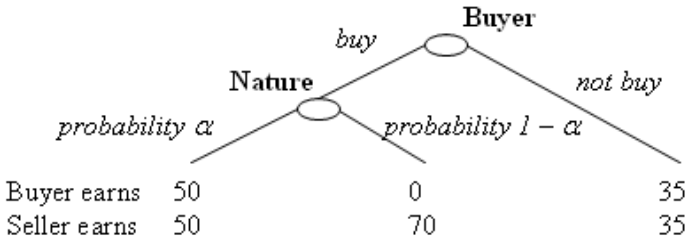


Figure 4: The random-shipping game

In their study, Bohnet and Zeckhauser (forthcoming) asked each buyer for the probability α such that he is indifferent between buying and not buying. They also asked each buyer in the basic trust game (where the sellers determine the shipping probability α endogenously) which probability of being matched with a trustworthy seller makes him indifferent between buying and not buying. By using an incentive-compatible experimental design, Bohnet and Zeckhauser could compare how much risk buyers are willing to take if the source of the risk is the morals of the sellers, as in the basic trust game, to the amount of risk they are willing to take in a lottery, as in the random-shipping game. In fact, Bohnet and Zeckhauser find that "individuals are much more willing to take risks when the outcome is due to chance, as opposed to an equivalent-odds situation where the outcome depends on whether another player proves trustworthy. Taking a chance on the latter risks incurring betrayal costs, costs shown to be above and beyond mere monetary losses."

A related effect has been observed by Blount (1995) in the context of bargaining games. The same offer was more acceptable to a responder when determined by a lottery than when determined by a (human) proposer. A plausible interpretation for these effects is based on a notion of procedural fairness that has been first put forward and tested by Bolton, Brandts and Ockenfels (2003). The idea is that an unfair outcome chosen by a fair lottery is more acceptable than the same unfair outcome chosen by another trader. This notion of procedural fair-

ness is not only consistent with Blount's bargaining experiment but also with Bohnet and Zeckhauser's trust game study. It is, however, inconsistent with both standard economic models and social preference models in behavioral economics that are all invariant to the fairness of procedures (see, however, Bolton et al. 2003, for a sketch of a model that extends the social utility approach to procedures).

3.2 The Nature of Trustworthiness

(a) Distributional aspects

Trust in shipping drives buying. But what drives shipping? As mentioned before, all social motives described so far—fairness, reciprocity and efficiency—may trigger shipping. Here we examine some empirical evidence on this issue.

Let's look first at the basic and the asymmetric trust game in Figures 1 and 2. Buying can, for both games, be equally interpreted as 'kind' behavior and thus, since the seller payoffs are unchanged, reciprocity models would predict no difference of shipping probabilities across the two games.[9] Furthermore, observe that efficiency-seeking behavior would also make no distinction, because an additional endowment to the buyer in the asymmetric trust game does not affect the amount of potential efficiency gains across outcomes. As explained before, however, fairness models predict that while some sellers may be willing to ship in the basic game, nobody is willing to ship in the asymmetric game. Thus, while reciprocity and efficiency models predict no change in seller behavior, fairness models predict less shipping in the asymmetric trust game.

We studied these games in classroom experiments.[10] While in the basic trust game 37 percent of the 30 sellers were willing to ship, in the asymmetric trust game only 7 percent shipped (see the numbers in the brackets in Figures 1 and 2). Thus, the data highly significantly ($p < .01$) reject reciprocity and efficiency as the predominant explanation of trustworthiness; fairness, on the other hand, is largely confirmed (though the two sellers who ship are inconsistent with the fairness motive but consistent with both other motives considered here).

Interestingly, even though far fewer sellers are willing to ship in the asymmetric game, significantly more buyers are willing to buy (27% in the basic trust game and 46% in the asymmetric trust game; $p < .05$). This suggests that fairness, and not trust in shipping, is the main driver for buying in the asymmetric trust game. Similar conclusions about the impact of distributive fairness preferences in trust and related games are drawn by a number of other experimental studies.[11]

---

[9] See, for example, Bolton and Ockenfels forthcoming. Since the intuition is quite plausible, we will make no attempt here to go more into the details of the reciprocity models' mechanics, which can be quite complex.

[10] In this and the other one-shot experiments, we have thirty independent observations, respectively, for each choice separately (buying and shipping). Each subject played only one of the studied games in only one role (buyer or seller). For sellers we applied the strategy method, that is each seller was asked whether he would ship if the buyer chose to buy. Half of all encounters were randomly chosen to be paid out according to the rules of the game. A payoff of ten in the game tree translated to 1 Euro cash.

[11] See Bolton/Ockenfels 2000 and forthcoming and the references cited therein. More re-

(b) Procedural aspect

To what extent do sellers condition their behavior on the history of play? The following game is the same as the basic trust game with the exception that the buying decision is made by a random procedure rather than by the buyer himself. From the view of distributional fairness and standard economic models, nothing has changed for the seller, who still has exactly the same options over payoff distributions. From the view of reciprocity models, however, since the buyer has no choice, buying cannot be attributed to the kindness of the buyer, and thus sellers have no reason to be 'kind in response'.

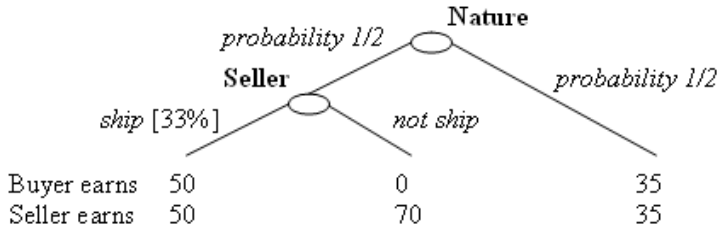


Figure 5: The random-buying game

In a classroom experiment study of the random-buying game, the shipping probability was with 33% only slightly (and not statistically significantly) smaller than in the basic trust game (Figure 5). This indicates that distributional concerns play a major role while the history of play, that is 'intentionality' and lottery procedures, only seem to play a minor role in determining trustworthiness.

This result too is in line with a number of earlier studies. In particular, the first experimental economics study of distributional versus reciprocity-based preferences in Bolton, Brandts and Ockenfels (1998) suggested, in the context of a sequential two-players cooperation game similar in the incentive structure as the one we study here, that the second mover's behavior too was invariant to the history of play and to the 'intentions' of the first mover. A number of subsequent studies (e.g., Brandts/Charness forthcoming; Güth et al. 2001) confirmed this invariance property of behavior. Charness and Rabin (2002) distinguish *negative reciprocity* games, games where other-regarding behavior has a punishment flavor to it (such as ultimatum and other bargaining games), from *positive reciprocity* games, games where other-regarding behavior has a reward flavor to it (such as dictator and trust games). Attribution-based behavior was found to be significant only in negative reciprocity games (also see Bolton et al. 1998; Offerman 2002). Some studies conclude that the response to fair or

---

cently, a couple of authors also found evidence for efficiency-seeking behavior in simple, non-strategic payoff allocation games—games that however cannot easily be related to trust problems (Charness/Rabin 2002; Engelmann/Strobel forthcoming; see Bolton/Ockenfels 2003, for a discussion of some of these results).

unfair behavior is substantially invariant (Blount 1995; Dufwenberg et al. 2001) or only weakly sensitive (Charness forthcoming) to whether the action is taken by a disinterested third party or by the actual game partner. Overall, it seems fair to conclude from the behavioral economics literature that, at least for the kind of trust games that we are considering here (positive reciprocity games), distributional fairness is a major motive in virtually all studies, while reciprocity and intentionality are sometimes found to be a secondary concern.[12]

Summing up, our basic trust game is an appropriate vehicle to study the motives underlying trust and trustworthiness among online traders. There is no trust without trustworthiness. Trustworthiness appears to be mainly driven by distributive fairness concerns. Risk, a key ingredient of trust, is treated differently according to the procedure by which the uncertainty is resolved. On the other hand, trustworthiness appears to be largely invariant to the history of play (e.g., intentions and lottery procedures).

4. Trust in Repeated Games: Combining Behavioral with Institutional Explanations

In expected monetary terms, the probability of a trustworthy seller needs to be at least 70% in order to make buying in the basic trust game profitable. In none of our one-shot games, was the shipping probability substantially greater than half the required amount. Thus, if trust rests solely on behavioral propensities, such as social preferences, there is little hope that trust among strangers can be stable.[13] However, when we embed the trust game in a repeated setting, the interaction of social behavior with *institutional design* such as reputation mechanisms can multiply the impact of the 'truly intrinsic' trustworthiness, because cleverly designed institutions can create strategic incentives to be trustworthy even for selfish and rational sellers. In this section we discuss these incentives, relate them to online market platforms and study their empirical content.

Suppose, for the moment, that the same seller and the same buyer play the basic trust game repeatedly, with an infinite time horizon, and so have no expectation of a stopping round of play. In such a setting, even if all traders are completely selfish and rational, there exist equilibria in which the buyer always buys and the seller always ships. The equilibria can be supported by trigger-strategies that call for a buyer, for instance, to trust as long as the seller ships. If the seller defects only once, he will never be trusted again. If future payoffs are sufficiently important, this gives an incentive to the seller to be trustworthy all the time, and thus for the buyer to trust all the time (Kandori 1992).

---

[12] We note that the fact that the implementation of a lottery affects behavior in the random-shipping game but not in the random-buying game is in line with Bolton et al.'s (2003) concept of procedural fairness. In the random-shipping game, an unfair outcome chosen by a fair lottery may be perceived as procedural fair; however, in the random-buying game an unfair outcome chosen by the seller cannot be justified as procedural fair within their model.

[13] While, of course, payoffs, framing and context are different on computer mediated market platforms compared to our simple classroom game, we believe that an analogous conclusion also holds for actual online market platforms.

There are, however, two problems with this kind of simple equilibria in our context. First, the trading horizon in online market platforms is typically finite. If either the buyer or the seller believes that there will be some upper boundary of items to be traded, cooperation among completely selfish, rational traders will unravel. Second, buying and shipping in infinite game equilibria do not capture trust and trustworthiness because, in equilibrium, there is no uncertainty about each others' behavior; in equilibrium, sellers can have a material incentive to ship, so there is no risk of being exploited.[14] Since this paper is concerned with trust (characterized by a risk of being exploited) in online markets (where traders trade a finite number of items), we prefer to study finitely repeated games. In finite games, however, trust may only emerge when there is some (possibly small) amount of 'truly intrinsic' trustworthiness within the seller population (Wilson 1985).[15] That there is intrinsic trustworthiness has been demonstrated in our experimental studies of the one-shot trust game.

4.1 Strangers' Market

Suppose the market transactions take place over a series of rounds. At the beginning of each round, a potential buyer is matched with a potential seller and they then play the basic trust game in Figure 2. In the simplest version of such a repeated setting, each game is played with a different transaction partner[16] and no information about trade outcomes leaks from one encounter to another one. This market is basically a sequence of one-shot games. Thus, because there is not sufficient intrinsic trustworthiness to make trust profitable, we hypothesize that all trading activities will collapse.

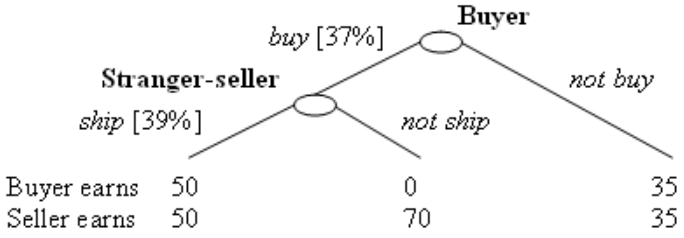


Figure 6: Average behavior in the Strangers market

[14] There are also more subtle equilibria in which cooperation in any given round is uncertain, but, in our view, trust is also not satisfactorily described as an equilibrium selection problem.

[15] If not, all trust and trustworthiness would unravel. We will not make an attempt to show the details of the equilibrium analysis of trust and trustworthiness in finitely repeated games in a world with fairness. The mechanics of these equilibria are relatively complex (see Bolton/Ockenfels 2004, for a theoretical and experimental treatment within a trust game environment). However, we note here that these equilibria in fact capture the risk of being exploited.

[16] In fact, empirical evidence suggests that buyers and sellers on Internet market platforms such as eBay only occasionally come back to the same trading partner (Resnick/Zeckhauser 2002).

In an experiment by Bolton, Katok and Ockenfels (forthcoming), we had subjects play in precisely this format, over 30 rounds. No buyer met the same seller more than once, so we call this experimental treatment the *Strangers market*. All interaction was computer-mediated and anonymous; subjects sat in cubicles in front of computers not knowing the true identity of their trading partners, capturing an important aspect of online trading. The rules and that all rounds were actually paid was common knowledge. Figure 6 shows the average buying and shipping (conditioned on buying) behavior over all rounds.

There is almost the same amount of trustworthiness as in the one-shot version of the trust game in Figure 2. This reflects that the Strangers market does not create additional incentives to be trustworthy as compared to the one-shot game. On the buyer side, there is, on average, more trust in the Strangers market than in the one shot version of the game, possibly reflecting the hope that repeated action may support more cooperation. But the dynamics reveal that buyers respond to the fact that, on average, this expectation was disappointed: They started out by trusting quite a lot but trust quickly collapsed. In fact, the percentage of last round trust was only 0.04%, much less than in the one shot-game, indicating that buying in the one shot game is mainly due to inexperience.

4.2 Reputation Market

From an economic theory perspective, reputation mechanisms of the sort we describe in the introduction can solve many of the trust problems associated with the Internet. In fact, simple theory suggests that information is the key prerequisite to the emergence of trust, because it allows buyers to avoid sellers who are known as fraudulent and to buy only from sellers who have been proven to be trustworthy in the past. Conditioning trust on the seller's history creates incentives for sellers to build up a reputation for being trustworthy, at least when the end of the market is not too close and maintaining a good reputation is still valuable. A reputation of being trustworthy can be developed and sustained even by completely rational and selfish sellers—as long as the probability of being matched with intrinsically trustworthy sellers is strictly positive. That intrinsic trustworthiness exists is demonstrated in our one-shot classroom experiment and in many other experiments. That it is not enough to sustain a trading platform that has no reputation system, however, is demonstrated in our experimental Strangers market. So does a feedback system help promoting trust and trustworthiness as suggested by theory?

In Bolton et al. (forthcoming) the Strangers market was compared to a *Reputation market*, played over 30 rounds, in which, as before, a buyer never met the same seller more than once. However, in this market we introduced a reputation system that, similar to eBay's feedback forum, informs buyers about all past actions of their current seller (see Duffy/Ochs 2003; Bohnet/Huck forthcoming, for related experimental work). Outside the lab, there are a couple of problems with feedback systems having to do with the voluntary provision and strategic manipulation of feedback, and manipulation of online identities etc. (see Dellarocas forthcoming; Bolton/Katok/Ockenfels forthcoming b; Ockenfels 2003).

In our experiments, however, feedback on the seller's past behavior is always shared and reliable (because it is not given by the buyers themselves but by the experimenter), and sellers had no way to change their online identity. This way, the experiment studies the impact of feedback information on trading behavior when an ideal, frictionless reputation mechanism is available.

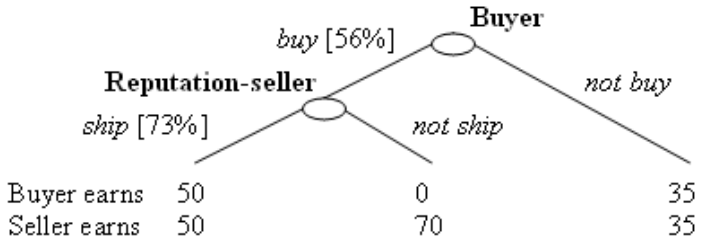


Figure 7: Average behavior in the Reputation market

On average, there is economically and statistically significantly more buying (56 vs. 37 %; $p < .05$) and shipping (73 vs. 39 %; $p < .01$) in the Reputation market than in the Strangers market. In fact, the shipping probability is slightly higher than the threshold of 70% for trust being profitable. As a consequence, the trade dynamics too look quite different than in the Strangers market; trading starts at about the same level as in the Strangers market but then the level remains stable until the very last rounds, when the strategic value of having a reputation for being trustworthy vanishes and virtually all cooperation collapses. We conclude that introducing a feedback system in a market with strangers has a strongly positive impact on trust, trustworthiness and trading efficiency. Both buyers and sellers respond strategically to the information provided.

4.3 Partners Market

The positive impact of electronic reputation mechanisms relative to a Strangers market has also been demonstrated in various field studies (see, e.g., Dellarocas forthcoming, for a recent survey). However, field studies cannot measure trust and trustworthiness on Internet markets relative to offline markets that are typically characterized by repeated interaction between the *same* trading partners.

Theoretically, offline partners-relationships should *not* do better than online markets that implement reputation mechanisms that reliably share all relevant information about the sellers' past behavior. That is, from a theoretical perspective matching is irrelevant for trust to emerge as long as buyers are sufficiently informed—either through their own experience or through the experience of third parties. For the seller to have an incentive to be trustworthy, he need only expect that a future buyer will punish or reward his behavior; whether these punishments or rewards come from the same or from different buyers is irrelevant. The buyer, to induce this trustworthiness, need only be equipped

with sufficient information about the sellers' histories; whether this information comes from one's own experience or from different sources is irrelevant. This is the basic message that derives from the game theoretic models: it is the information, not its source or its dissemination, that matters (Kreps et al. 1982; Ellison 1994; Bolton/Ockenfels 2004).

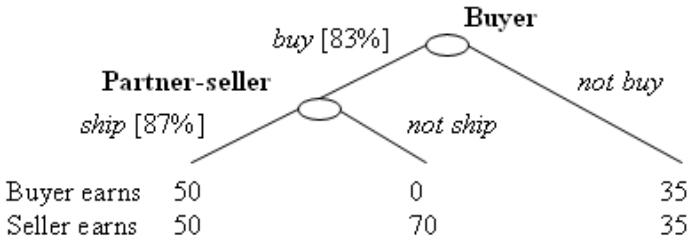


Figure 8: Average behavior in the Partners market

So, does the flow of reputation information make no difference? We tested this by comparing the Reputation market to a *Partners market* where the same two traders were matched over all 30 rounds, keeping everything constant including the anonymity of trading partners, communication channels, etc.

Our experiment demonstrates that the simple theory is quite misleading when it comes to the impact of matching and information flows (Figure 8). There is substantially and statistically significantly more buying (83 vs. 56 %) and shipping (87 vs. 73 %) in the Partners market than in the Reputation market. Thus, even an ideal reputation mechanism cannot substitute for a partners-relationship.[17]

5. Conclusions

Normative economic theory, assuming that rationality and selfishness are common knowledge, does not capture trust in our basic game, because there is either never trustworthiness (in a one-shot or finitely repeated version of the trust game) or there is no risk in 'trusting' (in simple equilibria of the infinitely repeated game). This all changes when traders are allowed to have social preferences. Then, uncertainty about one's seller morals opens the door for trust, reward, exploitation and reputation building—even in the one-shot basic trust game, but especially in the (finitely) repeated game.

In our experiments, sellers' intrinsic motivations to be trustworthy are not sufficient to sustain trade when not complemented by a feedback system. Translated to Internet market platforms, it seems likely that eBay or Amazon's market

[17] As we explain in Bolton et al. forthcoming, this is because a feedback mechanism exhibits a kind of public goods problem in that, unlike in the partners market, the benefits of trust and trustworthy behavior go to the whole community and are not completely internalized.

for used books would quickly collapse without a reputation system. On the other hand, a reputation system without any individual social concerns cannot work either. If no seller would ever ship in a one-shot encounter, and if this were common knowledge among buyers, any cooperation would unravel, because the promise of being trustworthy will always be broken. However, clever institutional design multiplies the positive impact of 'truly intrinsic' trustworthiness that exists even without such systems. It is the interaction of social preferences and reputation mechanisms that solves to a large extent the trust problem on Internet market platforms. The strategic aspects of this interaction is in important parts captured by modern behavioral economics.

At the same time, equilibrium theory and social preference models tend to underestimate the difficulties of promoting trust in anonymous online trading communities. For one, in one-shot games, economic theory does not capture the observation that the *source of risk*, independent of its stochastic properties, affects behavior. In terms of our basic trust game, the procedure by which the uncertainty of whether the item is shipped or not is resolved is critical for the buyers' decision to trust—even if the objective risk parameters are the same. The risk of being exploited by an untrustworthy seller makes buyers more cautious than (the same) risk that the item is accidentally lost by the postman. As a consequence, the trust problem is more severe than models of rational risk behavior would suggest.

Similarly, in repeated games, the role of information as provided by electronic reputation mechanisms is overrated by theory and the role of matching underrated. Information problems can theoretically be solved by appropriate information dissemination, but, in fact, the matching scheme limits what can be maximally reached in anonymous online trade communities. Economic theory does not capture the observation that the *source of information*, independent of its content, affects behavior. As a consequence, trust problems are more severe than models of rational processing of information would suggest.

We think these results suggest that a satisfactory explanation of trust and trustworthiness in Internet markets will require a model that grapples with both the mixed motives of the traders and how these motives inter-play with the institutional design of the market. In our experimental markets, trading behavior—trust and trustworthy behavior included—is clearly strategic

but nevertheless differs in important ways from existing strategic models. By the same token, any model that assumes that trust and trustworthy behavior is entirely a matter of social norms or morality will probably be unable to capture the important behavioral response traders have to other traders strategic options as well as their own.

Bibliography

Berg, J./J. Dickhaut/K. McCabe (1995), Trust, Reciprocity, and Social History, in: *Games and Economic Behavior 10*, 122–142

Blount, S. (1995), When Social Outcomes Aren't Fair: The Effect of Causal Attributions on Preferences, in: *Organizational Behavior and Human Decision Processes 63*, 131–144

Bohnet, I./S. Huck (forthcoming), Repetition and Reputation: Implications for Trust and Trustworthiness When Institutions Change, in: *American Economic Review (Papers and Proceedings)*

— /R. Zeckhauser (forthcoming), Trust, Risk and Betrayal, in: *Journal of Economic Behavior and Organization*

Bolton, G. E./J. Brandts/A. Ockenfels (2003), Fair Procedures: Evidence from Games Involving Lotteries, Working paper, Penn State University

— / — / — (1998), Measuring Motivations for the Reciprocal Responses Observed in a Simple Dilemma Game, in: *Experimental Economics 1*, 207–219

— /E. Katok/A. Ockenfels (forthcoming), How Effective are Online Reputation Mechanisms? An Experimental Investigation, in: *Management Science*

— / — / — (forthcoming b), Cooperation among Strangers with Limited Information about Reputation, in: *Journal of Public Economics*, College of Business Administration, Penn State University

— / — /R. Zwick (1998), Dictator Game Giving: Rules of Fairness versus Acts of Kindness, in: *International Journal of Game Theory 27*, 269–99

Bolton, G. E./A. Ockenfels (forthcoming), A Stress Test of Fairness Measures in Models of Social Utility, in: *Economic Theory*

— / — (2004), Reputation, Information and Matching: Some Theory and Evidence, Work in progress

— / — (2000), ERC: A Theory of Equity, Reciprocity and Competition, in: *American Economic Review 90*, 166–193

Brandts, J./G. Charness (forthcoming), Do Market Conditions Affect Preferences? Evidence from Experimental Markets, in: *Economic Journal*

Brosig, J./A. Ockenfels/J. Weimann (2002), The Effect of Communication Media on Cooperation, in: *German Economic Review 4*, 217–241

Charness, G. (forthcoming), Attribution and Reciprocity in a Experimental Labor Market, in: *Journal of Labor Economics*

— /M. Rabin (2002), Understanding Social Preferences with Simple Tests, in: *Quarterly Journal of Economics 117*, 817–869

Dellarocas, C. (forthcoming), The Digitization of Word-of-Mouth: Promise and Challenges of Online Reputation Mechanisms, in: *Management Science*

Duffy, J./J. Ochs (2003), Cooperative Behavior and the Frequency of Social Interaction, Working paper, University of Pittsburgh

Dufwenberg, M./G. Kirchsteiger (forthcoming), A Theory of Sequential Reciprocity, in: *Games and Economic Behavior*

— /U. Gneezy/W. Güth/E. van Damme (2001), Direct and Indirect Reciprocity: An Experiment, in: *Homo Oeconomicus 18*, 19–30

Ellison, G. (1994), Cooperation in the Prisoner's Dilemma with Anonymous Random Matching, in: *Review of Economic Studies 61*, 567–588

Engelmann, D./M. Strobel (forthcoming), Inequality Aversion, Efficiency, and Maximin Preferences in Simple Distribution Experiments, in: *American Economic Review*

Fehr, E./K. Schmidt (1999), A Theory of Fairness, Competition, and Cooperation, in: *Quarterly Journal of Economics 114*, 817–868

Frank, R. (1988), *Passions within Reason. The Strategic Role of the Emotions*, New York

Güth, W./H. Kliemt/A. Ockenfels (2001), Retributive Responses, in: *Journal of Conflict Resolution 45*, 453–469

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

Kreps, D. M/P. Milgrom/J. Roberts/R. Wilson (1982), Rational Cooperation in the Finitely Repeated Prisoners' Dilemma, in: *Journal of Economic Theory 27*, 245–252

Ockenfels, A. (2003), Reputationsmechanismen auf Internet-Marktplattformen, in: *Zeitschrift für Betriebswirtschaft 73*, 295–315

Offerman, T. (2002), Hurting Hurts More than Helping Helps, in: *European Economic Review 46*, 1423–1437

Rabin, M. (1993), Incorporating Fairness into Game Theory and Economics, in: *American Economic Review 83*, 1281–1302

Resnick, P./R. Zeckhauser (2002), Trust Among Strangers in Internet Transactions: Empirical Analysis of eBay's Reputation System, in: M. R. Baye (ed.), The Economics of the Internet and E-Commerce, *Advances in Applied Microeconomics 11*, Amsterdam, 127–157

Wilson, R. (1985), Reputations in Games and Markets, in: A. Roth (ed.), *Game Theoretic Models of Bargaining*, Cambridge, 27–62