

Kusterer, David; Bolton, Gary; Mans, Johannes

Conference Paper

Inflated Reputations Uncertainty, Leniency & Moral Wiggle Room in Trader Feedback Systems

Beiträge zur Jahrestagung des Vereins für Socialpolitik 2016: Demographischer Wandel -
Session: Experimental Economics: Cheating, No. C05-V1

Provided in Cooperation with:

Verein für Socialpolitik / German Economic Association

Suggested Citation: Kusterer, David; Bolton, Gary; Mans, Johannes (2016) : Inflated Reputations Uncertainty, Leniency & Moral Wiggle Room in Trader Feedback Systems, Beiträge zur Jahrestagung des Vereins für Socialpolitik 2016: Demographischer Wandel - Session: Experimental Economics: Cheating, No. C05-V1, ZBW - Deutsche Zentralbibliothek für Wirtschaftswissenschaften, Leibniz-Informationszentrum Wirtschaft, Kiel und Hamburg

This Version is available at:

<https://hdl.handle.net/10419/145794>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

INFLATED REPUTATIONS

UNCERTAINTY, LENIENCY & MORAL WIGGLE ROOM IN TRADER FEEDBACK SYSTEMS

Gary E. Bolton^{*}, David J. Kusterer[†] and Johannes Mans[‡]

29 July 2016

First draft: October 2013

Feedback systems associated with Internet markets are known to be subject to strategic manipulation that can create distortions in the reputation information provided to traders. The experiment we present suggests that distortions can emerge from sources that have heretofore been overlooked: the leniency and moral wiggle room that arise when there is uncertainty about the source of transaction problems. The control the laboratory affords permits us to separate the influence exerted by uncertainty per se from that implied by behavioral leniency. We observe that uncertainty about seller responsibility leads to leniency behaviors that reduce the informativeness of the feedback system, thereby diminishing the incentives for honest seller behavior. Under uncertainty, buyers pay about the same prices but get significantly less.

Keywords: Reputation; trust; leniency bias; electronic markets; experimental economics

JEL-Codes: C9; D4

^{*} Naveen Jindal School of Management, University of Texas at Dallas, Richardson, Texas 75080, gbolton@utdallas.edu.

[†] Department of Economics and Cologne Graduate School in Management, Economics and Social Sciences, University of Cologne, Albertus-Magnus-Platz, 50923 Köln, Germany, kusterer@uni-koeln.de, corresponding author.

[‡] Department of Economics and Cologne Graduate School in Management, Economics and Social Sciences, University of Cologne, Albertus-Magnus-Platz, 50923 Köln, Germany, mans@wiso.uni-koeln.de.

Financial support of the German Research Foundation (DFG) through the Research Unit “Design & Behavior (FOR 1371) is gratefully acknowledged. Kusterer would like to thank the DAAD for travel funds. We thank Bill Neilson, Axel Ockenfels, Dirk Sliwka, Achim Wambach, Christoph Feldhaus, Nicolas Fugger and Florian Goessl for helpful comments and suggestions. Informed consent of participants was obtained from subjects in accordance with the rule of the University of Cologne.

1 INTRODUCTION

‘Brag-and-moan’ is the norm on the Internet. Reviews of trade satisfaction tend to extremes: Very positive reviews are more prevalent than very negative reviews, with each of these more frequent than moderate ones. The factors behind this compression in reputation information, and how it influences the performance of markets constitute important questions for market design in a growing part of the economy. They also speak to certain gaps in our understanding of how reputation mechanisms work in general. In this paper, we investigate whether uncertainty about seller responsibility contributes to compression in feedback scores.

Internet feedback mechanisms are essential to the existence of electronic trading platforms such as Amazon, eBay, Etsy and Taobao, markets where many transactions take place between anonymous and geographically separated traders with no common trade history to build upon (e.g., Ba 2001; Resnick and Zeckhauser 2002; Dellarocas 2003; Levin 2013). A growing number of websites collect feedback ratings that influence transactions in offline markets as well, such as Airbnb, Google reviews, tripadvisor.com, and yelp.com.

Reputation systems inform others about issues of adverse selection, such as trader experience and professionalism, and moral hazard, such as whether a trader ships goods that fit the advertised description (Bar-Isaac and Tadelis, 2008). The concept behind these mechanisms is literally ancient (Greif, 1993): Word-of-mouth provides a link between past behavior and future payoffs; rational individuals need weigh both long-term as well as the short-term consequences when deciding on a course of action (e.g., Kreps and Wilson 1982; Wilson 1985; Milgrom, North, and Weingast 1990). The difference between traditional and electronic reputation systems is a matter of speed and scope: traditional word-of-mouth spreads sequentially, from acquaintance-to-acquaintance, whereas Internet word-of-mouth, once posted, is available for the simultaneous viewing of all traders, worldwide. As such, Internet word-of-mouth likely makes the reliability of reputation information an even bigger factor in market performance than it has been in the past.

Complaints about online transactions suggest that feedback profiles draw an overly optimistic picture (Bauerly, 2009; Dellarocas and Wood, 2008; Gregg and Scott, 2006; Rice, 2012).¹ As such, the upward compression in feedback makes it hard to distinguish between highly reliable trading partners and those who are merely good or mediocre. Much of the

¹ Numerous studies find a positive correlation between a seller’s reputation score and both the probability and price of sale. See for example Ba and Pavlou (2002), Bajari and Hortacsu (2003), Houser and Wooders (2006), Jin and Kato (2006) and Resnick et al. (2006).

early research into the upward compression of feedback centered on eBay during a period in which the marketplace had a two-way feedback system where buyer and seller rate one another. Ninety-nine percent of all ratings in the eBay two-way feedback system were positive (Kauffman and Wood, 2006; Resnick and Zeckhauser, 2002). Using structural estimation, Dellarocas and Wood (2008) found evidence that the actual risk of a dissatisfying transaction on eBay was significantly larger (21% for buyers and 14% for sellers in rare coin auctions). They conclude that satisfied and very unsatisfied traders are more likely to leave feedback than those mildly dissatisfied – the typical brag-and-moan pattern (also see Dellarocas, 2001). Studies examining the probability and timing of feedback giving suggested that dissatisfied buyers were afraid of retaliatory feedback and therefore chose to give either positive feedback or no feedback at all (Bolton et al., 2013; Dellarocas and Wood, 2008; Reichling, 2004).

Due in part to these considerations, eBay moved in 2008 to a one-way system (buyer rates seller only), effectively eliminating the possibility of trader retaliation. While this movement had positive effects on the system by increasing quality and reducing adverse selection (Klein et al., 2016; Saeedi et al., 2013), the mean detailed seller rating is 4.7 on a 5-point scale and more than 75% of the sellers have an average detailed seller rating of larger than 4.5 (Klein et al., 2009). In fact, the j-shaped brag-and-moan pattern is common to many one-sided feedback systems (Hu et al., 2009, 2006; Zervas et al., 2015).

Feedback retaliation is not possible in one-way systems, so there must be other factors behind the observed compression. The frequency of feedback giving in one-way systems is often high; for example on eBay, about 70% of traders left feedback both before and after the move to a one-way system (Bolton et al., 2013). However, the decision to remain silent depends on the transaction experience (Dellarocas and Wood, 2008). In a large field study on eBay, Nosko and Tadelis (2015) show that implementing a feedback measure they call the effective percent positive (EPP), which is the number of positive ratings divided by the number of transactions, not feedbacks, improves the prediction of buyer experiences by using the information contained in the transactions without a feedback. Sorting search listings by EPP leads to more satisfying transactions for buyers, indicating that the silence is non-random and provides information about the sellers. This dissatisfaction bias among those who choose not to report certainly plays an important role for compression, but it seems unlikely to be the entire story.

Efforts a seller makes towards a successful trade can be confounded with exogenous shocks. In posting an item for sale, sellers typically announce a description of their offering (condition, shipping time, etc.). Attributional uncertainty can arise due to the incomplete nature of seller descriptions, and differences between what a buyer expected given the announcement and the quality received might be due to the seller purposely misstating the quality, or to the ambiguity inherent in the phrase ‘good condition’. Shipping times might be due to seller decisions or to circumstances beyond the seller’s control.

There are reasons to suspect that attributional uncertainty will lead to trader rating leniency, albeit the suspicions rest on indirect evidence. Extensive research in personnel economics shows that when there is uncertainty, supervising managers tend to give more favorable ratings of employees than are justified by actual performance (Bol, 2011; Bretz et al., 1992; Landy and Farr, 1980; Moers, 2005; Prendergast, 1999; Prendergast and Topel, 1993; Saal et al., 1980; Sharon and Bartlett, 1969), or they refrain from giving feedback at all (Larson, 1986).² Golman and Bhatia (2012) provide a model in which the performance evaluation of the supervisor is lenient due to altruism and exhibits a centrality bias in order to discount for the noise in the signal.³ In another context, Ganzach and Krantz (1991) show when predicting future performance – e.g. predicting final GPA based on other test scores – higher uncertainty leads to more lenient predictions. In the U.S. judicial system, the standard for criminal conviction, innocent until proven guilty, is based on a preference for wrongful leniency over wrongful conviction; or as Benjamin Franklin put it, “it is better a hundred guilty persons should escape than one innocent person should suffer” (Bigelow, 1904). Rice (2012) investigates the influence of feedback uncertainty on simple trust games and finds that, with uncertainty, it is less likely for trustees to receive a poor rating for a given level of trustworthiness.

Still, it is easy to find potentially confounding factors that differentiate Internet markets from these other social situations. For instance, employer performance ratings are typically done in the context of a repeated, face-to-face relationship, whereas Internet

² This leniency is more pronounced when significant (monetary) decisions concerning the employee – such as pay raises or promotions – are tied to these ratings (Jawahar and Williams, 1997; Taylor and Wherry, 1951).

³ The model can be re-interpreted in terms of conflict avoidance (see Waung and Highhouse, 1997), where traders want to avoid complaints by the counterparty about unjustified ratings. There are other related theoretical papers in the literature on subjective performance evaluation, but in these models performance is observed without noise. Grund and Przemec (2012) study a model with two inequality-averse agents and a supervisor who cares for the agents’ payoffs, leading to lenient ratings. Giebe and Gürtler (2012) analyze optimal contract design with an inherently lenient supervisor in a three-tiered hierarchy with a principal, a supervisor, and an agent.

feedback ratings are typically given in the context of one-off, relatively anonymous transactions. A more direct test is necessary.

One of the challenges of testing the leniency hypothesis is separating leniency from any response feedback givers might have to exogenous uncertainty per se. For example, if the exogenous shock has a downward bias, so on average buyers receive less than sellers send, then traders might adjust their feedback scoring upward simply to account for the expected short fall; the feedback language, in effect, is adjusted to the (expected) shipment value. Exploiting the control the laboratory affords, we designed an auction environment with optional feedback in which we can directly manipulate whether the received quality is a true or noisy signal of seller effort. In the Baseline treatment the level of quality the buyer receives is equal to what the seller shipped. In the Uncertainty treatment shipments are subject to a random shock such that the expected quality of what is received closely approximates the level shipped. Absent a propensity for leniency, we expect possibly noisier feedback but no average difference in feedback across treatments (our null hypothesis). If, however, there is a propensity for leniency, then average feedback under Uncertainty should be biased upward relative to what we observe in the Baseline. The experiment also admits the reverse bias: a higher level of ‘moan’ feedback. While we judged this less likely, it is a plausible possibility given earlier findings that punishment for bad outcomes is an important motivator for feedback giving (Lafky, 2014).

While there is extensive research on the functioning of reputation systems, a general theory of how feedback in user-provided reputation systems is given is still missing (for binary feedback mechanisms, see Dellarocas, 2005). With our experiment, we provide data on factors other than reciprocity that influence the rating behavior of traders on online platforms. Our results are in line with the hypothesis that under uncertainty buyers give sellers the benefit of the doubt and leave more lenient ratings for less than advertised quality. Regarding silence, we observe that buyers in general are less likely to leave feedback ratings under uncertainty when the received quality differs from the previously announced quality. Overall, the reporting biases reduce the informativeness of the feedback system and make it difficult for buyers to differentiate between honest and dishonest sellers. Incentives for trustworthy behavior are reduced. Many sellers take advantage of the fact that the distortion factor disguises their true intentions and deceive buyers to a larger extent than in the Baseline treatment. Buyers, however, bid and pay about the same prices in Baseline and Uncertainty treatments. Overall, the strategic behavior of sellers significantly decreases buyer profits. Implications for market design and reputation studies are discussed in the Conclusion section.

2 EXPERIMENTAL DESIGN

In order to investigate the effect of uncertainty on feedback ratings and on (electronic) markets in general we implement an auction market in two treatments. One treatment introduces a factor that randomly distorts seller's shipped quality, while the other does not. In both treatments, market transactions take place over a series of periods, and in each period one seller and two buyers play the stage game outlined in Table 1.

In the first stage, the seller publicly announces a quality q_a from the interval (0%, 100%). The announcement corresponds to item descriptions sellers typically post on Internet market sites, and are the basis of buyer expectations for what will be received in a transaction. In addition, the seller privately chooses a quality q_s she is going to ship from the same interval at linear cost $c(q_s) = q_s$. In the second stage, buyers learn their valuation v_i , which is privately drawn from a uniform distribution of all integers between 100 and 300. Buyers also learn the announced quality q_a , the seller's feedback score and the number of feedback ratings the seller has received so far (given that the seller already received some feedback ratings). The feedback score is the arithmetic mean of all feedback ratings received until the current period. Buyers then submit their bids, with a minimum bid of 100 ECU (Experimental Currency Unit).

In the third stage, the buyer who submitted the higher of the two bids wins the auction and learns the received quality q_r . He pays a price p amounting to the second highest bid plus 1 ECU to the seller. In case both bidders state the same bid, the buyer who entered his bid first wins and pays his bid. If only one bidder submitted a bid, he wins and pays the minimum price of 100 ECU. The payoff for the winning bidder is his valuation v_i multiplied by the received quality q_r net of the price p (not including feedback costs described below): $\pi_{wb} = v_i * (q_r/100) - p$. The losing bidder receives a payoff of $\pi_{lb} = 0$. Seller's payoff is the price p less the costs for the shipped quality: $\pi_s = p - q_s$. If no bidder submits a bid, the product is not sold and the seller and both buyers receive a payoff of 0.

In the fourth stage, the winning bidder has the opportunity to leave a feedback rating for the seller on a five-point scale from 1 to 5, where 5 is the highest rating. In case a buyer posts a feedback rating his profit is reduced by 1 ECU. In the final stage, the buyer and the seller learn their respective payoffs and the seller also gets to know the feedback rating (in case the buyer submitted one) and the updated feedback score.

Stage	Seller	Bidders
1. Announcement	makes public announcement and privately determines shipped quality	
2. Auction		get to know announcement, seller's feedback score, and own valuation submit bid in sealed second-price auction
3. Transaction		get to know auction outcome winning bidder gets to know received quality
4. Feedback		winning bidder decides whether to leave costly feedback
5. Payoffs	gets to know received feedback if submitted $\pi_s = \text{Price} - \text{Shipped quality}$	$\pi_{wb} = \text{Valuation} \times \text{Received quality} - \text{Price}$ $\pi_b = 0$

Table 1: Overview of the experimental stage game.

The two treatments differ only in the relation between shipped and received quality. In the Baseline treatment, the received quality is equal to the shipped quality. We change this in the Uncertainty treatment where in 50% of all auctions, the buyers receive the shipped quality plus a random integer drawn from a normal distribution with mean 0 and standard deviation of 10.⁴ This shock happens randomly and neither the seller nor the buyer is informed whether quality has been changed or not. Hence, buyers cannot infer the sellers' responsibility when receiving something else than promised from the received quality with certainty. The shock provides a noisy signal of what the seller sent that is close to unbiased, up to a discrepancy due to the cutoffs implied by the admissible quality domain. Specifically, the bias is highest at the cutoffs: the expected shock to a shipment quality of 0 or 100 is, respectively, +0.6 or -0.6 quality units, while the expected shock to a shipment quality of 50 is 0.⁵ There is a 75% chance that the buyer will receive at least as much as was shipped and a 92% chance that the shortfall will be no more than 10 units.

The stage game was repeated for 45 periods. At the beginning of each period, subjects were randomly assigned the role of a seller or a bidder. We ensured that each subject was in the role of a seller in exactly 15 periods. In each period one seller was matched with two buyers and with the restriction that sellers did not meet the same buyer(s) in two consecutive periods. Each session consisted of 30 participants, which were assigned to 5 matching

⁴ Instructions to subjects provide an explanation of the random draw using a histogram of draws of a normal distribution with these parameters. We also provide examples of how likely specific values are in such a distribution. See Appendix C for a translated version of the instructions used in the experiment.

⁵ The expected value of the level of quality the buyer will receive when the seller ships quality q is $0.5*q + 0.5*(1.200 + 0.976q)$ where the value in parentheses is estimated using OLS.

groups. Subjects were re-matched with other subjects from their matching group only, such that there are 5 independent observations per session. Across matching groups, bidder valuations and the matching – including the sequence of roles⁶ – were held constant. At the beginning of the experiment subjects received an endowment of 1,000 ECU. Gains and losses were added to or deducted from this initial endowment.

The structure of the experiment builds on that used by Bolton, Greiner, and Ockenfels (2013) with the key differences that only buyers have the opportunity to leave feedback and sellers make a non-binding quality announcement at the beginning of each period. The added feature controls for buyer expectations.

We ran two sessions per treatment with 120 participants in total.⁷ The four sessions were run in the Cologne Laboratory for Economic Research (CLER) at the University of Cologne in November and December 2012. Subjects were recruited from the CLER's subject pool using ORSEE (Greiner, 2004) and the computerized experiment was run using z-Tree (Fischbacher, 2007). At the end of the experiment, the account balance was converted to euros (100 ECU = 1€) and paid out in cash. On average, subjects earned 21.28€ while sessions lasted for approximately 2 hours.

2.1 Hypotheses

As explained above, the exogenous shock is designed so that, absent a bias in feedback behavior, the introduction of uncertainty should not bias expected feedback scores. Specifically, while the introduction of uncertainty might lead to more noise in feedback giving, we take as our null hypothesis that average feedback given, along with the frequency of feedback giving, by a buyer who receives q_r when the seller announced quality q_a is the same across Baseline and Uncertainty treatments. Taking q_r/q_a as the measure of seller contract fulfillment, the *fill ratio*, we write the null hypothesis as

Hypothesis 0: Changes to feedback giving from the introduction of uncertainty largely reflect the added noise: The variance for feedback for $q_r/q_a < 1$ may rise but average feedback and the frequency of feedback giving is unchanged.

⁶ That is, subject 1 in the first matching group is in the role of seller (bidder) in the same periods as subject 1 in the second matching group, and so on.

⁷ The average age of the participants was 22.8 years, with 55.8% female. With regard to area of study: 46.7% studied at the Faculty of Management, Economics and Social Sciences, 17.5% at the Faculty of Arts and Humanities, 15% at the Faculty of Human Sciences, 13.3% at the Faculty of Mathematics and Natural Sciences, 4.2% at the Faculty of Medicine, and 2.5% at the Faculty of Law. One subject was not a registered student.

In line with the literature on leniency, the alternative hypothesis anticipates that with the introduction of uncertainty, a benefit of the doubt will exist either in the form of higher ratings or lower reporting in cases where buyers have a negative experience.

Hypothesis 1: Changes to feedback giving from the introduction of uncertainty reflect leniency: Average feedback given for $q_r/q_a < 1$ will be greater than in the Baseline or there is a reduced frequency of feedback given.

Hypothesis 1 implies that, under Uncertainty, feedback ratings will become compressed at the upper end; that is sellers who are very trustworthy and sellers that are somewhat less trustworthy will have more similar ratings under Uncertainty. So leniency at the individual level lowers the informativeness of the feedback system in the sense that highly rated sellers' trustworthiness is lower and exhibits higher variance.

Hypothesis 2: Under Uncertainty, feedback informativeness will be lower due to upward compression in ratings.

How might leniency affect market outcomes? Lower informativeness gives rise to moral wiggle room both in the sense that it creates pecuniary incentives for opportunistic seller behavior. Even sellers that are less strategic minded and more honest might be tempted because, in the present case, uncertainty about attribution makes one's actions less informative about one's true nature and so less damaging to one's self-image (Bénabou and Tirole 2011, Dana, Weber, and Kuang 2007 and Ockenfels and Werner 2012; see related results in ultimatum games Mitzkewitz and Nagel 1993, Güth, Huck, and Ockenfels 1996). We would then expect that, under Uncertainty, the lower expected value of the goods received, along with the greater variance (and so greater risk), would lead buyers to bid lower and less often, resulting in lower prices and lower sales volume.

Hypothesis 3: Under Uncertainty, sellers are more likely to send lower quality than announced. Buyers bid lower and pay lower prices.

3 RESULTS

3.1 A descriptive look at the data

Figure 1 provides a first look at how announced and shipped quality unfold over time. There are three observations to make. First, average announced quality, at about 86%, is very steady across periods and similar across treatments. Second, average shipped quality is lower than announced in all periods and across both treatments. Shipped quality falls off sharply in

the last 10 periods; this endgame effect being the first sign that feedback reputation is an important motivation for seller trustworthiness.⁸ Third, average shipped quality is lower in the Uncertainty treatment. Excluding the last 10 periods, shipped quality averages about 76% in the Baseline treatment but only 67% in the Uncertainty treatment.

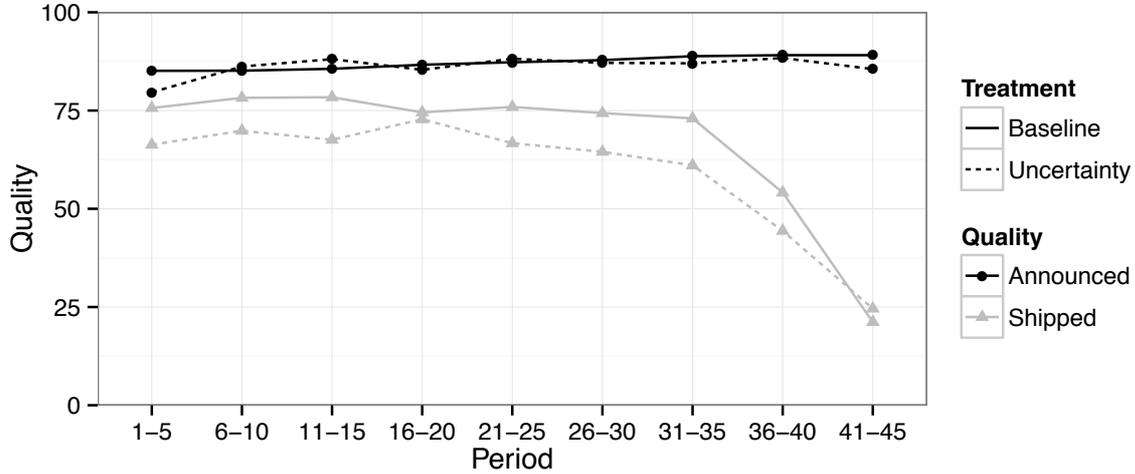


Figure 1: Average announced and shipped quality across period intervals.

We will see below that the feedback given to quality that falls short of the announcement depends on the extent of the shortfall. We classify the fill ratio into four categories, stipulated in Table 2. *Overfill* and *Fulfill* categories are self-explanatory. *Shortfill* describes shipped (received) quality that falls short of that announced by no more than 20%, while *Vshortfill* refers to falling short by more than 20%. In the Shortfill category, there can be meaningful uncertainty about seller responsibility. In the Vshortfill category, however, there is less uncertainty because deviations of more than 20% are very unlikely to be caused by the distortion factor alone. Also, because of the random distortion, buyer and seller in the Uncertainty treatment can perceive the fill ratio differently, so we classify by perspective. While results and analyses below are derived using the 20% cutoff, the main conclusions are robust to any cutoff factor in the range 10% to 40%.

Perspective: Seller	Fill ratio = shipped / announced quality
Buyer	Fill ratio = received / announced quality
Overfill	> 1
Fulfill	= 1
Shortfill	< 1 and ≥ 0.8
Vshortfill	< 0.8

Table 2: Classification of fill ratios for sellers and buyers.

⁸ Due to this endgame effect, we restrict all further analyses to the first 35 periods. Retaining the last ten rounds does no change the results qualitatively.

Figure 2 exhibits histograms of the seller fill ratio broken down by treatment.⁹ Both histograms are bimodal at fill ratios of one, where the seller fulfills the announced quality, and zero, where none of the announced quality gets shipped. Also observe the shift in the histogram when moving from the Baseline to the Uncertainty treatment. Most of the shift is accounted for by a displacement of Fulfills in favor of Shortfills. That is, in the Uncertainty treatment, Fulfills are observed less frequently, and fills that fall somewhat short more frequently than in the Baseline treatment. Also observe that some sellers ship more than they announced and that this is more frequent in Uncertainty than in Baseline. We discuss explanations for Overfills in Section 3.3.

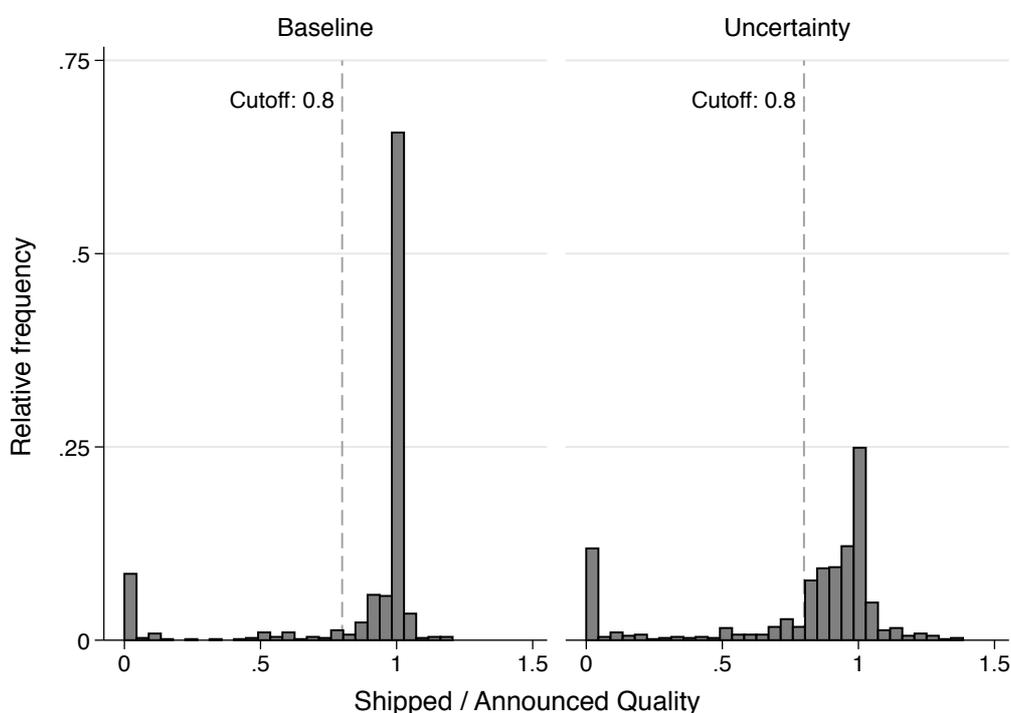


Figure 2: Histogram of seller fill ratios (= shipped / announced quality) across treatments (periods 1-35).

3.2 Feedback

One of our main hypotheses is that for instances in which the quality received is lower than the quality announced, buyer feedback ratings will be more lenient in the Uncertainty treatment than in the Baseline treatment. Figure 3 compares the average feedback rating across Baseline and Uncertainty treatments at different buyer fill ratios. Figure 3 also shows the share of silent transactions in which buyers did not leave a feedback rating. Also below, Table 3 reports the corresponding average feedbacks and rates of silence by fill category.

⁹ We exclude two auctions in which the seller announced a quality level of 0.

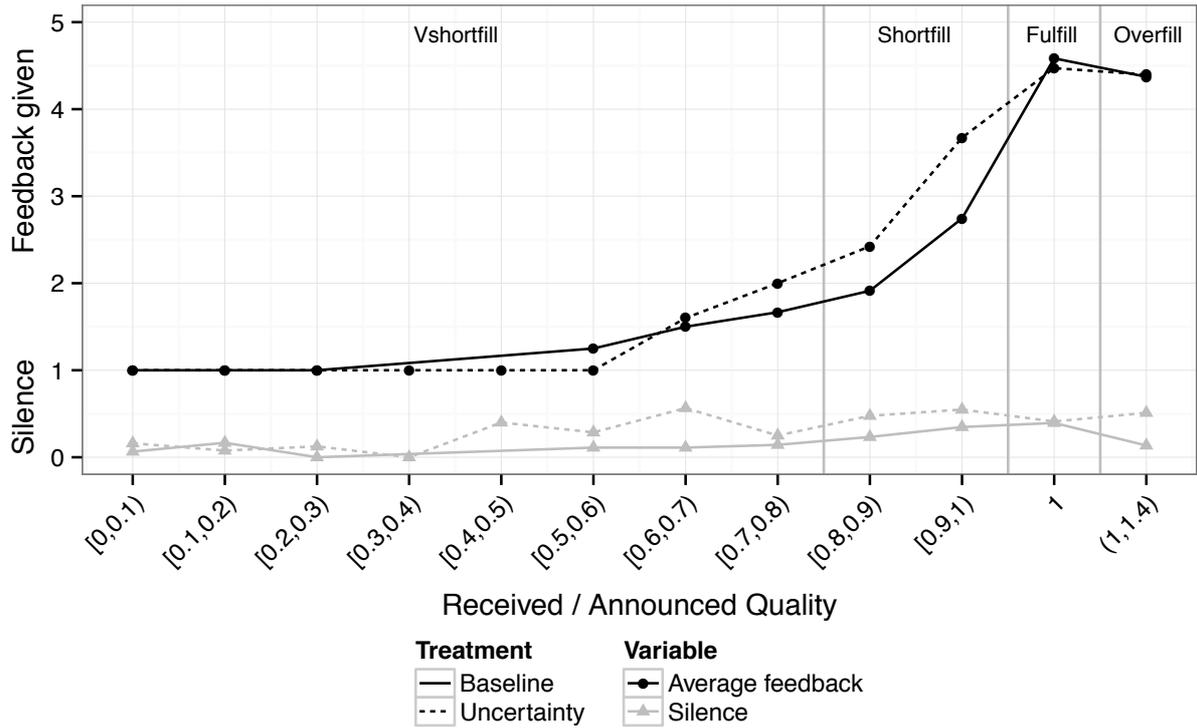


Figure 3: Average buyer feedback and frequency of not giving feedback (silence) for periods 1-35.

	Feedback			Silence		
	Baseline	Uncertainty	<i>p</i> -value	Baseline	Uncertainty	<i>p</i> -value
Overfill	4.37	4.40	0.68	0.14	0.51	0.06
Fulfill	4.58	4.47	0.73	0.39	0.41	0.76
Shortfill	2.47	3.02	0.01	0.31	0.51	0.04
Vshortfill	1.14	1.23	0.82	0.09	0.24	0.02
Overall	3.67	2.93	0.03	0.32	0.43	0.08

Table 3: Average feedback given and rate of silence in each seller classification group. The *p*-values are derived from two-tailed Mann-Whitney U tests with data aggregated on the matching group level for periods 1-35 (10 independent observations in each treatment).

Looking first at the feedback given: From Figure 3, the average feedback submitted for Overfill and Fulfill is similar between the two treatments. Buyers who receive as much as or more than promised submit average ratings of 4.4 to 4.6 in both treatments. Likewise, buyers who receive much less than announced, as represented by the Vshortfill classification, usually give the lowest possible rating in both treatments. For the Shortfill categories, however, buyers tend to leave higher ratings in the Uncertainty treatment compared to the Baseline, an average of 3.0 versus 2.5. These observations are statistically supported as reported in Table 3. The random effects Tobit estimates in Table 4 control for a number of additional factors, but nevertheless the same results hold.

Feedback rating	Overfill	Fulfill	Shortfill	Vshortfill
Uncertainty	0.078 (0.086)	-0.364 (-0.442)	1.269*** (3.673)	0.379 (0.754)
Received / Announced	4.930 (1.397)		16.131*** (8.249)	11.137*** (4.064)
Announced	0.104*** (3.762)	0.086*** (5.978)	0.007 (0.562)	0.047+ (1.912)
Price	0.010 (1.633)	0.005 (0.970)	-0.005 (-1.595)	0.027** (2.775)
Period	0.016 (0.709)	-0.005 (-0.339)	-0.014 (-1.354)	-0.103*** (-3.657)
Intercept	-9.875+ (-1.833)	-1.834 (-1.401)	-12.481*** (-6.543)	-12.789*** (-3.620)
<i>N</i>	119	303	197	186
Log likelihood	-121.9	-250.4	-292.2	-60.32

Table 4: Random effects Tobit regressions with submitted feedback ratings (1-5) as dependent variable for each category of seller fill ratio. Periods 1-35; *t* statistics in parentheses; + $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, * $p < 0.001$.**

Turning now to the frequency of feedback silence: The overall rate of feedback giving in Baseline and Uncertainty treatments is 68% and 57%, respectively. By way of comparison, the provision of one-sided (conventional) feedback on eBay is around 70% (Bolton et al., 2013). Figure 3 shows the frequency of silence in each treatment, broken down by the received-to-announced fill ratio. There is little difference in silence across treatments for the Fulfill category, where buyers receive as much as promised. In contrast, buyers who receive more (Overfill) or less than promised (Shortfill and Vshortfill) remain silent more often in the Uncertainty treatment. Hence, positive and negative surprises are reported less frequently. The statistical analysis reported in Table 3 confirms these observations. The random effects Probit models reported in Table 5 tell a similar story. The higher frequency of silence in Overfill was not predicted by our hypothesis. One explanation would be that buyers are more likely to remain silent when attribution is uncertain, not just in the case of when the quality falls short. Because positive (as well as negative) shocks can happen to quality, the attribution behind Overfill is ambiguous which may result in greater silence.

Our results are in line with Hypothesis 1. Buyers give more lenient feedback to sellers when what is received is less than announced and responsibility is unclear. In particular, if the received shortfall is less than 20% below what was announced, we observe leniency. The second part of Hypothesis 1 is also confirmed by the data. Silence is more frequent under Uncertainty in the case of Shortfills. Interestingly, this is also the case for Overfills,

suggesting that the silence hypothesis extends to circumstances where the buyer is pleasantly surprised and seller attribution is in doubt.

Silence	Overfill	Fulfill	Shortfill	Vshortfill
Uncertainty	1.356** (2.865)	0.041 (0.185)	0.675** (2.751)	0.449+ (1.834)
Announced	-0.024+ (-1.769)	-0.013* (-2.335)	-0.004 (-0.450)	0.028* (2.430)
Received / Announced	-2.665 (-1.596)		2.803+ (1.908)	1.199** (3.253)
Price	0.000 (0.007)	0.002 (1.269)	-0.005* (-1.994)	-0.016* (-2.176)
Period	0.006 (0.519)	0.009 (1.470)	-0.002 (-0.320)	-0.014 (-1.335)
Intercept	3.521 (1.462)	0.388 (0.791)	-2.142 (-1.465)	-2.090+ (-1.727)
<i>N</i>	198	503	363	230
Log likelihood	-111.8	-324.6	-232.3	-97.50

Table 5: Random effects Probit regressions for silence with 1 = no feedback given and 0 = feedback given for each category of seller fill ratio. Periods 1-35; *t* statistics in parentheses; + $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, * $p < 0.001$.**

3.3 Predictiveness of the feedback system

Perhaps the most important function of the feedback system is to distinguish honest sellers, who ship at least the level of quality they have announced, from sellers who ship less than announced. To compare how well feedback predicts honest sellers across treatments, we ran a Probit regression for each treatment, the dependent variable indicating whether the seller was honest or not,¹⁰ regressed on all the information that is available to the buyer before bidding: the seller's feedback score, the number of feedbacks the seller has received so far, the announced quality and the current period (see Table 12, Appendix A for the regressions). Based on these two regressions we calculate the predicted probabilities to encounter an honest seller along all feedback averages for the two treatments. The results appear in Figure 4.

¹⁰That is, whether she shipped at least what she announced.

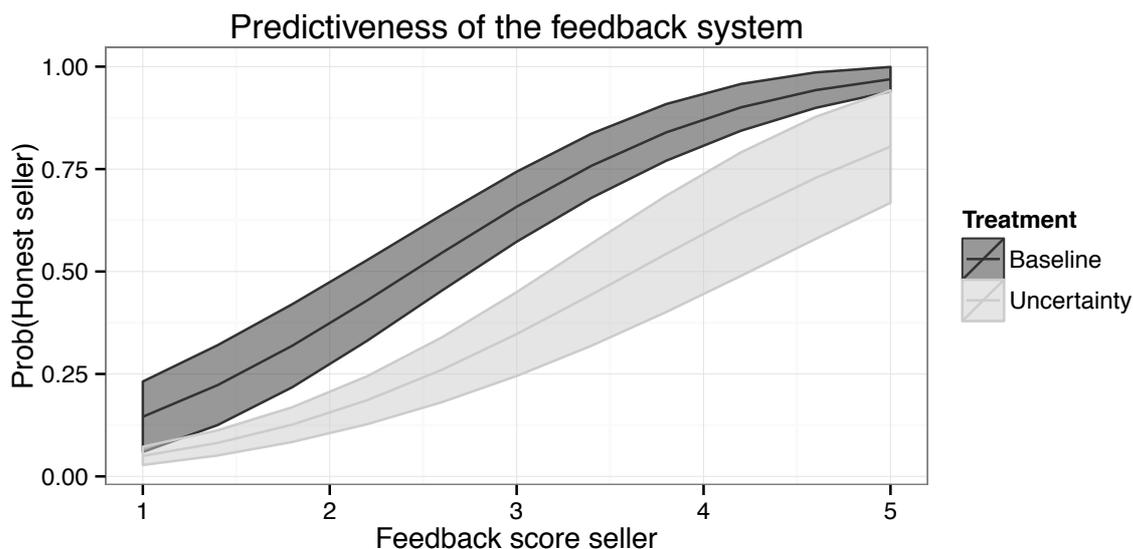


Figure 4: Predictions and 95% confidence intervals from Probit regressions for the probability of meeting an honest seller (for the regressions see Table 12 in Appendix A, data from periods 1-35).

Observe that the expected probability of meeting an honest seller is lower for all feedback scores when there is uncertainty about seller responsibility. In particular, observe that the predicted probability of meeting an honest seller is much lower and noisier for high feedback scores in the Uncertainty treatment. For example, a feedback score of 4 implies an 87% chance (confidence interval of 81 to 94%) of receiving at least as much as announced in Baseline but declines to a 59% chance (44 to 77%) in Uncertainty. A similar picture arises when looking at the best possible feedback score, where chances of honest quality still differ by 17 percentage points: 97% (94 to 100%) vs. 80% (67 to 94%). So predictions from feedback in the uncertainty system are less able to differentiate honest sellers from opportunistic sellers. This provides evidence for Hypothesis 2.

The null hypothesis for our experiment is that the introduction of uncertainty leads to perhaps noisier feedback but not biased feedback. To illustrate the predictiveness of the feedback system when feedback giving is noisy, we ran a Monte Carlo simulation with 1000 iterations of noisy feedback giving in the Baseline treatment. We mapped the uncertainty on the quality scale to the feedback scale and added draws from a normal distribution with mean 0 and standard deviation of 0.5 to each actual feedback rating submitted in the Baseline treatment.¹¹ Based on these new ratings, we recalculated sellers' feedback scores and

¹¹ The standard deviation of 0.5 accounts for the fact that the feedback scale has $\sim 1/20$ the range of the quality scale (1-5 compared to 0-100), which was shocked with draws from a normal distribution with a standard deviation of 10 in the Uncertainty treatment. Ideally, we would have run this simulation by adding noise to shipped quality, using the same draws we used in the Uncertainty treatment. However, as the feedback giving

predicted their honesty. The left panel of Figure 5 shows that neither the mean prediction nor the confidence interval changes noticeably. Unrealistically large levels of noise are needed to increase the confidence interval and flatten the curve until feedback scores become uninformative (see Figure 9 in Appendix B). We see that noisy feedback giving does not lead to the drop in predicted seller honesty in the Uncertainty treatment.

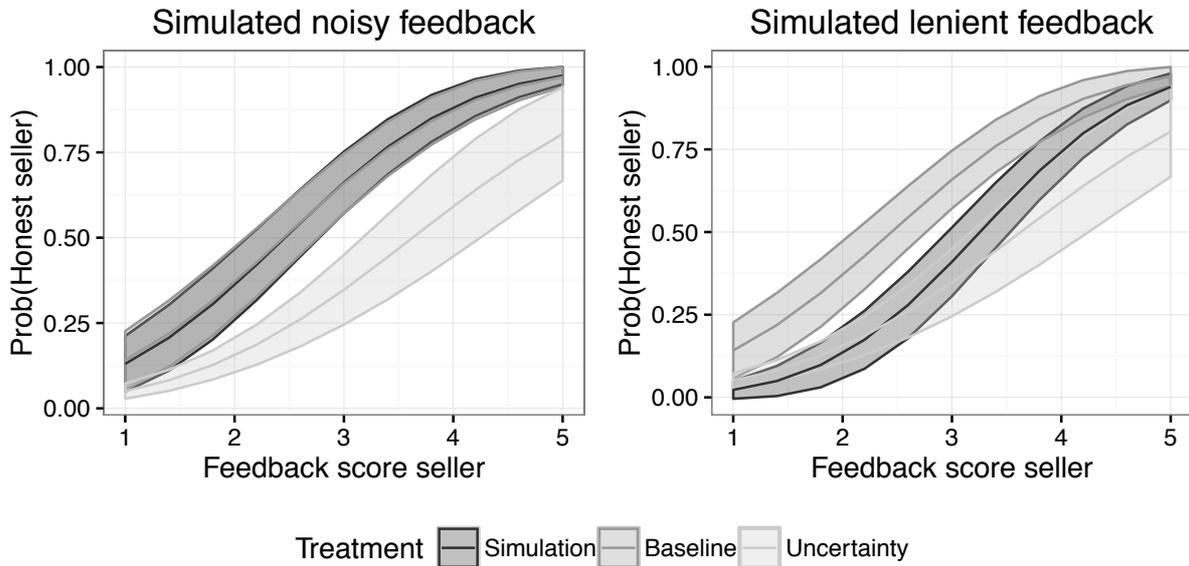


Figure 5: Monte Carlo simulations with 1000 iterations for noisy and lenient feedback in the Baseline treatment. Left panel: A random draw from a normal distribution with mean 0 and standard deviation of 0.5 was added to each feedback rating given (simulated ratings set to the closest admissible value). Based on these new ratings, seller feedback scores were recalculated and used to predict probabilities and 95% confidence intervals for seller honesty. The figure shows the averages over all iterations. Right panel: 0.5 was added to each feedback rating given. The same process as for the noisy feedback was used to produce the simulated average predicted seller honesty.

On the other hand, simulated lenient feedback shifts the curve downwards: The right panel of Figure 5 shows the results of a Monte Carlo simulation in which each actual feedback rating in the Baseline treatment was increased by 1 (simulations with alternative increase values appear in Figure 10, Appendix B). Predicted seller honesty for low-to-medium feedback scores decreases to a level comparable to what we observe in the Uncertainty treatment, while the decrease level is not as pronounced for sellers with very high feedback scores. This can be accounted for by the difference in seller honesty: In the Uncertainty treatment, sellers are less likely to ship what they announce. In the Baseline treatment, sellers with high but not perfect feedback scores are more honest such that the simulated leniency does not affect the prediction to a large extent. Hence, a combination of lenient feedback giving and the decrease

function of raters needed to generate new feedback ratings given the shocked quality is unknown, we directly shock the ratings.

in seller honesty can explain the drop in the prediction seen in Figure 4, while the exogenous shock on quality alone cannot.

Figure 6 provides an alternative way of looking at the information in Figure 4. Here, we can see the deviations from shipping the announced quality, given a seller’s feedback score. Observe that sellers with high feedback scores in the Uncertainty treatment are more likely to Shortfill than those in the Baseline treatment.

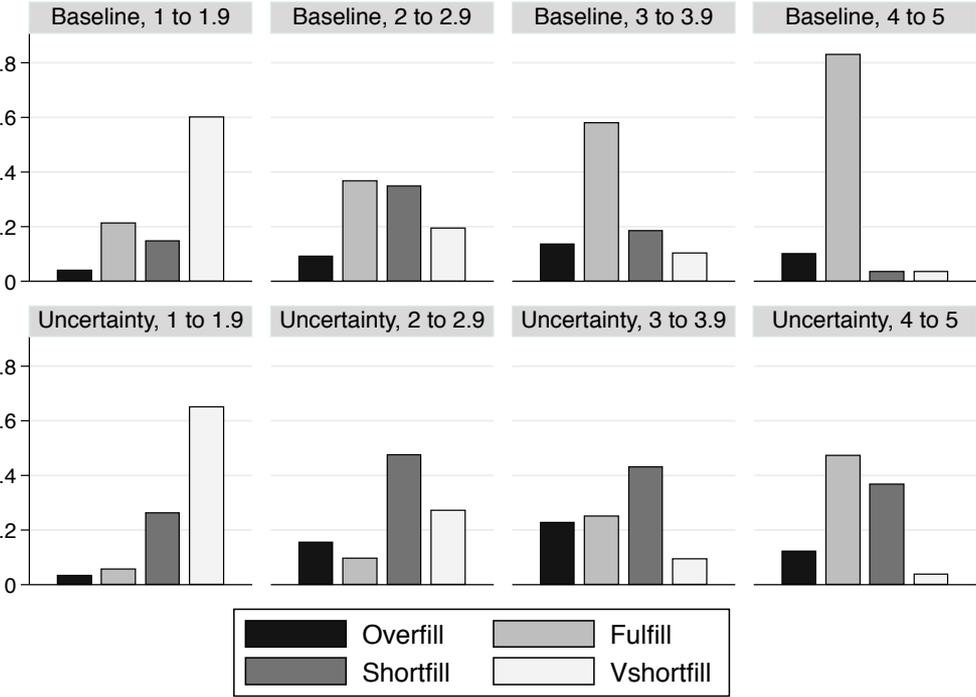


Figure 6: Frequency of fill ratios shipped for different levels of seller feedback scores, by treatment.

Figure 6 also provides insight into the nature of Overfill. We might have thought that those sellers most likely to overfill would be those with the highest feedback scores. In fact, the figure shows that overfilling is most prevalent among those with mediocre feedback scores. Plausibly, these sellers are overfilling in an attempt to curry favor with buyers and improve their feedback scores. In fact, as shown in Table 3 the rating frequency for the Baseline treatment is higher for Overfill than Fulfill, such that in this treatment, Overfilling increases the probability of getting a high feedback score.

3.4 Buyer use of the feedback system

The preceding discussion on the predictiveness of the feedback score has shown that conditional on the score, seller honesty decreases and the precision of this estimate becomes smaller as a function of lenient feedback giving and more dishonest seller behavior. This does

not necessarily imply a problem for the system if traders use a different feedback language where ratings are given according to a different standard, and traders adjust their expectations correspondingly. However, the following analysis shows that this is not the case.

A second important measure of the informational content of a feedback system is how well buyers use the information provided to form expectations about seller behavior. Table 6 offers three such measures: The squared and absolute prediction errors (shipped quality minus expected quality), along with the percentage of buyers who expect more than what was actually shipped. We define expected quality as a buyer’s bid divided by his valuation.

	Baseline		Uncertainty		<i>p</i> -value
	Mean	SD	Mean	SD	
Prediction error squared	692.85	1488.41	773.58	1347.23	0.496
Prediction error absolute	18.14	19.08	20.57	18.73	0.290
% Overexpectation	20.8%	4.1	34.0%	4.7	0.041
Observations	1127		1066		

Table 6: Subjective predictiveness: Prediction errors are calculated as shipped quality minus expected quality. Expected quality is a buyer’s bid divided by his valuation. We exclude subjects who do not submit a bid and those who submit bids larger than their valuation. Overexpectation is the percentage of buyers who expected more than they actually received. Two-tailed *p*-values from Mann-Whitney U tests with data aggregated on the matching group level for periods 1-35 (10 independent observations per treatment).

While both squared and absolute prediction errors are larger in the Uncertainty than in the Baseline treatment, the differences are not significant. These two measures, however, do not distinguish between overly optimistic and overly pessimistic predictions. In the Uncertainty treatment, 34% of subjects expected more than they received (% Overexpectation), which is significantly higher than 21% of subjects in the Baseline treatment. Hence, buyers do not make worse predictions in general under Uncertainty, but their expectations are too high more often. That is, relative to the Baseline treatment, buyers in the Uncertainty treatment do not fully adjust to the diminished informativeness of the feedback system, even though they experience a higher degree of disappointment as measured by the % Overexpectation variable.

To further investigate how buyers form expectations we ran panel Tobit regressions on expected quality and separately interact feedback score and announced quality with a treatment dummy (c.f. Table 13, Appendix A).¹² As one would expect a better feedback score and higher quality announcement significantly increase buyers’ expectations in both

¹² Also descriptively, we do not observe large treatment differences in terms of bidding behavior. The average bid in the Baseline and Uncertainty treatment is 154 and 146, respectively. The share of bidders submitting no bid is also similar across treatments (Baseline: 17%; Uncertainty: 19%).

treatments. However, the interaction effects with Uncertainty are not significant and thus we find no treatment differences of how buyers use the available information. The fact that a seller’s feedback score has no different effect under Uncertainty indicates that buyers fail to account for lenient ratings when interpreting feedback scores before submitting their bids. Consistent with this, we observe that prices are similar across treatments. On average, the final price is 128 ECU in the Baseline treatment and 123 ECU in the Uncertainty treatment.¹³ Panel Tobit regressions show that the seller’s feedback score and the announced quality have a significant positive effect on the final price (c.f. Model 1 Table 14, Appendix A) but do not indicate a treatment difference. Again, interaction effects in Models 2 and 3 with Uncertainty are not significant and thus do not suggest that feedback score or announcement are interpreted differently across treatments, contrary to the buyer part of Hypothesis 3.

3.5 Seller behavior

We hypothesized that sellers would take advantage of the random distortion that creates uncertainty about their choice of quality. As noted, the announcements made by sellers regarding quality are similar between treatments (Baseline: 86.6% vs. Uncertainty: 85.9%), while shipped quality is lower under Uncertainty (75.7% vs. 67%). Table 7 breaks out the difference in shipping behavior by fill ratio across treatments.

	Share of sellers			Shipped / announced	
	Baseline	Uncertainty	<i>p</i> -value	Baseline	Uncertainty
Overfill	9.0%	11.9%	0.59	1.05	1.11
Fulfill	61.2%	22.3%	<0.01	0	0
Shortfill	15.3%	40.1%	<0.01	0.92	0.90
Vshortfill	14.5%	25.8%	0.24	0.20	0.28

Table 7: Share of sellers in each fill ratio category and the respective average ratio of announced and shipped quality. Two-tailed *p*-values from Mann-Whitney U tests with data aggregated on the matching group level for periods 1-35 (10 independent observations per treatment).

The share of trades in which sellers ship at least as much as they announce is about twice as large in the Baseline treatment: 70.2% versus 34.2%. This is mostly due to the shift from the Fulfill to the Shortfill category. In 61.2% of all Baseline auctions, the shipped quality is equal to the announced quality, whereas this is only the case in 22.3% under Uncertainty; this difference is significant. In the Uncertainty treatment there are more sellers who short buyers by a small amount: 40.1% fall into the Shortfill category, whereas this

¹³ Selling probability is almost identical across treatments since the share of successful auctions is 93% in the Baseline treatment and 92% in the Uncertainty treatment.

happens in only 15.3% of all auctions when there is no random distortion of quality; also significant. Within the Shortfill category, the average level of deception is equally large in both treatments: in Baseline and Uncertainty, sellers classified as Shortfill on average ship about 9 percentage points less than promised.

Overall, our evidence is in line with the seller part of Hypothesis 3: Under Uncertainty, sellers strategically ship lower quality to increase their own profits at the expense of buyers. It is interesting to see that sellers display a high level of Shortfill already within the first five periods under Uncertainty (Figure 7). A well-functioning feedback system should be able to inform prospective buyers about seller trustworthiness and thereby educate sellers to provide high(er) quality. Figure 7 shows that this is not the case under Uncertainty: The share of Short- and Vshortfill sellers remains relatively stable over all periods. In contrast in the Baseline treatment, the initial share of honest (Ful- or Overfill) sellers is larger and increases over time. This suggests that the feedback system without noise is better able to educate sellers to fulfill their announcements.

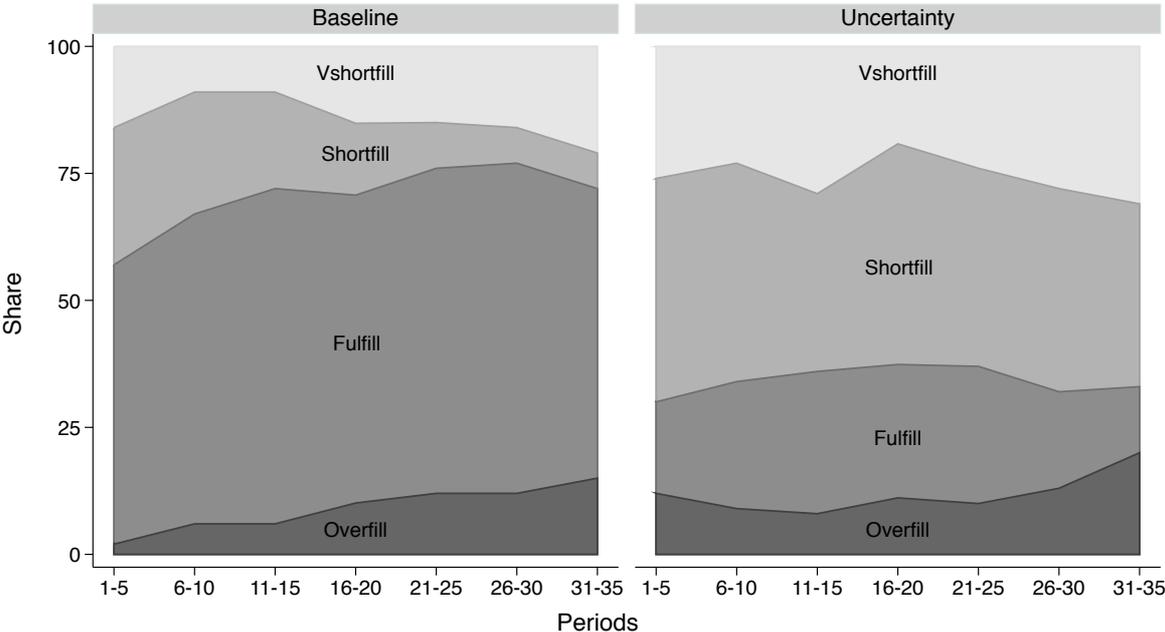


Figure 7: Share of sellers within the four fill ratio categories, by periods. Each data point represents the average share within this category in the respective five periods.

3.6 Sellers’ reaction to feedback

In order to closer investigate the disciplining effect of the feedback system, we analyze sellers’ reaction to feedback ratings. We test whether the likelihood that a short filling seller becomes honest in the next period depends on the change in his feedback score due to the current feedback rating and whether this reaction is different between treatments.

The two models in Table 8 show the results of random effects Probit regressions with a seller's trustworthiness (0 = dishonest and 1 = honest) in the *next* period as the dependent variable. The models are restricted to sellers who ship less than announced in the *current* period. To measure the effect of feedback ratings, we use the variable 'change in feedback score' which is the difference between the received feedback rating and the current feedback score. Thus, a positive (negative) value of the continuous variable 'change in feedback score' indicates that the received feedback was above (below) the seller's current feedback score and thus the score increases (decreases) in the following period.

Honest in next period	Model 1	Model 2
Uncertainty	-0.562** (-3.049)	-0.381* (-2.038)
Change in feedback score	-0.235** (-2.921)	-0.409*** (-3.641)
Feedback score	0.132 (1.574)	0.098 (1.187)
Shipped / announced	0.405 (1.496)	0.440 (1.635)
Seller profit	-0.002 (-1.091)	-0.003 (-1.233)
Period	-0.011 (-1.239)	-0.014 (-1.534)
Change X Uncertainty		0.294* (2.299)
Intercept	-0.936** (-3.103)	-0.949** (-3.211)
<i>N</i>	472	472
Log likelihood	-198.0	-195.3

Table 8: Dishonest sellers' reaction to feedback. Random effects Probit regression with dummy variable whether seller is honest in the *next* period (0 = not honest; 1 = honest). Observations are restricted to dishonest sellers in the *current* period. Periods 1-35; *t* statistics in parentheses; + $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, * $p < 0.001$.**

Model 1 shows that feedback ratings below the current feedback score significantly increase the likelihood that a dishonest seller becomes honest in the following period. The larger this difference, the higher the probability that the seller changes his behavior. The effects of uncertainty in both models confirm what we saw in Figure 7: in the Uncertainty treatments sellers in general are less likely to become more trustworthy. Furthermore, the interaction effect between the treatment dummy and the decrease in feedback score shows that the disciplining effect of bad feedback ratings is significantly lower under Uncertainty. When we compute the marginal effect of a change in a seller's feedback score this is also only

significant in the Baseline but not in the Uncertainty treatment. This means that sellers react less to bad feedback. A possible explanation could be that they expect that an unlucky draw of the distortion factor also provides an excuse for a low(er) feedback score so that buyers give the benefit of the doubt and assume sellers to be trustworthy despite a negative signal.

3.7 Does honesty pay?

As shown above, shortfalling sellers are more likely to get away with no negative feedback under Uncertainty, changing the incentives for honest behavior. In this regard, Model 1 of Table 9 shows that, in aggregate, being honest has a positive effect on seller profits. However, when we allow honest behavior to have different effects in the two treatments we observe that honest behavior pays significantly only in the Baseline treatment.¹⁴

	Model 1	Model 2
Uncertainty	1.789 (0.505)	9.371* (2.078)
Honest in last period	5.239* (2.048)	12.536*** (3.513)
Announced	0.901*** (8.861)	0.893*** (8.785)
Shipped	-0.769*** (-14.463)	-0.773*** (-14.592)
Period	0.305** (2.671)	0.286* (2.509)
Average seller fill ratio	48.616*** (6.297)	48.980*** (6.370)
Honest X Uncertainty		-13.780** (-2.923)
Intercept	-26.219* (-2.426)	-30.430** (-2.791)
<i>N</i>	1191	1191
Log likelihood	-5618.3	-5614.1

Table 9: Effect of honest behavior in last period on current seller profits. Random effects Tobit regression with seller profit as dependent variable. Periods where a seller did not sell his product are excluded. Periods 1-35; *t* statistics in parentheses; * $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, * $p < 0.001$.**

Looking at the marginal effects in Model 2, average profits of a previously deceiving seller under Certainty are 38 ECU whereas a previously honest seller in the same treatment earns 50.5 ECU. This is a significant increase of more than 30%. In contrast this honesty premium is not present in the Uncertainty treatment: earnings non-significantly decrease by 2.7% from 47.3 to 46 ECU. Overall these results indicate that uncertainty about seller responsibility

¹⁴ We control for the seller type in these regressions by including the seller-specific average seller fill ratio over all 35 rounds, i.e., the shipped quality divided by the announced quality. Using the share of periods where a seller was honest yields qualitatively similar results.

hampers the functioning of the feedback system and removes incentives for truthful seller behavior.

3.8 Market performance

Finally, we analyze how the feedback system affects market performance. Descriptive statistics in Table 10 show that efficiency, measured as the ratio of realized to maximum possible surplus, decreases from 68% to 58% under Uncertainty.¹⁵ The efficiency losses in both treatments are mainly due to sellers shipping less than maximum quality. Inferior quality accounts for roughly 75% of the efficiency losses in both treatments. Misallocation in the sense that the bidder with the lower valuation purchases the product causes the remaining efficiency losses. In addition, the level of shipped quality is lower under Uncertainty: 67% under Uncertainty and 76% under Certainty. However, both measures of market performance are only weakly significantly different across treatments when using panel Tobit regressions (c.f. Table 11).

	Baseline	Uncertainty	<i>p</i>-value
Efficiency in %	0.68 (0.33)	0.58 (0.34)	0.11
Shipped quality in %	75.72 (28.88)	66.96 (30.51)	0.17
Seller profit in ECU (if sold)	49.29 (38.23)	52.67 (34.56)	0.17
Buyer profit in ECU (if sold)	48.41 (66.17)	33.27 (68.46)	0.06
<i>N</i>	700	700	

Table 10: Descriptive statistics on market performance in periods 1-35 (standard deviation). Seller and buyer profits are based on successful trades i.e. when the product is sold (652 trades in the Baseline treatment and 642 in the Uncertainty treatment).

¹⁵ Realized surplus is the product of the winning bidder's valuation and the shipped quality net of the seller's costs. The maximum possible surplus is calculated by multiplying the larger of the two valuations with 100% quality minus 100 ECU seller's costs.

	Efficiency	Shipped	Seller profit	Buyer profit
Uncertainty	-0.130 ⁺ (-1.747)	-12.307 ⁺ (-1.888)	3.666 ⁺ (1.693)	-14.652 [*] (-2.067)
Announced	0.006 ^{***} (6.390)	0.539 ^{***} (6.634)	0.084 (0.953)	0.211 (1.159)
Period	-0.001 (-0.669)	-0.289 ^{***} (-3.294)	0.691 ^{***} (6.914)	-0.322 ⁺ (-1.767)
Intercept	0.178 ⁺ (1.812)	38.340 ^{***} (4.601)	29.478 ^{***} (3.778)	34.954 [*] (2.119)
<i>N</i>	1400	1400	1294	1294
Log likelihood	-830.8	-5720.8	-6418.4	-7265.0

Table 11: Random effects tobit regressions with different performance measures as dependent variables. Periods 1-35; *t* statistics in parentheses; ⁺ $p < 0.1$, ^{*} $p < 0.05$, ^{} $p < 0.01$, ^{***} $p < 0.001$.**

4 DISCUSSION AND CONCLUSION

We reported evidence on the influence of seller attributional uncertainty on the performance of a market that relies on a feedback system to prevent seller moral hazard and adverse selection. We find that, with attributional uncertainty, buyers show greater leniency towards sellers who provide value moderately less than advertised by giving them high ratings or remaining silent about their performance more frequently than they would if seller attribution were certain. The inflation of ratings introduced by leniency on the individual level then works its way up the information chain in the reputation system, hampering the predictiveness of sellers' feedback profiles. For example, under Uncertainty, a buyer is less likely to encounter an honest seller, even if the seller has a perfect feedback score. Hence, feedback profiles do a poorer job in helping buyers to discriminate between seller types. With the increase in moral wiggle room, sellers deliver less value under Uncertainty. Buyers fail to account for this reduction in the sense that the prices they pay a seller with a given feedback profile are about the same as under the more accurate feedback obtained when seller attribution is certain. As a result, buyers pay most of the cost of seller malfeasance under Uncertainty.

Overall, seller trustworthiness is significantly lower under Uncertainty as the number of sellers shipping at least as much as promised declines by over 50% from 70.3% to 34.3%. And sellers who receive bad ratings are less likely to change their behavior suggesting that they anticipate that the Uncertainty will provide a credible excuse for lower feedback scores. From the viewpoint of seller profits, the increase in deceptive behavior is rational. In the Baseline treatment, honest behavior leads to a significant increase of 33% in a seller's profits

in the following period. However, this is no longer the case when uncertainty disguises responsibility. Here, the honesty premium vanishes and profits insignificantly decrease by 2.7%. In short, the reputation system based on inflated ratings does not provide sufficient incentives for trustworthy seller behavior.

Less clear is the rationality of buyer behavior. In the Uncertainty treatment, the fraction of subjects who expected more than they received increases from 20% under Certainty to 34% under Uncertainty. This over-optimism leads to prices changing little across treatments. As a result, buyer profits fall 31% under Uncertainty. Why buyers do not learn to adjust to the less informative nature of the feedback system under Uncertainty (in contrast to sellers' considerable adjustment) is not clear. One potential explanation is the higher variability associated with using feedback under Uncertainty to forecast seller reliability on display in Figure 4. There is a large literature to show that variability in payoff feedback impedes learning about optimal actions (e.g., Bereby-Meyer and Roth 2006, Ockenfels and Selten 2014).

There is reason to believe that our results underestimate the true magnitude of feedback compression in field marketplaces. Participants in our experiment did not receive any personal information about each other and had no means of communicating. But in a typical transaction, the trading partners know each other's names, addresses and bank details and have exchanged various email messages. Such social communication can lead to a feeling of empathy (Andreoni and Rao, 2011), obligation (Malmendier and Schmidt, 2012), or social pressure (Malmendier et al., 2014). Reduced social distance can increase other-regarding behavior (Hoffman et al., 1996). Hence, it is conceivable that in real world interactions, the reporting rate of negative experiences is further decreased by social communication and closeness. There might even be a cross-effect with uncertainty: Fear of punishment for negative feedback that later turns out to be incorrectly founded.¹⁶

An immediate implication our study has for market design is that feedback system performance can be improved by reducing uncertainty about trader attribution in problematic trades, although for practical reasons the effectiveness of this remedy is likely limited. Looking at eBay, we observe actions to reduce uncertainty about seller responsibility. For example, shipping labels for parcels can be directly purchased via eBay and buyers automatically receive tracking numbers. While this reduces uncertainty about whether delays

¹⁶ For example, traders could worry about being the target of a lawsuit (<http://www.dailymail.co.uk/news/article-1265490/eBay-buyer-sued-defamation-leaving-negative-feedback-auction-site.html>).

are due to the seller or the postal service it does not help to clarify whether damages occurred before or during shipping. In a similar vein, eBay recently increased the number of images that can be included for free in a product listing. However, for items classified as ‘used’ the degree of signs of usage is still subject to interpretation. These examples illustrate how difficult it is to fully eliminate uncertainty about seller responsibility. (See Samak (2013) for some ideas on how to handle the problem of rating over heterogeneous good categories.)

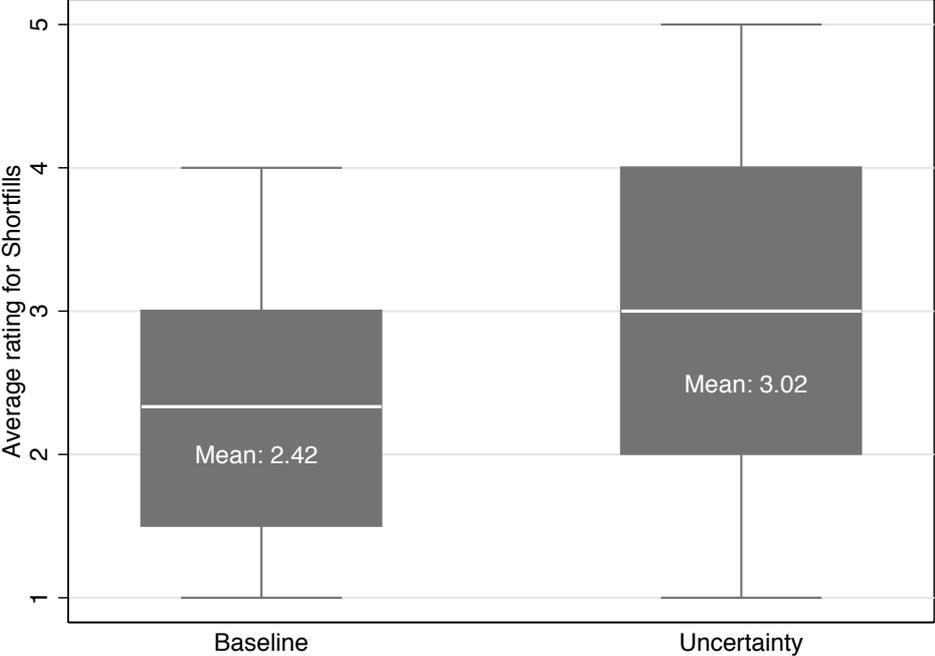


Figure 8: Boxplots for the distribution of individual average ratings for Shortfills in both treatments.

Another potential approach would be to put more weight on traders who are less likely to exhibit leniency under Uncertainty. To explore the potential of this approach, we went to our data and calculated for each subject the average rating for cases where he was shortfilled (not all subjects received qualities in the Shortfill range during the experiment, and of those who did, not all chose to leave feedback, so the following analysis is based on 55% of subjects in Baseline and 72% of subjects under Uncertainty). The boxplots in Figure 8 summarize the distributions of these averages. The average ratings under Uncertainty are higher and more spread out. However, there is still a sizeable portion of buyers under Uncertainty who rate Shortfills similarly ‘hard’ as raters in Baseline; specifically, 30.2% of the buyers under Uncertainty shown in Figure 8 on average leave ratings below 2.42, the mean of ratings in the Shortfill category in Baseline. Hence, one way to fix the compression of ratings would be to identify those buyers who are less prone to the leniency bias and place more weight on their ratings. This data also reinforces an earlier recommendation for

improving these systems: making efforts to get silent traders to report feedback (Dellarocas and Wood, 2008). The ‘harder’ raters under Uncertainty tend to be silent more often than their lenient counterparts (48% vs. 30.8%). We caution that these conclusions are based on rather small sample size; a more complete study need to be performed.

With regard to the broader implications for reputation systems, our findings regarding rating behavior under Uncertainty may also be relevant for credence goods markets where agents also have the possibility to exploit an informational advantage. Credence goods, such as medical treatment, car repair service or legal advice, are characterized by the fact that after the transaction or consumption a consumer can assess the derived utility from the good but still does not know whether the good or service provided by the agent was an adequate and efficient choice to solve the consumer’s initial problem. Hence, there is uncertainty about the responsibility of the trading partner, and consumers face a similar problem as in our setup. Studies of the influence of reputation systems on moral hazard in credence goods markets show that reputational information may have deterrent effects on agent’s fraudulent behavior (Dulleck et al., 2011; Grosskopf and Sarin, 2010; Mimra et al., 2013). In these studies, reputational information is either provided by repeated interactions or as exact history of agents’ past actions and not by voluntary and subjective feedback ratings submitted by consumers themselves. However, regarding medical aid, there are specific websites such as healthgrades.com or ratemydm.com gathering subjective ratings by consumers about their experiences with doctors. For lawyers and car mechanics similar Internet services now exist. A promising avenue for future research might be to investigate how subjective feedback works in general in credence goods markets and whether the inherent uncertainty also leads to leniency in feedback giving.

Finally, a common observation in the theoretical literature of reputation building (such as those referenced in the Introduction) is that reputation is just as effective at promoting cooperation between matched pairs who interact repeatedly as it is for strangers who interact just once, so long as the available information about past cooperation is equivalent. In the field, however, the reputation information available to stranger pairs is widely third-party in nature, with attributional uncertainty likewise a commonplace. Given this, our findings suggest that institutions that rely on matched pairs to facilitate cooperation are likely to be more effective at facilitating cooperation than otherwise equivalent institutions that rely on stranger pairs, since cooperating in the latter circumstance is more likely to be trust misplaced in an inflated reputation.

5 For Online Publication: APPENDIX A – SUPPLEMENTARY TABLES

Honest seller	Baseline	Uncertainty	Baseline	Uncertainty
Feedback score	0.630*** (9.824)	0.494*** (8.329)	0.732*** (8.655)	0.626*** (9.260)
# Feedbacks			-0.088 (-1.304)	-0.028 (-0.478)
Announced			-0.023* (-2.473)	-0.038*** (-4.872)
Period			0.015 (0.922)	-0.002 (-0.183)
Intercept	-1.471*** (-8.165)	-1.838*** (-8.892)	0.301 (0.365)	1.171 (1.889)
<i>N</i>	608	571	608	571
pseudo R^2	0.25	0.16	0.29	0.22
% correctly classified	78.9	71.3	82.9	72.5

Table 12: Probit regression on honest seller. Periods 1-35; t statistics in parentheses; + $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, * $p < 0.001$.**

Expected quality	Model 1	Model 2	Model 3
Uncertainty	-0.436 (-0.212)	0.074 (0.025)	8.591 (1.413)
Feedback score	5.344*** (16.783)	5.413*** (12.286)	5.370*** (16.873)
Announced quality	0.206*** (6.217)	0.205*** (6.197)	0.253*** (5.699)
Period	-0.025 (-0.632)	-0.026 (-0.650)	-0.025 (-0.628)
Feedback score X Uncertainty		-0.162 (-0.255)	
Announced X Uncertainty			-0.104 (-1.572)
Intercept	38.758*** (11.538)	38.560*** (10.832)	34.566*** (8.147)
<i>N</i>	2358	2358	2358
R ² overall	0.216	0.216	0.219

Table 13: Random-effects regression on expected quality. Observations where no bid was submitted are excluded. Periods 1-35; *t* statistics in parentheses; * $p < 0.05$, ** $p < 0.01$, * $p < 0.001$.**

Price	Model 1	Model 2	Model 3
Uncertainty	-1.482 (-0.306)	-7.615 (-0.892)	40.107 (1.832)
Feedback score	10.913*** (10.124)	10.034*** (6.806)	11.004*** (10.221)
Announced	0.724*** (6.016)	0.723*** (6.002)	0.923*** (5.844)
Period	0.431*** (3.412)	0.440*** (3.472)	0.430*** (3.411)
Feedback score X Uncertainty		1.838 (0.871)	
Announced X Uncertainty			-0.474 (-1.944)
Intercept	12.814 (1.098)	15.990 (1.309)	-4.968 (-0.336)
<i>N</i>	1075	1075	1075
Log likelihood	-4315.2	-4314.8	-4313.3

Table 14: Random effects Tobit regressions with price (100-300) as dependent variable. Periods 1-35; *t* statistics in parentheses; + $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, * $p < 0.001$.**

Shipped quality	Model 1	Model 2	Model 3
Uncertainty	-12.605 ⁺ (-1.688)	-12.307 ⁺ (-1.888)	28.977 ⁺ (1.923)
Period	-0.206 [*] (-2.350)	-0.289 ^{***} (-3.294)	-0.283 ^{**} (-3.229)
Announced		0.539 ^{***} (6.634)	0.774 ^{***} (6.844)
Announced X Uncertainty			-0.480 ^{**} (-3.006)
Intercept	83.348 ^{***} (15.117)	38.340 ^{***} (4.601)	17.983 ⁺ (1.686)
<i>N</i>	1400	1400	1400
Log likelihood	-5742.8	-5720.8	-5716.3

Table 15: Random effects tobit regressions with shipped (0-100) as dependent variable. Periods 1-35; *t* statistics in parentheses; ⁺ $p < 0.1$, ^{*} $p < 0.05$, ^{} $p < 0.01$, ^{***} $p < 0.001$.**

Efficiency	Model 1	Model 2	Model 3
Uncertainty	-0.134 (-1.572)	-0.130 ⁺ (-1.747)	0.059 (0.329)
Period	0.000 (0.200)	-0.001 (-0.669)	-0.001 (-0.639)
Announced		0.006 ^{***} (6.390)	0.007 ^{***} (5.412)
Announced X Uncertainty			-0.002 (-1.152)
Intercept	0.697 ^{***} (11.056)	0.178 ⁺ (1.812)	0.086 (0.679)
<i>N</i>	1400	1400	1400
Log likelihood	-851.3	-830.8	-830.1

Table 16: Random effects tobit regressions with efficiency (0-1) as dependent variable. Periods 1-35; *t* statistics in parentheses; ⁺ $p < 0.1$, ^{*} $p < 0.05$, ^{} $p < 0.01$, ^{***} $p < 0.001$.**

6 APPENDIX B – SUPPLEMENTARY FIGURES

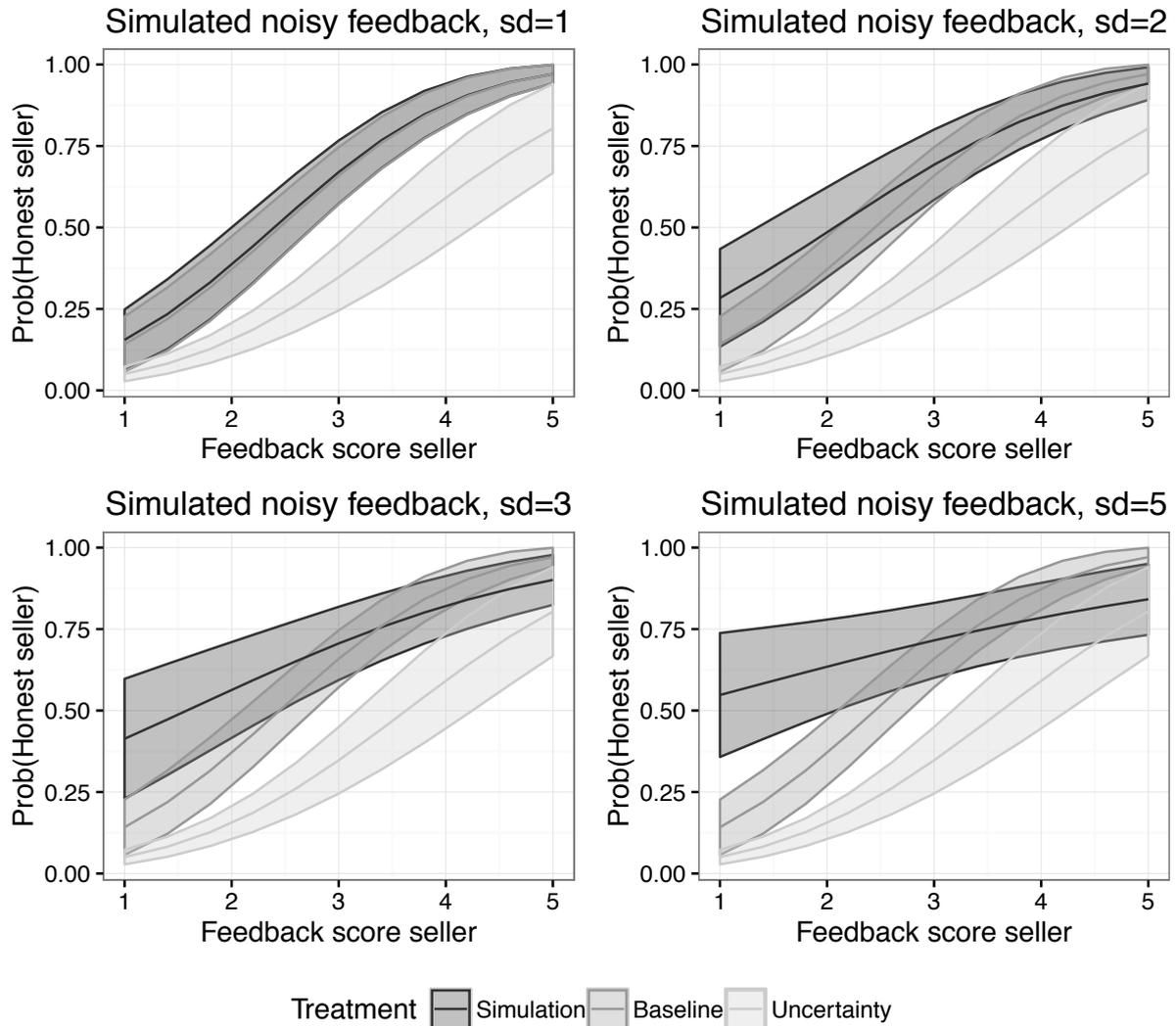


Figure 9: Monte Carlo simulation with 1000 iterations for noisy feedback in the Baseline treatment. Existing feedbacks were shocked with random draws from a normal distribution with mean 0 and standard deviation of 1 (2, 3, 5). Simulated ratings out of the rating scale were set to the closest admissible value. These noisy feedback ratings were used to calculate new seller feedback scores, based on which seller honesty was predicted. This process was iterated 1000 times, and the average predicted probabilities and upper and lower bounds for the confidence intervals are shown in the figure.

Figure 9 shows the predicted seller honesty for simulations with varying standard deviations. It can be seen that smaller standard deviations below 2 do not affect the prediction or increase the confidence interval to a large extent. The precision of the forecast decreases for standard deviations above 2, but very large noise is needed for the feedback score to become uninformative about seller honesty. However, even with a standard deviation of 5 (which equals the range of the feedback scale), the prediction does not decrease as in the Uncertainty treatment.

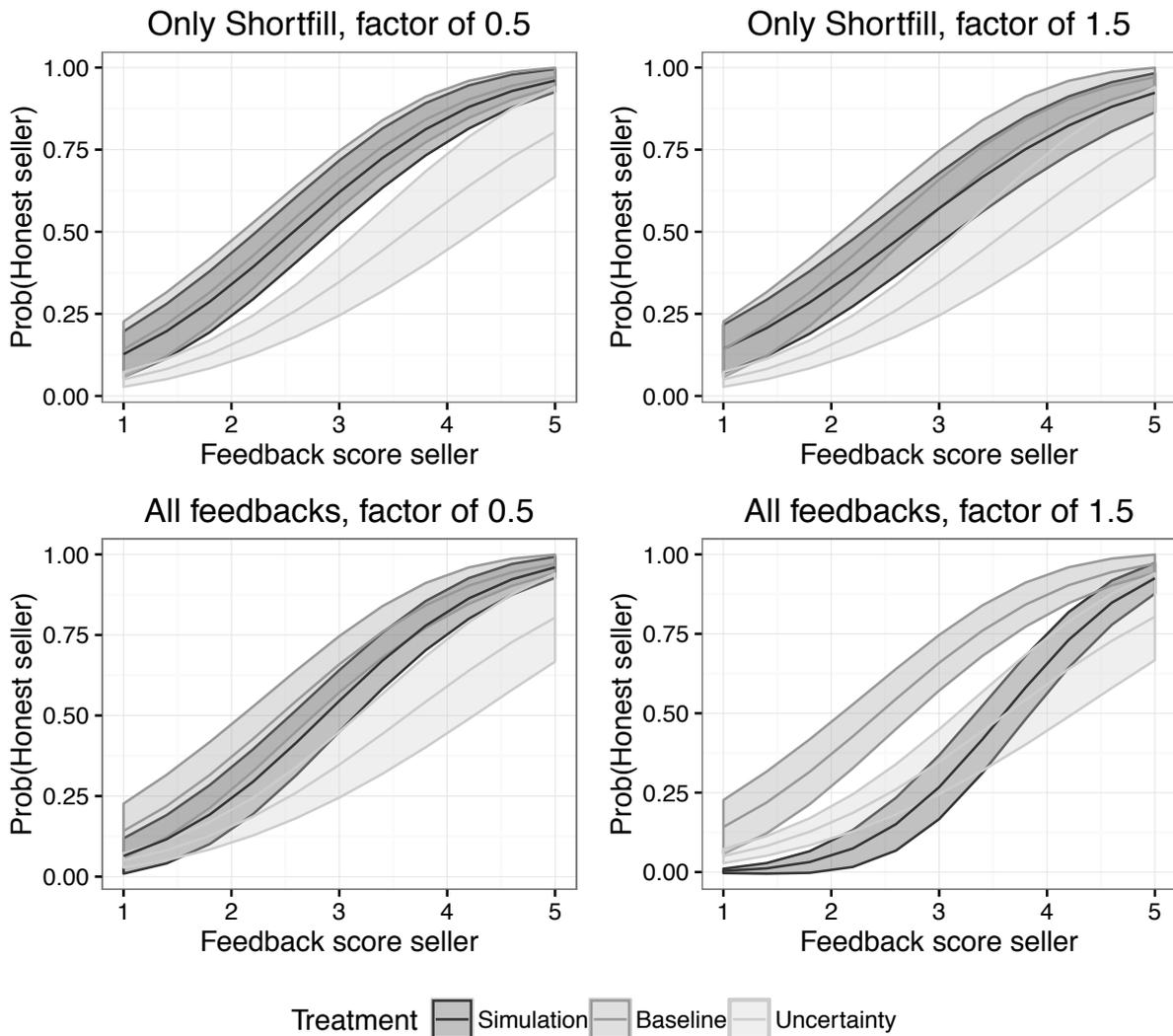


Figure 10: Monte Carlo simulation with 1000 iterations for lenient feedback in the Baseline treatment. For the upper panels, a factor of 0.5 (1.5) was added to all feedback ratings given for shortfilling sellers. For the lower panels, the same factor of 0.5 (1.5) was added to all feedback ratings given. Simulated ratings out of the rating scale were set to the closest admissible value. Based on these simulated ratings, seller feedback scores were recalculated and used to predict probabilities and upper and lower bounds for seller honesty. The figure shows the averages over all iterations.

Figure 10 shows further simulations for lenient feedback. The upper panels give the results where only the ratings for shortfilling sellers were increased by either 0.5 (the average difference in ratings between the Baseline and the Uncertainty treatments in this category) or 1.5. Due to the low number of transactions in this category in the Baseline treatment (6.18 percent), the effect on predicted seller honesty is small, but the direction is downward as in the Uncertainty treatment. The lower panels show the results for the same leniency factors added to all feedback ratings. Predicted seller honesty decreases for all levels of feedback scores but the highest. This can be accounted for by the comparatively high levels of honesty for sellers with a less-than-perfect feedback score in the Baseline treatment.

7 APPENDIX C – INSTRUCTIONS

Instructions (Baseline)

Welcome and thank you for participating in this experiment. Take the time to read carefully the instructions. If you have any questions, please raise your hand and one of the supervisors will come to help you.

You can earn money in this experiment. The specific amount depends on your decisions and the decisions of other participants. In the experiment we use ECU (Experimental Currency Unit) as the monetary unit. All participants will be endowed with an amount of 1000 ECU. Profits during the experiment will be added to this account losses will be deducted. At the end of the experiment, the balance of the account will be converted from ECUs into Euros, and paid out in cash. The conversion rate is 100 ECUs are worth 1 Euro.

From now on until the end of the experiment, please do not communicate with other participants. If you do not comply with this rule we have to exclude you from the experiment and all payments.

The experiment is repeated for 45 periods. Participants are matched into groups of three. In each group, one participant is the **seller**, the other two participants are **bidders**. At the beginning of each period, the role and the group of each participant are newly randomly determined.

In each period, the seller offers one good which, if shipped in 100% quality, costs him 100 ECUs. Each of the bidders is assigned a valuation for the good, which lies between 100 and 300 ECUs. The valuation represents the value of the good for the winning bidder if he/she receives it in 100% quality (more about quality will be said below). The valuations of the two bidders will be newly randomly drawn in each period. When drawing a valuation, every integer value between 100 and 300 has the same probability to be selected.

Each period consists of four stages:

1. In the Announcement stage the seller **publicly and non-bindingly** (i.e., without commitment) announces a quality level he/she is going to deliver after the auction and **privately and bindingly** (i.e., with commitment) decides about the actual quality of the good he/she will ship.
2. In the Auction stage the two bidders may bid for the item offered by the seller. The bidder who submits the highest bid will win the auctioned good.
3. In the Transaction stage the seller receives the price, which has to be paid by the winning bidder, and the winning bidder receives the good in the previously determined actual quality.
4. In the Feedback stage the winning bidder may give feedback on the transaction, which is then made available to traders as average feedback rating in later periods.

In the following we explain the procedures of the four stages in detail.

7.1.1 Announcement stage

In the first stage of each period, sellers enter the **announced** quality and the **shipped** quality. The announced quality is non-binding and is made public to the two bidders in the same group **before** they submit their bids in the following Auction stage. The shipped quality is binding and is only revealed to the winning bidder, **but not until** the Transaction stage. The quality must be an integer between 0% and 100%. Each quality percent costs the seller 1 ECU. Thus, the costs for the seller for shipping the good are 0 ECU if the quality is 0%, 100 ECU if the quality is 100%, and $Quality * 1 ECU$ for intermediate values of quality. In case the product is not sold, the seller does not incur any costs.

7.1.2 Auction stage

In the second stage of each period, each bidder may submit a maximum bid for the good. On the bidding screen, the bidders see the following information: The **average feedback rating** of the current seller and the **number of feedbacks** this seller received in previous periods, the **announced quality**, and his own **valuation** in the current period. The average feedback rating is the average of all feedback ratings this seller received in previous periods. Furthermore, there is a hypothetical profit calculator where bidders can enter hypothetical prices and quality levels. The calculator displays the hypothetical profit for the entered values given the bidder's valuation in the current period.

1. If you want to participate in the auction, please submit a maximum bid. Your maximum bid is the maximum amount you are willing to pay for the offered good. Your maximum bid must be at least 100 ECUs, which is the minimum price, and must not exceed the current amount on your account. If you do not want to participate in the auction in the current period, click the "No bid" button.
2. The bidder who submits the highest maximum bid wins the auction. The price the winning bidder has to pay is equal to the second highest bid plus 1 ECU.

Exceptions:

- If only one bidder submits a bid, the price is equal to 100 ECU.
 - If both maximum bids are the same, the bidder who has submitted his/her bid first wins the auction. In this case, the price is equal to the maximum bid of the winning bidder.
 - If no bidder submits a bid, the product is not sold.
3. You may think of the bidding system as standing in for you as a bidder at a live auction. That is, the system places bids for you up to your maximum bid, but using only as much of your bid as is necessary to maintain your highest bid position. For this reason, the price cannot exceed the second highest bid plus 1 ECU.

The winner of the auction must pay the price to the seller and proceeds to the Transaction stage. The losing bidder earns a profit of 0 ECU in this period. In case the product is not sold, the seller and both bidders earn a profit of 0 ECU in this period.

7.1.3 Transaction stage

The seller receives the price and the winning bidder receives the good in the previously determined actual quality. The actual value of the good for the winning bidder equals the

quality of the good times his/her valuation for the good. Thus the actual value of the good for the buyer is 0 ECU if the quality is 0%, and equal to his/her valuation if the quality is 100%.

In equations:

The payoff in ECU for the seller in this period equals:

$$\text{Seller's Payoff} = \text{Auction price} - (\text{Quality} * 1 \text{ ECU})$$

The payoff in ECU for the winning bidder in this period is:

$$\text{Winning Bidder's Payoff} = [(\text{Quality} / 100) * \text{Valuation}] - \text{Auction price}$$

7.1.4 Feedback stage

After the Transaction stage the winning bidder decides whether or not he/she wants to submit a feedback on the transaction. Submitting a feedback costs 1 ECU. The feedback rating allows the winning bidder to give feedback on the following scale:

“Please rate the transaction on a five point scale (1 is the lowest rating and 5 is the highest rating).”

After the Feedback stage the period ends and a new period with newly matched groups begins as described above.

If you have any further questions, please raise your hand and one of the supervisors will come to help you.

Instructions (Uncertainty)

Welcome and thank you for participating in this experiment. Take the time to read carefully the instructions. If you have any questions, please raise your hand and one of the supervisors will come to help you.

You can earn money in this experiment. The specific amount depends on your decisions and the decisions of other participants. In the experiment we use ECU (Experimental Currency Unit) as the monetary unit. All participants will be endowed with an amount of 1000 ECU. Profits during the experiment will be added to this account losses will be deducted. At the end of the experiment, the balance of the account will be converted from ECUs into Euros, and paid out in cash. The conversion rate is 100 ECUs are worth 1 Euro.

From now on until the end of the experiment, please do not communicate with other participants. If you do not comply with this rule we have to exclude you from the experiment and all payments.

The experiment is repeated for 45 periods. Participants are matched into groups of three. In each group, one participant is the **seller**, the other two participants are **bidders**. At the beginning of each period, the role and the group of each participant are newly randomly determined.

In each period, the seller offers one good which, if shipped in 100% quality, costs him 100 ECUs. Each of the bidders is assigned a valuation for the good, which lies between 100 and 300 ECUs. The valuation represents the value of the good for the winning bidder if he/she receives it in 100% quality (more about quality will be said below). The valuations of the two bidders will be newly randomly drawn in each period. When drawing a valuation, every integer value between 100 and 300 has the same probability to be selected.

Each period consists of four stages:

1. In the Announcement stage the seller **publicly and non-bindingly** (i.e., without commitment) announces a quality level he/she is going to deliver after the auction and **privately and bindingly** (i.e., with commitment) decides about the actual quality of the good he/she will ship.
2. In the Auction stage the two bidders may bid for the item offered by the seller. The bidder who submits the highest bid will win the auctioned good.
3. In the Transaction stage the seller receives the price, which has to be paid by the winning bidder, and the winning bidder receives the good. The received quality may be different from the shipped quality. In each period and for each seller, there is a 50% probability that a random number is added to the shipped quality. This random number can either be positive or negative. On average this random number is zero. At the end of instructions we will explain in more detail how this random number is drawn.
4. In the Feedback stage the winning bidder may give feedback on the transaction, which is then made available to traders as average feedback rating in later periods.

In the following we explain the procedures of the four stages in detail.

7.1.5 *Announcement stage*

In the first stage of each period, sellers enter the **announced** quality and the **shipped** quality. The announced quality is non-binding and is made public to the two bidders in the same group **before** they submit their bids in the following Auction stage. The shipped quality is binding and determines the costs for the seller. With a probability of 50% a positive or negative random number is added to the shipped quality. This equals the received quality, which is only revealed to the winning bidder, **but not until** the Transaction stage. The quality must be an integer between 0% and 100%. Each quality percent costs the seller 1 ECU. Thus, the costs for the seller for shipping the good are 0 ECU if the quality is 0%, 100 ECU if the quality is 100%, and $Quality * 1 ECU$ for intermediate values of quality. In case the product is not sold, the seller does not incur any costs.

7.1.6 *Auction stage*

In the second stage of each period, each bidder may submit a maximum bid for the good. On the bidding screen, the bidders see the following information: The **average feedback rating** of the current seller and the **number of feedbacks** this seller received in previous periods, the **announced quality**, and his own **valuation** in the current period. The average feedback rating is the average of all feedback ratings this seller received in previous periods. Furthermore, there is a hypothetical profit calculator where bidders can enter hypothetical prices and quality levels. The calculator displays the hypothetical profit for the entered values given the bidder's valuation in the current period.

1. If you want to participate in the auction, please submit a maximum bid. Your maximum bid is the maximum amount you are willing to pay for the offered good. Your maximum bid must be at least 100 ECUs, which is the minimum price, and must not exceed the current amount on your account. If you do not want to participate in the auction in the current period, click the "No bid" button.
2. The bidder who submits the highest maximum bid wins the auction. The price the winning bidder has to pay is equal to the second highest bid plus 1 ECU.

Exceptions:

- If only one bidder submits a bid, the price is equal to 100 ECU.
 - If both maximum bids are the same, the bidder who has submitted his/her bid first wins the auction. In this case, the price is equal to the maximum bid of the winning bidder.
 - If no bidder submits a bid, the product is not sold.
3. You may think of the bidding system as standing in for you as a bidder at a live auction. That is, the system places bids for you up to your maximum bid, but using only as much of your bid as is necessary to maintain your highest bid position. For this reason, the price cannot exceed the second highest bid plus 1 ECU.

The winner of the auction must pay the price to the seller and proceeds to the Transaction stage. The losing bidder earns a profit of 0 ECU in this period. In case the product is not sold, the seller and both bidders earn a profit of 0 ECU in this period.

7.1.7 Transaction stage

The seller receives the price and the winning bidder receives the good. With 50% probability the received quality is equal to the shipped quality and with the counter-probability of 50% the received quality is equal to the shipped quality plus the positive or negative random number. The actual value of the good for the winning bidder equals the quality of the good times his/her valuation for the good. Thus the actual value of the good for the buyer is 0 ECU if the quality is 0%, and equal to his/her valuation if the quality is 100%.

In equations:

The payoff in ECU for the seller in this period equals:

$$\text{Seller's Payoff} = \text{Auction price} - (\text{shipped Quality} * 1 \text{ ECU})$$

The payoff in ECU for the winning bidder in this period is:

$$\text{Winning Bidder's Payoff} = [(\text{received Quality} / 100) * \text{Valuation}] - \text{Auction price}$$

7.1.8 Feedback stage

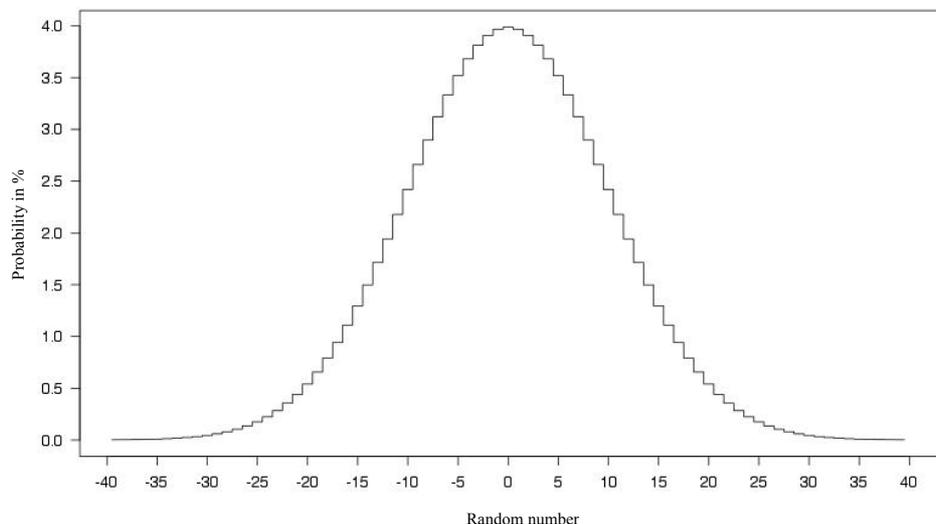
After the Transaction stage the winning bidder decides whether or not he/she wants to submit a feedback on the transaction. Submitting a feedback costs 1 ECU. The feedback rating allows the winning bidder to give feedback on the following scale:

“Please rate the transaction on a five point scale (1 is the lowest rating and 5 is the highest rating).”

After the Feedback stage the period ends and a new period with newly matched groups begins as described above.

Random number

As explained before, the received quality equals the shipped quality determined by the seller plus or minus a random distortion term. This random distortion term takes on only integer values and is drawn in such a way that on average it equals zero and negative and positive values are equally likely. In the figure you see for each value between -40 and 40 how likely it is that the distortion term equals this value.



The figure reveals that smaller distortions (positive as well as negative) occur more often than larger ones and values aperiod 0 occur most often. The probability that the distortion is exactly equal to zero is about 4%. Loosely speaking this means that in about 4 of 100 cases the distortion term will be exactly equal to 0. The area below the line displays the probability that the distortion term falls in a particular range. For example, the probability that the noise term is in between -15 and 15 is about 88%.

In 50% of the cases (in 50 of 100 cases) the distortion term will be between -7 and 7 .

In 75% of the cases (in 75 of 100 cases) the distortion term will lie between -12 and 12 .

In 95% of the cases (in 95 of 100 cases) the distortion term will lie between -20 and 20 .

For participants with knowledge of statistics: the distortion terms are drawn from a normal distribution with mean 0 and standard deviation 10. It does not matter if this does not mean anything to you: it only matters that you understand "qualitatively" how often different values of the distortion term occur.

There is a very small probability that the noise term is smaller than -40 : in 3 of the 100.000 cases the value is smaller than -40 . Likewise, there is a very small probability that the noise term is greater than 40: in 3 of the 100.000 cases the noise term is greater than 40 (you cannot infer this from the figure).

Each seller's distortion term is independently determined in the way described above. This means that the noise term in a seller's signal is (very likely) different from the noise terms in the signal of the other sellers. It also means that a noise term in the one period does not depend on the noise terms in any other period.

Because quality cannot be lower than 0% or higher than 100%, the sum of the shipped quality and the distortion term is capped at 0 (100) if it is lower (higher) than 0 (100).

If you have any further questions, please raise your hand and one of the supervisors will come to help you.

8 BIBLIOGRAPHY

- Andreoni, J., Rao, J.M., 2011. The power of asking: How communication affects selfishness, empathy, and altruism. *Journal of Public Economics* 95, 513–520.
- Ba, S., 2001. Establishing online trust through a community responsibility system. *Decision Support Systems* 31, 323–336.
- Ba, S., Pavlou, P.A., 2002. Evidence of the Effect of Trust Building Technology in Electronic Markets: Price Premiums and Buyer Behavior. *MIS Quarterly* 26, 243–268.
- Bajari, P., Hortacsu, A., 2003. The Winner's Curse, Reserve Prices, and Endogenous Entry: Empirical Insights from eBay Auctions. *The RAND Journal of Economics* 34, 329–355.
- Bar-Isaac, H., Tadelis, S., 2008. Seller reputation. *Foundations and trends in microeconomics* 4, 273–351.
- Bauerly, R.J., 2009. Online auction fraud and eBay. *Marketing Management Journal* 19, 133–143.
- Bénabou, R., Tirole, J., 2011. Identity, morals, and taboos: Beliefs as assets. *Quarterly Journal of Economics* 126, 805–855.
- Bereby-Meyer, Y., Roth, A.E., 2006. The speed of learning in noisy games: Partial reinforcement and the sustainability of cooperation. *American Economic Review* 96, 1029–1042.
- Bigelow, J., 1904. *The works of Benjamin Franklin*.
- Bol, J.C., 2011. The Determinants and Performance Effects of Managers' Performance Evaluation Biases. *Accounting Review* 86, 1549–1575.
- Bolton, G.E., Greiner, B., Ockenfels, A., 2013. Engineering Trust - Reciprocity in the Production of Reputation Information. *Management Science* 59, 265–285.
- Bretz, R.D., Milkovich, G.T., Read, W., 1992. The Current State of Performance Appraisal Research and Practice: Concerns, Directions, and Implications. *Journal of Management* 18, 321–352.
- Dana, J., Weber, R., Kuang, J., 2007. Exploiting moral wiggle room: experiments demonstrating an illusory preference for fairness. *Economic Theory* 33, 67–80.
- Dellarocas, C., 2005. Reputation Mechanism Design in Online Trading Environments with Pure Moral Hazard. *Information Systems Research* 16, 209–230.

- Dellarocas, C., 2003. The Digitization of Word of Mouth: Promise and Challenges of Online Feedback Mechanisms. *Management Science* 49, 1407–1424.
- Dellarocas, C., 2001. Analyzing the economic efficiency of eBay-like online reputation reporting mechanisms, in: *Proceedings of the 3rd ACM Conference on Electronic Commerce*. ACM, New York, NY, USA, pp. 171–179.
- Dellarocas, C., Wood, C.A., 2008. The Sound of Silence in Online Feedback: Estimating Trading Risks in the Presence of Reporting Bias. *Management Science* 54, 460–476.
- Dulleck, B.U., Kerschbamer, R., Sutter, M., 2011. The Economics of Credence Goods: An Experiment on the Role of Liability, Verifiability, Reputation, and Competition 101, 530–559.
- Fischbacher, U., 2007. z-Tree: Zurich Toolbox for Ready-made Economic Experiments. *Experimental Economics* 10, 171–178.
- Ganzach, Y., Krantz, D.H., 1991. The psychology of moderate prediction: II. Leniency and uncertainty. *Organizational Behavior and Human Decision Processes* 48, 169–192.
- Giebe, T., Gürtler, O., 2012. Optimal contracts for lenient supervisors. *Journal of Economic Behavior & Organization* 81, 403–420.
- Golman, R., Bhatia, S., 2012. Performance evaluation inflation and compression. *Accounting, Organizations and Society* 37, 534–543.
- Gregg, D.G., Scott, J.E., 2006. The Role of Reputation Systems in Reducing On-Line Auction Fraud. *International Journal of Electronic Commerce* 10, 95–120.
- Greif, A., 1993. Contract Enforceability and Economic Institutions in Early Trade: The Maghribi Traders' Coalition. *The American Economic Review* 83, 525–548.
- Greiner, B., 2004. The Online Recruitment System ORSEE 2.0 - A Guide for the Organization of Experiments in Economics (Unpublished Work).
- Grosskopf, B., Sarin, R., 2010. Is Reputation Good or Bad? An Experiment. *American Economic Review* 100, 2187–2204.
- Grund, C., Przemeczek, J., 2012. Subjective performance appraisal and inequality aversion. *Applied Economics* 44, 2149–2155.
- Güth, W., Huck, S., Ockenfels, P., 1996. Two-Level Ultimatum Bargaining with Incomplete Information: an Experimental Study. *The Economic Journal* 106, 593–604.
- Hoffman, E., McCabe, K., Smith, V.L., 1996. Social Distance and Other-Regarding Behavior

- in Dictator Games. *The American Economic Review* 86, 653–660.
- Houser, D., Wooders, J., 2006. Reputation in Auctions: Theory, and Evidence from eBay. *Journal of Economics & Management Strategy* 15, 353–369.
- Hu, N., Pavlou, P.A., Zhang, J., 2009. Overcoming the J-shaped Distribution of Product Reviews. *Communications of the ACM* 52, 144–147.
- Hu, N., Pavlou, P.A., Zhang, J., 2006. Can Online Reviews Reveal a Product’s True Quality? Empirical Findings and Analytical Modeling of Online Word-of-Mouth Communication, in: J. Feigenbaum J.Chuang, D.P. (Ed.), *Proceedings of the 7th ACM Conference on Electronic Commerce*. ACM Press, New York, pp. 324–330.
- Jawahar, I.M., Williams, C.R., 1997. Where all the children are above average: The performance appraisal purpose effect. *Personnel Psychology* 50, 905–925.
- Jin, G.Z., Kato, A., 2006. Price, quality, and reputation: evidence from an online field experiment. *The RAND Journal of Economics* 37, 983–1005.
- Kauffman, R., Wood, C., 2006. Doing their bidding: An empirical examination of factors that affect a buyer’s utility in Internet auctions. *Information Technology and Management* 7, 171–190.
- Klein, T.J., Lambertz, C., Spagnolo, G., Stahl, K.O., 2009. The Actual Structure of eBay’s Feedback Mechanism and Early Evidence on the Effects of Recent Changes 1. *International Journal of Electronic Business* 7, 301–320.
- Klein, T.J., Lambertz, C., Stahl, K.O., 2016. Market Transparency, Adverse Selection, and Moral Hazard. *Journal of Political Economy*.
- Kreps, D.M., Wilson, R., 1982. Reputation and imperfect information. *Journal of Economic Theory* 27, 253–279.
- Lafky, J., 2014. Why do people rate? Theory and evidence on online ratings. *Games and Economic Behavior*.
- Landy, F.J., Farr, J.L., 1980. Performance rating. *Psychological Bulletin* 87, 72–107.
- Larson, J.R., 1986. Supervisors’ performance feedback to subordinates: The impact of subordinate performance valence and outcome dependence. *Organizational Behavior and Human Decision Processes* 37, 391–408.
- Levin, J., 2013. The Economics of Internet Markets, in: Acemoglu, D., Arellano, M., Dekel, E. (Eds.), *Advances in Economics and Econometrics: Tenth World Congress*. Cambridge

- University Press, New York, pp. 48–75.
- Malmendier, U., Schmidt, K., 2012. You Owe Me. NBER Working Paper Series.
- Malmendier, U., te Velde, V.L., Weber, R., 2014. Rethinking Reciprocity. *Annual Review of Economics* 6, 849–874.
- Milgrom, P.R., North, D.C., Weingast, B.R., 1990. The Role of Institutions in the Revival of Trade: The Law Merchant, Private Judges, and the Champagne Fairs. *Economics & Politics* 2, 1–23.
- Mimra, W., Rasch, A., Waibel, C., 2013. Price competition and reputation in credence goods markets: Experimental evidence (Unpublished Work).
- Mitzkewitz, M., Nagel, R., 1993. Experimental results on ultimatum games with incomplete information. *International Journal of Game Theory* 22, 171–198.
- Moers, F., 2005. Discretion and bias in performance evaluation: the impact of diversity and subjectivity. *Accounting, Organizations and Society* 30, 67–80.
- Nosko, C., Tadelis, S., 2015. The Limits of Reputation in Platform Markets: An Empirical Analysis and Field Experiment (unpublished).
- Ockenfels, A., Selten, R., 2014. Impulse balance in the newsvendor game. *Games and Economic Behavior* 86, 237–247.
- Prendergast, C., 1999. The Provision of Incentives in Firms. *Journal of Economic Literature* 37, 7–63.
- Prendergast, C., Topel, R., 1993. Discretion and Bias in Performance Evaluation. *European Economic Review* 37, 355–365.
- Reichling, F., 2004. Effects of reputation mechanisms on fraud prevention in eBay auction (Unpublished Work).
- Resnick, P., Zeckhauser, R., 2002. Trust among strangers in Internet transactions: Empirical analysis of eBay's reputation system. *Advances in Applied Microeconomics* 11, 127–157.
- Resnick, P., Zeckhauser, R., Swanson, J., Lockwood, K., 2006. The value of reputation on eBay: A controlled experiment. *Experimental Economics* 9, 79–101.
- Rice, S.C., 2012. Reputation and Uncertainty in Online Markets: An Experimental Study. *Information Systems Research* 23, 436–452.
- Saal, F.E., Downey, R.G., Lahey, M. a., 1980. Rating the ratings: Assessing the psychometric

- quality of rating data. *Psychological Bulletin* 88, 413–428.
- Saeedi, M., Shen, Z., Sundaresan, N., 2013. The Value of Feedback: An Analysis of Reputation System (Unpublished Work).
- Samak, A.C., 2013. An experimental study of reputation with heterogeneous goods. *Decision Support Systems* 54, 1134–1149.
- Sharon, A.T., Bartlett, C.J., 1969. Effect of instructional conditions in producing leniency on two types of rating scales. *Personnel Psychology* 22, 251–263.
- Taylor, K., Wherry, R.J., 1951. A Study of Leniency in Two Rating Systems. *Personnel Psychology* 4, 39–47.
- Waung, M., Highhouse, S., 1997. Fear of Conflict and Empathic Buffering: Two Explanations for the Inflation of Performance Feedback. *Organizational Behavior and Human Decision Processes* 71, 37–54.
- Wilson, R., 1985. Reputation in games and markets, in: Roth, A.E. (Ed.), *Game-Theoretic Models of Bargaining*. Cambridge University Press, Cambridge, UK.
- Zervas, G., Proserpio, D., Byers, J., 2015. A First Look at Online Reputation on Airbnb, Where Every Stay is Above Average.