COLOGNE GRADUATE SCHOOL
in Management, Economics and Social Sciences

Vol 6 | No 4

CGS WORKING PAPER

Inflated reputations
Leniency & moral wiggle room in trader feedback systems

Gary E. Bolton (UT Dallas)
David J. Kusterer (University of Cologne)
Johannes Mans (University of Cologne)

Cologne Graduate School
in Management, Economics
and Social Sciences
Albertus-Magnus-Platz
50923 Köln
www.cgs.uni-koeln.de

University of Cologne

INFLATED REPUTATIONS
LENIENCY & MORAL WIGGLE ROOM IN TRADER FEEDBACK SYSTEMS

Gary E. Bolton*, David J. Kusterer† and Johannes Mans‡

20 May 2015

First draft: October 2013

Reputation systems associated with Internet markets are known to be subject to strategic manipulation. The experiment we present suggests that this manipulation can extend to factors that have heretofore been overlooked: the leniency and moral wiggle room that arise from uncertainty about the source of transaction problems. Uncertainty about seller culpability leads to 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.

‡ 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, corresponding author.

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, 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 Cologne University.

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. Overall, most traders have very high feedback scores. The factors behind this upward compression in reputation information, and how this influences the performance of markets constitute important questions for market design in a growing part of the economy. The same questions speak to certain gaps in our understanding of how reputation mechanisms work in general.[1]

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 (Gregg and Scott 2006; Dellarocas and Wood 2008; Bauerly 2009; Rice

[1] While our focus here is on high end, 'brag' feedback, the 'moan' phenomenon is important too. See Lafky 2014 for an innovative experiment that also explores the motives for negative rating.

2012).[2] 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 early research into the upward compression of feedback centered on eBay during a period in which the marketplace had a two-way feedback system, in which buyer and seller rate one another. Ninety-nine percent of all ratings in the eBay two-way feedback system was positive (Resnick and Zeckhauser 2002; Kauffman and Wood 2006). 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). An important conclusion of their study is that satisfied and very unsatisfied traders were more likely to leave feedback than those who were 'mildly' unsatisfied. 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 (Reichling 2004; Dellarocas and Wood 2008; Bolton, Greiner, and Ockenfels 2013).

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 (Saeedi, Shen and Sundaresan 2015), 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).[3] In fact, the j-shaped brag-and-moan pattern is common to many one-sided feedback systems (Hu, Pavlou, and Zhang 2006; Hu, Pavlou, and Zhang 2009; Zervas, Proserpio, and Byers 2015).

Feedback retaliation is not possible in one-way systems, so there must be other factors behind the observed compression. The frequency of 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). So while a dissatisfaction bias among those who choose not to report, while it might well play a role here, it seems unlikely to be the entire story. More likely, some traders are reporting a higher satisfaction level than the true level.

The guiding hypothesis behind our investigation is that uncertainty about a trader's culpability for a problematic trade leads to upward feedback compression. Attributional

---

[2] 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).

[3] Buyers received the additional option to give 'detailed seller ratings' on four predefined categories (item description, communication, shipping time, and shipping charges). One year later, eBay restricted sellers to give only positive ratings.

uncertainty of this sort is commonplace in trading environments: One reason is that some actions are unobservable: A late or non-delivery might be simply due to the parcel service or it might be due to a lazy or fraudulent seller. A second reason is imprecisions in language: Disappointment over the quality of an item sent by a seller might be due to uncertainties in a phase like ‘good condition’ or if might be due to the seller purposely overstating the quality.

Reasons to believe that attributional uncertainty will lead to trader rating leniency can be found in a number of sources. With regard to field observation: 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 (Bretz, Milkovich, and Read 1992; Bol 2011; Landy and Farr 1980; Moers 2005; Prendergast and Topel 1993; Prendergast 1999; Saal, Downey, and Lahey 1980; Sharon and Bartlett 1969), or they refrain from giving feedback at all (Larson 1986).[4] 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). With regard to laboratory observation: 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 feedback ratings are typically given in the context of one-off, relatively anonymous transactions. Also, trust games lack the pricing mechanisms of markets that might help separate seller types. A more direct test is necessary. Conducting such a test in the field is complicated, however, because quantifying uncertainty about seller culpability is difficult.

With these issues in mind, we designed a laboratory auction environment with optional feedback in which we are able to directly manipulate whether the received quality is a true signal of a seller’s effort. In the Baseline treatment, the level of shipped quality by the seller always remains unchanged while in the Uncertainty treatment, in half of the auctions a

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

random positive or negative distortion factor with expected value of zero is added, making a seller's true effort uncertain. Hence, more lenient feedback cannot be attributed to reduced expectations due to the uncertainty.

Leniency if observed – in the form of overly positive ratings or increased silence – might reduce feedback informativeness on who is an honest seller. Contingent on this being so, we hypothesized that attributional uncertainty would lead to greater opportunism on the part of sellers, and lower offered prices on the part of the buyers. Leniency in the face of uncertainty about culpability creates moral wiggle room for strategic sellers to exploit by shipping somewhat less than they promised (in our experiment, the greater the difference between what is shipped and what is promised, the lower the uncertainty about culpability). The aforementioned research into strategic retaliation in two-way feedback systems finds that relatively low informativeness is accompanied by lesser seller reliability and lower auction transaction prices (Bolton, Greiner, and Ockenfels 2013). 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). Also, in ultimatum game experiments, proposers are willing to exploit receiver uncertainty by offering less (Mitzkewitz and Nagel 1993; Güth, Huck, and Ockenfels 1996).

Our results provide evidence 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, these 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 and hence, 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 a 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 average and the number of feedback ratings the seller has received so far (given that the seller already received some feedback ratings). The feedback average 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): $\pi_b = v_i \times q_r - p$. The losing bidder receives a payoff of 0. Seller's payoff is the price p less the costs for the shipped quality: $\pi_s = p - 100q_s$. 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 average, 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 π_s = Price − 100 × Shipped quality	π_{wb} = Valuation × Received quality − Price $\pi_{lb} = 0$

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.[5] This distortion happens randomly and neither the seller nor the buyer is informed whether quality has been changed or not. Hence, buyers cannot infer the sellers' intentions from the received quality with certainty. On average, a buyer receives what a seller ships, but the seller could also have sent a higher or a lower quality.

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 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 roles[6] – were held constant. At the

[5] We restrict the random draw to integer values because sellers are limited to submitting integer quality levels as well. Instructions to subjects provide an explanation of the random draw using a graph of a normal distribution. We also provide examples of how likely specific values are in such a distribution. See Appendix B for a translated version of the instructions used in the experiment.

[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.

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

In line with the literature on leniency and moral wiggle room discussed in the introduction, we expect a benefit of the doubt to be present in buyers' ratings in one-sided feedback systems in Internet markets. We formulate the following hypothesis:

Hypothesis 1a: Under uncertainty, when culpability is unclear, buyers who receive less than announced will leave higher feedback ratings.

Besides giving overly positive feedback, leniency in ratings may also manifest in not reporting negative experiences by submitting no feedback at all. In a theoretical analysis of rating leniency in feedback systems, Dellarocas (2001) assumes that when there is noise buyers refrain from punishing sellers with bad ratings but rather prefer to remain silent when quality is "slightly bad but not too bad" (ibid, p. 173). As mentioned earlier, Dellarocas and Wood (2008) empirically investigate how different reporting probabilities conditional on the transaction experience and the trading partner's submitted feedback introduce distortion into the feedback system. In contrast to us, they assume that traders either give accurate or no feedback but never leave better or worse ratings that do not coincide with the actual received

[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.

quality. With the help of this simplifying assumption the authors show that mildly satisfied traders have a probability to provide feedback of less than 3% and thus such experiences are not recorded in feedback profiles. Overall, taking this silence bias into account, they estimate that the actual probability to make a 'mildly' satisfying ("neutral" in terms of eBay feedback) experience is significantly lower (21% for buyers and 14% for sellers) than the almost exclusively positive submitted feedbacks suggest (Resnick and Zeckhauser 2002; Kauffman and Wood 2006). Following these results, we expect that with imperfect information leniency is not only introduced by higher ratings but also by the omission of negative experiences.

Hypothesis 1b: Under uncertainty, when culpability is unclear, buyers who receive less than announced will be more likely to remain silent, leaving no feedback rating.

Together, hypotheses *1a* and *b* imply 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 sellers with high ratings deliver lower expected quality and with higher variance.

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

Lower informativeness gives rise to moral wiggle room both in the sense that it creates pecuniary incentives for opportunistic seller behavior and in the sense that is implied by Bénabou and Tirole's (2011) theory of self-interest. 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.

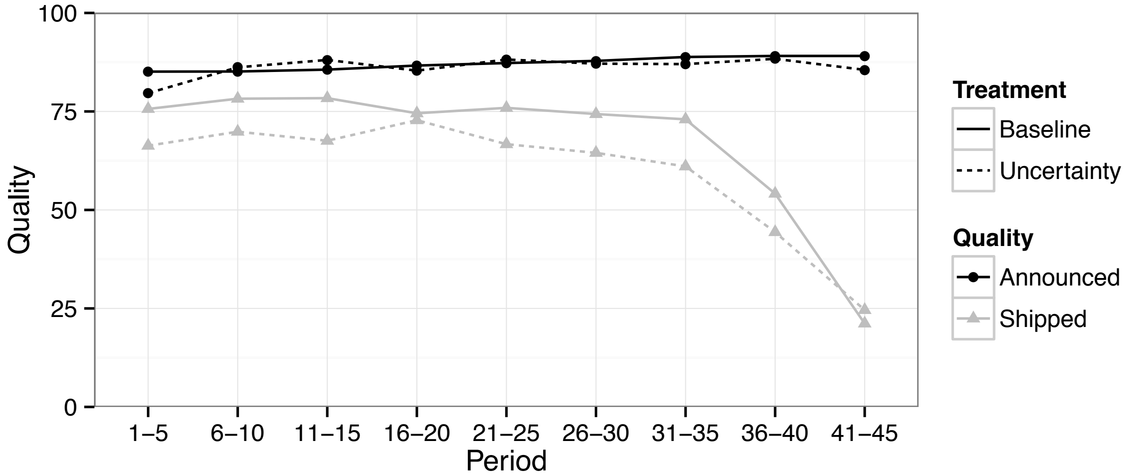


Figure 1: Average announced and shipped quality across period intervals.

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 culpability. 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 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 announce and that this is more frequent in Uncertainty than in Baseline. We discuss explanations for these Overfills when we analyze seller responses to the feedback system, in Section 3.3.

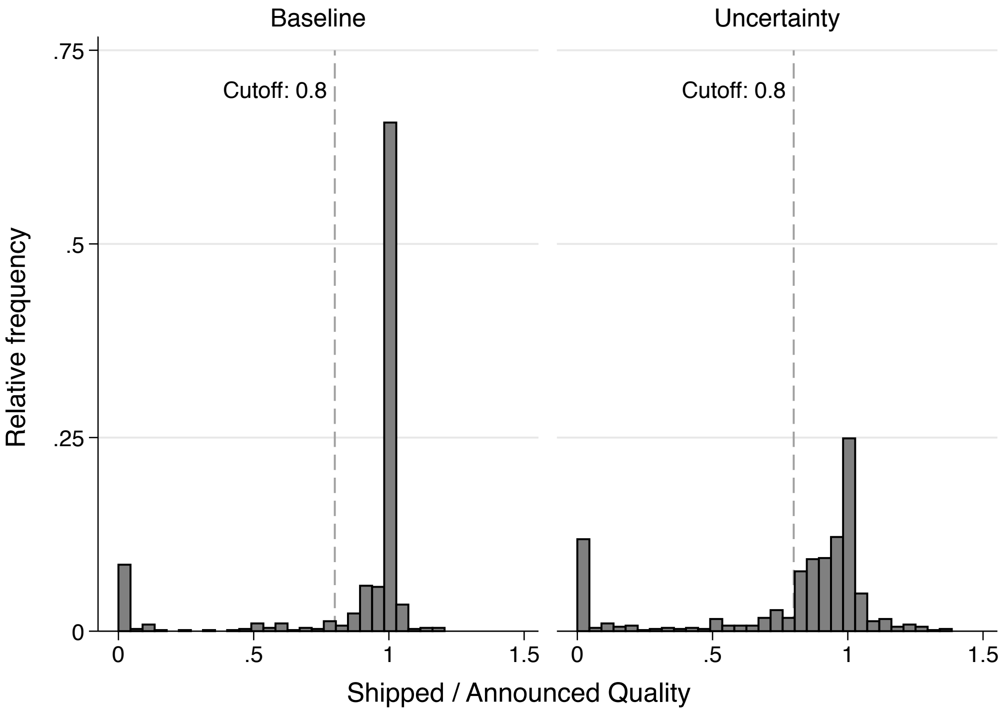


Figure 2: Histogram of fill ratios (= shipped / announced quality) across treatments (periods 1-35).

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 fill ratios. Figure 3 also shows the

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

share of silent transactions in which buyers did not leave a feedback rating. Also below, Table 3 reports the corresponding average feedbacks given and rates of silence by fill category.

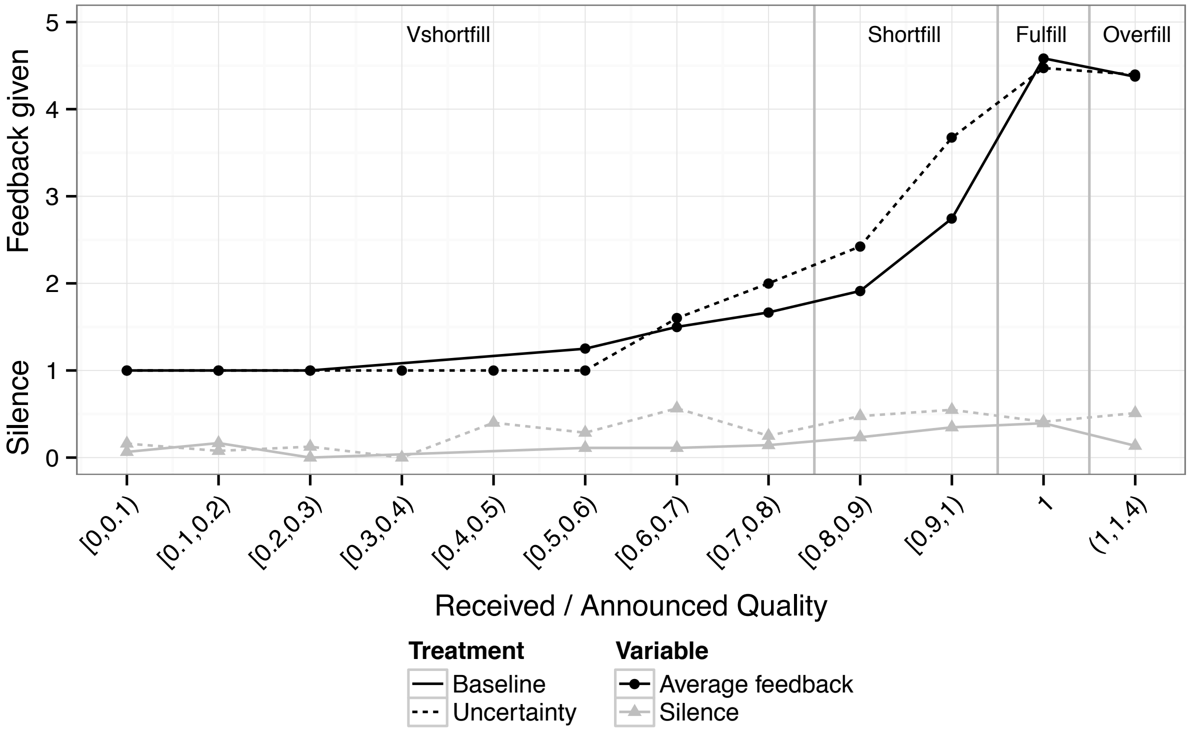


Figure 3: Average buyer feedback and frequency of not giving feedback (silence) for periods 1-35.

	Feedback			**Silence**		
	Baseline	Uncertainty	*p*-value	Baseline	Uncertainty	*p*-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.[10]

Feedback rating	Overfill	Fulfill	Shortfill	Vshortfill
Uncertainty	0.078	-0.364	1.269***	0.379
	(0.086)	(-0.442)	(3.673)	(0.754)
Received / Announced	4.930		16.131***	11.137***
	(1.397)		(8.249)	(4.064)
Announced	0.104***	0.086***	0.007	0.047^{+}
	(3.762)	(5.978)	(0.562)	(1.912)
Price	0.010	0.005	-0.005	0.027**
	(1.633)	(0.970)	(-1.595)	(2.775)
Period	0.016	-0.005	-0.014	-0.103***
	(0.709)	(-0.339)	(-1.354)	(-3.657)
Intercept	-9.875^{+}	-1.834	-12.481***	-12.789***
	(-1.833)	(-1.401)	(-6.543)	(-3.620)
N	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 trustworthiness. 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 detailed seller ratings on eBay is around 50% (Bolton, Greiner, and Ockenfels 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

[10] Because we do not expect a homogenous effect of uncertainty on feedback content and provision over all fill ratios, we run the same regression for each fill ratio category separately. This enables us to investigate the influence of uncertainty on feedback ratings in each of these categories.

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.

Silence	**Overfill**	**Fulfill**	**Shortfill**	**Vshortfill**
Uncertainty	1.356**	0.041	0.675**	0.449^{+}
	(2.865)	(0.185)	(2.751)	(1.834)
Announced	-0.024^{+}	-0.013*	-0.004	0.028*
	(-1.769)	(-2.335)	(-0.450)	(2.430)
Received / Announced	-2.665		2.803^{+}	1.199**
	(-1.596)		(1.908)	(3.253)
Price	0.000	0.002	-0.005*	-0.016*
	(0.007)	(1.269)	(-1.994)	(-2.176)
Period	0.006	0.009	-0.002	-0.014
	(0.519)	(1.470)	(-0.320)	(-1.335)
Intercept	3.521	0.388	-2.142	-2.090^{+}
	(1.462)	(0.791)	(-1.465)	(-1.727)
N	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. Periods 1-35; *t* statistics in parentheses; $^{+}$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Our results are in line with Hypothesis 1a. Buyers give more lenient feedback to sellers when what is received is less than announced and culpability is unclear. In particular, if the received shortfall is less than 20% below what was announced, we observe leniency. Hypothesis 1b 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.

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,[11] regressed on all the information that is available to the buyer before bidding: the seller's feedback average, the number of feedbacks the seller has received so far,

[11] To classify sellers' trustworthiness we use the ratio between announced and shipped quality such that the results are not biased by the distortion factor.

the announced quality and the current period (c.f. Table 12 in 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.

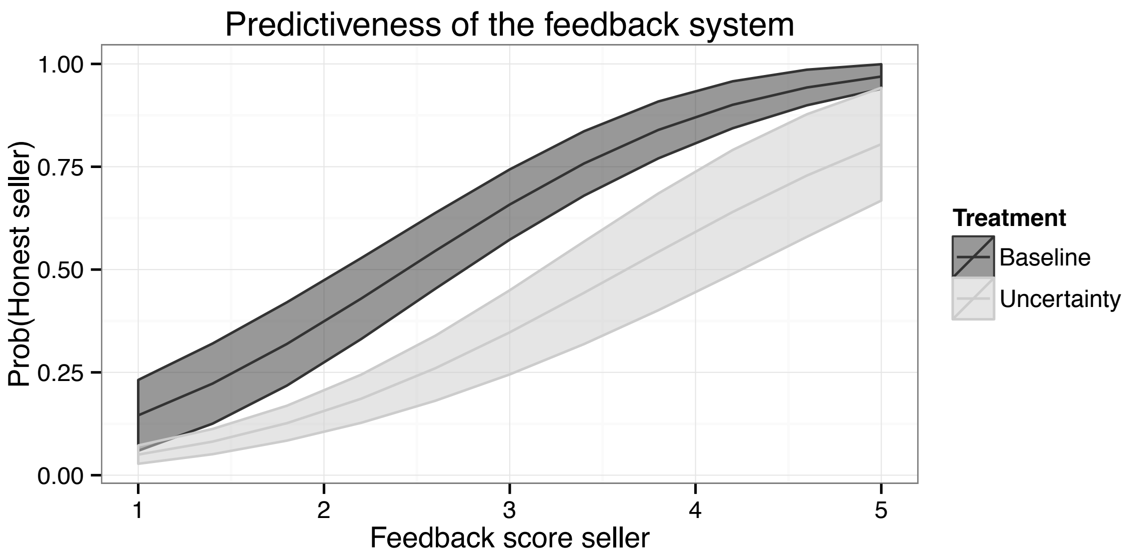


Figure 4: Predictions and 95% confidence intervals from Probit regressions for the probability of meeting an honest seller (for the regressions see Table 12 in Appendix A, data from periods 1-35).

It can be seen that the expected probability of meeting an honest seller is lower for all feedback averages when there is uncertainty about seller intentions. In particular, observe that the predicted probability of facing an honest seller is much lower and noisier for high feedback averages in the Uncertainty treatment. That is, even when a buyer observes an average feedback of 4 or higher it is less likely and less certain that the respective seller is honest. For example, a feedback average 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 average, 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.

Figure 5 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 average feedback score. Observe that sellers with high average feedback scores in the Uncertainty treatment are more likely to Shortfill than those in the Baseline treatment.

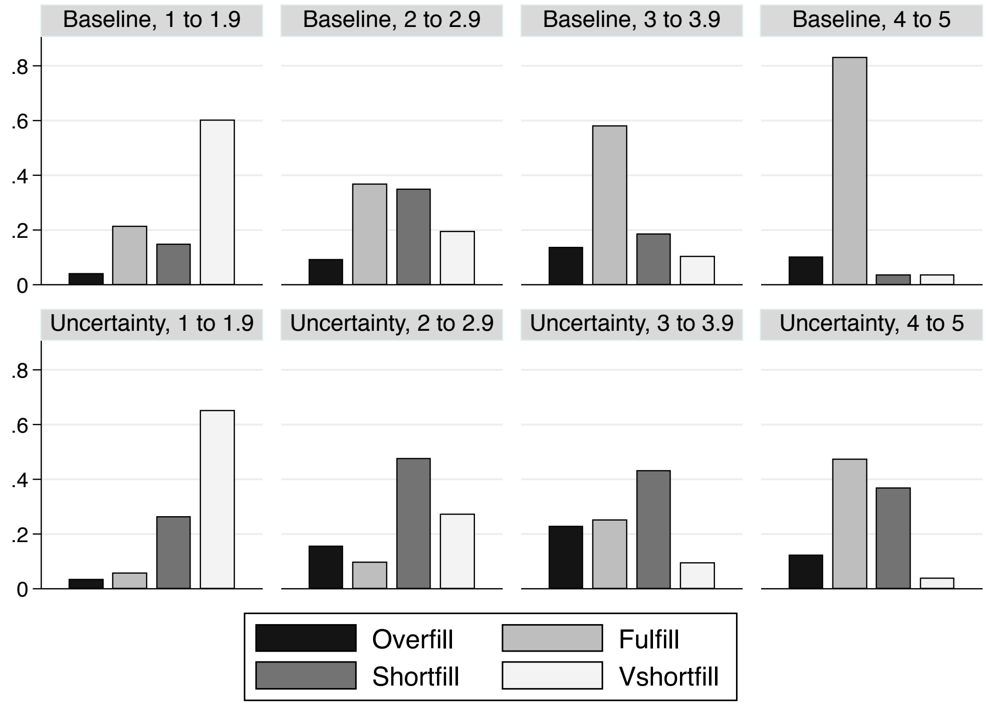


Figure 5: Frequency of fill ratios shipped for different levels of seller feedback averages, by treatment.

Figure 5 also provides insight into the nature of Overfullfillment. We might have thought that those sellers most likely to overfill would be those with the highest average 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. However, this strategy does not work under uncertainty since the rating frequency is lower for Overfill than for Fulfill.

3.4 Buyer use of the feedback system

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		
	Mean	SD	Mean	SD	*p*-value
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, the fraction of subjects who expected more than they received (% Overexpectation) increases to 34%, significantly higher than 21% in the Baseline treatment. Hence, buyers do not make worse predictions in general under uncertainty, but they have too high expectations 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 average and announced quality with a treatment dummy (c.f. Table 13 in Appendix A).[12] As one would expect a better average feedback and higher quality announcement significantly increase buyers' expectations in both 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 average has no different effect under uncertainty indicates that buyers fail to account for lenient ratings when looking at feedback averages 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 average and the announced quality

[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%).

[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.

have a significant positive effect on the final price (c.f. Model 1 Table 14 in 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 average 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 not very different 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	p-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 between Fulfill and Shortfill categories. 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 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, we find clear evidence for 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 6). 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 6 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.

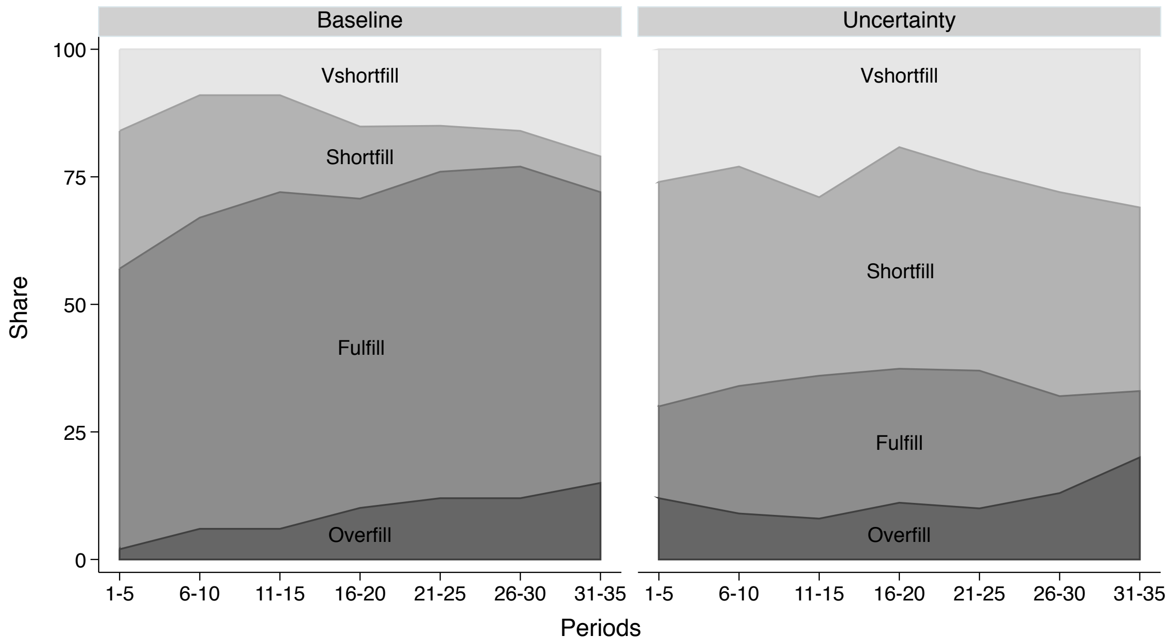


Figure 6: Share of sellers within the four fill ratio categories, by periods. Each data point represents the average share within this category in the respective five periods.

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 average 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 average. 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 average and thus the average increases (decreases) in the following period. Model 1 shows that feedback ratings below the current feedback average 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 already saw in Figure 6: in the Uncertainty treatments deceiving sellers in general are less likely to become more trustworthy. Furthermore, the interaction effect between the treatment dummy and the decrease in feedback average 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 average this is also only significant in the Baseline but not in the Uncertainty treatment.

Honest in next period	**Model 1**	**Model 2**
Uncertainty	-0.562^{**}	-0.381^{*}
	(-3.049)	(-2.038)
Change in feedback average	-0.235^{**}	-0.409^{***}
	(-2.921)	(-3.641)
Feedback average	0.132	0.098
	(1.574)	(1.187)
Shipped / announced	0.405	0.440
	(1.496)	(1.635)
Seller profit	-0.002	-0.003
	(-1.091)	(-1.233)
Period	-0.011	-0.014
	(-1.239)	(-1.534)
Change X uncertainty		0.294^{*}
		(2.299)
Intercept	-0.936^{**}	-0.949^{**}
	(-3.103)	(-3.211)
N	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; $^{+}$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

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) average feedback 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.

Seller profit	**Model 1**	**Model 2**
Uncertainty	0.839	8.087
	(0.194)	(1.565)
Honest in last period	8.735***	15.752***
	(3.444)	(4.392)
Announced	0.772***	0.765***
	(7.577)	(7.517)
Shipped	-0.567***	-0.569***
	(-13.151)	(-13.255)
Period	0.380***	0.362**
	(3.306)	(3.157)
Honest X uncertainty		-13.194**
		(-2.760)
Intercept	8.741	4.794
	(0.910)	(0.493)
N	1191	1191
Log likelihood	-5637.9	-5634.1

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 effect in Model 2, average profits of a previously deceiving seller under certainty are 36.6 ECU whereas a previously honest seller in the same treatment earns 52.4 ECU. This is an increase of more than 40%. In contrast this honesty premium is much smaller in the Uncertainty treatment: earnings increase only by 5.8% from 44.7 to 47.3 ECU. Overall these results indicate that uncertainty about seller intentions seriously hampers the functioning of the feedback system and thus incentives for truthful seller behavior are no longer given.

3.8 Market performance

Finally, we analyze how the feedback system affects market performance. Descriptive statistics in Table 10 show that efficiency decreases from 68% to 58% under uncertainty.[14]

[14] Efficiency is measured as the ratio of realized and maximum possible surplus. 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.

The efficiency losses in both treatments are mainly due to sellers shipping less than maximum quality.[15] In addition, 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).

The biggest change across marketplaces has to do with market surplus captured by buyers. Comparing the two treatments, buyer profits significantly decrease by 31% from 48 ECU in the Baseline treatment to 33 ECU in the Uncertainty treatment (c.f. Table 11). In the same respect, seller profits increase by 7%. Taken together, while under certainty sellers and buyers on average receive almost identical shares of the total profit, the marketplace with uncertainty disadvantages buyers since their share accounts only for 38%.

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

Table 10: Descriptive statistics on market performance 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).

	Efficiency	**Shipped**	**Seller profit**	**Buyer profit**
Uncertainty	-0.130^{+}	-12.307^{+}	3.666^{+}	-14.652^{*}
	(-1.747)	(-1.888)	(1.693)	(-2.067)
Announced	0.006^{***}	0.539^{***}	0.084	0.211
	(6.390)	(6.634)	(0.953)	(1.159)
Period	-0.001	-0.289^{***}	0.691^{***}	-0.322^{+}
	(-0.669)	(-3.294)	(6.914)	(-1.767)
Intercept	0.178^{+}	38.340^{***}	29.478^{***}	34.954^{*}
	(1.812)	(4.601)	(3.778)	(2.119)
N	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; $^{+}$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

[15] 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.

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. We find that 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 43% in a seller's profits in the following period. However, this is no longer the case when uncertainty disguises intentions. Here, the honesty premium accounts only for an insignificant increase of around 6%. 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 (see Bereby-Meyer and Roth 2006 for an example in the context of cooperative games). A recent paper by Ockenfels and Selten (2014) which provides a model for this behavioral principle and applies it to data obtained from games that require players to forecast product demand.

There is reason to believe that our results underestimate the true magnitude of feedback compression in field marketplaces. In a typical transaction, the trading partners know each other's names, addresses and bank details and have exchanged various email messages. This social communication can lead to a feeling of empathy (Andreoni and Rao 2011), obligation (Malmendier and Schmidt 2012), or social pressure (Malmendier, te Velde, and Weber 2014). Reduced social distance also can increase reciprocal behavior (Hoffman, McCabe, and Smith 1996). Hence, it is conceivable that in real world interactions, the reporting rate of negative experiences is further decreased by social communication and closeness. Given that participants in our experiment did not receive any personal information about each other and had no means of communicating, our study likely provides a lower bound for the effect of leniency in real world marketplaces where social distance is reduced by various forms of communication.

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 culpability. 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 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 culpability. (See Samak (2013) for some ideas on how to handle the problem of rating over heterogeneous good categories.)

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 boxplot in Figure 7 summarizes the distribution 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 7 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 Woods 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.

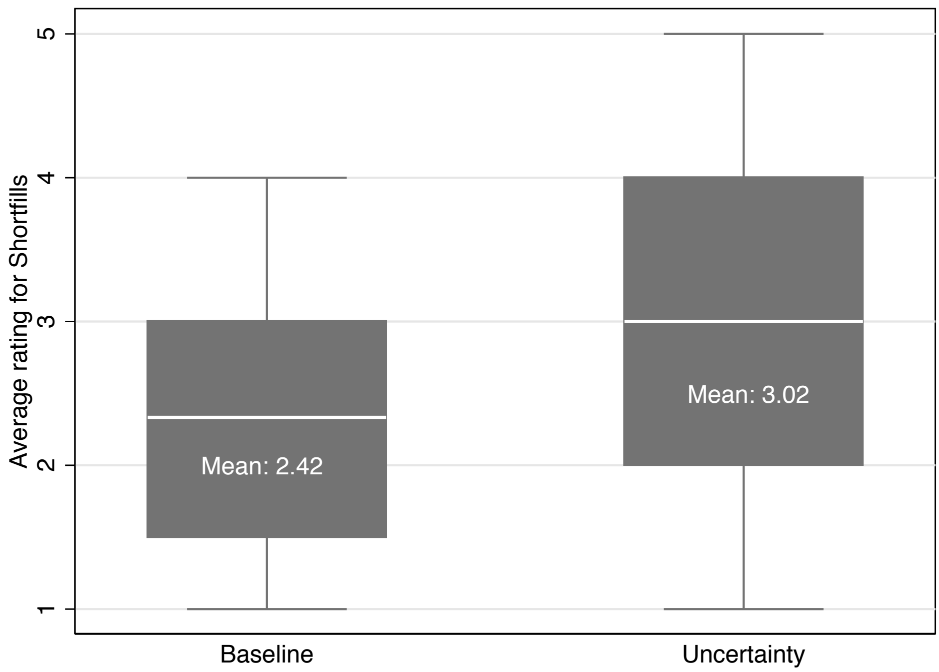


Figure 7: Boxplots for the distribution of individual average ratings for Shortfills in both treatments.

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 intentions 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 (Grosskopf and Sarin 2010; Dulleck, Kerschbamer, and Sutter 2011; Mimra, Rasch, and Waibel 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 ratemymd.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 average	0.630***	0.494***	0.732***	0.626***
	(9.824)	(8.329)	(8.655)	(9.260)
# Feedbacks			-0.088	-0.028
			(-1.304)	(-0.478)
Announced			-0.023*	-0.038***
			(-2.473)	(-4.872)
Period			0.015	-0.002
			(0.922)	(-0.183)
Intercept	-1.471***	-1.838***	0.301	1.171
	(-8.165)	(-8.892)	(0.365)	(1.889)
N	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	0.074	8.591
	(-0.212)	(0.025)	(1.413)
Feedback average	5.344^{***}	5.413^{***}	5.370^{***}
	(16.783)	(12.286)	(16.873)
Announced quality	0.206^{***}	0.205^{***}	0.253^{***}
	(6.217)	(6.197)	(5.699)
Period	-0.025	-0.026	-0.025
	(-0.632)	(-0.650)	(-0.628)
Feedback average X uncertainty		-0.162	
		(-0.255)	
Announced X uncertainty			-0.104
			(-1.572)
Intercept	38.758^{***}	38.560^{***}	34.566^{***}
	(11.538)	(10.832)	(8.147)
N	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	-7.615	40.107
	(-0.306)	(-0.892)	(1.832)
Feedback average	10.913^{***}	10.034^{***}	11.004^{***}
	(10.124)	(6.806)	(10.221)
Announced	0.724^{***}	0.723^{***}	0.923^{***}
	(6.016)	(6.002)	(5.844)
Period	0.431^{***}	0.440^{***}	0.430^{***}
	(3.412)	(3.472)	(3.411)
Feedback average X uncertainty		1.838	
		(0.871)	
Announced X uncertainty			-0.474
			(-1.944)
Intercept	12.814	15.990	-4.968
	(1.098)	(1.309)	(-0.336)
N	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^{+}	-12.307^{+}	28.977^{+}
	(-1.688)	(-1.888)	(1.923)
Period	-0.206^{*}	-0.289^{***}	-0.283^{**}
	(-2.350)	(-3.294)	(-3.229)
Announced		0.539^{***}	0.774^{***}
		(6.634)	(6.844)
Announced X uncertainty			-0.480^{**}
			(-3.006)
Intercept	83.348^{***}	38.340^{***}	17.983^{+}
	(15.117)	(4.601)	(1.686)
N	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	-0.130^{+}	0.059
	(-1.572)	(-1.747)	(0.329)
Period	0.000	-0.001	-0.001
	(0.200)	(-0.669)	(-0.639)
Announced		0.006^{***}	0.007^{***}
		(6.390)	(5.412)
Announced X uncertainty			-0.002
			(-1.152)
Intercept	0.697^{***}	0.178^{+}	0.086
	(11.056)	(1.812)	(0.679)
N	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 – 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.

6.1.1 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.

6.1.2 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.

6.1.3 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:

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

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

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

6.1.4 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.

6.1.5 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.

6.1.6 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.

6.1.7 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:

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

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

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

6.1.8 Feedback stage

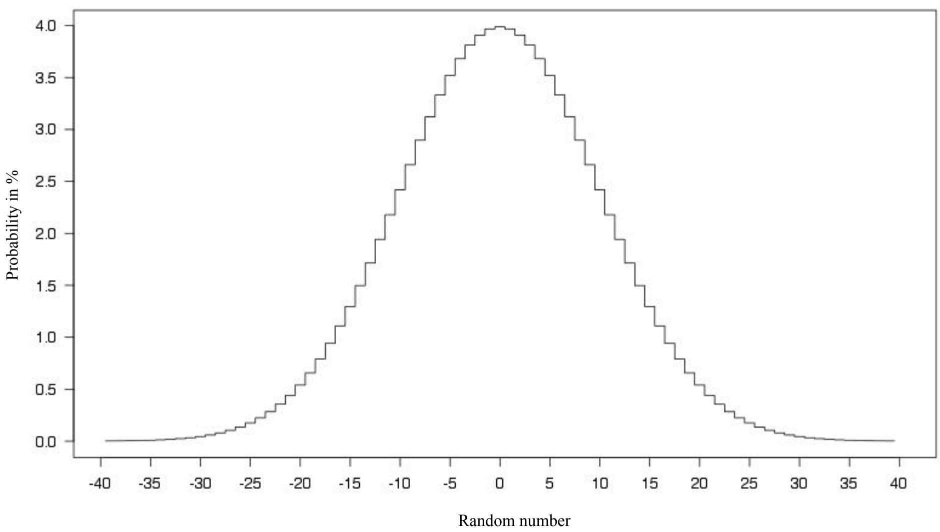
After the Transaction stage the winning bidder decides whether or not he/she wants to submit a feedback on the transaction. Submitting a feedback costs 1 ECU. The feedback rating allows the winning bidder to give feedback on the following scale:

"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.

7 BIBLIOGRAPHY

Andreoni, James, and Justin M. Rao. 2011. "The Power of Asking: How Communication Affects Selfishness, Empathy, and Altruism." *Journal of Public Economics* 95 (7-8): 513–520.

Ba, Sulin. 2001. "Establishing Online Trust through a Community Responsibility System." *Decision Support Systems* 31 (3): 323–336.

Ba, Sulin, and Paul A. Pavlou. 2002. "Evidence of the Effect of Trust Building Technology in Electronic Markets: Price Premiums and Buyer Behavior." *MIS Quarterly* 26 (3): 243–268.

Bajari, Patrick, and Ali Hortacsu. 2003. "The Winner's Curse, Reserve Prices, and Endogenous Entry: Empirical Insights from eBay Auctions." *The RAND Journal of Economics* 34 (2): 329–355.

Bar-Isaac, Heski, and Steven Tadelis. 2008. "Seller Reputation." *Foundations and Trends in Microeconomics* 4 (4): 273–351.

Bauerly, Ronald J. 2009. "Online Auction Fraud and eBay." *Marketing Management Journal* 19 (1): 133–143.

Bénabou, Roland, and Jean Tirole. 2011. "Identity, Morals, and Taboos: Beliefs as Assets." *Quarterly Journal of Economics* 126 (2011): 805–855.

Bereby-Meyer, Yoella, and Alvin E. Roth. 2006. "The Speed of Learning in Noisy Games: Partial Reinforcement and the Sustainability of Cooperation." *American Economic Review* 96 (1995): 1029–1042.

Bigelow, John (ed.). 1904. *The Works of Benjamin Franklin*.

Bol, Jasmijn C. 2011. "The Determinants and Performance Effects of Managers' Performance Evaluation Biases." *Accounting Review* 86 (5): 1549–1575.

Bolton, Gary E, Ben Greiner, and Axel Ockenfels. 2013. "Engineering Trust - Reciprocity in the Production of Reputation Information." *Management Science* 59 (2): 265–285.

Bretz, Robert D, George T Milkovich, and Walter Read. 1992. "The Current State of Performance Appraisal Research and Practice: Concerns, Directions, and Implications." *Journal of Management* 18 (2): 321–352.

Cabral, Luis, and Ali Hortacsu. 2010. "The Dynamics of Seller Reputation: Evidence from eBay." *Journal of Industrial Economics* 58 (1): 54–78.

Dana, Jason, Roberto Weber, and Jason Kuang. 2007. "Exploiting Moral Wiggle Room: Experiments Demonstrating an Illusory Preference for Fairness." *Economic Theory* 33 (1): 67–80.

Dellarocas, Chrysanthos. 2001. "Analyzing the Economic Efficiency of eBay-like Online Reputation Reporting Mechanisms." In *Proceedings of the 3rd ACM Conference on Electronic Commerce*, 171–179. EC '01. New York, NY, USA: ACM.

———. 2003. "The Digitization of Word of Mouth: Promise and Challenges of Online Feedback Mechanisms." *Management Science* 49 (10): 1407–1424.

Dellarocas, Chrysanthos, and Charles A Wood. 2008. "The Sound of Silence in Online Feedback: Estimating Trading Risks in the Presence of Reporting Bias." *Management Science* 54 (3) (March): 460–476.

Dulleck, By Uwe, Rudolf Kerschbamer, and Matthias Sutter. 2011. "The Economics of Credence Goods: An Experiment on the Role of Liability, Verifiability, Reputation, and Competition" 101 (2): 530–559.

Eaton, David H. 2007. "The Impact of Reputation Timing and Source on Auction Outcomes." *The B. E. Journal of Economic Analysis & Policy* 7 (1).

Ederington, Louis, and Michael Dewally. 2006. "Reputation, Certification, Warranties, and Information as Remedies for Seller-Buyer Information Asymmetries: Lessons from the Online Comic Book Market." *The Journal of Business* 79 (2): 693–729.

Fischbacher, Urs. 2007. "Z-Tree: Zurich Toolbox for Ready-Made Economic Experiments." *Experimental Economics* 10 (2): 171–178.

Ganzach, Yoav, and David H Krantz. 1991. "The Psychology of Moderate Prediction: II. Leniency and Uncertainty." *Organizational Behavior and Human Decision Processes* 48 (2) (April): 169–192.

Gregg, Dawn G, and Judy E Scott. 2006. "The Role of Reputation Systems in Reducing On-Line Auction Fraud." *International Journal of Electronic Commerce* 10 (3): 95–120.

Greif, Avner. 1993. "Contract Enforceability and Economic Institutions in Early Trade: The Maghribi Traders' Coalition." *The American Economic Review* 83 (3): 525–548.

Greiner, Ben. 2004. "The Online Recruitment System ORSEE 2.0 - A Guide for the Organization of Experiments in Economics." Working Paper.

Grosskopf, Brit, and Rajiv Sarin. 2010. "Is Reputation Good or Bad? An Experiment." *American Economic Review* 100 (5): 2187–2204.

Güth, Werner, Steffen Huck, and Peter Ockenfels. 1996. "Two-Level Ultimatum Bargaining with Incomplete Information: An Experimental Study." *The Economic Journal* 106 (436): 593–604.

Hoffman, Elizabeth, Kevin McCabe, and Vernon L Smith. 1996. "Social Distance and Other-Regarding Behavior in Dictator Games." *The American Economic Review* 86 (3): 653–660.

Houser, Daniel, and John Wooders. 2006. "Reputation in Auctions: Theory, and Evidence from eBay." *Journal of Economics & Management Strategy* 15 (2): 353–369.

Hu, Nan, Paul A Pavlou, and Jennifer Zhang. 2006. "Can Online Reviews Reveal a Product's True Quality? Empirical Findings and Analytical Modeling of Online Word-of-Mouth Communication." In *Proceedings of the 7th ACM Conference on Electronic Commerce*, edited by D.Pennock J. Feigenbaum J.Chuang, 324–330. New York: ACM Press.

Hu, Nan, Paul A Pavlou, and Jie Zhang. 2009. "Overcoming the J-Shaped Distribution of Product Reviews." *Communications of the ACM* 52 (10): 144–147.

Jawahar, I M, and Charles R Williams. 1997. "Where All the Children Are above Average: The Performance Appraisal Purpose Effect." *Personnel Psychology* 50 (4): 905–925.

Jin, Ginger Zhe, and Andrew Kato. 2006. "Price, Quality, and Reputation: Evidence from an Online Field Experiment." *The RAND Journal of Economics* 37 (4) (December): 983–1005.

Kauffman, Robert, and Charles Wood. 2006. "Doing Their Bidding: An Empirical Examination of Factors That Affect a Buyer's Utility in Internet Auctions." *Information Technology and Management* 7 (3): 171–190.

Klein, Tobias J, Christian Lambertz, Giancarlo Spagnolo, and Konrad O Stahl. 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 (3): 301–320.

Kreps, David M, and Robert Wilson. 1982. "Reputation and Imperfect Information." *Journal of Economic Theory* 27 (2): 253–279.

Lafky, Jonathan. 2014. "Why Do People Rate? Theory and Evidence on Online Ratings." *Games and Economic Behavior* (March).

Landy, Frank J, and James L Farr. 1980. "Performance Rating." *Psychological Bulletin* 87 (1): 72–107.

Larson, James R. 1986. "Supervisors' Performance Feedback to Subordinates: The Impact of Subordinate Performance Valence and Outcome Dependence." *Organizational Behavior and Human Decision Processes* 37 (3): 391–408.

Levin, Jonathan. 2013. "The Economics of Internet Markets." In *Advances in Economics and Econometrics: Tenth World Congress*, edited by Daron Acemoglu, Manuel Arellano, and Eddie Dekel, 48–75. New York: Cambridge University Press.

Livingston, Jeffrey A. 2005. "How Valuable Is a Good Reputation? A Sample Selection Model of Internet Auctions" 87 (3): 453–465.

Lucking-Reiley, David, Doug Bryan, Naghi Prasad, and Daniel Reeves. 2007. "Pennies from eBay: The Determinants of Price in Online Auctions" 55 (2): 223–233.

Malmendier, Ulrike, and Klaus Schmidt. 2012. "You Owe Me." *NBER Working Paper Series* (18543).

Malmendier, Ulrike, Vera L te Velde, and Roberto Weber. 2014. "Rethinking Reciprocity." *Annual Review of Economics* 6 (1): 849–874.

McDonald, Cynthia G., and V. Carlos Slawson. 2002. "Reputation in an Internet Auction Market." *Economic Inquiry* 40 (3): 633–650.

Melnik, Mikhail I, and James Alm. 2003. "Does a Seller's eCommerce Reputation Matter? Evidence from eBay Auctions." *The Journal of Industrial Economics* 50 (3) (March 27): 337–349.

Milgrom, Paul R, Douglass C North, and Barry R Weingast. 1990. "The Role of Institutions in the Revival of Trade: The Law Merchant, Private Judges, and the Champagne Fairs." *Economics & Politics* 2 (1): 1–23.

Mimra, Wanda, Alexander Rasch, and Christian Waibel. 2013. "Price Competition and Reputation in Credence Goods Markets: Experimental Evidence."

Mitzkewitz, Michael, and Rosemarie Nagel. 1993. "Experimental Results on Ultimatum Games with Incomplete Information." *International Journal of Game Theory* 22 (2): 171–198.

Moers, Frank. 2005. "Discretion and Bias in Performance Evaluation: The Impact of Diversity and Subjectivity." *Accounting, Organizations and Society* 30 (1): 67–80.

Ockenfels, Axel, and Reinhard Selten. 2014. "Impulse Balance in the Newsvendor Game." *Games and Economic Behavior* 86: 237–247.

Prendergast, Canice. 1999. "The Provision of Incentives in Firms." *Journal of Economic Literature* 37 (1): 7–63.

Prendergast, Canice, and Robert Topel. 1993. "Discretion and Bias in Performance Evaluation." *European Economic Review* 37: 355–365.

Reichling, Florian. 2004. "Effects of Reputation Mechanisms on Fraud Prevention in eBay Auction."

Resnick, Paul, and Richard Zeckhauser. 2002. "Trust among Strangers in Internet Transactions: Empirical Analysis of eBay's Reputation System." *Advances in Applied Microeconomics* 11: 127–157.

Resnick, Paul, Richard Zeckhauser, John Swanson, and Kate Lockwood. 2006. "The Value of Reputation on eBay: A Controlled Experiment." *Experimental Economics* 9 (2): 79–101.

Rice, Sarah C. 2012. "Reputation and Uncertainty in Online Markets: An Experimental Study." *Information Systems Research* 23 (2) (June): 436–452.

Saal, Frank E., Ronald G. Downey, and Mary a. Lahey. 1980. “Rating the Ratings: Assessing the Psychometric Quality of Rating Data.” *Psychological Bulletin* 88 (2): 413–428.

Saeedi, Maryam, Zeqian Shen and Neel Sundaresan, "The Value of Feedback: An Analysis of Reputation System," Ohio State University, working paper.

Samak, Anya C. 2013. “An Experimental Study of Reputation with Heterogeneous Goods.” *Decision Support Systems* 54 (2): 1134–1149.

Sharon, Amiel T, and C J Bartlett. 1969. “Effect of Instructional Conditions in Producing Leniency on Two Types of Rating Scales.” *Personnel Psychology* 22 (3): 251–263.

Taylor, K, and Robert J Wherry. 1951. “A Study of Leniency in Two Rating Systems.” *Personnel Psychology* 4 (1): 39–47.

Wilson, Robert. 1985. “Reputation in Games and Markets.” In *Game-Theoretic Models of Bargaining*, edited by Alvin E Roth. Cambridge, UK: Cambridge University Press.

Zervas, Georgios, Davide Proserpio, and John Byers. 2015. “A First Look at Online Reputation on Airbnb, Where Every Stay Is Above Average.” Working Paper.