Chris Snijders/Richard Zijdeman

Reputation and Internet Auctions: eBay and Beyond

Abstract: Each day, a countless number of items is sold through online auction sites such as eBay and Ricardo. Though abuse is being reported more and more, transactions seem to be relatively hassle free. A possible explanation for this phenomenon is that the sites' reputation mechanisms prevent opportunistic behavior. To analyze this issue, we first summarize and extend the mechanisms that affect the probability of sale of an item and its price. We then try to replicate the results as found in four recent papers on online auctions. Our analyses reveal that (1) it makes sense to differentiate between 'power sellers' and the less regular users, (2) there are variables that have an effect on sales that are often not controlled for, (3) one should carefully consider how reputation is operationalized, (4) neglecting heteroscedasticity in the data can have serious consequences, and (5) there is some support indicating that effects differ across auction sites.

1. Introduction

1.1 Trust and Online Auctions

Thousands of individuals sell and purchase items from one another on the Internet. These people meet each other on virtual market places or auction sites, such as eBay.com, ePier.com, or qxl.co.uk. Chances are that you are one of them, but in case you do not know how it works, the procedure roughly runs as follows: sellers promote goods they want to sell by giving a product description on an auction site. Buyers can compete with each other through an auction system to purchase the promoted product. The winner pays, usually upfront, the money to the seller and then the seller ships the item. Anybody with access to the internet is a potential seller or buyer on an auction site.

An important aspect of these online C2C markets is that consumers can trade without direct physical interaction. In other words, consumers can engage in trade with others whom they do not meet. In a way, this means that the market gets closer to what a 'perfect market' is supposed to be. Unfortunately, this key aspect of online auctions is its Achilles' heel in trading as well. Lack of physical interaction means that careful checking of products and trading partners is hardly possible. Perhaps an even larger drawback is the time lag between purchasing and receiving goods. At least one party has to send in his good, either the item or the money, without strong guarantees of getting something in return. Nevertheless, some major online auction sites are quite successful. For instance,

in 2001 eBay.com counted 29.4 million registered users (van Swol 2001), while in 2003 eBay claimed to have 12 million items on sale in 18,000 categories on any given day. How can this be? To analyze this situation, previous researchers (e.g. Ba/Pavlou 2002; Diekmann/Wyder 2002) have already mentioned the obvious underlying trust issue. A buyer places trust by bidding on an auctioned item, whereas a seller may honor trust by shipping the auctioned item, or may abuse trust by either shipping an item of lower quality than advertised, or by not shipping the item at all (we disregard the related but smaller trust issues on the seller's side, such as buyers who win an auction but do not pay).

Information often plays an important role in trust problems. In online auctions, two questions are typically of interest to the buyer. Can I trust this buyer? and Can I trust this item? If a buyer is uncertain enough about either one—the buyer's intentions or the product's quality—a potentially mutually profitable sale may not materialize (Akerlof 1970). Therefore, providing a buyer with information on the quality and value of a product and on the 'kind' of seller one is dealing with may help to decrease the trust problem, as auction sites obviously recognize.

At least two mechanisms are in place that may provide a seller with such information: through an extensive product description the seller can try to reduce uncertainty about the quality of the good, and a reputation system exists as a measure of the reliability of the seller. Online auction sites provide a mechanism for their users to evaluate each other. Although not all reputation mechanisms are alike, all do provide information on a seller's past behavior on the auction site. For instance, after a completed auction on eBay.com, buyer and seller can (but need not) evaluate each other by 'leaving feedback' about the other party. Negative feedback is represented by a score of '-1', '0' resembles neutral feedback, while '+1' represents positive feedback. When a user first registers at eBay.com he receives a reputation score of zero. Over time, buyers and sellers leaving feedback leads to the reputation score increasing or decreasing according to the behavior in auctions the user participated in. A more extensive elaboration on eBay's reputation system is available at: http://pages.ebay.com/services/forum/feedback.html.

Different reputation mechanisms have emerged on different auction sites. Most, if not all reputation mechanisms are at least partly quantitatively based. Every review left by a party results in feedback, and the summation of all points is a person's 'reputation score' or 'feedback rating'. In principle the reputation mechanisms can differ with respect to the kinds of feedback that can be given: positive, neutral, or negative. eBay's feedback mechanism—at least at the time of the data collection for this paper—allows positive, negative, and neutral feedback. The reputation score is calculated as the difference between the number of (unique) positive comments and the number of (unique) negative comments. Except for quantitative comments, most larger auction sites such as eBay and QXL allow for small bits of qualitative reputation as well. Seller and buyer can provide a couple of lines of text regarding the party they dealt with. What the reputation mechanisms have in common is that they are all acting as measurements of the trustworthiness of the people participating in the auction.

1.2 Mixed Findings in the Literature

A number of recent papers have considered the effects of reputation on sales and the results are typically mixed. In fact, the most straightforward effect one would expect, a positive reputation having a positive effect on both the probability of sale and the price, is not always found. Some researchers even come up with highly unexpected effects, such as a negative reputation having a *positive* effect on an auction's end price (Eaton 2002; Kauffman/Wood 2000). Dellarocas (2003) provides the most recently published overview; many working papers on the topic are available from the Reputations Research Network (http://databases.si.umich.edu/reputations).

Lee, Im, and Lee (2000) give three possible reasons for the variation in the results found in previous papers by for example Kauffman and Wood (2000) and Ba and Pavlou (2002). First, they mention the differences in the types of products under study. Although they are not very concrete about how this might matter, a likely explanation is that the risk involved in purchasing an item depends on a number of product characteristics, combined with the assumption that a buyer's 'need' to deal with a seller of high reputation differs for different levels of risk. A buyer's risk is larger for items of larger value, so for the larger value items reputation might matter, whereas it has no impact when one considers items of smaller value. The need for information on a seller's trustworthiness may also depend on a product's condition. Ba and Pavlou (2002) believe that the fact that they did not find any significant effects for negative reputation is because they considered used and refurbished products, as opposed to only new products. The probability that an item malfunctions is lower for new products than for used products, the risk of purchasing a broken item is therefore higher for used products than for new products. Accordingly, buyers may be more sensitive to a seller's reputation when purchasing a used item.

A similar argument holds for the complexity of a product. As Buskens mentioned, "complex products leave more opportunities for hidden defects" (Buskens 1999, 132). The more features a product has, the more features might be malfunctioning, increasing a buyer's risk of purchasing a product that is not perfect. Therefore, buyers may have greater need for information on the trustworthiness of the seller when buying more complex products than when buying less complex products. The second factor that might cause mixed results is the way the independent variable of reputation is operationalized. Eaton (2002), for instance, reports that he finds no significant effects for negative reputation unless negative reputation is being dichotomized. Other researchers log-transform the reputation score, include or exclude the negative comments in their measurement, and so on. Third, the way in which data was collected could contribute to the mixed results. While some researchers control for variables concerning product condition, others control for transaction related variables like shipment costs and means of payment, whereas still others control for hardly any other possible intervening variables. Finally, Resnick, Zeckhauser, Swanson, and Lockwood (2002) point to the effects of the different statistical methods that have been

used. Most researchers have used OLS regressions; some have used Heckman-like selection models instead.

1.3 Research Questions

To better understand behavior in online auctions, we employ the following strategy. First, we try to systematize and extend the theory as found in previous papers on reputation in online auctions. We then try to replicate the results of four of those studies using auction data we collected ourselves. In our data we tried to include all of the variables that were included in previous studies, and added some that we considered of possible interest based on our theoretical account. We then test the hypothesized effects using these data. The following research questions are at the core of this study: What is the effect of seller reputation on sales? What is the effect of product description on sales? What other factors affect sales in online auctions? Do effects of reputation differ across auction sites?

The latter question needs some additional explanation. Since most research on online reputation has been conducted on eBay.com, it is wise to think about the extent to which results are 'eBay-specific'. There are indeed some arguments that suggest that effects of reputation may differ across auction sites. For instance, one alternative site we consider is ePier.com. This is a general auction site, largely similar to eBay, but with a different message. ePier.com claims to be "waiting to set internet free" and for this reason charges no listing fees (what you have to pay to list your item) and smaller final value fees (what you have to pay the auction site if the item is sold). In general, ePier.com tries to create a different identity than eBay.com, portraying an image of eBay as the big commercial auction site and of ePier.com as a 'happy community' where one comes to exchange goods with friends. In such a place, if it exists, a reputation system would add no value.

2. Auctions as Trust Problems

Let us first make the resemblance between auctions and trust issues more concrete. Coleman (1990) mentions four elements that define trust situations between a trustor and a trustee. The first is the fact that placing trust provides a trustee with resources, that he otherwise would not have received. The second element is that a trustor is better off placing trust than not placing trust, provided that the trustee is trustworthy. If the trustee on the other hand is not trustworthy, the trustor is worse off placing trust than when not placing trust. Third, Coleman stresses that the placement of trust by a trustor is voluntarily and without formal safeguards. Finally, there should be a time lag between a trustor placing trust and a trustee honoring trust. As one can easily see, auction sales can be considered a trust problem according to this definition.

Figure 1 shows the trust problem between seller and buyer in online auctions (the Trust Game: Dasgupta 1988; Kreps 1990). After a seller has opened the auction, a buyer can choose to either place a bid (cooperate) or not (defect).

If a buyer chooses to defect, she has no chance of either receiving a potential gain (R_b), nor a potential loss (S). Her payoff is then lower than the potential gain, but higher than the potential loss. A seller will try to encourage a buyer to place trust, because if a buyer defects, the seller will receive the lowest payoff possible: P_s. However, if a buyer places trust, she is uncertain whether the seller will honor or abuse trust. The chance of receiving payoff R_b depends on the probability that a seller is trustworthy. Although R_b is the highest possible payoff for a buyer, by cooperating, a buyer runs the risk of receiving the lowest possible payoff S, which means that after a buyer has paid the seller, the seller defects and ships an item of lower quality than advertised or does not ship the item at all.

Basically, if T is larger than R_s, the seller has an incentive to abuse trust. There can be reasons, however, why abuse by the seller is not rewarding after all, for instance because the seller can anticipate that abuse would lead to negative feedback on the auction site, which could frustrate future sales. The 'effective T', so to speak, could then be smaller than R_s. Though the seller is aware of his effective T, the buyer is uncertain about this value, and can try to infer something about it, for instance based on the seller's reputation score. Typically, the buyer is also uncertain about the value of R_b. Again, the buyer can try to decrease this uncertainty before buying. Note that instead of explicitly modeling future trade as new Trust Games, or uncertainty through moves of nature, we choose to define the rewards of future trade and uncertainty implicitly in the values of the payoffs (cf. Raub/Snijders 2001).
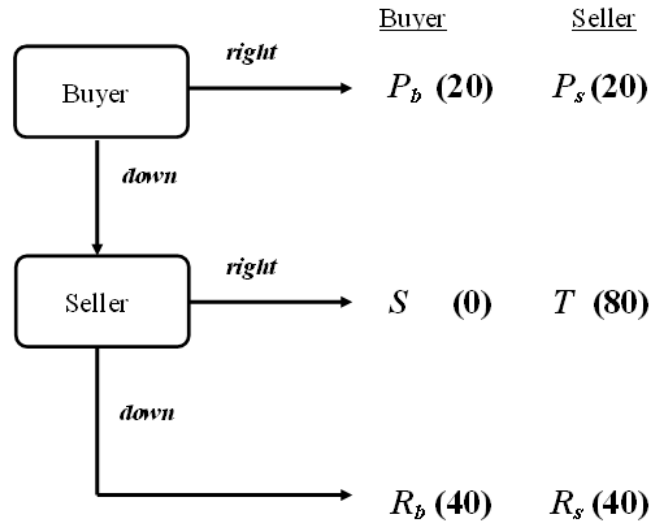

Figure 1: Trust Game. $S < P_b < R_b$ and $P_s < R_s < T$. Example payoffs between brackets.

3. Explaining Trust through Seller Characteristics: Reputation

Although seller characteristics other than reputation could well be of influence on sales and are included in our data collection and analyses (for instance, one could imagine that it helps if a seller has a label 'seller is shopowner' in his auctions), we first focus solely on the seller's reputation.

Reputation: Basic Argument

It is straightforward to show how, at least in theory, reputation can help solving the trust problem. One could do with three assumptions:

1. a winning buyer always provides feedback on a seller's behavior after each completed auction;

2. a winning buyer always provides feedback honestly;

3. a seller's behavior is stable over time, in the sense that his past behavior is a strong predictor of his present behavior.

If a seller succeeds in auctioning a product he receives a reputation rating as indicated by the first assumption. From the second assumption one can conclude that a seller's reputation is an accurate overview of the behavior of that specific seller. The third assumption allows to make inferences of the kind "if this is a high reputation seller, it is likely that my trust will be honored".

Let us now examine to what extent these three assumptions are realistic. Judging from the high reputation scores and the large number of sellers with a (positive) reputation, it is certainly true that buyers (and sellers) provide feedback. However, Resnick and Zeckhauser report that on eBay.com only a little more than 50 percent of the buyers provide feedback and warn that it may be " ... dissatisfied customers [who] are substantially less likely to give feedback" (Resnick/Zeckhauser 2002, 3). There are at least two reasons why it is harder to give negative than positive feedback. First of all, giving negative feedback comes with a certain risk. A seller who is sanctioned by a buyer, could retaliate by giving negative feedback to that buyer. In addition, by giving negative feedback a buyer admits that she has been 'outsmarted' by a seller, which is to a certain degree loss of face for the buyer. The second assumption does not seem to be realistic as well. Those who want to give negative feedback when positive feedback would have been appropriate can do so. Lee et al. have already pointed out that the subjectivity of ratings can indeed lead to unfairly low or high ratings, though others have argued that this is a minor issue "a large volume of accumulated information about a seller might converge to a certain point that represents the level of trustworthiness of the seller" (Lee et al. 2000, 4). There is also an incentive to provide positive feedback regardless of the behavior of the other party. The reason for this is that providing negative feedback regarding a bad seller or buyer is a collective good. The negative feedback is useful for others who may want to engage in future transactions

with the seller who cheated you, but comes at a potential cost to the one who gives it. One runs the risk of retaliation with negative comments, and could in fact be missing out on positive feedback one could have received in case of a positive review. The final assumption stated that a seller's behavior is stable over time, or at least an accurate assessment of a seller's honesty. As we argue later, the argument given by Lee et al. (2000) that the volume of accumulated information converges to the true level of trustworthiness has a double edge. For a seller, the incentive to earn a positive rating decreases as his reputation increases because one negative rating hardly affects a positive score of 1,000. Moreover, a seller might just be waiting and increasing his positive score in interactions of smaller value, only to deceive when stakes are high (in several fraud cases this seems to have been the strategy that was used). In sum, one might expect that a seller's behavior will not deviate for long periods of time, but that especially high reputation sellers have incentives to defect when the occasion presents itself.

Although the argument supporting an effect of reputation on sales is straight-forward and compelling ("higher reputation goes with more and better sales"), there are clear objections that raise doubts as to whether the conditions under which auctions are run actually support a useful implementation of this argument.

Will the real Reputation please rise?

There are many ways in which one could incorporate the reputation score, and researchers have used different ways to measure what they feel is the most adequate representation of the seller's 'real reputation' (whatever that may be). Some have used the total number of (unique) positive feedback, others count only the (unique) negative feedback, and some use ratios such as the percentage of all feedback that is positive. Given that there is a wide range of positive reputation ratings—on eBay.com there are sellers with reputation scores well over 1,000—it may be a good idea to transform the scores. Based on the plausible idea that the added value of an extra 'reputation point' decreases as reputation increases, many researchers have considered not the reputation score, but its logarithm (Ba/Pavlou 2002; Diekmann/Wyder 2002; Resnick/Zeckhauser 2002; Houser/Wooders 2000). Including the neutral comments increases the number of ways to operationalize reputation even more.

Moreover, on eBay and other large auction sites the feedback profile does not only provide an overview of all feedback ratings for a seller. It also provides an overview of feedback ratings separately for the last week, the last month, and the past six months. This allows users to subjectively value ratings differently depending on the time it has been since the ratings have been given. There are at least two arguments why information on the age of feedback ratings could influence the weight of each singular rating. First of all, because people's behavior might change over time, more recent past behavior could be a better predictor of near future behavior than less recent past behavior. This would imply that sellers with high positive feedback scores would have a smaller benefit from their oldest feedback ratings than from their more recent feedback ratings (which ac-

tually goes against the argument that the 501st positive rating adds less than the 11th). Another implication of that idea would be that larger time gaps between feedback ratings have a negative effect on the influence of feedback ratings itself. Related to that, one can imagine buyers are able to discern patterns in a seller's behavior. For example, consider a seller with a feedback score of 87 consisting of 90 positives and 3 negatives, which is not too bad a score. Then if the negatives were collected in the three most recent transactions, this seems a reasonable indicator of 'a seller gone sour' and could influence buyer behavior. Upfront we do not have very compelling thoughts as to which of the above is the best measurement, so we will treat this as an empirical question and get back to this issue when analyzing the data. We do note, however, that researchers often pick one without much argumentation, which may have given rise to the inconsistent results across researchers.

4. Explaining Trust through Buyer and Product Characteristics

Basically, the elements of a sale on an auction site are the following: a buyer buys a product from a seller on an auction site (through a product description). Logically, one can expect differences with regard to trust that relate to precisely these different emphasized elements. Some characteristics of buyers may affect probability of sale, some items are perhaps more easily sold, some sellers are more likely to sell, and some auction sites might generate higher prices than others. In this section we consider these separate elements and review and elaborate on the arguments that may affect sales in online auctions.

Buyer Characteristics

Not all buyers have the same need for information to estimate the probability of a successful sale. Although buyer characteristics are hardly mentioned in the literature, whether trust is placed does not solely depend on the trustworthiness of the seller, but also on the trustfulness (the degree to which a person is willing to trust) of the buyer (Snijders 1996). When a buyer bids at an auction site for the first time, she has no experience, for instance, regarding the effectiveness of the site's reputation mechanism to keep fraudulent sellers out. As a buyer successfully engages in more auctions, past experiences reduce the uncertainty about the probability of a positive outcome. Therefore, a more experienced buyer (with positive experiences) will have a smaller need to reduce uncertainty through information from reputation and product description than a buyer who has no experiences. We therefore hypothesize that more experienced buyers are willing to pay higher prices, and that effects of reputation and product description on sales are smaller for experienced buyers (with positive experiences).

Product Characteristics

Part of the risk a buyer runs in online auctions is caused by the characteristics of the product. As the buyer's trust problem increases, she will demand more safeguards from a seller (cf. Snijders/Buskens 2001). In other words, a buyer's need for certainty increases as the risk of the transaction increases. Assuming that reputation indeed reduces uncertainty about a seller's trustworthiness, then a buyer's demand for reputation will increase as the level of risk involved in the purchase increases. We consider three characteristics of products that influence the risk involved for the buyer: the value of a product, it's condition, and it's complexity.

Product value is the product characteristic that is most often controlled for in previous research. For instance, in Eaton (2002) one of the main reasons to choose Paul Reed Smith Guitars as the object of study was the relatively high price (typically over 1,000 US Dollar), combined with the fact that this is a good with a quality that is difficult to assess. The monetary risk a buyer runs is higher for high value products than for low value products. We therefore hypothesize that the effect of reputation on sales increases as product value increases. A second product characteristic that influences a buyer's risk is whether the product is used or new. Ba and Pavlou (2002) point out that for used or even for refurbished products, "[...] the quality variance might increase significantly, which means the risk level for the buyer also increases". Besides the objective variance in quality, subjective interpretation of the description of the quality matters as well. For example, the interpretation of a description 'in reasonable condition' is ambiguous, whereas the description 'item is unused' leaves no room for discussion. Given that purchasing used products is more uncertain than purchasing new products, one would expect reputation to have larger effects for used and refurbished products than for new products. Third, the complexity of the product is also likely to influence a buyer's risk. Complex products leave more opportunities for defects that are hidden in the description, or can at least be more easily denied afterwards ("I did not know that option 23 in menu 4 did not work"). The uncertainty about the value of the good, and thereby the value of R_b, decreases as the complexity of the product increases. We therefore hypothesize that for complex products, the effect of reputation on sales should be larger.

5. The Combination of Seller and Product: Product Descriptions

The most basic information that is displayed in (almost) all product descriptions in auctions is a product's brand name and type. In fact, this is one way for the seller to convey quality and functionality of the product. In a similar context, Resnick and Zeckhauser (2002) refer more generally to standardization as one way of signaling information. Besides brand and type, one can try to reduce the inherent trust problem by including pictures. As Eaton puts it "[t]o reduce uncertainty concerning the condition of a product a seller can post pictures

of an item for sale" (Eaton 2002, 8). It makes sense to distinguish between two types of pictures: pictures taken from some catalogue, and home made pictures. Sometimes sellers add pictures to their description that are taken from the website of the original manufacturer (e.g., a picture of a Nokia phone, as it appears on Nokia's website). Such pictures can give a good idea about what the product looks like in general and might be more useful than just referring to the product type, especially when the number of different products with a similar type number is large (Nokia 5200, 3100, 3300). Such a picture does not give any idea about the product's actual condition, of course, which is why it can make sense to include home made pictures (i.e., pictures made by the seller) as well. The pictures can likewise signal information about the seller when the picture reveals something unrelated to the product itself, but possibly related to characteristics of the seller. Adding home made pictures showing a phone in pristine state but in a really messy room might do the seller more harm than good. Although they may prove to be very important, differences between kinds of home made pictures are (empirically) beyond the scope of this research. However, we do consider the difference between home made and catalogue pictures. In accordance with our earlier argumentation we hypothesize that adding a picture improves sales, and adding a home made picture helps more than adding a catalogue picture (given that the picture is of reasonable quality). In addition, and in line with Ba and Pavlou (2002) and the argument that a buyer's risk is higher for used products, we expect that the beneficial effect of pictures is larger for used products than for new products.

In the previous paragraph we already pointed out that for certain information derived from the product description it is unclear whether the information concerns the product or the seller. For example, some sellers mention that the original packaging is included with the sale. In principle, the packaging has little or no value, but one could argue that a person who keeps a product's original packaging is a person taking more than average care in handling his products. The interpretation of such information is anything but straightforward and largely beyond the scope of this paper. An interesting example is a seller who indicates that a product has been slightly scratched, thereby giving the impression that he is honest by telling that the product is not in mint condition, which implicitly suggests that, besides the scratch that is mentioned, nothing is wrong with the product (or otherwise the seller would have told you). Obviously, at the heart of this problem lies the fact that the product description is not provided by some outside objective source but by the seller of the good. This is in itself already problematic for the buyer, since the seller is the person with whom the buyer has the trust problem to begin with. In principle, this implies that all hypothesized effects of characteristics of the product description are conditional on the buyer trusting the description. For this reason one could expect that the effect of characteristics of the product description are stronger for sellers with higher reputations. Unfortunately, it is just as easy to argue the opposite. For if a seller has a low reputation score, reputation cannot be used to reduce the trust problem, and only product information can help to reduce it. If this argument is correct, characteristics of the product description only

have an effect when reputation scores are low, since these are the only cases in which there still is a trust problem to solve. Since we see no particular reason to favor one or the other reasoning, we leave the effect of product description as an empirical issue.

6. Contextual Factors: Type of Auction and Site Characteristics

We now discuss some arguments rarely mentioned or measured in the literature that could partially explain mixed results found in previous research for the effects of reputation. The general argument is that a number of characteristics might have had an effect on sales, but have previously not always been controlled for. The first part of this section concentrates on auction characteristics such as whether or not the seller uses a 'reserve price' (i.e. a minimum selling price), and how long the auction period lasts. The second part focuses on site characteristics such as the ratio of demand and supply for a good and a site's reputation.

Auction-type Characteristics

Besides reputation and product information some researchers claim that other auction characteristics also influence sales, or at least correlate with the probability of sale and the end price. Since most of these characteristics are generally not controlled for, they might also form an explanation for the different findings with regard to the effects of reputation as found in previous research. This subsection discusses five auction characteristics that could affect sales.

The first auction characteristic we consider are the available means of payment for a buyer. Eaton suggests that "[...] escrow services may actually send a negative signal to prospective buyers [especially if] a seller has negative feedback" (Eaton 2002, 18). Diekmann and Wyder report that the higher the reputation of a seller, the more likely that he will demand payment in advance (Diekmann/Wyder 2002, 20). Since payment in advance is more risky to a buyer, this could have a negative impact on sales and thus obscure an effect of reputation if it is not controlled for. The second and third characteristic are the bid increment and the reserve price. Strictly speaking the size of the bid increment should not matter: everybody bids up to his or her maximum, and the only difference with larger bid increments would be that bidding stops a bit sooner, which could lead to a somewhat lower price. Nevertheless Diekmann and Wyder find positive effects for bid increment on end price (and no effects on probability of sale). Perhaps there are other issues at stake, such as the subjective idea that when bid increments are larger, you are more likely to think that the good is valuable. Similar arguments apply to the minimum bid and to (sometimes secret) reserve prices. Lower reserve prices attract more potential buyers, thereby increasing the number of bidders. A fourth and perhaps a somewhat underrated characteristic that could explain differences in reputation effects are the shipment costs. In the (working) papers published up to April 2004, the size of the effects of reputation on end price are no larger than a few percent of the product's value

(Melnik/Alm 2002; Houser/Wooders 2000; Eaton 2002; Dewan/Hsu 2001). In our study, shipment costs ranged between roughly eight and twenty US Dollars. Estimating the mean price of all completed auctions in the research on 200 US dollars, the difference between the highest and lowest shipment costs are about six percent of the mean auction price. This implies that, given that the effects of reputation from previous research are rather small in size, forgetting to look for the shipment costs might prove a bigger mistake than forgetting to consider the seller's reputation.

Finally, the auction duration could have an effect on the probability of sale or on an auction's end price. The longer an auction lasts, the more time people have to notice the product and be interested, which increases the probability of sale and the end price. On the other hand, a longer auction likewise implies there is a larger probability that potential bidders find an attractive other item on offer, so that the net effect is not clear. There is some evidence though that most of the bidding occurs close to the end of the auction. An extreme example of this is the bidder strategy 'sniping' (Roth/Ockenfels 2000). The strategy comes down to trying to squeeze in your bid as close to the auction deadline as possible to avoid other bidders from overbidding you and thus avoiding a price war amongst bidders. Roth and Ockenfels (2000) show that this strategy is used more often on eBay.com than on Amazon.com since Amazon uses a system with an open ending rule for the auction duration. (The auction does not close until a certain amount of time after the highest bid has been placed.) Since most previous research has been conducted on eBay.com one might argue that this strategy does not affect the mixed results found for reputation effects. However, Roth and Ockenfels (2000) report that even within eBay.com there is a large difference between categories of products with respect to the prevalence of sniping.

Site Characteristics

Apart from auction characteristics, site characteristics can also influence the effects of reputation on sales. Each auction site consists of a number of sections, for instance cars, real estate, electronics and art. Judging from the differences in the total number of auctions, not all sections enjoy the same amount of activity. Differences in the number of sellers and buyers within a certain auction section interfere with effects of reputation. First, demand for and supply of products is likely to influence sales, for obvious standard economic reasons. Some interference can also be expected from differences in mean scores of reputation across sites or across sections within sites. For instance, on eBay.com a reputation score of +50 is not bad, but a +50 on ePier.com is outstanding. These differences are all hard to control for. Measuring demand and supply for, say, a mobile phone is hard to do, because it is not clear what the reference category is (all identical phones? all similar phones? all phones?).

Reputation mechanisms differ across sites, and even within sites over time. Some mechanisms only monitor negative reputation, while others distinguish positive, neutral and negative comments. A more subtle difference can be found between the reputation mechanisms of ePier.com and eBay.com. While ePier.com generates a reputation score from the total number of ratings, eBay.com

only takes unique ratings into account (so that 10 positive feedback comments, all from person A to person B only count as +1 for B). Such deviations in reputation mechanisms and thus in the way reputation is calculated, may cause differences in reputation effects. Even if two sites have the same reputation mechanism, effects of reputation could differ per site because of other structural differences between sites. This becomes clear when one looks at the way in which auction sites advertise themselves in the 'about us'-section (examples below are taken from the sites in 2003).

For example, eBay.com clearly markets their large number of members and auctions, trying to convey that eBay.com is where the action is.[1] The Dutch QXL site, Ricardo.nl, does not display any figures about the number of auctions, but refers to two tests, one by a Dutch computer magazine, the other by an 'independent research bureau' using a 'research instrument developed by the University of Amsterdam'. Both tests call Ricardo.nl the best Dutch auction site, although it is not clear what that means. A somewhat different approach is chosen by ePier.com. Apart from an 'about us'-section, they have a special section called 'eBay vs ePier'. This section shows a price comparison for insertion and final value fees between eBay.com and ePier.com. EPier.com does not charge insertion fees and charges lower final value fees than eBay.com for all item values (at least at the time of data collection). EPier.com is trying to position itself as the friendly alternative to eBay.com, perhaps hoping that users with anti-eBay sentiments organize themselves in ePier.com and make it flourish. In any case, in principle this kind of marketing could lead to users identifying themselves with the auction community. A strong belief in the honesty of other members of an auction site and a strong moral imperative to treat other community members honestly could then make a reputation system superfluous. Table 1 provides an overview of our hypotheses.

7. Data and Replications

We start by trying to replicate the results of four earlier conducted researches, albeit based on newly collected data. Our choice of papers to a large extent reflects the papers that were available to us at the start of our research; meanwhile several other (working) papers have been published (see Dellarocas 2003 for a recent overview). We chose to try to replicate Eaton (2002), Lee et al. (2000), Ba and Pavlou (2002), and Diekmann and Wyder (2002). Eaton (2002) analyzes the effect of reputation and product description on sales of Paul Reed Smith guitars. Lee et al. (2000) focus on the effects of negative feedback on the difference between retail and auction price for both new and used monitors

---

[1] Interestingly, it is not that obvious that a larger site is necessarily a good thing for a potential user. A potential buyer knows there is a large probability that what he wants to buy is actually for sale, but also knows there is a lot of competition from other buyers. A potential seller knows there are lots of potential buyers, but also that there are a lot of rival sellers. Perhaps the reason for eBay.com to advertise with size of the site is that it wants to claim that the auction system works there, and on a grand scale.

Table 1	Overview of hypotheses	
	Probability of sale	End Price
Reputation of seller	+	+
Buyer is experienced		+
Interaction with reputation:		
Value of the good	+	+
Good is used	+	+
Complexity of the good	+	+
Picture included[a]	+	+
Picture is home made[a]	+	+
Description is included[a]	+	+

[a] Possibly, interactions with reputation exist. An argument can be made that these effects only surface when reputation is high (because this is the only case when a picture can be trusted) or only when it is low (because this is the only case in which one needs a picture to compensate for a lack of reputation).

Table 1: Other variables mentioned in the main text are variables about which we do not have any specific hypotheses.

and printers. Ba and Pavlou (2002) examine the effects of feedback on what they call 'price premium' for 18 different products, such as Pokemon Gold cards, camcorders, music and software, and include (the log of) positive and negative feedback to study the effect of seller reputation on price premium. Diekmann and Wyder (2002) study the effects on sales of reputation and different auction characteristics, such as auction duration and bid increment. They use auctions of the Nokia 8310 from the Swiss QXL-site Ricardo.ch.

Our data has been gathered (by hand) from November 2002 until February 2003 from four different auction sites. Two auction sites are Dutch: eBay.nl and Ricardo.nl. The other two auction sites are US-based: eBay.com and ePier.com. The number of cases varies between the four auction sites: 111 from eBay.nl, 125 from Ricardo.nl, 103 from eBay.com and 79 from ePier.com, adding up to a total of 418 cases from 266 unique sellers. We chose to consider auctions of different types of Nokia and Motorola mobile phones. Mobile phones are popular on auction sites and although standardized products, they have the potential to induce trust problems, given their vulnerability and complexity. Data were collected on both new and used phones, but only data on phones in working condition have been gathered. Also both simlocked and simlock free phones are in the data; sometimes accessories were included in the sale.

Besides 'standard' variables concerning reputation, auction duration, number of bids, and number of pictures, we collected other possible intervening variables

Table 2   Overview of the variables collected in this and other studies.

SELLER Variable	Author	PRODUCT DESCRIPTION Variable	Author	AUCTION Variable	Author
Reputation (#)	[E][*]	Picture (#)	[E][*]	Price Premium	[L][D][*]
Negative Reputation (d)	[E][*]	Picture (d)	[L][E][*]	Auction Duration	[D][*]
Negative Reputation (#)	[L][*]	Home made Picture(d,#)	[*]	Shipment Cost	[D][*]
Membership: Duration	[*]	Warranty (length)	[*]	Bid Increment	[D][*]
Comments (#)	[*]	Warranty (d)	[*]	Price	[E][L][P][*]
Comments past 7 days	[*]	Color (d)	[*]	Net Price	[D]
Comments past month	[*]	Function Desc. (d)	[*]	Bids (#)	[D][*]
Comments past 6 months	[*]	Reason of Sale (d)	[*]	Bidders (#)	[*]
Qualitative Reputation	[*]	Condition (d)	[*]	Buy now ($)	[*]
Removed from auction	[*]	Original Box (d)	[*]	Buy now purchase (d)	[*]
Auctions Deleted	[*]	Manual (d)	[*]		
Stars (#,color)	[*]	Url of product (d)	[*]		
Online store (d)	[*]	Phone nr./email of seller	[*]		
New Seller-tag	[*]				
ID-tag	[*]				
Me-tag	[*]				

BUYER Variable	Author	CONDITION Variable	Author	PAYMENT Variable	Author
Membership; Duration	[*]	Refurbished (d)	[L]	Pay in advance	[D]
Reputation	[*]	Used (d)	[*][L]	Cash & Pickup	[D]
Feedback (#)	[*]	Condition	[*]	Cash & Delivery	[D]
				Cash	[*]
				Wire & Delivery	[*] [D]
				Pay with CC	[*][E] [D]
				Escrow Service	[E]

ACCESSORIES Variable	Author
Charge Device	[*]
Covers	[*]
Headset	[*]
Beltclip	[*]
Accessories (d)	[*]

Capitals between brackets indicated whether a variable was included in the focal study, with [E]=Eaton(2002), [L]=Lee,Im,Lee(2001),[P]=Ba & Pavlou(2002),[D]=Diekmann & Wyder(2002), and [*]=Snijders & Zijdeman.
(d)=dummy variable,(#)=variable represent number of ... ($)=price in US Dollar or Euro.

such as miscellaneous seller properties, characteristics of the product description and the means of payment. Table 2 gives an overview of the variables used in the four papers we try to replicate, and of our own data.

Seller reputation (when defined as number of unique positive feedback comments minus number of unique negative feedback comments) in our data is as follows: -1(3 sellers), zero (88 sellers), +1 (19 sellers), +2 (20 sellers), and a maximum reputation score of +1032 (1 seller). The mean of the seller reputation is 58. Reputation scores differ between auctions sites, as can be seen in Table 3. EBay.com is by far the site with the largest reputation scores and ePier.com by far the site with the lowest.

Table 3	Overview of reputation scores (number of unique positive minus number of unique negatives) across auction sites.			
Site	mean	median	75-percentile	maximum
Ricardo.nl	15.6	1	10	519
eBay.nl	42.0	3	27	867
eBay.com	131.6	40	145	1032
ePier.com	2.3	0	0	87

A picture is included in 80% of the auctions; a home made picture is included in 25% of the auctions. The buyer's are generally a bit less experienced than the sellers: 2 buyers with a reputation of -2, three with -1, 92 with zero, and a mean of about +13. Our auctions end in a sale in 51% of the cases, with prices ranging from about 10-25 Euro for the oldest and crappiest phones up to 500-900 Euro for the most expensive models.

First a few general remarks on the process of replication. We have tried to copy the research methods of each paper as much as possible. To keep matters tractable, we tried to replicate at most two analyses from each paper, the ones that we thought were at the heart of the issue in the particular paper. Space constraints do not allow us to discuss the separate replications in detail, but a detailed account of the way in which the replications were run can be found in Zijdeman (2003). Here we summarize our replication results. The first thing to note is that the results from our replication vary considerably from the original results, and it seems unlikely to us that all differences can be attributed to differences in the way in which we performed the replications. A summary of the findings of the replication of previous results is shown in Table 4.

These meager results show that the results on the effect of reputation are not only different across papers, but also across replications. Apparently, there is still much to be learned here. Some remarks on the most conspicuous replication results are in order. First, in some analyses on the probability of sale, seller reputation has a negative sign. In the replication of Ba and Pavlou this negative

Table 4	Overview of the number of successfully replicated effects, compared to the total number of effects aimed to replicate (hence, 2:4 means "2 out of 4 replicated").				
	Probability of Sale			End price	
Observations from:	Eaton	Diekmann et al.	Eaton	Lee et al.	Ba et al.
Ebay.com only	2:4	1:4[a]	2:4	2:5	-
All Sits	3:4	1:4	3:4	1:5	4:5

[a] Observations from both Dutch sites were used, not from eBay.com.

effect of reputation is even significant. It seems quite unlikely that the perceived trustworthiness of the seller decreases with increasing reputation. One possible explanation for this finding is that as a seller's reputation increases, his auction behavior changes in a way that has a negative influence on the probability of sale, and there is some evidence for this in the data. For instance, sellers with high reputation scores are more often 'power sellers', who steadily offer their goods on sale for relatively high prices, in the expectation that in the end someone will buy at this higher level. A related argument is that sellers with high reputation scores could be trying to receive higher prices as a bonus for their perceived trustworthiness. In any case, this is something to reckon with in the extended analyses on our data below.

A second interesting finding is that the effects of reputation on an auction's end price are so diverse. For instance, Ba and Pavlou found a significant positive effect for negative reputation (!) on end price, but we could not replicate this result. A likely reason for this positive effect is that there is a high correlation between positive and negative feedback (in our data 0.85). Positive and negative feedback can then hardly be distinguished and may actually both measure something like 'eBay experience'.

8. Analyses on the Mobile Phone Data

In the analyses of our mobile phone data we intend to test our hypotheses, as well as control for possible effects of additional variables that we collected. First we turn to our measurement of reputation.

Reputation

Resnick and Zeckhauser (2002) report three ways in which reputation has been modeled in previous research. As we mentioned before, in some papers only negative feedback is used to represent a seller's reputation. In other papers the feedback rating or the 'net score' (positive feedback minus negative feedback, most of the times based on only the unique comments) is used as an indicator for a seller's reputation. Finally, in several papers the logarithm of feedback or the logarithms of positive and negative feedback are used to operationalize seller reputation. When comparing the results of previous research the question arises to what degree these measurements of reputation correspond. Table 5 shows the correlations in our data between the measurements of reputation based on

observations from all four auction sites. The table shows that correlations between different ways of operationalizing reputation are high. When we use the observations from each auction site separately, the correlation matrices based on the observations of eBay.com and eBay.nl show similar results, and the correlation matrix based on observations from ePier.com even shows somewhat higher correlation coefficients.

	Seller reputation	# positive comments	# negative comments	Net reputation	Log positive reputation	Log negative reputation
Table 5	Correlation matrix for several indicators of reputation based on observations from all four auction sites (N=418).					
# positive comments	0.99*					
# negative comments	0.84*	0.85*				
Net reputation	0.99*	0.99*	0.83*			
Log positive reputation	0.74*	0.74*	0.54*	0.74*		
Log negative reputation	0.90*	0.90*	0.87*	0.90*	0.77*	
Negative reputation (dummy)	-0.07	-0.07*	-0.10*	-0.07*	0.23*	0.049

* indicates significance at $p = 0.05$ level.

Besides the total number of positive, neutral and negative feedback, the four auction sites also provide information on the number of feedbacks received in the last week, the last month (4 weeks on ePier.com) and the last 6 months. A comparison of these numbers could, in principle, give some insight into whether the effects of more recent feedback differ from the effects of less recent feedback. Again the strong correlations prevent such conclusions. First of all, 'feedback in the last month' is necessarily correlated with 'feedback in the last 6 months', simply because one is a subset of the other. We therefore subtracted the feedback from the last week from the feedback from the last month. A similar procedure was used for the feedback from the last month and the last half year, to get separate measures for feedback in the last week, the three weeks before that, and the five months before that. Even then the correlations remain strong, and using logarithms or comparing negative and positive feedback does not make these strong correlations disappear. The smallest correlation coefficient we found was between positive and negative feedback in the last week (r=0.375); most other correlations between different ways of measuring reputation are above 0.75.

Though previous papers have operationalized reputation in various ways, our data show strong correlations between different measurements of reputation. On the one hand this suggests that the different ways of operationalizing reputation are not likely to be the reason for the different results in different papers. On the other hand this finding is problematic because it makes it hard to disentangle

precisely what drives the effects of reputation on sales. In our analyses we chose to use the reputation score as used by eBay.com, the number of unique positive feedback items minus the number of unique negative feedback items.

Characteristics of Buyer, Product, and Seller

The only buyer characteristic in the analyses is the reputation rating of the winning buyer, as a measure of buyer experience. We hypothesized that more experienced buyers (with positive experiences) perceive a smaller risk – for instance because over time they start to believe that the auction system actually works—and are therefore willing to pay higher prices, as compared to buyers with less experience. Another argument, with opposite implications, is that with experience buyers learn to better evaluate what is a bargain and what not. This would instead suggest that 'rookie buyers' are more likely to overpay than seasoned eBay professionals.

We also argued that there may be a difference between pictures copied from a website and home made pictures. The number of catalogue pictures included in an auction ranges from one to seven, except for a single outlier with 14 pictures. The number of home made pictures included in an auction ranges from one to six, except for two outliers from a single seller with 18 home made pictures. After removing the outliers the total number of pictures correlates almost equally with the number of catalogue pictures and the number of home made pictures, both 0.65. There is a small negative correlation, -0.13, between the number of internet pictures and home made pictures. We use both variables in the analyses.

Our Table 2 showed a number of variables representing elements in the product description. For example, whether the color of the phone was given and whether a function description of the phone was given. Although we controlled for all of these variables in various analyses, few of them turned out to influence sales, only <function description included> and <original packaging included> sometimes did. Therefore, only these two have been included in the analyses as controls. We start our analyses with a base model including seller reputation, the number of unique positive feedback items minus the number of unique negative feedback items, to measure reputation. With respect to the product description we start with whether or not one or more pictures were included (<picture (dummy)>), the total number of pictures (<pictures (number)>), and whether home made pictures were included (<home made pictures (dummy)>). Buyer experience was measured with the buyer's reputation score (<buyer experience>). We also include whether a good was used (<used>) and whether the good was in a good condition (<good condition>), ranging on a five-point scale from poor ('1') to new ('5'). The complexity of goods, which we hypothesized to have an effect on the effect of reputation on sales, hardly varies in our data so we do not include it. We also control for the four different kinds of auctions (eBay.com, eBay.nl, ePier.com, and Ricardo.ch), for the different types of phones, and for the date (at the end of the auction), because prices of mobile phones decrease over time. More elaborate models include several variables related to the characteristics of the auction itself such as means of payment, kinds of bid increments, shipment costs, etc.

Explaining the Probability of Sale

There is clustering in the data, because some of the auctions are from the same sellers. We therefore correct the standard errors of our logistic regression model based on Huber's procedure for clustered data (Huber 1981, as implemented in Stata). The base model with all observations shows a significant negative effect of seller reputation but the coefficient is very small: 10 extra reputation points lead to an decrease in the probability of sale of 0.5 percentage points. We think this negative effect largely arises because of high volume sellers using a strategy offering goods against relatively high reserve prices. If we include a dummy variable measuring whether the reputation score is larger than 50, then this variable has a substantial and significant negative effect (-1.06, p=0.035), the difference is more than 20 percentage points, which indeed corroborates the idea that high reputation sellers ('power sellers') tend to be more willing to wait until a buyer with a high need for the good comes along. It does not help much if a seller offers a headset or some other accessory, as can be seen from the absence of an effect of the number of items for sale. Including a picture helps (about a 20 percentage points increase if you have one included), and there is some indication that including more pictures increases the probability of sale further (p=0.06), but whether or not pictures are home made does not seem to matter. Moreover, whether the phone was used or new also has no effect. Compared to eBay.com the probability of selling a phone on any of the other auction sites is significantly lower. We find that during our data collection period, phones sell less well as time progresses. Note that we did not include the buyer reputation in this model.

If we add the interaction of reputation with the site dummies, we find that the effects of reputation differ across auction sites, though the differences are not that large. Only on ePier.com the effect of reputation is positive (but not significantly different from zero) It is important to realize that sellers' reputation scores are higher on eBay.com than on the other sites. If we confine the analysis to the cases where reputation scores are smaller than 75, the differences between sites disappear. When we include the interactions of reputation with other variables, we find only one interaction effect of reputation that turns out to have a statistically significant effect: the positive effect of a picture increases with increasing reputation (or: the positive effect of reputation increases when a picture is present). For some other variables we have relatively large numbers of missing values (for instance for shipment costs and condition of item). Separate analyses show that neither of these show a significant relationship with probability of sale.

Explaining End Price

Next we consider the analyses on end price, using the same independent variables as in the previous analysis on probability of sale. The only exception is that we can now also include buyer reputation, as a measure for a buyer's experience in auctions. Obviously, we only consider auctions where the phone was eventually sold. As in the previous analyses, we control for the different

Table 6	Logit regression on whether a phone was sold.	
Variable	**Effect**	**Std. error**
Seller reputation	-0.0018**	0.0008
Number of items	0.0488	0.1316
Home made picture included (dummy)	0.1642	0.3038
Picture included (dummy)	0.8364**	0.3659
Number of pictures included	0.1136*	0.0611
Function description included (dummy)	-0.0115	0.3141
Original box included	-0.0093	0.1504
Product is used (dummy)	-0.0946	0.3430
Ricardo.nl	-2.4622***	0.5195
ePier.com	-3.9085***	0.6293
eBay.nl	-2.5561***	0.4676
Auction's end date	-0.0048***	0.0012
Phone_category1	1.5937**	0.7494
Phone_category2	-0.2269***	0.4222
Phone_categroy4	-0.3189	0.5660
constant	77.0800	19.1986

*: $p < .1$, **: $p < .05$, ***: $p < .001$. Wald chi^2(15) = 75.17, Prob.>chi^2 = 0.0000, standard error adjusted for clustering on sellers (N=416).

types of phones in the data. Closer inspection of the data shows that one of the assumptions of standard regression models, homoscedasticity, does not hold in our data. Loosely translated, the assumption that is violated in our data is that the regression line should fit about equally well for different values of the independent variables. For instance, we find that the model predicts much better for the cases where the seller has high reputation. We therefore estimate a model assuming multiplicative heteroscedasticity as in Harvey (1976).[2] Though this may appear to be a mere technicality, the results will show that precisely this issue, using standard regression models where this is not appropriate, is an important one.

The basic idea of this alternative estimation method is that in the regression analyses one includes not only the predictors for the mean value of the dependent variable, but also predictors for its variance. The technical details are as follows. The estimated model is $y_i = E[y_i] + e_i$ with $E[y_i] = b_0 + b_1 x_{1i} + \ldots + b_k x_{ki}$, as in the usual regression model, but in addition we model $logVar[e_i] = a * E[y_i](g_0 + g_1 z_{1i} + \ldots + z_{mi})$. The e_i are assumed to be standard normal. The b's, g's, and a are to be estimated. Roughly stated, the model is appropriate in cases

---

[2] We thank Jeroen Weesie, who implemented the estimation of this model in Stata.

where it is reasonable to assume that the heteroscedasticity varies smoothly with predictors g.

In our case the heteroscedasticity problem arises at least to some extent because the model fits better for less expensive phones. This is not so strange: it is easier to predict selling prices for phones of 100 dollars than it is to predict prices of phones of 500 dollars, simply because there is more room for error around 500 (taking the log of end price as the dependent variable does not solve this issue). In our base model, we also include the reputation score in the prediction of the variance in the residuals (the variable z_1 then equals the seller's reputation score) and whether a product was used (the variable z_2 equals whether or not a product was used). The disadvantage is that we are now no longer able to control for clustering at the seller level, but given that the differences between the results with and without this option in previous analyses seem rather small, we will take that for granted. Table 7 shows the results of our analyses.

Table 7	Regression of end price, assuming multiplicative heteroscedasticity.		
Variable		Effect	Std. error
Mean	Seller reputation	0.0662*	0.03458
	Number of items	-10.6199**	4.6258
	Home made picture included (dummy)	-3.7688	9.4568
	Picture included (dummy)	26.1201**	12.6432
	Log of Number of pictures included	4.6822	10.6617
	Buyer reputation	-0.1143	0.0977
	Function description included (dummy)	30.3045***	10.2619
	Original box included	-4.5530	7.6250
	Product is used (dummy)	-59.1183***	14.6938
	Ricardo.nl	-50.1290***	12.5275
	ePier.com	27.0364	18.4878
	eBay.nl	-21.5475	14.3296
	Auction's end date	0.2461***	0.0515
	Phone_category1	-51.6195***	10.37952
	Phone_category2	-8.5019	17.64102
	Phone_category4	48.1995	32.64756
	constant	-3641.094***	802.9132
Variance	Seller reputation	-0.0035***	0.0010
	Product is used (dummy)	2.3786	0.6198
	constant	2.7273	1.2824
Alpha	constant	0.0276	0.0060

*: $p < .1$, **: $p < .05$, ***: $p < .001$. Multiplicative Heteroscedastic Regression, estimator: mle/Newton-Raphson, Wald $\text{chi}^2(19) = 2444.905$, Prob.>$\text{chi}^2 = 0.0000$, $\text{corr}^2(y, yh) = 0.33$, $\text{vwcorr}^2(y, yh) = 0.57 (N = 206)$

In our base model, we now find that there *is* a positive effect of reputation

on end price. This is all the more noteworthy because without controlling for heteroscedasticity we do *not* find such an effect. On average, every 10 additional reputation points yield an extra 60 dollar cents. People attending our presentation in Bielefeld at the *Trust and community on the Internet* conference of August 2003 might remember that we did not control for heteroscedasticity in the analyses we presented there and subsequently concluded—wrongly, as it turns out—that no effect of reputation exists. In fact, in none of the papers we considered in which no effect of reputation was found, heteroscedasticity is given any serious attention. For Diekmann and Wyder (2002) the issue is not likely to play a role, since they consider only one specific (new) phone and therefore a relatively homogenous set of end prices, and in fact they do find an effect of reputation (comparable in size). Reputation also plays a role in the variance component. It turns out that the variance in the residuals decreases with increasing reputation scores (controlling for the other independent variables, that is). This is again an indication that auctions from 'power sellers' are different from auctions of others. The image that pops up on the basis of our results is that the power sellers do not mind putting their goods up for auction as often as is necessary to get their wanted price, and that eventually their goods are indeed sold for around these prices. If we include the square of reputation in the predictors for the residuals, we get a significant quadratic effect of reputation with a maximum around reputation score 290. Besides the effect of reputation on the residuals, we also find the expected positive effect of the mean of end price (a=0.028, z=4.63): variance in the residuals increases with price.

The other independent variables show the following results. Including a picture increases the value with on average 26 dollars, but adding more pictures does not help, and whether the pictures include home made ones is also irrelevant. Adding a function description increases the value with 30 dollars (p=0.003), and when a phone is not new this decreases the end price with on average 59 dollars. Prices on Ricardo.nl are lowest (an average difference of 50 dollars), followed by eBay.nl (21 dollar lower end prices). eBay.com and ePier.com have the highest end prices, with a statistically not significant edge for ePier.com of 27 dollars.

When we add interactions of reputation with site dummies, we find that the positive effect of reputation is largely due to the eBay.com cases. For ePier.com and eBay.nl the net effect of reputation is even slightly negative. This is another reason to be cautious about generalizing the usual eBay.com results. Adding other interactions of reputation shows small effects: the effect of reputation slightly decreases when pictures are used, and the effect of reputation slightly increases when used goods are considered. Several variables hardly ever approached significance or had too many missing values to run meaningful analyses and are not included in the analyses we present here: the interaction of seller reputation with buyer reputation, the means of payment, the size of the bid increment, reserve price, and shipment costs.

When we compare these results with our hypotheses, we see that the following hypothesized effects are supported by the data: the probability of sale increases with the inclusion of a picture and more pictures is better, and the probability of sale varies across auction sites. The end price is higher when the seller has

a higher reputation, when a picture is included, when a function description is included, and when the item is new. Hardly any of the hypothesized interaction effects with reputation emerged.

9. Summary and Conclusion

The general aim of this paper was to gain some insight into the mixed results found in previous research on the effects of reputation in online auctions. First we summarized and extended the arguments in previous papers on reputation effects, using the Trust Game as a metaphor for online interaction. We then tried to replicate four recent papers using data gathered from four different auction sites and also analyzed this data somewhat more extensively using the additional variables we collected.

The results of the replications are rather different from the original results, not only with respect to the effects of reputation. The only replication that seems to work out well is that of the analyses by Eaton (2002), although also there not all results match. We considered several reasons for these differences, but none of these seem completely satisfactory. The first possible reason for differences between the papers was that there are considerable differences as to which variables are included in the analyses. This is why we took the trouble to collect (almost) all variables as used in these papers, plus a large set of potentially useful others. Moreover, the papers differed in the way in which reputation was operationalized. Our analyses show that neither of these possible reasons for differences between papers can explain away the differences completely. Especially the different reputation measures are so highly correlated that it does not seem to matter much which measurement of reputation one takes. Another potential reason for differences is that the products under study varied across papers. In particular whether goods are used versus new, or standard versus complex, might be of influence. For this we also do not find much corroboration in our data. However, we do find strong evidence suggesting that in previous papers some problems may have gone unnoticed with respect to heteroscedasticity in the data (when considering the end price). When analyzed with ordinary least squares techniques, which implies ignoring the heteroscedasticity, our data show no effect of reputation on end price, just as most of the other auction papers do. When we do control for heteroscedasticity, we find a positive and significant, albeit small, effect of reputation. To replicate this result we collected an additional set of auction data, consisting of 1152 eBay.com auctions of five kinds of Ipod mp3-players (only end price, reputation, and a few product characteristics were collected). In these data we indeed found similar results: end price significantly increases with reputation, but this is found only when one controls for heteroscedasticity in the data; otherwise one finds no effect.

Hence, with respect to our original hypotheses we can say that there is some evidence in our data that reputation helps sales, in the sense that it may get you a higher price. A high reputation does not help to get the goods sold. For this latter finding there is in fact another explanation, namely that high reputation

sellers tend to sell only for higher prices (through establishing a high minimum price). Some additional corroboration for this explanation comes from the fact that the variance in end price is smaller for sellers with a high reputation. We also found some variables that have thus far not often been included in the analyses. Adding a picture to the product description helps: goods are sold more often and sold goods sell for higher prices when a picture is included. We find no evidence of extra benefits from the picture being home made. Moreover, there is evidence of higher end prices when a function desription of the product is included. No differences between experienced and inexperienced buyers are found. The experienced buyers neither buy at lower prices (there is some evidence of this in the data, but not significantly so) nor do they value the effect of reputation differently than less experienced buyers. Whereas it does not matter whether a good is used or new for the probability of sale, used goods obviously sell at lower prices. The variance in end prices is larger for used goods than for new ones.

We have only scratched the surface of what could be going on on auction sites and it certainly is a surprising finding that both model replication as well as deduction of hypotheses that can stand a comparison with the data is so difficult. Especially since predictions in auction research should be where social science excels: virtually all the information that is known to the user of an auction site is available to the researcher, making it a setup under quasi-experimental conditions. Certainly auctions should be easier to analyze than topics where the only data one can use are based on field studies in noisy circumstances, as so often is the case. Apparently, it is not so simple. We conclude with some suggestions for researchers who wish to analyze auction and reputation data. First, include other variables besides sale, end price, and reputation in your data collection. Especially whether pictures are included and whether a function description is included has an impact, but so might other variables. At this point we are still not sure enough about which issues matter most, and additional empirical evidence is necessary to get some guidance with respect to theory development. Second, be sure to consider the different ways in which reputation can be operationalized and check for large correlations between them. For instance, what you think is an effect of negative reputation could be an effect of positive reputation (given that these tend to correlate strongly). Third, think carefully about whether you want to analyze auctions of different goods simultaneously. There is evidence that this leads to heteroscedasticity, which in turn can produce inadequate results when analyzed with standard regression analyses. You should check for heteroscedasticity in any case, since there is evidence that other independent variables, such as the reputation score of the seller, correlate with the fit of the model. Fourth, think about analyzing the data, or perhaps even designing theory, separately for the occasional user and the professional user of an auction site. There is strong evidence that both groups show different behavior. Finally, some extra attention should be paid to the differences with respect to the auction sites themselves. Although we do not have enough data to make strong claims here, there are differences that are worthy of further investigation. Probability of sale and end prices differ between

the auction sites we considered (eBay.com, eBay.nl, Ricardo.nl, and ePier.com) and it is not clear whether these differences would disappear if one would control for supply and demand characteristics. Perhaps more important is that the site on which the effect of reputation seemed to work the strongest was eBay.com. The emphasis of researchers on eBay.com is logical, since it is by far the largest auction site, but given our results some caution as to whether these results carry over to other sites is in order.

Bibliography

Akerlof, G. A. (1970), The Market for 'Lemons': Quality Uncertainty and the Market Mechanism, in: *Quarterly Journal of Economics 84*, 488–500

Ba, S./P. A. Pavlou (2002), Evidence of the Effect of Trust Building Technology in Electronic Markets: Price Premiums and Buyer Behaviour, in: *MIS Quarterly 26*, 43–268

Buskens, V. (1999), *Social Networks and Trust*, Amsterdam

Coleman, J. S. (1990), *Foundations of Social Theory*, Cambridge

Dasgupta, P. (1988), Trust as a Commodity, in: Gambetta, D. (ed.), *Trust. Making and Breaking Cooperative Relations*, Oxford, 49–72

Dellarocas, C. (2003), The Digitization of Word-of-Mouth: Promise and Challenges of Online Reputation Mechanisms (March), MIT Sloan Working Paper No. 4296-03. http://ssrn.com/abstract=393042

Dewan, S./V. Hsu (2001), Trust in Electronic Markets: Price Discovery in Generalist Versus Specialty Online Auctions, Working paper (http://databases.si.umich.edu/reputations)

Diekmann, A./D. Wyder (2002), Reputation und Vertrauen bei Internet-Auktionen, in: *Kölner Zeitschrift fur Soziologie und Sozialpsychologie 54*, 674–693

Eaton, D. H. (2002), Valuing Information: Evidence from Guitar Auctions on eBay, Working paper, Department of Economics, University of Arizona (http://databases.si.umich.edu/reputations)

Harvey, A. C. (1976), Estimating Regression Models with Multiplicative Heteroscedasticity, in: *Econometrica 44*, 461–465

Houser, D./J. Wooders (2000), Reputation in Auctions: Theory and Evidence, Working paper, Department of Economics, University of Arizona (http://databases.si.umich.edu/reputations)

Huber, P. J. (1981), *Robust Statistics*, New York

Kauffman, R. J./C. Wood (2000), Running up the Bid: Modeling Seller Opportunism in Internet Auctions. Working paper, presented on America's Conference on Information Systems (AMCIS), Long Beach

Kreps, D. M. (1990) Corporate Culture and Economic Theory, in: J. E. Alt/K. A. Shepsle (eds.), *Perspectives on Positive Political Economy*, Cambridge, 90–143

Lee, Z./I. Im/S. J. Lee (2000), The Effect of Negative Buyer Feedback on Prices in Internet Auction Markets, in: *Proceedings of the 21st International Conference on Information Systems*, Brisbane, 286–287

Melnik, M. I./J. Alm (2002), Does a Seller's eCommerce Reputation Matter?, in: *Journal of Industrial Economics 50*, 337–349

Raub, W./C. Snijders (2001), A Reluctant Match: Models for the Analysis of Trust in Durable Two Party Relations, in C. Castelfranchi/Y.-H. Tan (eds.), *Trust and Deception in Virtual Societies*, Amsterdam, 195–220

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. Volume 11 of Advances in Applied Microeconomics*, Amsterdam, 127–157

— / — /J. Swanson/K. Lockwood (2002), The Value of Reputation on eBay: A Controlled Experiment. Working paper, available online from the Reputations Research Network (http://databases.si.umich.edu/reputations)

Roth. A. E./A. Ockenfels (2000), Last Minute Bidding and the Rules for Ending Second-Price Auctions: Theory and Evidence from a Natural Experiment on the Internet, in: *American Economic Review 92*, 1093–1103

Snijders, C. (1996), *Trust and Commitments*, Thela Thesis, Amsterdam

— /V. Buskens (2001), How to Convince Someone that You can be Trusted? The Role of 'Hostages', in: *Journal of Mathematical Sociology 24*, 355–383

van Swol, L. M. (2001), EBay and the Return of the Town Square: How eBay uses Reputational Gossip to Build Trust, Working paper, Evanston

Zijdeman, R. L. (2003), *Reputation and Internet Auctions. EBay and Beyond*, Master's Thesis, Utrecht University, ISCORE paper no. 214