Trust in Markets
–
Essays on Feedback Systems and Advice Giving

Inauguraldissertation
zur Erlangung des Doktorgrades
der Wirtschafts- und Sozialwissenschaftlichen Fakultät
der Universität zu Köln

2015

vorgelegt von
Diplom-Volkswirt Johannes Mans
aus Lahr im Schwarzwald

Referent: Prof. Dr. Axel Ockenfels

Koreferent: Prof. Dr. Dirk Sliwka

Tag der Promotion: 2.10.2015

to M(m)

Contents

Chapter 1

Introduction

"The advantage of humankind of being able to trust one another penetrates into every crevice and cranny of human life."
(John Stuart Mill 1848)

Trust is the lubricant of economic interactions where uncertainty and ignorance make us vulnerable to opportunistic behavior (Arrow 1974). To this effect, trust explains why we are willing altogether to put some of our resources into the hands of others without knowing with certainty that they will actually behave in our best interest. Following Rousseau et al. (1998), this brief description illustrates two universal characteristics considered in the psychological, sociological as well as economic conceptualization of trust.[1] First, in order for trust to be relevant, a decision situation needs to involve *risk* in the sense that the desired outcome is uncertain. Trust would not be necessary if every contingency could be fixed in a legally binding agreement. Second, this voluntarily taken risk crucially *depends on others' behavior*. Thus, trust implicitly means to rely on another person with the confidence that this will be to our advantage. More specifically, if the trusted person indeed turns out to be trustworthy, we are better off than without placing trust. Vice versa, if our trust is exploited, we are worse off than in the first place

[1] Although the concept of trust has been intensivley studied in many disciplines of the social sciences, it lacks a general, unified definition. Accordingly, Williamson (1993) states that "'trust' is a term with many meanings" (p. 453). As an explanation, McKnight and Chervany (2001) argue that each discipline approaches trust from a different angle through their own 'disciplinary lens' and therefore might fail to take into consideration important aspects outside this domain-specific focus. For an excellent cross-disciplinary overview on the various definitions of trust see Rousseau et al. (1998) and also McKnight and Chervany (2001).

(Fehr 2009). Therefore, placing trust implies a positive assessment of others' trustworthiness (Coleman 1990; Williamson 1993).

On markets, where individuals come together to engage in economic interactions, the risk of being exploited emerges from informational asymmetries between trading partners. For this reason, markets frequently suffer from adverse selection and moral hazard and are limited in their efficiency. In a worst-case scenario, this may even lead to complete market failure (Akerlof 1970). These issues arise in traditional offline markets but even more so on the Internet in online markets where typically anonymous and geographically distant traders interact only once (Resnick and Zeckhauser 2002; Dellarocas 2003). Irrespective of online or offline settings, establishing trust among traders is of crucial importance in order to encourage people to trade despite the possibility of being deceived. In this regard, the central topic of this thesis is to investigate the impact electronic reputation systems can have on creating a trusting and trustworthy market environment.

A frequently used mechanism to establish trust and foster cooperation is reputation building (Kreps and Wilson 1982; Wilson 1985; Milgrom and Roberts 1982; Milgrom, North, and Weingast 1990). In its most basic form, reputation is simply the history of an individual's past behavior and therefore can be used to assess his/her trustworthiness. In this regard, reputation provides a link between past behavior and future expected profits, so that rational individuals not only have to incorporate the short-term consequences but also the long-term effects on their reputation when optimizing their decisions (Wilson 1985). Depending on whether the transaction risk stems from adverse selection or moral hazard, reputation promotes trust among traders in two different ways. In settings of adverse selection, reputation serves as a signaling device by providing information about a trading partner's unknown, innate abilities. In contrast, reputation systems deter moral hazard by acting as a sanctioning device. Here, untrustworthy behavior is punished when traders refuse to do business with those who have a bad reputation. Such a punishment mechanism deters opportunistic behavior if the short-term profits from cheating are lower than the long-term profits from cooperation.

Reputation information can either be based on repeated interactions so that traders establish a common trading history or on a publicly available, exact summary of one's past actions.[2] An alternative and more indirect way to build up a reputation is word-to-mouth. Here, self-reported experiences of previous interaction partners provide the reputation information for the assessment of a trader's trustworthiness. Such third-party information is an ancient concept which, for example, has already been used in the eleventh century by Mediterranean merchants to control their overseas trading agents (Greif 1993). The digital counterpart to such traditional word-of-mouth are community-based feedback systems which enable members to publicly report any kind of (mis)conduct within the community (Ba 2001; Resnick and Zeckhauser 2002; Dellarocas 2003).

For the evolution of online market platforms, these user-based feedback systems have been – and still are – of crucial importance. Dellarocas (2003) argues that many of these marketplaces would not have come into existence without the implementation of such feedback systems. Meanwhile, the scalability of electronic reputation systems provides online markets with a competitive advantage compared to offline markets because the provision, storage, aggregation, and dissemination of reputation information are much easier online than offline (Bolton, Katok, and Ockenfels 2004). This development paved the road for the rise of what is now called the 'peer-to-peer' or 'sharing' economy.[3] In brief, this term subsumes online market platforms that provide people with the opportunity to recirculate, rent, and share their goods and to exchange services and expertise (Albinsson and Perera 2012). The origins of this economy go back to the mid 1990s when eBay started its business as an online marketplace for consumer-to-consumer and business-to-consumer trade.[4] By now, with more than 150 million user

---

[2] Bolton, Katok, and Ockenfels (2004), for example, show in an experimental trust game that reputation based on repeated interaction as well as based on histories of past behavior increases trust, trustworthiness, and transaction efficiency in comparison to a market without any reputation mechanism. However, all these positive effects are larger for repeated interactions than for public histories.

[3] In 2013, the size of the global sharing economy was valued at $26 billion and estimations project an increase to more than $110 billion within the next few years (Botsman and Rogers 2010; Cannon and Summers 2014). Similarly, according to Forbes Magazine, participants on this collaborative consumption generated an additional income of more than $3.5 billion in 2013, increasing by more than 25% within a year (Geron 2013).

[4] Other examples for businesses associated with the sharing economy which rely on feedback systems to establish trust on their marketplace are Airbnb (accommodation rental), Etsy (marketplace for handmade and vintage items), Lyft (carpooling), RelayRides (peer-to-peer car rental), DogVacay (dog-kennel services), TaskRabbit (agency for errands), LendingClub (peer-to-peer money lending), Neighborgoods (rent and share consumer goods), etc. For an introduction to the sharing economy see Botsman and Rogers (2010).

accounts, marketplaces in more than 30 countries and a net revenue of $17.9 billion, the former online pioneer eBay is one of the worldwide leading e-commerce businesses (eBay Enterprise 2015). The success of eBay is often ascribed to its sophisticated and consistently refined reputation system (Dellarocas 2003). Therefore, eBay's feedback system has served as the real-world test bed for a series of economic studies to investigate the effectiveness of online reputations. A consistent finding across studies is that a trader's reputation indeed has a positive effect on the probability to sell the offered product and also on the final price (Ba and Pavlou 2002; Bajari and Hortacsu 2003; Cabral and Hortacsu 2010; Eaton 2007; Ederington and Dewally 2006; Houser and Wooders 2006; Jin and Kato 2006; Livingston 2005; Lucking-Reiley et al. 2007; McDonald and Slawson 2002; Melnik and Alm 2003; Resnick and Zeckhauser 2002; Resnick et al. 2006). Such user-generated feedback is not only decisive for a trader's sales on eBay but also for the success or failure of a product (Chevalier and Mayzlin 2006; Li and Hitt 2008; Zhu and Zhang 2010) or even for a restaurant's customer demand (Luca 2011).

Besides the interpretation of feedback ratings, it is also important to understand how these ratings are given in the first place. There are two major issues with the provision of feedback. First, feedback ratings constitute a public good. Once submitted, they are publicly and free of cost available to all members. For this reason, there is the threat of under-provision because not all users might be willing to invest the time and effort to leave feedback. This public good characteristic can lead to biased reputation information if there is self-selection in the sense that only certain types of traders submit feedback or only specific trading experiences receive a rating. The second issue emerges from the subjective nature of feedback ratings and refers to the truthfulness of ratings. Because transaction experiences are only privately observed, it is impossible to assess how accurate and justified the content of a feedback rating is. Thus, from an engineering point of view, the two major challenges when designing a feedback system are to ensure that sufficient ratings are provided and to induce honest, unbiased ratings (Dellarocas 2003; Miller, Resnick, and Zeckhauser 2005). To provide a sound scientific basis for the further development of feedback mechanisms, recent empirical and experimental studies investigated which social motivations and cognitive constraints affect the provision of feedback and how to design new feedback systems to eliminate reporting biases. For instance, in two-sided feedback systems, i.e. systems where both parties rate each other, the timing of feedback provision and the differences in reporting probabilities when rating first or second suggest that reciprocity and the fear of retaliatory feedback are a major source for biased feedback information (Bolton, Greiner, and Ockenfels 2013; Jian, Mackie-Mason, and Resnick 2010; Klein et al. 2006; Masclet and Pénard 2012). Using structural estimation, Dellarocas and Wood (2008) show that slightly dissatisfying experiences are largely underreported and thereby the actual trading risks cannot be inferred from the sellers' feedback score. One-sided feedback or simultaneous feedback systems can help to alleviate the issue of reciprocity (Bolton, Greiner, and Ockenfels 2013; Masclet and Pénard 2012).[5] Furthermore, results indicate that traders in general follow a brag-and-moan rating scheme, i.e. extreme experiences – negative or positive – are reported more frequently than mediocre outcomes (Anderson 1998; Hu, Pavlou, and Zhang 2006; 2009; Lafky 2014).[6]

Other factors apart from self-selection based on transaction experience also influence traders' willingness to leave feedback ratings. Lafky (2014) demonstrates that rating costs have a negative effect on the provision of feedback, especially for mediocre levels of quality. Furthermore, he is able to disentangle the sanctioning and signaling function of feedback ratings and thereby shows that the provision of feedback is indeed driven by a desire to reward and punish transaction partners for their behavior but also by a concern for advising other traders. In this regard, the provision of movie ratings can be encouraged by making the uniqueness and value of an individual's contribution more salient (Beenen et al. 2004; Ludford et al. 2004). According to Wang (2010), prolific raters are also motivated by concerns about their social image within the online community. Moreover, social comparison processes affect individuals' feedback behavior. Chen et al. (2010) show that information about the median user's rating efforts encourages below median raters but discourages above median raters.

---

[5] Within the last years, eBay gradually implemented a one-sided feedback system. First, in 2007, 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 positve ratings. Empirical studies showed that these modifications not only increase informativeness within the system but also improve seller trustworthiness (Klein et al. 2009; Klein, Lambertz, and Stahl 2013; Ye, Gao, and Viswanathan 2010).

[6] Hu, Pavlou, and Zhang (2006; 2009) also describe this brag-and-moan pattern as a j-shaped distribution of ratings: few negative, almost none mediocre and many positive ratings.

This thesis contributes to the stream of literature on the functioning and effectiveness of reputation systems based on self-generated ratings. More specifically, three of the four chapters focus on the provision and interpretation of feedback ratings. In this regard, the overarching goal of this thesis is to identify further sources for reporting biases, analyze how traders interpret this potentially biased information and, finally, how the interaction of these two factors impairs the main function of feedback systems, namely to foster trust and trustworthiness among traders and thereby to enhance market efficiency. Based on the results of these studies, it is possible to derive implications for the design of real-world reputation systems. The last chapter then takes a slightly different perspective on trust. We analyze how the social relationship between individuals affects the willingness to exploit others' trust by giving misleading, selfish advice about private information.

Besides the focus on trust in markets, the used experimental methodology provides a link between the four studies of this thesis. A crucial advantage of experimental studies over empirical studies is that the actual transaction experience, i.e. the quality of the good or service, can be observed. As previously discussed, the transaction experience is of major importance in the rating process and is associated with potential sources of rating biases. In real-world online settings, however, the experience is private information and – if at all – difficult to verify. For this reason, observational studies conducted on, e.g., eBay suffer from an omitted variable bias. Even restricting on standardized goods cannot eliminate this issue because additional factors, which may also make for a satisfactory transaction experience (such as product description and presentation, communication, shipping time, etc.) are difficult to measure objectively (Resnick et al. 2006). In contrast, the experimental approach provides the opportunity to vary the variable of interest while all other factors, such as transaction experience, can be controlled for or held constant (Falk and Heckman 2009). Overall, this thesis comprises four experimental studies.[7] The following paragraphs give a brief overview on the research questions in each chapter, sketch the experimental design and summarize the central results.

[7] Financial support of the German Research Foundation (DFG) through the Research Unit "*Design & Behavior – Economic Engineering of Firms and Markets*" (FOR 1371) for all four projects presented in this thesis is gratefully acknowledged.

The first project in chapter 2 *"Inflated Reputations – Leniency and Moral Wiggle Room in Trader Feedback Systems"* investigates whether uncertainty about the attribution of transaction problems gives rise to reporting biases.[8] Such uncertainty mainly emerges for two reasons. First, traders often have diverging quality standards and thus, dissatisfying quality can be owed to the buyer's different perception of the item description or to the seller concealing or coloring quality flaws on purpose. Second, there are also factors beyond a seller's control that have a negative effect on the transaction experience. For example, late or non-delivery can be simply due to the parcel service rather than to a deceptive seller. For these reasons, it can be difficult for a buyer to infer a seller's true intentions even after the transaction. From personnel economics it is well known that such uncertainty introduces a so-called leniency bias into performance appraisals. When observed performance is only a noisy signal of the employee's actual effort, managers tend to give the benefit of the doubt and assign more favorable ratings than justified by the observed performance (e.g., Landy and Farr 1980). We hypothesize that this 'in dubio pro reo' approach also extends to market environments when traders rate their transaction partners.

In order to investigate whether this leniency bias also occurs in feedback systems, we use an experimental auction setup in which two buyers compete for a good of endogenous and initially unknown quality offered by a seller. In this environment, we manipulate between treatments whether the quality buyers receive is equal to the quality shipped by the seller or whether a distortion factor randomly increases or decreases the shipped quality. Our results provide strong evidence that under uncertainty buyers give sellers the benefit of the doubt and leave more lenient ratings for less than advertised quality. Furthermore, buyers are in general less likely to leave feedback under uncertainty when the received quality differs from the previously announced quality. These reporting biases reduce the informativeness of the feedback system, and thus make it more difficult for buyers to distinguish trustworthy from fraudulent sellers. This

---

[8] The study in chapter 2 is joint work with Gary E. Bolton and David J. Kusterer and is based on Bolton, Kusterer, and Mans (2014). All authors were equally involved in generating the idea. David Kusterer and I programmed and conducted the experiment. Statistical analyses were carried out by David Kusterer and myself receiving feedback from Gary Bolton. We all contributed equally to the writing of this draft. We thank Bill Neilson, Axel Ockenfels, Dirk Sliwka, Ben Greiner, Peter Werner, Christoph Feldhaus, Nicolas Fugger, Florian Goessl and participants at the ESA European Conference 2012 in Cologne and the ESA World Meeting 2013 in Zurich for helpful comments and suggestions.

in turn diminishes the incentives for trustworthy behavior. As such, sellers exploit the fact that exogenous distortions disguise their true efforts and deceive buyers to a larger extent under uncertainty. Overall, the increase in deceptive behavior leads to welfare losses, which are borne solely by the buyers. In summary, our findings highlight that leniency due to uncertainty substantially decreases the effectiveness of reputation systems in terms of informing other buyers but also in terms of sanctioning seller behavior. Eliminating potential sources of uncertainty should therefore be an important goal of online market platforms.

The third chapter *"Long vs. Short-Memory Feedback Systems"* examines whether traders endogenously adjust how to give and interpret ratings to specific features of the feedback system in place.[9] More specifically, we compare two feedback systems which differ only in the number of ratings displayed to the buyer(s). While the first system includes all previously submitted ratings, the second only reveals the most recent feedback rating. We compare these systems in an adverse selection experiment without strategic interaction where buyers trade with two different types of computerized sellers. Here, feedback ratings serve as signals and help buyers to optimize their purchase decisions by improving their beliefs about sellers' types.

We argue that theoretically either of the two systems can provide buyers with the same amount of information about seller types. This requires, however, that buyers in each case coordinate on a specific feedback language, i.e. on a homogeneous way to give and interpret feedback. In the feedback system that displays all previous ratings of a seller (long-memory), it is required that buyers report the received quality. Based on the full history of ratings, subsequent buyers are able to form a belief about the seller's type. In case only the most recent feedback rating is displayed (short-memory), this reporting feedback language cannot be fully informative because only the most recent signal about a seller's type would be passed on. Instead, buyers need to update their belief according to the received quality and submit the resulting posterior belief as feedback

[9] The study in chapter 3 is joint work with Gary E. Bolton, David J. Kusterer, and Axel Ockenfels. All authors were equally involved in generating the idea. David Kusterer and I programmed and conducted the experiment. Statistical analyses were carried out by myself receiving feedback from all other authors. The current draft was written by myself with comments from David Kusterer. We thank Ben Greiner, Mattia Nardotto, Peter Werner, Christoph Feldhaus, Nicolas Fugger, Florian Goessl and Christopher Zeppenfeld for helpful comments and suggestions. We are also thankful to Kevin Breuer for his help with programming, testing and organizing the experiments.

rating. The following buyer then uses this rating as his prior belief. Hence, when buyers coordinate on the appropriate feedback language, both systems provide the same information and assist buyers equally to distinguish between seller types.

However, our results show that subjects are able to coordinate only on the appropriate feedback language when all previous feedbacks are displayed. In both systems, many feedback ratings are equal to the received quality, which is only fully informative in the long-memory feedback system. Transmitting posterior beliefs via feedback ratings is rarely used when only the most recent rating is available. For this reason, informativeness of ratings is lower in the short-memory feedback system and buyers struggle to identify good sellers. Hence, buyers under the short-memory feedback system forego potentially beneficial trades with good sellers and, overall, have significantly lower profits than their counterparts with full feedback history.

Based on these findings, we derive that long-memory outperform short-memory feedback systems in terms of provided information and trading efficiency. A potential explanation is that in our setup the appropriate feedback language of a long-memory system is more intuitive and also less prone to rating errors. Furthermore, our results emphasize how important it is that raters use a consistent feedback language, in terms of feedback giving as well as feedback interpretation.

The starting point for the study *"The Influence of Social Identity and Trading Frequency on the Provision of Feedback"* presented in the fourth chapter is the empirical observation that in international transactions on eBay the probability to give feedback is substantially lower than in national transactions.[10] Feedback provision decreases from 72% to 50% for buyers and sellers, and thus the share of auctions with no rating is twice as large in international than in national trades (44% vs. 22%). In order to find explanations for this feedback difference, we analyze experimentally whether social identity considerations and individual trading frequency have an influence on the provision of feedback. We argue that social identity plays a role

---

[10] The study in chapter 4 is single-authored. I thank Axel Ockenfels and Gary E. Bolton for their significant contributions to the initial idea, the experimental design and the hypotheses. Also, I would like to thank Gary E. Bolton for his hospitality while conducting the experiments at the University of Texas at Dallas. In addition, I received very helpful comments from Bettina Rockenbach, Ben Greiner, Mattia Nardotto, Peter Werner, Christoph Feldhaus, Florian Goessl, David Kusterer, Anne Schielke, Tobias Stangl. I am also thankful to Michael Cristescu, Owen Ma and Oliver Baker for their help with programming, testing and organizing the experiments.

because in national transactions trading partners are perceived as ingroup members while in international trades trading partners are perceived as outgroup members. In line with recent findings from economics and social psychology (e.g., Chen and Li 2009), we expect that a shared social identity strengthens the reward but weakens the punishment motive in the provision of feedback. Accordingly, in ingroup transactions, satisfied traders are more likely to leave a feedback rating while unsatisfied buyers remain silent more often. Trading frequency refers to how often an individual interacts on a particular market. This market participation relates to the public good nature of feedback ratings. Traders with a high trading frequency profit more from feedback information and disciplined trading partners on their frequently visited (home) market. They benefit more from the feedback public good and therefore also are more concerned to establish a norm of feedback provision among traders on their home market. Trading frequency might thus help to explain the observed feedback differences because, on eBay, buyers and sellers typically trade more often on their national than on an international market.

In order to test the influence of social identity and trading frequency on feedback provision, we use a repeated trust game with feedback option for the buyer (trustor). Sessions are conducted simultaneously in laboratories at the University of Cologne and the University of Texas at Dallas. To manipulate social identity, we match subjects within and across universities. Moreover, we create two separate markets to vary individuals' trading frequency. While sellers always stay on one market, buyers switch between markets. Thereby, each buyer is assigned a home market, where he conducts most of his trades and an away market, which he visits less frequently. Overall, we have a 2x2 within-subject design, which allows separating the effects of social identity and trading frequency on the provision of feedback.

The results show that a shared social identity increases the probability that extreme – positive or negative – transaction experiences are reported. In contrast, buyers' trading frequency on a particular market does not seem to affect the willingness to leave a feedback rating. The content of the feedback rating depends only on the transaction experience and is not influenced by social identity or trading frequency. Also trust and trustworthiness do not differ between ingroup and outgroup transactions.

Taken together, these findings suggest that social identity might be one reason why we observe more feedback ratings in national than in international transactions on eBay. Thus, in online marketplaces, emphasizing the community feeling to create a shared social identity among users can be beneficial, because higher feedback provision for extreme outcomes sets proper incentives for trustworthy seller behavior.

The last chapter of this thesis *"Who Do You Lie to? Social Identity and the Costs of Lying"* takes a different perspective on trust in markets.[11] Here, we analyze whether the social relationship between individuals affects the willingness to exploit others' trust by giving false advice. In particular, we examine whether social identity influences the propensity to lie about private information in order to increase one's own profit. Recent experimental studies have shown that the tendency to lie is heterogeneous among individuals (e.g., Gneezy 2005; Sutter 2009; Fischbacher and Föllmi-Heusi 2013). When facing the decision whether to tell a lie or not, people take their own gains from lying as well as the costs of the lie to others into account. Besides monetary consequences, situational (e.g., Gino, Ayal, and Ariely 2009) and contextual (e.g., Lundquist et al. 2009) factors as well as individual characteristics of the potential liar (e.g. Dreber and Johannesson 2008) affect the propensity to lie. However, it has not been studied yet how the relationship between the liar (sender) and the victim of the lie (receiver) influences lying behavior in an economic setting. An important aspect of social relationships is whether people belong to the same social group and thus share a common social identity. Social identity has been shown to influence social preferences and thereby leads to favoritism towards ingroup members (Chen and Li 2009).

We argue that a shared social identity of sender and receiver increases the sender's aversion to lie by raising two types of costs: the *allocative* and the *social* costs of the lie. Allocative costs should be larger in ingroup interactions because social preferences are stronger and thus losses to the receiver are weighted more heavily. In addition to these monetary consequences, senders might be more reluctant to tell a lie to someone who is

[11] The study in chapter 5 is joint work with Christoph Feldhaus and is based on Feldhaus and Mans (2014). Both authors contributed equally to the idea of this project, the design and organization of the classroom experiment, the statistical analyses, and the writing of the draft. We thank Axel Ockenfels, Matthias Sutter, Peter Werner, David Kusterer and Christopher Zeppenfeld for helpful comments and suggestions.

a member of the same social group, because they feel more obliged to live up to social norms (*social* costs of lying) in closer relationships.

In contrast to our hypotheses, our experimental results from a modified three-person sender-receiver game do not provide evidence that social identity affects lying behavior. While across all treatments about half of the participants send a dishonest message, we do not observe differences in lying behavior towards ingroup and outgroup members: neither with respect to allocative nor in terms of social costs. Hence, in our experiment lying behavior is robust to social identity manipulations.

In summary, the four studies highlight the role of trust and trustworthiness for economic interactions, especially on online markets. In this regard, reputation systems based on user-generated feedbacks can help to establish trust and foster cooperation among traders. However, we have shown that the provision as well as the interpretation of feedback information can be biased for several reasons and that these feedback patterns have to be taken into account when designing feedback systems. The following chapters of this thesis present each of the four research studies in detail.

Chapter 2

Inflated Reputations

Leniency and Moral Wiggle Room in Trader Feedback Systems

2.1 Introduction

'Brag-and-moan' is the norm on the Internet.[12] Reviews of trade satisfaction tend to extremes: Very positive and very negative reviews are more frequent than those of more moderate content. This compression in reported opinion raises questions. What factors cause the compression and how does it influence the performance of markets?

Internet feedback mechanisms are a common feature of modern marketplaces. These mechanisms are arguably essential to the existence of electronic trading platforms – e.g., Amazon, eBay, Etsy and Taobao – because online transactions typically take place between anonymous and geographically separated traders who have no common trade history to build upon (e.g., Ba 2001; Resnick and Zeckhauser 2002; Dellarocas 2003). They are increasingly important to non-electronic transactions as well.[13]

The concept behind these markets is literally ancient (Greif 1993): Word-of-mouth provides a link between past behavior and future profits, so that rational individuals

---

[12] This chapter is joint work with Gary E. Bolton and David J. Kusterer. All authors were equally involved in generating the idea. David Kusterer and I programmed and conducted the experiment. Statistical analyses were carried out by David Kusterer and myself receiving feedback from Gary Bolton. We all contributed equally to the writing of this draft. We thank Bill Neilson, Axel Ockenfels, Dirk Sliwka, Ben Greiner, Peter Werner, Christoph Feldhaus, Nicolas Fugger, Florian Goessl and participants at the ESA European Conference 2012 in Cologne and the ESA World Meeting 2013 in Zurich for helpful comments and suggestions. Financial support of the German Research Foundation (DFG) through the Research Unit "*Design & Behavior – Economic Engineering of Firms and Markets*" (FOR 1371) is gratefully acknowledged.

[13] Numerous websites collect ratings for offline services such as Google reviews, tripadvisor.com, yelp.com, ratemymd.com, ratemyprofessors.com, ratemymechanic.us, ratedpeople.com.

need to incorporate not only the short-term consequences but also the long-term effects on their reputation when optimizing their decisions (e.g., Kreps and Wilson 1982; Wilson 1985; Milgrom, North, and Weingast 1990). When successful, reputation systems inform others about issues of adverse selection, e.g., trader experience and professionalism, and moral hazard, e.g., whether a trader ships goods that fit the advertised description. Perhaps the major difference between traditional and electronic versions of the system is that, in the traditional systems, word-of-mouth spreads sequentially, from acquaintance-to-acquaintance, whereas Internet word-of-mouth immediately gets posted to the world. The speed-and-scale difference arguably amplifies the influence reputation has on modern markets. Therefore, Internet word-of-mouth makes trader reputation a bigger factor in market performance than in the past.

The study we present here focuses on the 'brag', the upward compression part of the brag-and-moan phenomenon. While numerous empirical studies find that a seller's reputation has a positive effect on his price premium and also the probability to sell the product,[14] there are a large number of complaints about fraud in online transactions suggesting that feedback profiles draw an overly optimistic picture (Gregg and Scott 2006; Bauerly 2009; Rice 2012). This is not to say that 'moan', downward compression, is unimportant. Lafky (2014), using an innovative experimental design, explores the factors behind this phenomenon.

Previous research work, centered on the eBay marketplace, reveals how strategic manipulation can cause upward compression in two-way feedback systems (i.e., systems where buyer and seller rate one another). Ninety-nine percent of all ratings in the eBay two-way feedback system were positive (Resnick and Zeckhauser 2002; Kauffman and Wood 2006). Using structural estimation, Dellarocas and Wood (2008) provide evidence that the actual risk of a dissatisfying transaction on eBay is significantly larger (21% for buyers and 14% for sellers in rare coin auctions). This is due to the fact that reporting probabilities depend on the level of satisfaction such that satisfied and very unsatisfied traders are more likely to leave feedback than those who are 'mildly' unsatisfied. Additionally, the timing of feedback giving and the differences

[14] See for example Ba and Pavlou (2002), Bajari and Hortacsu (2003), Cabral and Hortacsu (2010), Eaton (2007), Ederington and Dewally (2006), Houser and Wooders (2006), Jin and Kato (2006), Livingston (2005), Lucking-Reiley et al. (2007), McDonald and Slawson (2002), Melnik and Alm (2003), Resnick and Zeckhauser (2002), and Resnick et al. (2006).

in reporting probabilities when feedback is given first or second indicate that dissatisfied buyers are afraid of retaliatory feedback and therefore either prefer to give positive or no feedback at all (Reichling 2004; Dellarocas and Wood 2008; Bolton, Greiner, and Ockenfels 2013). This kind of distortion compresses feedback in the upward direction, and can lead to uncertainty about who is and who is not a trustworthy trade partner and this, in turn, can impede market efficiency (Bolton, Greiner, and Ockenfels 2013).

Due in part to these considerations, in 2007 eBay moved from a two-way feedback system to a one-way system (buyer rates seller), effectively eliminating the possibility of trader retaliation.[15] Nevertheless, in the new system, the mean detailed seller rating is 4.7 on a 5-point scale and more than 75% of the sellers have an average detailed seller rating of larger than 4.5 (Klein et al. 2009). In fact, the brag-and-moan pattern is common to many one-sided feedback systems; for example, Hu, Pavlou, and Zhang (2006) show that product reviews on Amazon often have a j-shaped, bi-modal distribution, indicating that buyers are likely to either brag or moan about a product and tend to leave extreme ratings (also see Hu, Pavlou, and Zhang 2009). There must be additional factors other than retaliation at play that lead to biased feedback ratings.

The experiment we report in this paper examines whether uncertainty about sellers' culpability for a problematic trade can cause feedback to be compressed in an upward direction. Uncertainty about attribution arises in trades mainly for two reasons. First, sellers and buyers often have diverging quality standards. Thus, dissatisfying quality might be due to the different perceptions of the item description or to the seller concealing or coloring quality flaws on purpose. Second, there are also factors that influence quality negatively but are beyond a seller's control, for example late or non-delivery might be simply due to the parcel service rather than a lazy or fraudulent seller. Hence, even after receiving the purchased object, it might be difficult for the buyer to assess the seller's true effort.

There are a number of reasons to believe that buyers, as well as sellers, will react to this attributional uncertainty, in ways that would lead to upward compression in feedback.

[15] 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.

In personnel economics, when supervisors are asked for subjective performance appraisals of their subordinates, observed performance is often only a noisy signal of the employee's actual effort. In this context, it has been shown that under uncertainty managers tend to give favorable ratings that are higher than is justified by the actual performance (Berger, Harbring, and Sliwka 2013; 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 refrain from giving feedback at all (Larson 1986). 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). In a similar vein, 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. There are reasons to believe that this bias is more socially general than employer-employee relationships: Judicial judgments are also prone to errors of either too harsh judgments, when an innocent defendant is found guilty, or too lenient judgments, when guilty individuals get away without punishment. Normative expressions such as "in dubio pro reo" or "innocent until proven guilty" or in the words of Benjamin Franklin: "it is better a hundred guilty persons should escape than one innocent person should suffer" (Bigelow 1904) imply a preference for (misplaced) leniency over wrongful convictions when there is uncertainty about the defendant's guilt. We might then expect this kind of leniency under uncertainty to extend to market environments when buyers evaluate trading partners.

Both forms of leniency – overly positive ratings and the omission of negative ratings – lead to a compression of feedback scores at the upper end of the rating scale and therefore reduce the informativeness of the reputation system. This biased feedback provision and content plausibly hinders discriminating between honest and dishonest sellers and disciplining deceptive sellers and thus provides (moral) wiggle room for self-interested opportunistic behavior. In experiments, ultimatum game proposers take advantage of the receiver's uncertainty or ignorance (Mitzkewitz and Nagel 1993; Güth, Huck, and Ockenfels 1996). In addition to the pecuniary incentives, the theory of moral wiggle room implies that even honest minded sellers will be tested because the uncertainty makes one's actions less informative about one's true nature and so less damaging to one's self-image (Bénabou and Tirole 2011). Accordingly, Dana, Weber, and Kuang (2007) show in a dictator game that, when it is not revealed to recipients whether a dictator himself or a random computer draw made an unfair decision, the fraction of fair shares decreases substantially (also see Ockenfels and Werner 2012).

A critical question is whether traders in the system adjust to changes in the informativeness in the system. Previous work finds that trader behavior can be quite sensitive to changes in feedback informativeness; for example, in two-way systems in which sellers can retaliate negative feedback, the relatively low informativeness is accompanied by lower buyer auction bids, and so lower transaction prices, and lesser seller reliability (Bolton, Greiner, and Ockenfels 2013). On the other hand, environmental uncertainty can slow down learning (Bereby-Meyer and Roth 2006) so it is plausible that a market environment that includes attributional uncertainty may impede traders' ability to find the best course of action.

Because uncertainty about seller culpability is difficult to capture in field data sets, we designed a laboratory auction environment 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 random positive or negative distortion factor with expected value of zero is added and thereby disguises sellers' true effort. An earlier study by Rice (2012) investigates the influence of uncertainty on feedback in the context of a simple trust game, in circumstances where the random distortion of the amount sent by the trustee is always negative and feedback giving is mandatory. The results show that with uncertainty it is less likely for trustees to receive a poor rating conditional on the level of trustworthiness. We investigate an auction environment in which feedback giving is optional and the random distortion can be either negative or positive and is zero in expectation, so that more lenient feedback cannot be attributed to reduced expectations due to the uncertainty. We examine how bidding behavior and market efficiency are affected by uncertainty and reporting biases due to lenient ratings and lower reporting frequency.

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. Overall, this fraudulent behavior significantly decreases buyer profits.

2.2 *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 2.1.

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_{wb} = Valuation × Received quality − Price
	π_s = Price − 100 × Shipped quality	π_{lb} = 0

Table 2.1: 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.[16] 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,

[16] 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.

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[17] – were held constant. At the beginning of the experiment subjects received an endowment of 1,000 ECU. Gains and losses were added to or deducted from this initial endowment.

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

We ran two sessions per treatment with 120 participants in total.[18] 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.

[17] 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.

[18] 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.

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, if 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 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, if 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.

Hypothesis 3:

Under uncertainty, sellers are more likely to send lower quality than announced.

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.

2.3 *RESULTS*

A descriptive look at the data

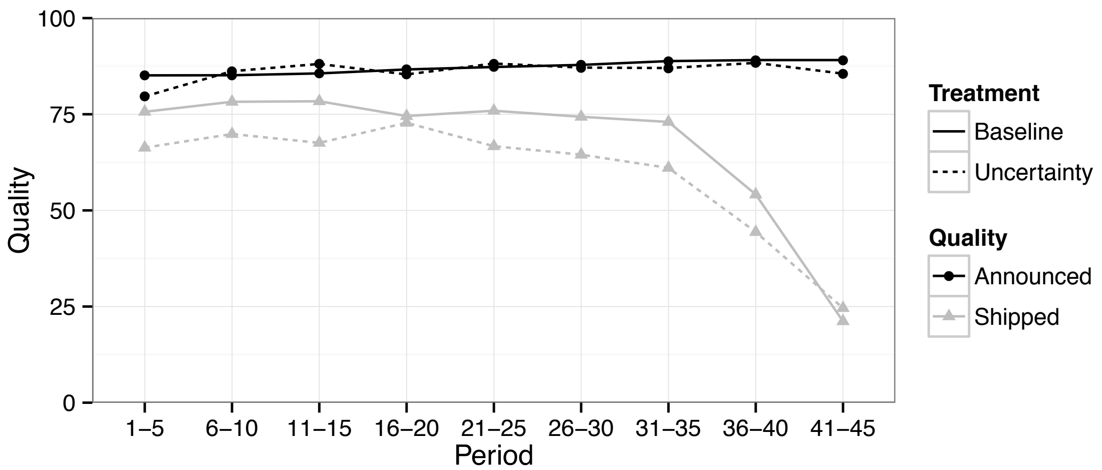


Figure 2.1: Average announced and shipped quality across period intervals.

Figure 2.1 provides a first look at how announced and shipped quality unfold over time and across treatments. There are three observations to make. First, average announced quality is very steady across periods and similar across treatments, at a rate of about

86%. 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.[19] 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.

In what follows below, we will see that the feedback given to quality that falls short of the announcement depends on the extent of the shortfall. Due to the random distortion factor in the Uncertainty treatment, buyer and seller can perceive the fill ratio – ratio of announced-to-shipped quality – differently, so we need to classify by perspective. It will be useful to classify the fill ratio into four categories, stipulated in Table 2.2. *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, uncertainty about seller culpability is relevant because such minor deviations could be caused by the seller as well as by the distortion factor. 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. While results and analyses below are derived using the .8 cutoff, the main conclusions are robust to any cutoff factor in the range .6 to .9.

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 2.2: Classification of fill ratios for sellers and buyers.

Figure 2.2 exhibits histograms of the fill ratio broken down by treatment.[20] 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.

[19] 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.

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

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 0.

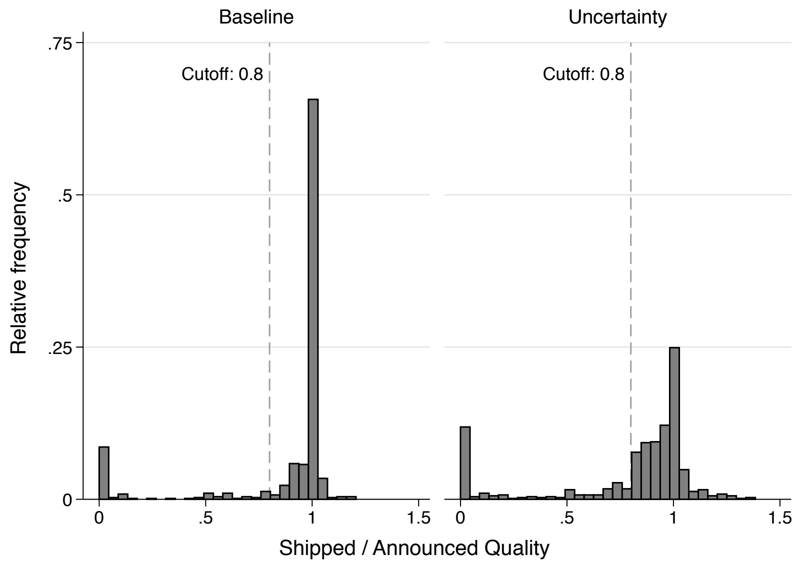


Figure 2.2: Histogram of fill ratios (= shipped / announced quality) across treatments (periods 1-35).

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 2.3 compares the average feedback rating across Baseline and Uncertainty treatments at different fill ratios. Figure 2.3 also shows the share of silent transactions in which buyers did not leave a feedback rating. Also below, Table 2.3 reports the corresponding average feedbacks given and rates of silence by fill category.

Looking first at the feedback given: Figure 2.3 shows that 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 2.3. The random effects Tobit estimates in Table 2.4 control for a number of additional factors, but nevertheless the same results hold.[21]

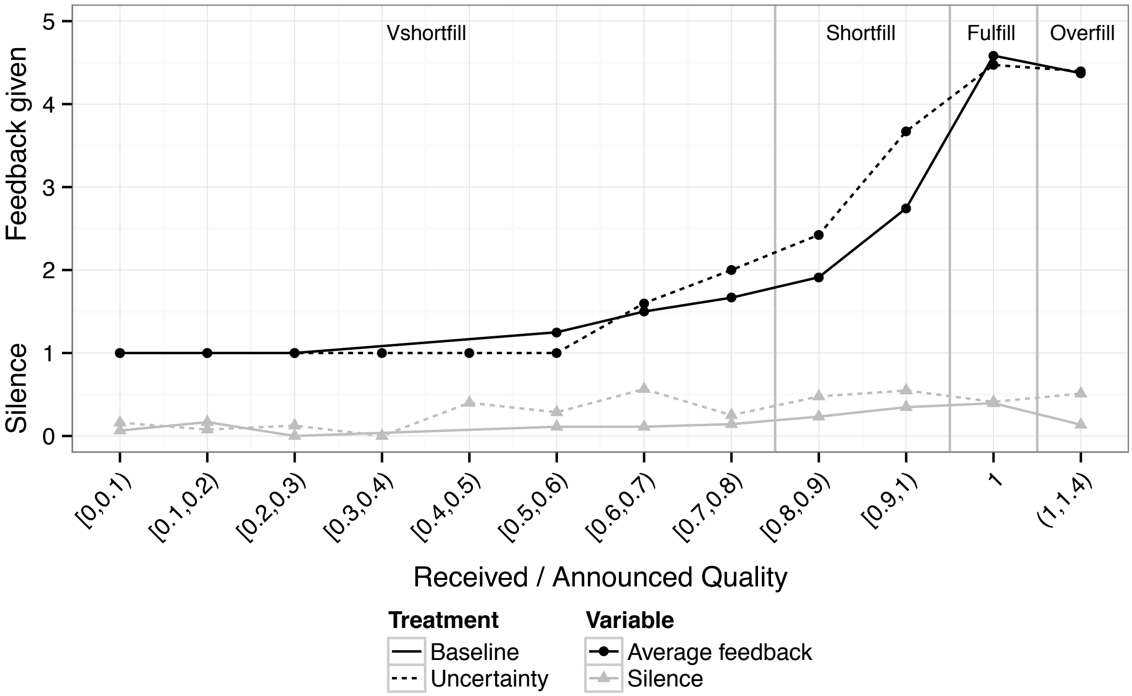


Figure 2.3: Average buyer feedback and frequency of not giving feedback (silence) for periods 1-35.

	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 2.3: 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).

[21] 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. As an additional robustness check we ran the regressions for feedback ratings and silence also using Tobit models with clustering on the matching group level. The results do not change qualitatively (see Table 2.12 and Table 2.13 in Appendix A).

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)
Log likelihood	-121.9	-250.4	-292.2	-60.32
N	119	303	197	186

Table 2.4: 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 2.3 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 2.3 confirms these observations. The random effects Probit models reported in Table 2.5 tell a similar story. The higher frequency of silence in Overfill was not predicted by our hypothesis. One explanation would be that buyers are more likely to remain silent when attribution is uncertain, not just in the case of when the quality falls short. Because positive (as well as negative) shocks can happen to quality, the attribution behind Overfill is ambiguous which may result in greater silence.

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)
Log likelihood	-111.8	-324.6	-232.3	-97.50
N	198	503	363	230

Table 2.5: 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, greater silence extends to Overfills, suggesting that the moral wiggle room hypothesis extends to circumstances where the buyer is pleasantly surprised and seller attribution is in doubt.

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,[22] 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, the announced quality and the current period

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

(c.f. Table 2.14 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 2.4.

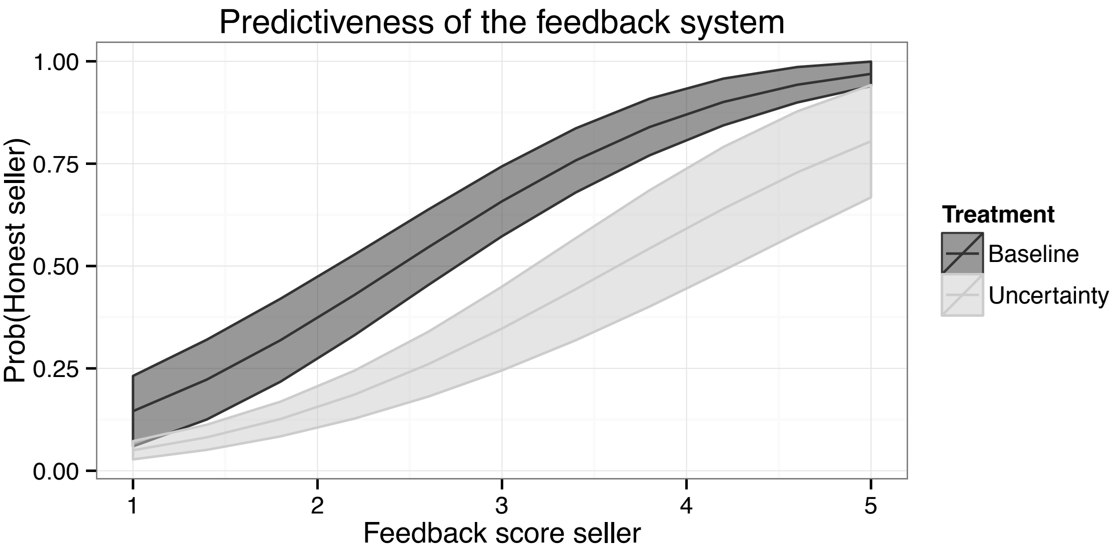


Figure 2.4: Predictions and 95% confidence intervals from Probit regressions for the probability of meeting an honest seller (for the regressions see Table 2.14 in Appendix A, periods 1-35).

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 19 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 2.5 provides an alternative way of looking at the information in Figure 2.4. Here, we can see the deviations from shipping 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.

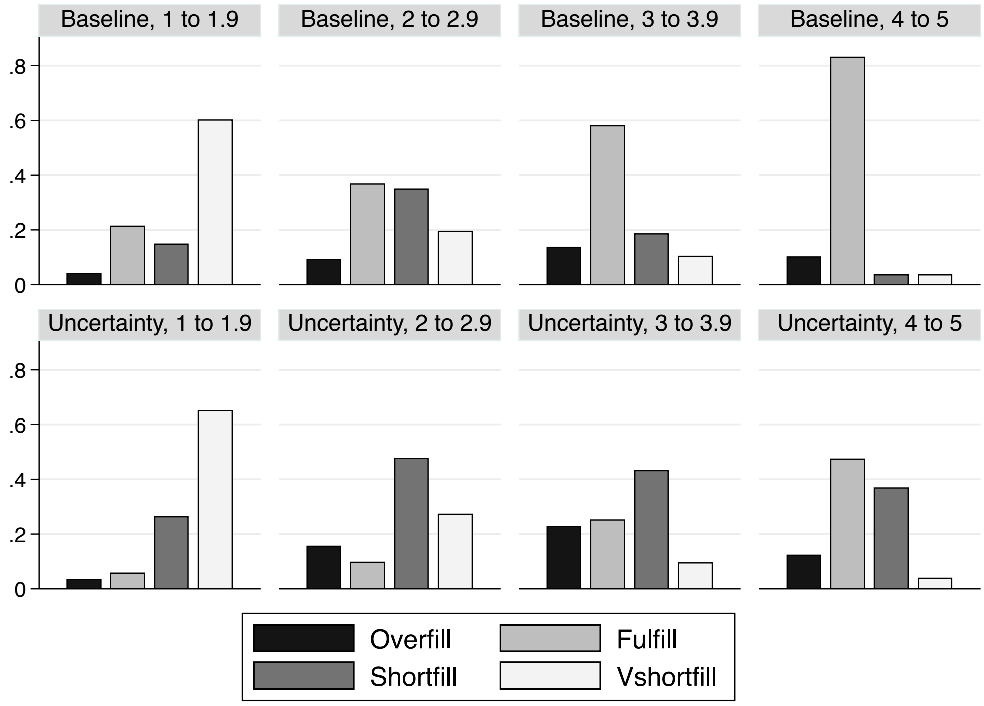


Figure 2.5: Frequency of fill ratios shipped for different levels of seller feedback averages, by treatment. Periods 1-35.

Figure 2.5 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 2.3, 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.

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 2.6 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 2.6: 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 20% 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 2.15 in Appendix A).[23] 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. As a consequence, we observe that prices are also similar across treatments. On average, the final price is 128 ECU in the

[23] 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%).

Baseline treatment and 123 ECU in the Uncertainty treatment.[24] Panel Tobit regressions show that the seller's feedback average and the announced quality have a significant positive effect on the final price (c.f. Model 1 Table 2.16 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.

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 2.7 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 2.7: 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

[24] 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.

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 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 2.6). 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 2.6 shows that this is not the case with the lenient feedback ratings given 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.

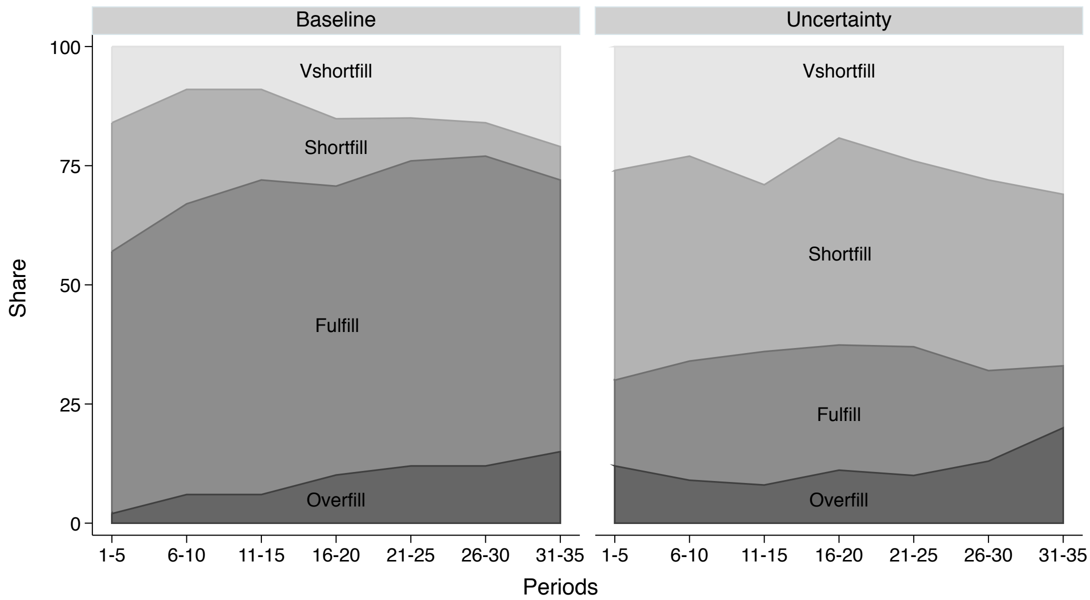


Figure 2.6: Share of sellers within the four fill ratio categories, by periods. Each data point represents the average share within this category in the respective five periods.

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 2.8 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 2.6: in the Uncertainty treatments deceiving sellers in general are less likely to become more trustworthy.

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)
Log likelihood	-198.0	-195.3
N	472	472

Table 2.8: 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; $^{+}$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

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. 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.

Does honesty pay?

As we have seen before, dishonest sellers are more likely to get away with no negative feedback under uncertainty. Thereby, the feedback system might not provide sufficient incentives for honest behavior. In this regard, Model 1 of Table 2.9 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 (Model 2).

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)
Log likelihood	-5637.9	-5634.1
N	1191	1191

Table 2.9: Effect of honest behavior in last period on current seller profits. Random effects Tobit regression with seller profit as dependent variable. A robustness check using Tobit regressions with clustering on the matching group level can be found in Table 2.18. Periods where a seller did not sell his product are excluded. Periods 1-35; *t* statistics in parentheses; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Looking at the marginal effect, 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.

Market performance

Finally, we analyze how the biased feedback system affects market performance. Descriptive statistics in Table 2.10 show that efficiency decreases by 10 percentage points from 68% to 58% under uncertainty.[25] In general, efficiency losses in both treatments are mainly due to sellers shipping less than maximum quality.[26] 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 2.11).

	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 2.10: 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).

[25] 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.

[26] 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.

As a consequence of this reduction in delivered quality, the marketplace under uncertainty is less attractive for 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 2.11). 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%.

	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)
Log likelihood	-830.8	-5720.8	-6418.4	-7265.0
N	1400	1400	1294	1294

Table 2.11: Random effects tobit regressions with different performance measures as dependent variables. Robustness checks using Tobit regressions with clustering on the matching group level can be found in Table 2.19. Periods 1-35; *t* statistics in parentheses; $^{+}$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

2.4 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 lead 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 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 overoptimism 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 2.4. There is a large literature to show that variability in payoff feedback impedes learning about optimal actions (e.g., Bereby-Meyer and Roth 2006); for example, see 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 requires 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.)

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.

2.5 *Appendix A – Supplementary tables*

Feedback rating	**Overfill**	**Fulfill**	**Shortfill**	**Vshortfill**
Uncertainty	0.612	-0.294	1.217**	0.379
	(0.741)	(-0.392)	(3.211)	(0.842)
Received / Announced	-0.468		15.824***	11.137***
	(-0.122)		(7.315)	(5.493)
Announced	0.092*	0.066**	0.009	0.047**
	(2.193)	(2.681)	(0.641)	(3.307)
Price	0.001	0.003	-0.005+	0.027**
	(0.186)	(0.425)	(-1.831)	(3.252)
Period	0.006	0.009	-0.012	-0.103***
	(0.182)	(0.333)	(-0.794)	(-4.251)
Intercept	-2.172	-0.112	-12.437***	-12.789***
	(-0.299)	(-0.056)	(-4.534)	(-5.105)
Log likelihood	-132.0	-278.1	-299.5	-60.32
N	119	303	197	186

Table 2.12: Tobit regressions with submitted feedback ratings (1-5) as dependent variable for each category of seller trustworthiness. Standard errors clustered on matching group level. Periods 1-35; *t* statistics in parentheses; $^{+}$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Silence	**Overfill**	**Fulfill**	**Shortfill**	**Vshortfill**
Uncertainty	1.201***	0.076	0.622**	0.449+
	(3.700)	(0.417)	(2.731)	(1.909)
Announced	-0.026*	-0.015**	-0.012	0.028***
	(-2.040)	(-2.601)	(-1.376)	(3.809)
Received / Announced	-2.301		2.111	1.199**
	(-1.515)		(1.346)	(2.600)
Price	0.001	0.001	-0.006*	-0.016***
	(0.390)	(0.613)	(-2.058)	(-3.542)
Period	0.004	0.008	-0.000	-0.014
	(0.386)	(1.497)	(-0.006)	(-1.472)
Intercept	3.219	0.691	-0.716	-2.090**
	(1.513)	(1.202)	(-0.447)	(-2.584)
Log likelihood	-117.1	-332.3	-237.9	-97.50
N	198	503	363	230

Table 2.13: Probit regressions for silence with 1 = no feedback given and 1 = feedback given. Standard errors clustered on matching group level. Periods 1-35; *t* statistics in parentheses; $^{+}$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

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)
pseudo R^2	0.25	0.16	0.29	0.22
% correctly classified	78.9	71.3	82.9	72.5
N	608	571	608	571

Table 2.14: Probit regression on honest seller. Standard errors clustered on matching group level. 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)
R^2 overall	0.216	0.216	0.219
N	2358	2358	2358

Table 2.15: 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)
Log likelihood	-4315.2	-4314.8	-4313.3
N	1075	1075	1075

Table 2.16: Random effects Tobit regressions with price (100-300) as dependent variable. Periods 1-35; *t* statistics in parentheses; $^{+}$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Honest in next period	**Model 1**	**Model 2**
Uncertainty	-0.525***	-0.362*
	(-3.686)	(-2.351)
Change in feedback average	-0.229+	-0.409***
	(-1.954)	(-4.003)
Feedback average	0.122	0.094
	(1.328)	(1.186)
Shipped / announced	0.388	0.434
	(1.263)	(1.463)
Seller profit	-0.002	-0.003
	(-1.009)	(-1.254)
Period	-0.012	-0.014
	(-1.490)	(-1.585)
Change X uncertainty		0.301*
		(2.217)
Intercept	-0.930***	-0.949***
	(-4.422)	(-4.300)
Log likelihood	-198.4	-195.5
N	472	472

Table 2.17: Dishonest sellers' reaction to feedback. Probit regressions 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. Standard errors clustered on matching group level. Periods 1-35; *t* statistics in parentheses; $^{+}$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Seller Profit	Model 1	Model 2
Uncertainty	1.885	7.402
	(0.466)	(1.377)
Honest in last period	9.584***	14.886***
	(4.096)	(4.488)
Announced	0.780***	0.772***
	(5.669)	(5.376)
Shipped	-0.500***	-0.499***
	(-7.959)	(-7.959)
Period	0.391*	0.378*
	(2.293)	(2.207)
Honest X uncertainty		-9.888*
		(-2.324)
Intercept	2.217	-0.749
	(0.185)	(-0.062)
Log likelihood	-5650.9	-5648.6
N	1191	1191

Table 2.18: Effect of honest behavior in last period on current seller profits. Tobit regressions with seller profit as dependent variable. Periods where a seller did not sell his product are excluded. Standard errors clustered on matching group level. Periods 1-35; *t* statistics in parentheses; $^{+}$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

	Efficiency	Shipped	Buyer profit	Seller profit
Uncertainty	-0.127^{+}	-11.868^{+}	-14.735*	3.681^{+}
	(-1.747)	(-1.867)	(-2.185)	(1.669)
Announced	0.010***	0.841***	0.488*	0.083
	(4.620)	(4.093)	(2.282)	(0.935)
Period	-0.001	-0.344*	-0.327	0.689***
	(-0.631)	(-1.996)	(-1.245)	(4.118)
Intercept	-0.115	12.950	11.540	29.606***
	(-0.642)	(0.771)	(0.609)	(4.235)
Log likelihood	-910.6	-5805.6	-7276.5	-6418.6
N	1400	1400	1294	1294

Table 2.19: Tobit regressions with different performance measures as dependent variable. Standard errors clustered on matching group level. Periods 1-35; *t* statistics in parentheses; $^{+}$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

2.6 *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.

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.

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.

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:

*Seller's Payoff = Auction price – (Quality * 1 ECU)*

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

*Winning Bidder's Payoff = [(Quality / 100) * Valuation] – Auction price*

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.

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.

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.

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:

*Seller's Payoff = Auction price – (shipped Quality * 1 ECU)*

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

*Winning Bidder's Payoff = [(received Quality / 100) * Valuation] – Auction price*

Feedback stage

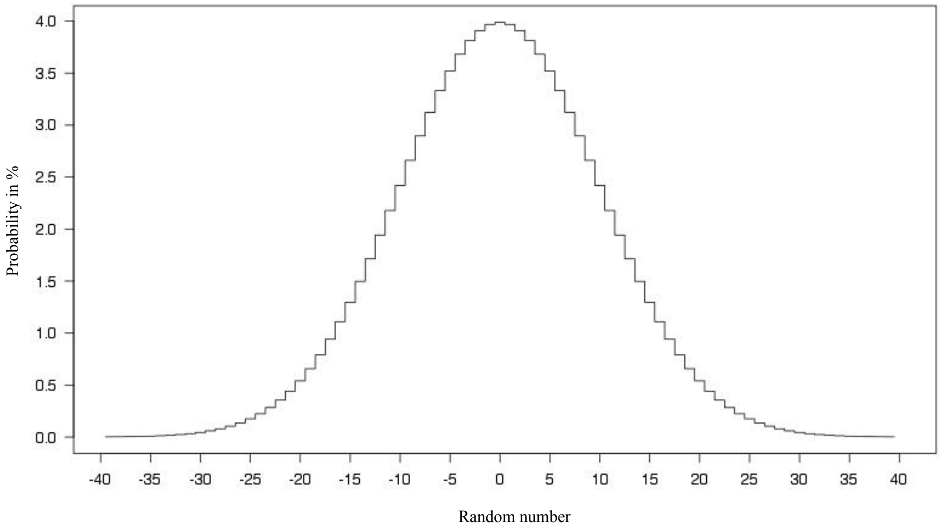
After the Transaction stage the winning bidder decides whether or not he/she wants to submit a feedback on the transaction. Submitting a feedback costs 1 ECU. The feedback rating allows the winning bidder to give feedback on the following scale:

"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 of 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.

2.7 APPENDIX C – SCREENSHOTS

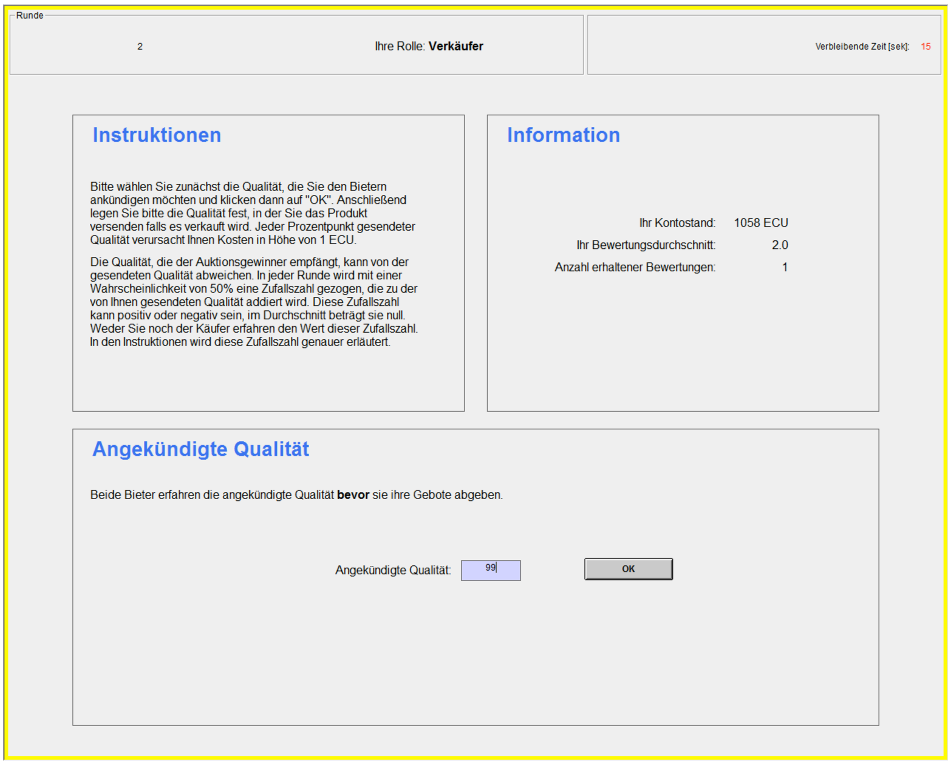


Figure 2.7: Sellers' quality announcement.

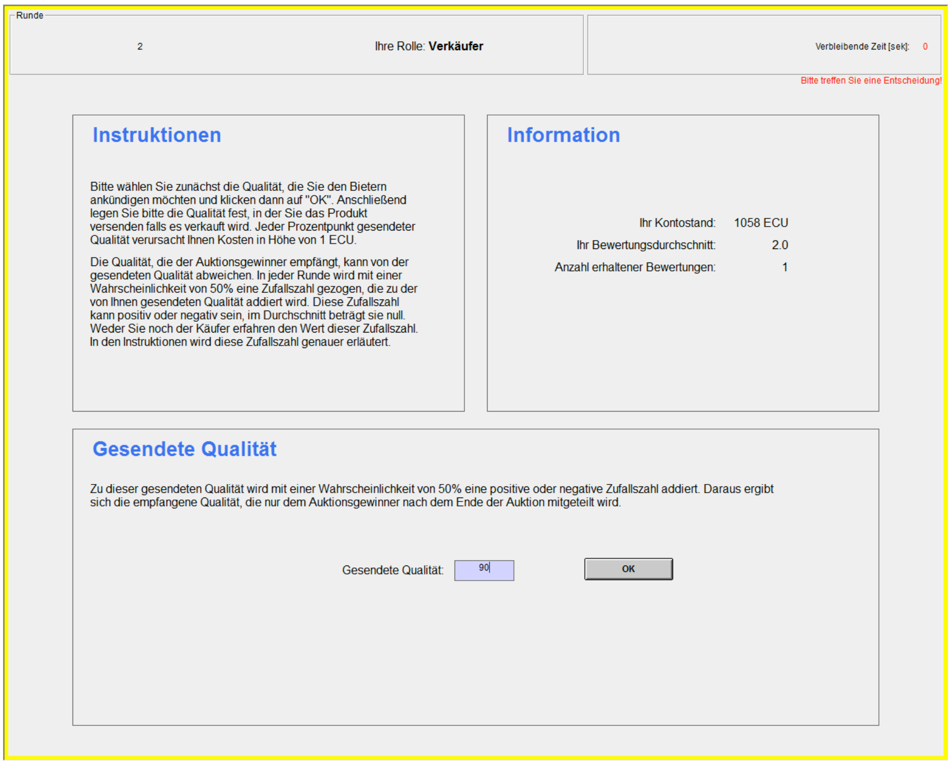


Figure 2.8: Seller's quality decision.

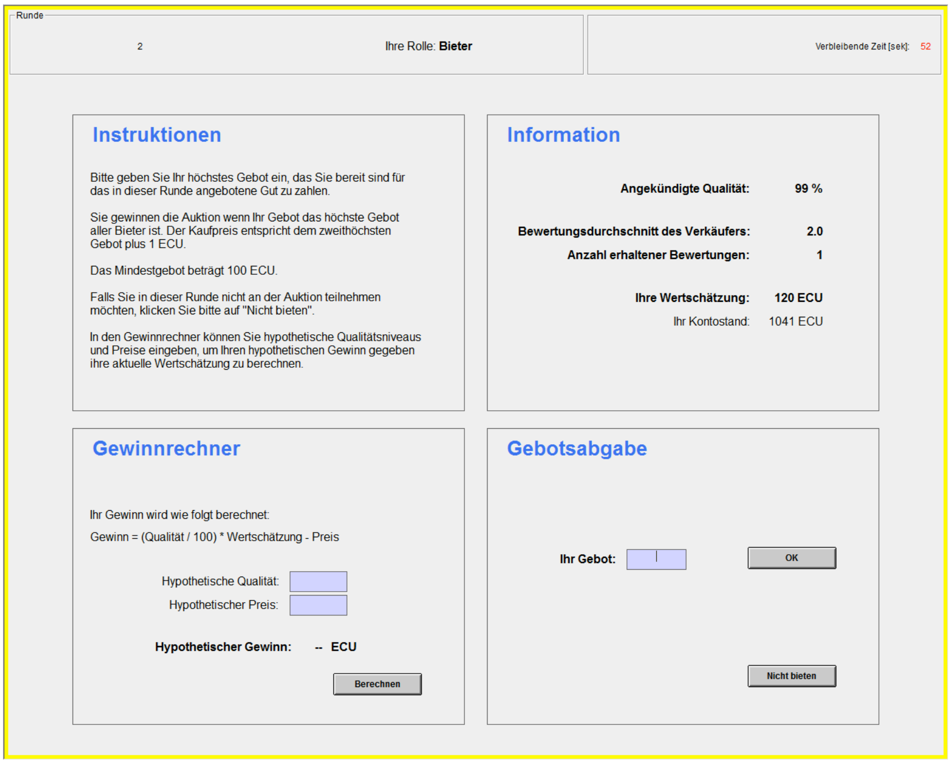


Figure 2.9: Bidding stage.

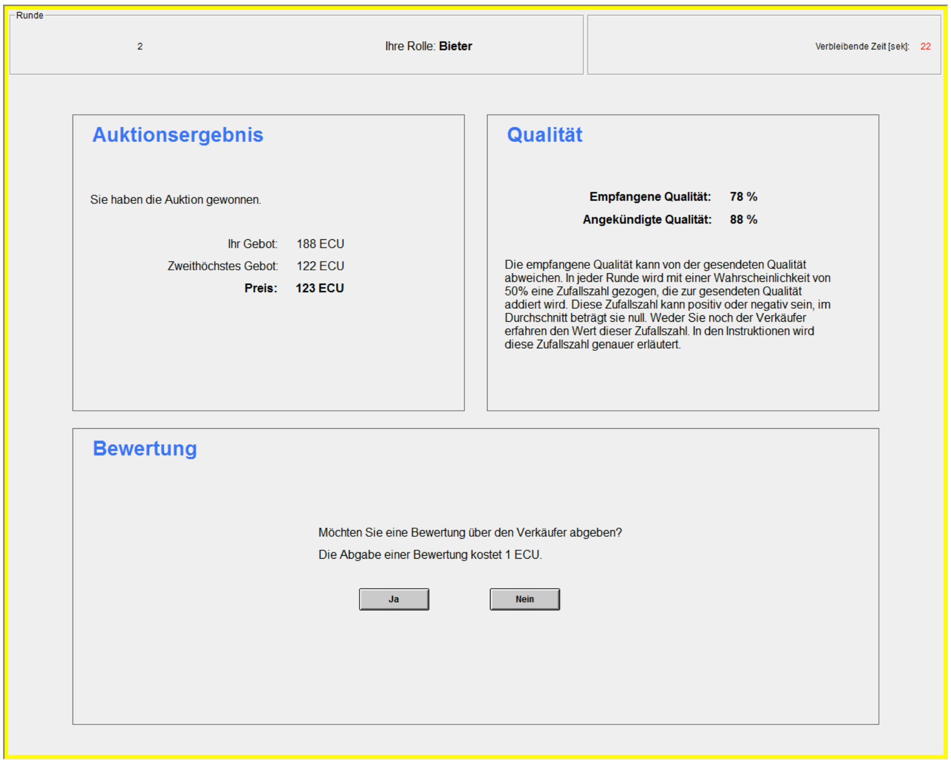


Figure 2.10: Feedback stage.

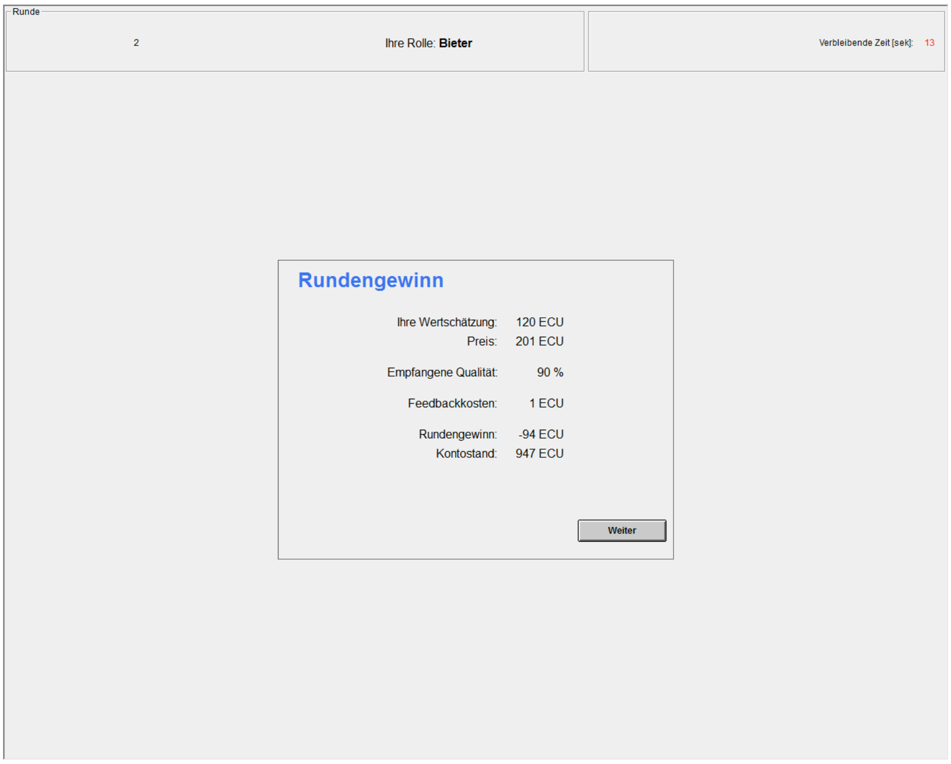


Figure 2.11: Info stage buyer.

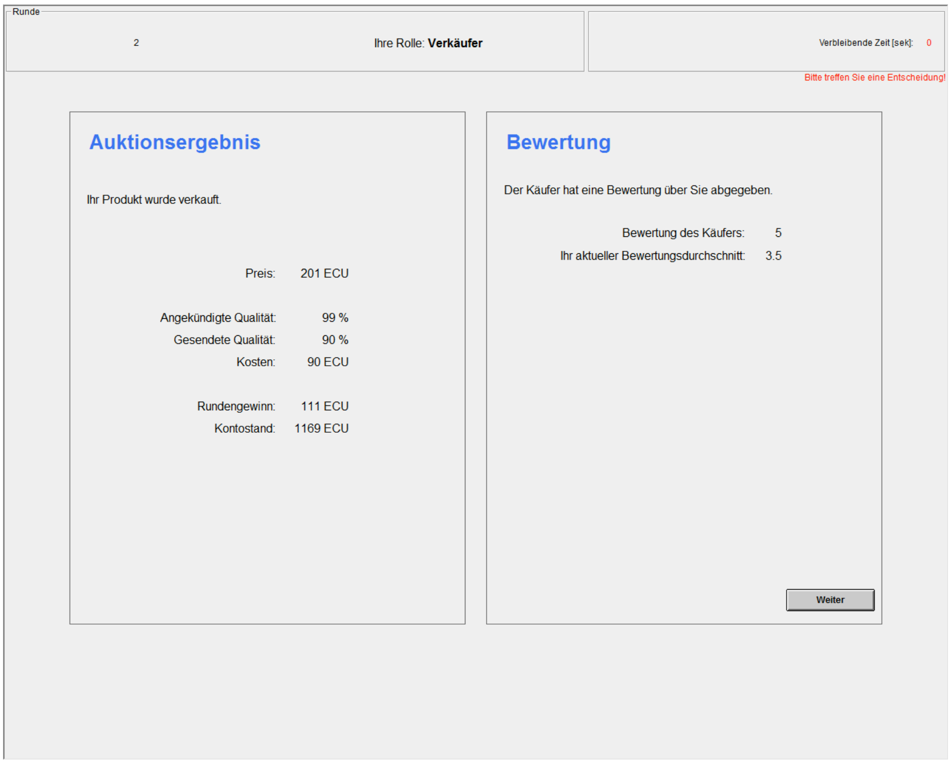


Figure 2.12: Info stage seller.

Chapter 3

Long vs. Short-Memory Feedback Systems

3.1 Introduction

Feedback ratings are ubiquitous on the Internet.[27] We rate trading partners on eBay, physicians on ratemds.com, mechanics on ratemymechanic.com, restaurants, bars and hotels on yelp.com or tripadvisor.com, all kinds of consumer goods on amazon.com, and many more. Hence, for almost all daily economic interactions – online as well as offline – people have the possibility to leave feedback to publicly report their experience with a purchased good or engaged service. Thereby, feedback systems help to mitigate issues of adverse selection and moral hazard in these markets by providing a signaling and sanctioning mechanism (Dellarocas 2005a). Hence, online reputation systems play a key role in establishing trust and incentivizing trustworthy behavior on these market platforms (Resnick and Zeckhauser 2002; Dellarocas 2003).[28]

However, it has been shown that feedback ratings can be biased for strategic as well as psychological reasons and thus do not necessarily reflect a trader's actual experience

---

[27] This chapter is joint work with Gary E. Bolton, David J. Kusterer, and Axel Ockenfels. All authors were equally involved in generating the idea. David Kusterer and I programmed and conducted the experiment. Statistical analyses were carried out by myself receiving feedback from all other authors. The current draft was written by myself with comments from David Kusterer. We thank Ben Greiner, Mattia Nardotto, Peter Werner, Christoph Feldhaus, Nicolas Fugger, Florian Goessl and Christopher Zeppenfeld for helpful comments and suggestions. We are also thankful to Kevin Breuer for his help with programming, testing and organizing the experiments. Financial support of the German Research Foundation (DFG) through the Research Unit "*Design & Behavior – Economic Engineering of Firms and Markets*" (FOR 1371) is gratefully acknowledged.

[28] Many empirical studies find evidence that buyers take feedback information into consideration and thus a good reputation increases sellers' probability to sell the product and also its price (Ba and Pavlou 2002; Bajari and Hortacsu 2003; Cabral and Hortacsu 2010; Eaton 2007; Ederington and Dewally 2006; Houser and Wooders 2006; Jin and Kato 2006; Livingston 2005; Lucking-Reiley et al. 2007; McDonald and Slawson 2002; Melnik and Alm 2003; Resnick and Zeckhauser 2002; Resnick et al. 2006).

(e.g. Dellarocas and Wood 2008; Bolton, Greiner, and Ockenfels 2013; Bolton, Kusterer, and Mans 2014; Rice 2012). According to Cabral (2012), a crucial advantage of online markets in comparison to offline markets is that in the former, market engineers have a higher degree of flexibility to change and adjust rules and mechanisms. In spirit of this economic engineering approach, design modifications of the feedback system can help to mitigate these biases. As a consequence, several design improvements have been studied in field as well as laboratory studies (Masclet and Pénard 2012; Bolton, Greiner, and Ockenfels 2013; Fradkin et al. 2014).

In this paper we take a slightly different perspective on the design of reputation systems. We investigate whether a modification of the feedback system is able to change the feedback language, i.e. the way traders use feedback ratings to pass on information about their experience to other traders. More specifically, we compare two extreme versions of feedback systems which differ only in the number of ratings displayed to the following buyer(s). The first reputation system includes all previously submitted ratings about a given trader, and thus resembles the design used on many online market platforms nowadays. In contrast, the second feedback system only shows the most recent feedback rating about a respective trader. One potential advantage of this reduced feedback system is that it requires users to process fewer pieces of information since all previous experiences should be condensed in one – the most recent – feedback. Recent research by Wolf and Muhanna (2011) demonstrates that buyers who are confronted with a detailed history of feedback ratings are prone to judgmental biases similar to those known from the psychological literature on judgment and decision making (Kahneman, Slovic, and Tversky 1982; Griffin and Tversky 1992). In particular, buyers systematically overweight the proportion of positive ratings but insufficiently take into account the statistical reliability, i.e. the overall number of ratings. Hence, in their assessment of seller trustworthiness, buyers tend to be too trusting towards sellers who have good but few ratings and thus are vulnerable to fraud from these seller types. A feedback system that provides only the most recent feedback does not allow for such an overweighting bias in feedback interpretation. Moreover, a short feedback history appears to be more similar to former word of mouth reputation building in traditional offline markets where not all previous recommendations might be available to all potential traders. Imagine, an initial buyer gives a recommendation about a particular seller to a second buyer, who then in turn is asked for a recommendation by a third buyer. Because the third buyer does not know the initial buyer, the recommendation of the second buyer should not only reflect his own personal experience but also the experience of the initial customer in order to give a precise assessment of the seller's trustworthiness. In this regard, traditional reputation networks rely more heavily on direct, bilateral interactions among traders to share the experiences they made personally but also to pass on earlier recommendations they received from previous customers.

We compare these two feedback systems in an adverse selection experiment without strategic interaction, where buyers trade with two different types of computerized sellers. These good and bad seller types differ in their exogenous probability to ship high or low quality products. Ex ante, buyers do not know which type of seller they are facing. Thus, feedback ratings of earlier buyers are helpful to improve beliefs about sellers' types and thereby advise buyers to optimize their purchase decisions. Using exogenous seller types, we focus on the signaling and information provision function of feedback systems and set aside the additional sanctioning aspect in situations including seller moral hazard.[29]

Both feedback systems rely on Bayesian updating and we argue that, theoretically, either of them can provide buyers with the same amount of information about seller types. This requires, however, that buyers coordinate on a common feedback language, i.e. on a homogeneous way to give and interpret feedback. This refers to how the signal (received quality) is converted into a feedback rating so that private information about sellers is passed on to following customers.

In a feedback system that displays all previous ratings of a seller (long-memory), an intuitive feedback language is that buyers simply report the quality received. Based on the full history of signals, subsequent buyers are then able to update their belief about the seller accordingly. Therefore, over time, buyers obtain more precise expectations about the seller's quality and can adjust their bidding behavior. To this end, buyers should interpret feedback ratings as signals about quality, use them to update their belief about the seller type, and then give feedback also in form of a signal about quality.

[29] Thereby, we also eliminate potential effects of the feedback system on the supplied quality. Additionally, social preferences between buyers and sellers cannot play a role.

In case only the most recent feedback rating is displayed (short-memory) such a simple reporting feedback language cannot be fully informative because only the most recent signal about a seller's type is passed on. In order to sustain all previous information, it is necessary that buyers coordinate on a different feedback language. One possible way to maintain informativeness is that after a purchase buyers update their belief about a seller's type according to the received quality and submit the resulting posterior belief as feedback rating. The following buyer subsequently uses this rating as his prior belief and makes a corresponding bidding decision. Thereby, if feedback ratings reflect posterior beliefs exactly, a short-memory feedback system can achieve the same degree of informativeness as when all previous ratings are displayed. Hence, both feedback systems are able to assist buyers equally in distinguishing between different seller types.

As shown, both feedback languages rely on Bayesian updating. However, they crucially differ with regard to the point when the updating occurs, and the way private information is passed on to future traders, i.e. how feedback is given. In long-memory feedback systems with the reporting feedback language, updating based on all previous ratings occurs *before* the purchase decision (prior belief) and feedback is given in the form of an additional signal. In contrast, in short-memory feedback systems with the updating feedback language, buyers update their belief *after* the purchase based on the received quality and the previous feedback rating. The resulting posterior belief is then posted as feedback rating for the following buyer. From an engineering perspective, the question is whether buyers are indeed able to endogenously coordinate on a common and informative feedback language in either of the two systems. In order to investigate this question, we test these two feedback systems in an experimental adverse selection setup that eliminates all strategic considerations and focuses solely on the aspect of information provision. At first, we shed light on how feedback is given in these two feedback systems and whether buyers are able to efficiently pass on their private information to other traders. Second, we also gain insight how feedback ratings are interpreted in order to differentiate between good and bad sellers and to optimize purchase decisions. Finally, we are also interested in comparing how the two different feedback systems influence overall market performance. Therefore, as the feedback information works its way through the system, we analyze how differences between the long- and short-memory feedback systems regarding feedback giving and interpretation influence trading efficiency in terms of buyer profits.

Our results show that only in the treatment where all previous feedbacks are displayed subjects are able to coordinate on an appropriate feedback language. In both treatments, many feedback ratings are equal to the received quality, which is only fully informative in the long-memory feedback system. Transmitting updated beliefs via feedback ratings is rarely used in the treatment where only the most recent rating is available. For this reason, the informativeness of ratings is lower in the short-memory feedback system and buyers struggle to identify good sellers. Overall, buyers under the short-memory feedback system forego potentially beneficial trades with good sellers and have significantly lower profits than their counterparts with full feedback history.

The paper proceeds with a short discussion of the related literature. Afterwards, we briefly analyze our adverse selection setup with computerized sellers We then discuss how long- and short memory feedback system may affect buyers' feedback languages. Following, we describe the experimental design. The next section presents our results on how the two systems influence feedback giving and interpretation. Finally, the last section summarizes our study and provides possible explanations why buyers in short-memory feedback systems were not able to apply and coordinate on the updating feedback rule.

3.2 *RELATED LITERATURE*

With this paper we contribute to the recent literature on the design of feedback mechanisms. Many empirical and experimental studies investigated potential shortcomings of current feedback systems, which give rise to biased feedback ratings, and tested new mechanisms to overcome these issues. One crucial disadvantage of two-sided feedback systems – where buyers and sellers have the possibility to leave a rating – is that the feedback process between traders is characterized by reciprocity (Bolton, Greiner, and Ockenfels 2013; Klein et al. 2006; Jian, Mackie-Mason, and Resnick 2010). This, in turn, causes dissatisfied traders to conceal negative experiences and submit no rating because they fear retaliatory feedback by their trading partner. Using structural estimations, Dellarocas and Wood (2008) show that silence or non-ratings cause a substantial bias in feedback ratings and disguise the true risks associated with trading. They suggest to include the number of non-rated transactions into a seller's feedback profile because these non-ratings also carry important information for future traders.

Bolton, Greiner, and Ockenfels (2013) experimentally investigate two alternative design modifications to reduce the reciprocity bias in feedback ratings. Adding a one-sided feedback component as well as making the feedback double-blind has positive effects on the informativeness of the feedback system and increases traded quality and market efficiency. Similarly, Masclet and Pénard (2012) find that simultaneous ratings or sequential ratings with a fixed rating order enhance trust and trustworthiness.

To eliminate reciprocity in feedback ratings, eBay introduced a one-sided feedback system by adding optional detailed seller ratings for buyers and restricting sellers to give only positive ratings. Recent studies analyzed these design modifications. These changes not only had a positive effect on feedback informativeness but also led to more trustworthy seller behavior, i.e. reduced moral hazard on the seller side (Klein et al. 2009; Klein, Lambert, and Stahl 2013; Ye, Gao, and Viswanathan 2010).

On *Airbnb* – a platform for accommodations – ratings are also biased due to reciprocity and the fear of retaliation. In a field experiment on *Airbnb*, Fradkin et al. (2014) find that offering coupons for ratings increases feedback provision and helps to encourage dissatisfied guests to submit a rating. Similar effects can be obtained by introducing a simultaneous, blind rating process. Furthermore, there is evidence that closer and direct social interaction between traders further increases biases in ratings because negative information are even more likely to be omitted.

Li and Xiao (2014) investigate a design change where sellers can choose to offer a rebate to buyers if they leave a rating after the transaction. This rebate option works as a screening device for good sellers and thereby trading efficiency increases with the probability that the seller offers a rebate. Cabral and Li (2012) take this rebate scheme to a field test on eBay. Their results show that rebates effectively increase the provision of feedback but unfortunately lead to more biased ratings as compensated buyers are even more likely to leave a positive rating even for slower delivery. In addition, offering a rebate does not attract more or higher bids and thus has no effect on sales prices.

Overall, with the rise of online markets, different design modifications of reputation systems have been investigated to elicit unbiased feedback ratings and to increase trust and trade. This paper takes a slightly different approach and examines whether and how the design of a feedback system influences traders' feedback language and whether it is indeed sufficient to display only the most recent feedback rating. In such a feedback system, buyers need to change their feedback language from reporting the received quality to reporting their updated belief about the seller's type.

With regard to Bayesian updating based on information provided by other subjects, our paper is also related to the literature on social learning (Banerjee 1992; Bikhchandani, Hirshleifer, and Welch 1992; Welch 1992; Bikhchandani, Hirshleifer, and Welch 1998). The exogenous, computerized seller types in our experiment resemble the classical urn setup as used in experimental studies on information cascades (e.g., Anderson and Holt 1997). In this classical inference problem each subject of a series of individuals sequentially draws a ball from one of two urns. Both urns are equally likely and both contain balls of two different colors. The amount of balls of each color is different in the two urns and thus is an informative signal. After drawing a ball each subject has to guess which urn has been selected and receives a fixed payment for a correct guess. His guess but not his private signal is then observed by all subsequent individuals and therefore serves as social information. There are two crucial differences between the information transmission process in social learning theory and our feedback mechanism. First, the sequence of receiving a private signal and making a payoff-relevant decision is reversed. In our setup buyers first make the purchase decision based on the information received from earlier traders and then – if they decide to buy the product – observe their own private signal about the seller's type. Second, the payoff-relevant decision and the transmitted information are disentangled. Feedback is endogenous and buyers are free to give any rating they find appropriate given the received quality.[30] Especially the latter difference is important for our investigation because we are mainly interested in how different reputation systems affect buyers' feedback language and how they map observed quality into feedback ratings. Nevertheless, we can learn from experiments on social learning theory that in such an urn setup people have difficulties to apply Bayes' rule and update their beliefs incorrectly. In several laboratory

[30] Disentangling a buyer's action and his feedback also eliminates the threat of information cascades in which later subjects simply follow earlier subjects and ignore their own signal (see e.g. Bikhchandani, Hirshleifer, and Welch 1992).

experiments it has been found that the majority of subjects are overconfident and overweight their private signal instead of making rational decisions based on Bayesian updating (Huck and Oechssler 2000; Nöth and Weber 2003; Celen and Kariv 2004; Spiwoks, Bizer, and Hein 2008). These findings are of particular interest for our study as in both feedback systems buyers have to apply Bayes' rule. However, the difference between the two feedback languages lies in the point at which updating is necessary. With the reporting feedback rule in the long-memory feedback system updating has to be done before the purchase decision and thus, it incorporates only signals of earlier customers. Instead, with the updating feedback language in the short-memory feedback buyers update their belief after the private signal is observed. If the tendency to overweight the private signal is also relevant in the provision of feedback, this should only have a negative effect on feedback informativeness when beliefs are given as feedback ratings.

3.3 *THEORETICAL FRAMEWORK AND HYPOTHESES*

The basic situation that we are interested in resembles a standard adverse selection case.[31] In each period, each buyer is matched with a seller who offers a product of unknown quality Q. Buyers know that there are two types of sellers, good sellers G and bad sellers B. However, they do not know with which type they are matched. The prior probability to meet one of the two seller types is identical $\Pr\{G\} = \Pr\{B\} = 0.5$, and this is common knowledge. Quality can either be high H or low L. Good sellers are more likely to sell a product of high than of low quality, such that $\Pr\{H \mid G\} = \alpha > \Pr\{L \mid G\} = 1 - \alpha$, with α being an informative but noisy signal, for which holds that $0.5 < \alpha < 1$. The probabilities of bad sellers are reversed and thus, they are more likely to ship low quality, such that $\Pr\{H \mid B\} = 1 - \alpha < \Pr\{L \mid G\} = \alpha$. Hence, the expected quality of a good seller is $E[Q \mid G] = \alpha H + (1 - \alpha)L$. Analogously, the expected quality of a bad seller is $E[Q \mid B] = (1 - \alpha)H + \alpha L$. The unconditional expected quality

[31] According to Dellarocas (2005b) on- and offline trading environments are often simultaneously characterized by issues of adverse selection and moral hazard with one of these issues typically being more pronounced. We analyze short- and long-memory feedback systems in a moral hazard context in a separate project. For a brief discussion of this study see Appendix A.

then is $E[Q] = \beta E[Q \mid G] + (1-\beta)E[Q \mid B]$ with β representing a buyer's prior belief that the seller is of the good type.[32] According to Bayes' rule, this belief is defined as

$$\beta(\gamma) = \frac{\alpha^{\gamma}}{\alpha^{\gamma} + (1-\alpha)^{\gamma}}$$

where γ refers to the difference between the number of public ratings (signals) indicating high quality and the number of public ratings (signals) indicating low quality. Thus, γ resembles a seller's feedback score as it is used on eBay, for example.[33] If a buyer purchases the good his profit is the product of his valuation *v* for the good and the received quality net of the price *p*:

$$\pi = vQ - p \, .$$

For convenience, a buyer's valuation is set to one in this model. In order to purchase the offered product the buyer submits a maximum bid, which is then compared to a randomly drawn price *p*. For each seller-buyer pair, the price *p* is drawn independently. A buyer only purchases the product if his bid matches or exceeds this price. If the bid is below the randomly drawn price, the product is not sold and the buyer earns nothing.[34] This procedure corresponds to a Becker-DeGroot-Marschak (BDM) mechanism and ensures that buyers have an incentive to bid their true valuation (Becker, DeGroot, and Marschak 1964). In our setup, in which valuation is set to one, a buyer's optimal bidding strategy is to submit a bid *M* corresponding to the quality he expects based on his current beliefs about the seller's type, $M(\beta) = \beta E[Q \mid G] + (1-\beta)E[Q \mid B] = E[Q]$.[35] From this it directly follows that the optimal bids of a rational, risk-neutral buyer range between the minimum expected quality $E[Q \mid B] = (1-\alpha)H + \alpha L$, which is the case if

---

[32] In the following, when we talk about beliefs this always refers to buyers' beliefs that the seller is of the good type.

[33] This simple feedback score rule (difference between high and low signals) only applies because we have a symmetric situation with Pr{H|G} = Pr{L|B}. Therefore, both signals are equally informative (c.f. Anderson and Holt 1997).

[34] In this model and the following experiment, sellers do not take any strategic decisions since product quality is solely determined by chance via the type-specific probabilities for high and low quality. Therefore, we do not analyze sellers' profits but rather focus only on buyers' profits.

[35] The argument given here is the same as in Vickrey auctions. Buyers do not have an incentive to bid less than the expected quality because they then may not buy for some values of *p* although the valuation exceeds the price on average and thus they forego potential profits. At the same time increasing the bid above the expected quality does not pay off because prices of the additionally purchased products on average exceed expected quality and also the valuation.

the buyer believes that he faces a bad seller for sure, and the maximum expected quality $E[Q \mid G] = \alpha H + (1-\alpha)L$, which is the case if the buyer believes that he faces a good seller for sure. We assume that the random price is drawn from a uniform distribution also ranging between $E[Q \mid B]$ and $E[Q \mid G]$.

In this setting, a feedback system is helpful to buyers as it provides information about the sellers' past behavior and thus, they can be used to form beliefs about sellers' types. Over time, buyers are able to derive more accurate expectations about the unknown quality of the product on sale and thereby increase their expected profit. Assuming that the same buyer-seller pair can meet in the future and that feedback ratings are costless, buyers have an incentive to give informative feedback, i.e. a feedback that helps to differentiate between good and bad sellers. The reason is that buyers may profit from their own submitted feedback as there is the chance that they are matched with the same anonymous, unidentifiable seller in the future again.[36] This incentive to provide informative feedback ratings and thereby to improve beliefs about seller types holds for either of the two feedback languages, which both rely on a Bayesian updating process. The crucial difference is at which point the beliefs are updated and how exactly feedbacks need to be given in order to be fully informative.

In the first reputation system, a seller's feedback history contains all individual feedback ratings from previous buyers. If all previous buyers reported their received quality the current buyer is able to determine γ based on the feedback history.[37] Hence, the buyer derives the prior belief that the seller is of the good type from the existing ratings and adjusts his bidding decision accordingly. After receiving the product, the buyer leaves an additional feedback rating by also simply reporting the observed quality, which – together with all previous ratings – can then be used by the next buyer to form a new prior belief. In this sense, each buyer individually forms a belief based on all provided feedback ratings. For this reason, a reporting feedback language is only

[36] In our setup, there is no competition among buyers because in each round every seller is matched with one buyer only. Thus, there is no incentive to provide false information about sellers' types in order to influence other buyers bids.

[37] For such a reporting feedback language it is required that the rating scale allows to mirror the observed quality. For example, with binary quality a binary feedback rating scale – e.g. positive or negative – is sufficient to convey all information in feedback ratings to form accurate Bayesian beliefs.

fully informative if buyers see the entire history of feedback ratings, such that each additional rating further improves the precision of the belief about the seller's type.

In contrast, a feedback system that displays only the most recent rating does not allow for such a 'discrete' updating process because valuable information would get lost.[38] Here, not only the observed quality is private information but also the feedback rating of the previous buyer. In order to sustain all available information about a seller, buyers need to adapt their feedback language and give ratings that reflect not only their own transaction experience but also the information about earlier transaction experiences contained in the feedback rating they observed. One possible feedback language to circumvent a loss of information is that buyers post feedback ratings not in form of signals about the received quality but instead in form of an updated belief. Buyers should update their belief according to the observed product quality and communicate this posterior belief via the feedback system. The following buyer directly uses this feedback rating as his prior belief to make an informed bidding decision. In case of a purchase, the latter buyer also updates his prior belief based on the observed quality and provides a corresponding rating. Depending on the received quality the feedback rating R can be formalized as

$$R_t = \begin{cases} \dfrac{\alpha R_{t-1}}{\alpha R_{t-1} + (1-\alpha)(1-R_{t-1})} & \text{if } Q_t = H \\ \dfrac{(1-\alpha) R_{t-1}}{(1-\alpha) R_{t-1} + \alpha(1-R_{t-1})} & \text{if } Q_t = L \end{cases} \tag{1}$$

where $R_{t\text{-}1}$ is the feedback rating which the buyer observed himself before the purchase or – if no rating has been given so far – the initial probability to meet a good seller $\Pr\{G\}$. Thereby, analogous to additional signals in the long-memory feedback system, the belief about a seller becomes more precise with every additional updating step. However, in order to convey exact beliefs it is necessary that feedback ratings can be given on a continuous scale between 0 and 1. Taken together, instead of a 'discrete' updating process where every buyer uses all previous ratings to build a belief on his own, they need to switch to a more 'continuous' updating procedure. In this procedure

---

[38] If only one rating is available γ could either be equal to 1, if the rating indicates high quality, or -1, if the rating indicates low quality. Hence, the belief would be equal to α in the former and to 1-α in the latter case.

the posterior belief of the preceding buyer is adopted first, is then updated based on the observed quality and is finally transmitted as feedback rating. Such an updating feedback language guarantees that buyers obtain the same beliefs as if ratings simply report the received quality. Thus, both feedback systems can be equally informative. Generally, in both feedback systems buyers could also coordinate on other, fully informative feedback languages. These are discussed in greater detail after the experimental design has been introduced.

Following these theoretical considerations we want to compare a long-memory feedback system in which all ratings are displayed to a short-memory feedback system in which only the most recent rating is transmitted. As we have seen, both feedback systems can supply buyers with the same amount of information about sellers but differ in the way this information is generated and communicated via feedback ratings. This difference refers to the feedback language or, more specifically, whether private information about a seller is passed on as a signal about quality or as a belief about seller's type. We want to explore in which feedback system buyers are better able to identify and coordinate on a common, fully informative feedback language. We are interested in the way feedback is given in the two systems and whether buyers deviate coordinate on the respective feedback language. Furthermore, we also want to compare how helpful short- and long-memory feedback systems are to buyers to differentiate between good and bad sellers and to optimize bidding behavior. Finally, as the feedback information works its way through the system, we are also interested in how potential differences between the two feedback systems in terms of feedback giving and interpretation have an effect on buyer profits.

3.4 Experimental design

In our experimental setup we have matching groups of 12 players consisting of 6 computerized sellers – three of each type – and 6 human participants in the role of buyers. In each period, buyers and sellers are randomly matched into pairs but with the restriction that the same pair does not interact in two consecutive periods. Each seller offers a product of unknown quality Q to the buyer. As mentioned before, there are good and bad sellers and buyers do not know which type they are dealing with. Quality ranges between 0 and 100 but it still represents a binary signal about seller types as we define two separate intervals: values below 50 are classified as low quality L and values of 50 and above are considered as high quality H. For good sellers α is equal to 0.75 which means that with a probability of 75% the value of the quality is drawn uniformly from the high quality range $H \sim U(50,100)$ and with 25% from the low quality range $L \sim U(0,49)$. For bad sellers, probabilities are reversed: values from the high (low) quality range have a likelihood of 25% (75%). In any case, quality only takes integer values. Hence, in expectation, a good seller ships quality of 62.375 whereas a bad seller ships 37.125. The unconditional expected quality is 49.75. Buyers have an individual valuation $v \sim U(100,300)$ for the offered product. The valuation is randomly drawn and disclosed to buyers at the beginning of each period. In addition, a buyer also gets feedback information about the seller he is matched with in the current period. With this information, buyers submit their maximum bid M and buy the product if their bid matches at least the randomly drawn price $p \sim U(0,200)$.[39] Hence, a buyer's profit is

$$\pi = \begin{cases} v \times \dfrac{Q}{100} - p & \text{if } M \geq p \\ 0 & \text{otherwise} \end{cases}$$

After the buyer has received the product and observed its quality he has to leave a feedback rating R at no cost.

In line with our earlier considerations, our two treatment variations relate to the way these ratings are displayed to future buyers. In Treatment '*All*', a long-memory feedback system is in place and buyers observe each individual rating whereas in treatment '*Last*', feedbacks are short-memory and buyers only get to know the most recent feedback rating. As additional information in this treatment, buyers learn the number of periods in which the seller did not sell his product.[40] In both treatments, we elicit these feedback ratings by asking the same feedback question: "Please rate the

---

[39] The BDM mechanism is incentive compatible and hence a rational and risk-neutral buyer submits a bid equal to his expected valuation for the product. His expected valuation is his assigned valuation v times the expected quality in percent. The expected quality depends on buyer's belief β to be matched with a good seller: $E[Q] = 37.125 + 25.25\beta$ Therefore, bids should only range between 37.125 – if the buyer knows for sure that he is matched with a bad seller, i.e. $\beta = 0$ – and 62.375 if $\beta = 1$.

[40] We do so because buyers in the long-memory feedback system can infer this information from the full feedback history immediately.

seller!". Buyers leave a feedback on a continuous scale between 0 and 100.[41] Thereby, we ensure that both feedback languages – reporting quality and updated belief (in percent) – can be expressed within the rating scale.

With a continuous rating scale it is of course possible for buyers to use an updating feedback language in the long-memory feedback system, too. Other feedback languages can also be fully informative in this system. It suffices that all buyers coordinate on a language in which ratings clearly distinguish between quality from the high and low quality interval. For example, a consistent and commonly used rating scheme of 99 for low quality and 100 for high quality also provides all necessary information about the seller. However, we would argue that the most intuitive and focal way to provide feedback in a long-memory feedback system is a reporting feedback language that maps the received quality one to one into a rating. In the short-memory feedback system, an alternative feedback language could be to use a weighted average to map quality into feedback: $R = (nR_{t-1} + Q) / n + 1$, where n is the seller's number of sold products which is equivalent to the total number of signals received by all preceding buyers. With increasing number of sold products this weighted average converges to the true average quality of a good or bad seller and thus reveals a seller's type. However, such a weighted average converges linearly while beliefs about seller types converge exponentially and thus faster. Therefore, giving beliefs as feedback is a more efficient way to identify sellers' types than to provide a weighted average.

Overall, the experiment lasted for 60 periods. We drew valuations, prices, quality levels and buyer-seller-matching in advance so that these values and the order of trading partners were constant across all matching groups.[42] For each treatment we ran one session with five independent matching groups leading to 3600 buyer observations in total. The sessions took place at the Cologne Laboratory for Economic Research (CLER) and lasted for 2 hours on average. Participants were recruited from the CLER's subject pool using ORSEE (Greiner 2004) and the computerized experiment was coded in z-Tree (Fischbacher 2007). At the beginning of each session, the instructions were

[41] In fact, we allowed feedback ratings with up to seven numbers after the decimal point so that beliefs could be given as precise as possible.

[42] Thereby, in each matching group identical sellers – three bad and three good – are present.

handed out to the participants.[43] At the end of the experiment, the account balance was converted to euros (100 ECU = 1€) and paid out in cash. Including an initial endowment of 600 ECU subjects earned 22.55€ on average.

3.5 Results

At first, we look at the descriptive results. In Treatment *All* with full feedback history buyers submit higher maximum bids than buyers in *Last,* who observe just the most recent rating (103.27 vs. 93.27). These average bids are not significantly different according to a Mann-Whitney-U test (MWU; $p = 0.12$).[44] Nevertheless, these higher average bids result in a six percentage points higher purchasing probability in Treatment *All*. Additionally, a full feedback history seems to enable buyers to detect good seller types and thereby, they purchase higher quality. This means that, conditional on buying the product, quality is weakly significantly larger in Treatment *All* (MWU; $p = 0.08$). Overall, differences regarding quality and purchase frequency result in significantly higher profits per period (MWU; $p = 0.03$). This demonstrates that buyers who receive information about all previous feedback ratings are able to make more profitable bidding decisions.

	All	***Last***	**MWU**
Bid	103.27	93.27	$p = 0.12$
	(53.62)	(58.58)	
Product bought	0.54	0.48	$p = 0.08$
	(0.50)	(0.50)	
Quality | Buy	54.55	52.93	$p = 0.08$
	(29.15)	(28.77)	
Round profit	29.02	25.98	$p = 0.03$
	(61.09)	(58.70)	
N matching groups	5	5	
N individual	1800	1800	

Table 3.1: Descriptive statistics: means with standard deviations in parentheses. Mann-Whitney-U tests are based on means aggregated on matching group level.

[43] A translated version of the instructions can be found in Appendix B.

[44] Unless indicated otherwise all statistical tests used in this paper are two-tailed Mann-Whitney-U tests based on averages of independent matching groups.

In order to compare the performance of our two feedback systems, we investigate at first how buyers give ratings and whether they use different languages to map quality into feedback in the two treatments. As mentioned earlier, in the treatment where all previous feedbacks are displayed, the easiest way to provide future buyers with full information is by simply reporting the received quality as a feedback rating (reporting feedback). In the treatment in which the information consists of only the most recent feedback rating as information, this straightforward feedback language is no longer successful in identifying good and bad sellers. Here, reporting the received quality would provide just one signal about the seller's recently delivered level of quality, and all previous information would be lost.[45] In contrast, with such a short-memory feedback system in place, buyers should interpret the observed feedback as the posterior belief of the previous buyer and use it as their own prior belief for their bidding decision. If the bid is successful and the buyer receives the product, he should then update his prior belief based on the received level of quality and share this posterior belief as feedback rating with future buyers (updating feedback). In this way, the loss of information could be avoided and buyers should be able to identify good and bad sellers as quickly as in the full feedback treatment.

Feedback giving

In Table 3.2 we report the share of cases in which a submitted feedback matches the reporting or updating feedback language for both treatments. In the treatment where all feedbacks are displayed, half of the submitted ratings exactly match the received quality and thus, they are perfectly in line with the informative feedback language. Adding a tolerance of +/- 5 to account for rounding or typing errors, this share increases to 69.1%. For Treatment *Last*, the updating feedback language is much less prominent as only 2.1% of all feedback ratings are equal to the updated belief based on the received quality and the previous rating and only 16% are within the tolerance range.[46] This non-

[45] With only the most recent signal available, a buyer would only be able to update his belief that he is facing a good type to either 75% if the most recent feedback is 50 or higher and to 25% otherwise.

[46] We calculate the 'correct' posterior beliefs based on Equation 1. We assume that buyers always use the last feedback as prior belief even if its value does not belong to the set of possible values for beliefs. E.g. with our parameters feedback ratings between 26 and 49 as well as between 51 and 74 should not occur because the first updated belief increases (decreases) from 50 to 75 (25) when a high (low signal) is observed.

use of the updating feedback language exists already at the beginning of the experiment. According to the updating feedback rule, the first feedback rating a seller receives should be either 25 if the buyer received low quality or 75 if the buyer received high quality. However, none of the sellers in Treatment *Last* gets a first feedback rating of 25 or 75. Instead, many subjects also use the reporting feedback language in Treatment *Last* as almost 40% (54.4% with +/- 5 tolerance) of the feedback ratings simply mirror the received quality. Although this feedback language is not able to convey all information, it seems to be a more intuitive way for subjects to rate sellers.[47]

Feedback language	***All*** exact (+/- 5)	***Last*** exact (+/- 5)
Feedback = Quality	50.5% (69.1%)	39.4% (54.4%)
Feedback = Belief	0.3% (6.7%)	2.1% (16.0%)
Observations	968	870

Table 3.2: Used feedback strategies in both treatments.

Figure 3.1 confirms our findings. In both treatments ratings cluster around the diagonal if we plot quality against feedback (see Figure 3.1, upper Panels 1 and 2) but this is not the case for updated beliefs and feedback in Treatment *Last*. In fact, no such clear relationship between updated beliefs and feedback ratings can be observed (lower Panels 3 and 4). Taken together, these results suggest that with a long-memory feedback system the majority of buyers use an efficient feedback language such that no information gets lost. In contrast, when only the most recent feedback rating is displayed, buyers have severe difficulties to coordinate on the appropriate feedback language as they do not give ratings as updated beliefs. Instead, many buyers simply report their quality whereby only one quality signal is passed on to the following buyer and all previous information gets lost.

[47] For our second treatment with only the most recent feedback, we tested also other feedback rules like the weighted average as described above and the simple average, i.e. mean of observed feedback rating and received quality. However, for both of these rules the shares of feedback ratings which are perfectly in line with these rules are also only around 2%.

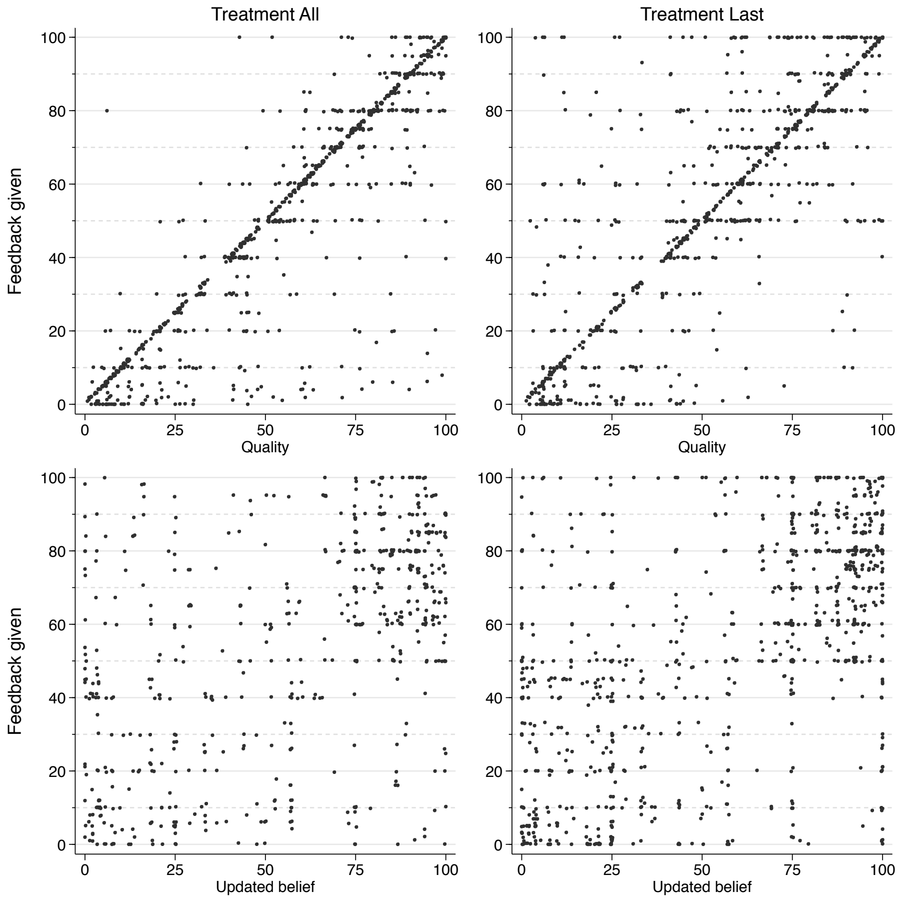


Figure 3.1: Scatterplots of feedback strategies in both treatments.

Regression results confirm our findings that reporting received quality is the prominent feedback rule in both treatments (see Table 3.3). For Treatment *All* received quality has a significant, positive impact on the feedback rating and the coefficient is close to 1 indicating an almost perfect linear relationship between quality and feedback. For Treatment *Last*, we first include the posterior belief based on the observed rating and the received quality as independent variables. Theoretically, the posterior belief should also be mapped 1-to-1 into a feedback but its coefficient is significantly smaller than 1. Including quality as an independent variable further reduces the influence of beliefs on the submitted feedback rating in the second treatment (Model 3). In the last model we use observations from both treatments and test whether quality has a different influence in *All* than in *Last*. The interaction effect between quality and treatment is not significant and thus we find no evidence that quality has a different influence on feedback ratings in the short-memory than in the long-memory system.

Overall, our descriptive, graphical and regression analysis clearly show that in a long-memory feedback system buyers indeed rely on the fully informative feedback rule and report the received quality. In contrast, in a feedback system that displays only the most recent feedback rating, very few feedbacks are in line with the updating feedback rule. Instead, reporting quality is again used very frequently. In this treatment, however, this feedback language leads to only the most recent quality signal being passed on and all previous signals getting lost. As a consequence, information contained in the short-memory feedback system should be lower than in the long-memory feedback system and therefore it should be more difficult for buyers to identify good and bad sellers.

Feedback	***All***	***Last***	***Last***	***All & Last***
Quality	0.950^{***}		0.766^{***}	0.950^{***}
	(22.339)		(15.124)	(23.558)
Belief in %		0.478^{***}	0.177^{***}	
		(15.651)	(4.339)	
Last				8.239^{*}
				(2.572)
Quality x *Last*				-0.085
				(-1.564)
Period	0.025	0.006	-0.002	0.013
	(0.754)	(0.182)	(-0.066)	(0.529)
Constant	-1.384	26.235^{***}	2.936	-1.065
	(-0.621)	(9.353)	(1.261)	(-0.590)
R^2 within	0.864	0.283	0.715	0.780
R^2 between	0.229	0.604	0.335	0.208
R^2 overall	0.787	0.301	0.663	0.707
Wald test Quality = 1	$p = .24$		$p < .001$	
Wald test Belief = 1		$p < .001$	$p < .001$	
N	968	870	870	1838

Table 3.3: Random-effects regression with feedback as dependent variable. Standard errors clustered on matching group level. *t* statistics in parentheses; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

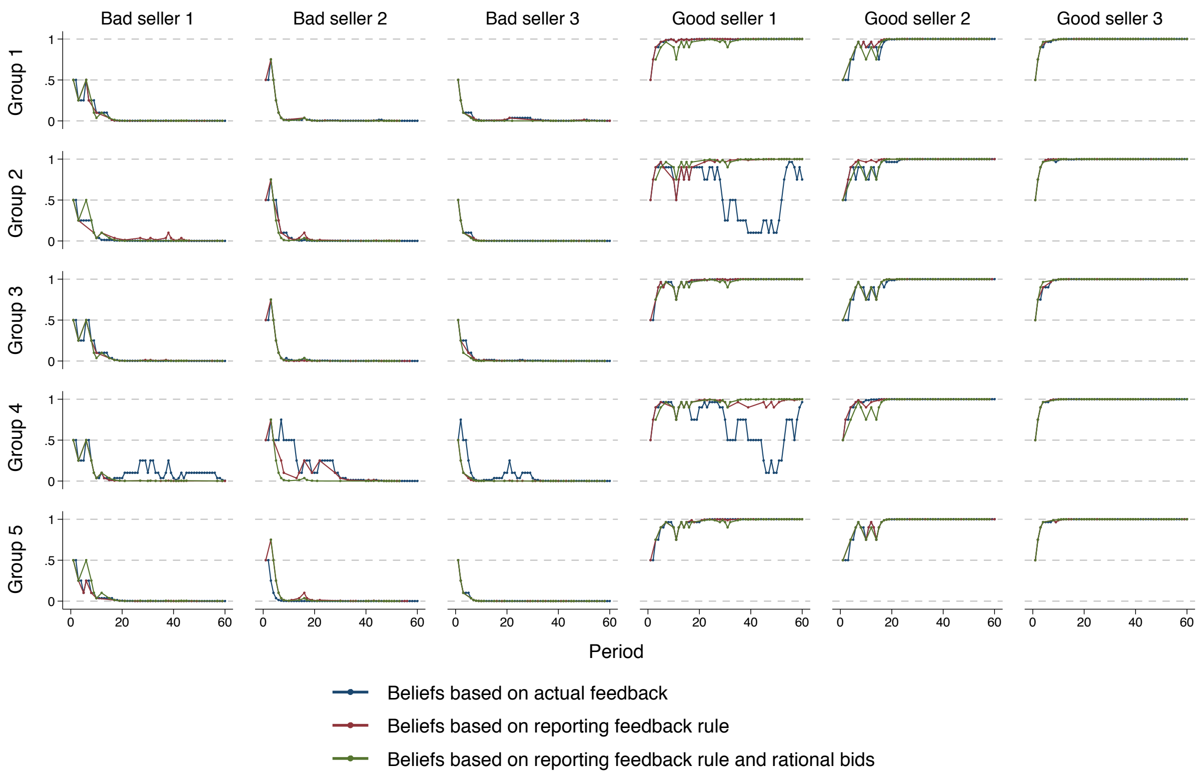
Feedback informativeness

In order to assess the informativeness of the feedback ratings given in both treatments, we investigate how the prior beliefs about sellers' types converge to their true value – either 0 for low sellers or 1 for high sellers.[48] For Treatment *All*, these priors can be derived from the difference between the number feedbacks indicating high quality (≥

[48] Prior beliefs are buyers' beliefs about sellers' types at the beginning of the period after they have seen previous feedback rating(s) but before they make their bidding decision.

50) and the number of feedbacks indicating low quality (< 50). For Treatment *Last*, the only feedback rating displayed should reflect the prior belief. Besides these beliefs based on the actual given feedback ratings, we additionally calculate two different benchmarks for each treatment: 1) how beliefs would converge if all buyers who purchased the product had used the fully informative feedback language (reporting quality in Treatment *All* and updating feedback in Treatment *Last*) and 2) how beliefs would converge if buyers had followed the fully informative feedback language *and* also had submitted rational, risk-neutral bids.[49]

Figure 3.2: Convergence of beliefs in Treatment *All*

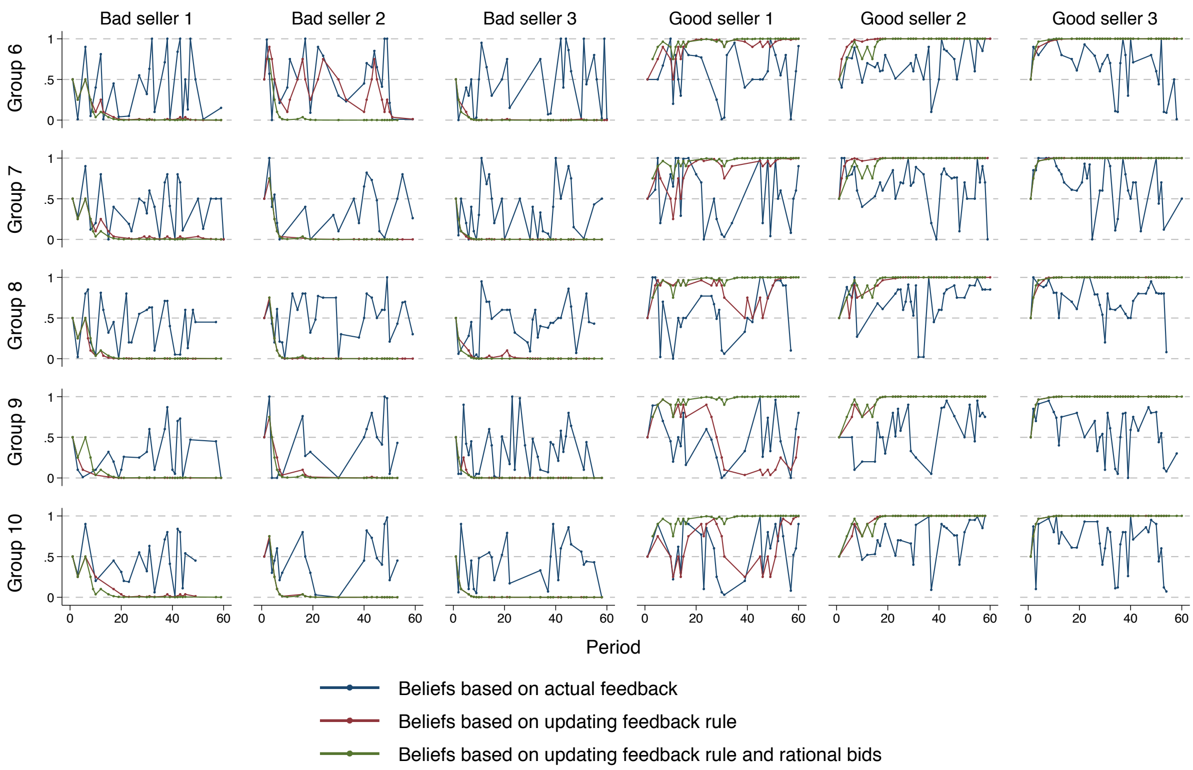


In Figure 3.2 and 3.3, each row shows the theoretical prior beliefs over time for sellers from one matching group. In addition, since computerized sellers are constant across matching groups, each column displays identical sellers but in different matching groups. Looking at Figure 3.2, we observe that, in a long-memory feedback system, the information contained in the ratings is sufficient to correctly identify types of most

[49] Rational, risk-neutral bids are important because it affects the number of purchases and thereby also the number of feedbacks (signals) within the feedback system. Here, rational and risk-neutral means that buyers form rational expectations about quality based on the prior belief and then submit a maximum bid in height of their valuation times the expected quality. This benchmark is identical in both treatments (green line).

sellers within the first third of the experiment. Only for few sellers the two benchmarks perform better and beliefs converge faster to the true value. In contrast, in the feedback systems with only the most recent feedback, beliefs do not converge at all; neither for bad or for good sellers (see Figure 3.3).

Figure 3.3: Convergence of beliefs in Treatment *Last*



This highlights that in the second treatment buyers do not use updated beliefs but rather simply report the received quality or give feedback in a different way. The red line shows that an updating feedback strategy would have been successful to identify most of the sellers. Furthermore, as indicated by the green line, rational bidding would have also helped to further improve the accuracy of beliefs about some sellers.[50]

Overall, these results suggest that, theoretically, information contained in the feedback ratings is large enough, such that buyers in the first treatment should be able to distinguish between good and bad sellers. Because feedbacks are not given in line with

---

[50] Because buyers in Treatment *Last* bid too cautious, they purchase less often and therefore they receive fewer signals. For the four sellers where the red and green line show the largest differences (bad seller 2 in matching group 6, good seller 1 in matching group 8, 9, 10) a rational buyer would have bidden high enough to buy the product in 144 of 240 transactions (60%) whereas actual buyers purchase only in 108 of 240 transactions (45%).

the optimal updating rule, the informativeness of the short-memory feedback system is substantially lower and thus, buyers should have difficulties to distinguish between seller types. Therefore, in the following section, we will examine how buyers use feedbacks in the two systems to make their bidding decision and how well their bid matches to the seller's type.

Feedback interpretation

In our adverse selection setup the main purpose to provide feedback is to help future buyers – including oneself – to identify good and bad sellers and to optimize future bidding decisions. For a rational and risk-neutral buyer, optimal bidding means to submit a bid equal to their expected valuation. The expected valuation is their private valuation *v* multiplied with the quality they expect. The expected quality only depends on a buyer's belief and – with our parameters – is given by $E[Q(\beta)] = 37.125 + 25.25\beta$. Following these considerations, dividing a buyer's bid through his valuation provides an estimate of the quality he expects to receive from the current seller. Rational expectations should range between 37.125 – the average quality of a bad seller – and 62.375 – the average quality of a good seller. Theoretically, with the information contained in the feedback system, buyers should be able to form accurate beliefs about their current trading partner and thereby also about the expected quality.

Figure 3.4 shows the distribution of buyers' expectations about quality for good and bad sellers in both treatments.[51] Looking at the upper panels, it can be seen that in Treatment *All* average expected qualities are close to the theoretical values for good and bad sellers. In addition, both distributions peak around the true expected quality. In fact, mean expected quality is significantly larger when matched with a good than with a bad seller (MWU; p = .01). Thus, using the information contained in the long-memory feedback system, buyers seem to be able to differentiate between good and bad sellers.

Regarding our second treatment, both distributions have their peak around the average expected quality of a bad seller but distributions of mean expected qualities are weakly

[51] For both treatments only around one third of expected qualities are between 37.125 and 62.375 (in *All* 31.8% and in *Last* 39.8%). In *All* 31.6% are below 37.125 and 32.0% exceed 62.375. In *Last* 38.4% are below 37.125 and 21.7% exceed 62.375.

significantly higher for good sellers (MWU; p = .05).[52] However, expected qualities for good and bad sellers differ only by 9.3 in Treatment *Last* in comparison to 24.1 in Treatment *All*. This difference between expectations for good and bad sellers is significantly larger in Treatment *All* (MWU; p = .01). Hence, due to the higher informativeness of the long-memory feedback system buyers seem to be better able to differentiate between good and bad sellers and therefore have on average more accurate expectations.

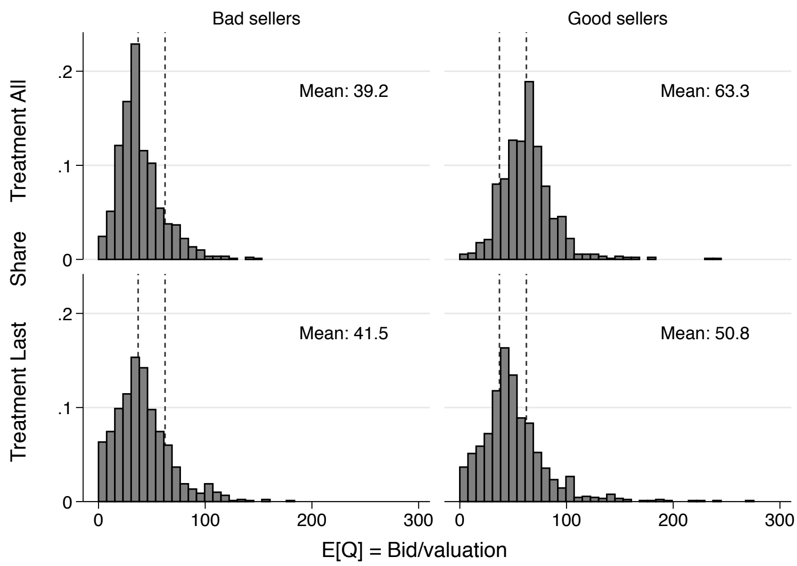


Figure 3.4: Histogram of expected quality (in percent) for bad and good sellers in both treatments. Dashed vertical lines indicate average quality of a bad seller (37.125) and average quality of a good seller (62.375).

Next, we want to test how buyers use the observed feedback ratings in terms of prior beliefs when making their bidding decision. For this reason, we test how prior beliefs – calculated based on the available feedback information – influence the expected quality. Table 3.4 shows the results of random-effects regressions on expected quality with prior beliefs as independent variable for both treatments. As mentioned earlier, the theoretical relationship between expected quality and prior belief is given by $E[Q(\beta)] = 37.125 + 25.25\beta$. Wald tests show that for the first treatment the coefficient of the prior belief and the constant are not significantly different from the theoretical

[52] Comparing the mean expected quality for bad sellers across treatments shows no significant difference (MWU; p = .46). With regard to good seller types mean expected quality is significantly larger in Treatment *All* (MWU; p = .03).

values. In contrast, for Treatment *Last*, the coefficient for the prior belief is significantly larger while the constant is significantly smaller than the theoretical values. These findings show that in a long-memory feedback system, buyers' bidding behavior is – at least on average – consistent with the theoretical predictions while this is no longer the case in the short-memory feedback system.

Expected quality	***All***	***Last***
Prior belief	24.495***	32.258***
	(12.532)	(15.078)
Constant	39.260***	30.883***
	(28.501)	(19.945)
R^2 within	0.00682	0.106
R^2 between	0.847	0.387
R^2 overall	0.226	0.127
Wald test belief = 25.25	$p = .70$	$p < 0.01$
Wald test constant = 37.125	$p = .12$	$p < 0.001$
N	1800	1800

Table 3.4: Random-effects regression with expected quality as dependent variable. Standard errors clustered on matching group level. Theoretical prior beliefs are calculated based on the number of feedbacks indicating high and low quality in Treatment *All*. For Treatment *Last*, these beliefs are simply the most recent feedback rating. *t* statistics in parentheses; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Purchasing probabilities and buyers' profits

As discussed before, when only the most recent feedback rating is displayed buyers are less able to distinguish between good and bad sellers and submit significantly lower bids when matched with a good seller compared to the treatment in which all previous feedback ratings are available. As the BDM price is randomly drawn, this change in bidding behavior has a direct effect on buyer's probability to purchase the offered product. Figure 3.5 shows that buyers in both treatments are equally likely to buy from bad sellers. However, in Treatment *All* buyers are significantly more likely to purchase from good sellers than in Treatment *Last*.

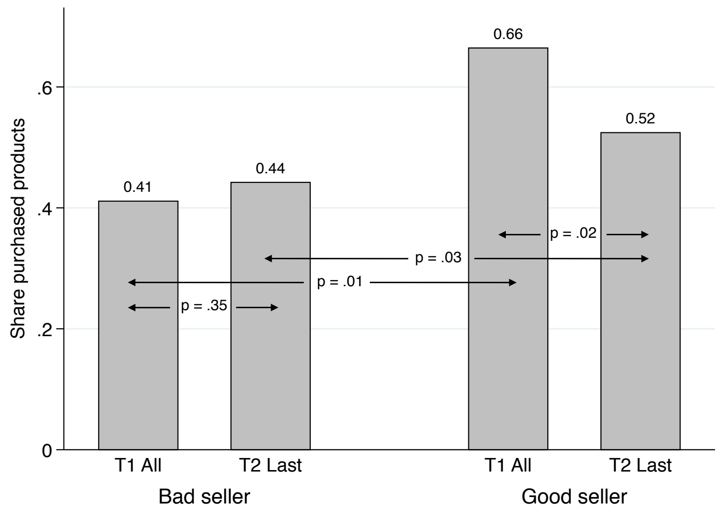


Figure 3.5: Share of purchased products from good and bad sellers in both treatments. *P*-values are derived from two-sided Mann-Whitney *U* tests.

Hence, as a consequence of their more cautious bidding behavior, buyers in Treatment *Last* forego potential profits from trading with good sellers. These missed opportunities are also reflected in buyers' overall profits, which are significantly larger in Treatment *All* (17.47 €) than in Treatment *Last* (15.64 €; MWU; p = .03).[53]

3.6 Discussion and Conclusion

In this paper we compare two different feedback systems; a long-memory feedback system where all previous ratings from earlier buyers about a seller are displayed and a short-memory feedback system where only the feedback rating of the most recent buyer is displayed. This design modification of sellers' feedback profiles should have an effect on how buyers give and interpret feedback ratings. While in the first system, it is sufficient that buyers simply report the received quality to achieve fully informative feedback, in the latter one way to keep all information in the system is to report updated beliefs about a seller's type. Thus, we are interested in whether subjects are able to coordinate on a fully informative feedback language given the implemented feedback system.

[53] In addition to their earnings from trading, all subjects received a show-up fee of 600 ECU = 6 €.

Our results show that in the treatment where all previous feedbacks are displayed, buyers are able to coordinate on the optimal feedback rule as most feedback ratings mirror exactly the received quality. In the short-memory feedback system, however, many traders also apply this feedback language instead of reporting an updated belief. Only very few feedback ratings are in line with Bayesian updating. This non-optimal feedback language leads to a loss of information because only the most recent signal about a seller is passed on from one buyer to the next. As a consequence, buyers in the short-memory feedback system have more difficulties to differentiate between good and bad sellers in comparison to buyers who receive all previous information. The difference between average bids in Treatment *All* for good and bad sellers is more than twice as large as in the Treatment *Last* (24.1 vs. 9.3). Furthermore, regression results suggest that the most recent rating is not interpreted as a prior belief when building expectations about quality and submitting bids. In contrast, in the long-memory system, theoretical prior beliefs derived from all available feedback ratings have the expected influence on expected quality and bidding behavior. Overall, these differences between treatments result in a significantly lower probability to buy from good sellers in Treatment *Last* (52%) than in Treatment *All* (66%). Thereby, buyers in the short-memory feedback system forego potentially beneficial trades and have significantly lower profits than their counterparts with full feedback history.

One possible explanation for the prominence of the reporting feedback rule in the short-memory feedback system might be that simply transmitting the received quality is more intuitive and less cognitive demanding than calculating an exact posterior belief. Furthermore, most real-world online reputation systems rely on cumulative feedback histories. Therefore, buyers in our experiment might be already more familiar with simply reporting quality. An additional advantage of the long-memory feedback system is that the reporting feedback language is less prone to rating errors than the updating feedback language. Because succeeding buyers interpret ratings as individual signals of a seller's type it suffices that feedback and quality are from the same interval, i.e. quality below (above) 50 receives a rating of below (above) 50. In fact, almost all feedbacks (94.3%) in Treatment *All* correspond to this more simplified rule. For updating feedback no such simplification exists, as there is only one updated belief that exactly reflects the currently available information. For this reason, minor deviations from the optimal feedback language do not necessarily lead to biased beliefs in the first but immediately do so in the second treatment. Thus, it might be easier for subjects to coordinate on the reporting feedback rule in long-memory feedback systems. Furthermore, even if one subject deviates from the reporting feedback rule in the long-memory feedback system, it is just one biased rating out of many while in the short-memory feedback system one biased rating will affect all future ratings.

In our experiment we have used the same question to elicit reporting and updating feedback. Future research may attempt in the short-memory feedback system to push subjects into the direction of updating feedback by asking explicitly for a buyer's beliefs about the seller's type. An alternative idea for improvement would be to design a short-memory feedback mechanism that elicits reporting feedback but then automatically calculates the current belief and passes it on to upcoming traders. Such a system would combine the intuitive feedback rule of the long-memory feedback system with the 'all information in one rating' advantage of the short-memory system.

3.7 APPENDIX A – LONG VS. SHORT-MEMORY FEEDBACK SYSTEMS IN MORAL HAZARD SETTINGS

Adverse selection and moral hazard often characterize online trading environments. In some settings both forms of these informational asymmetries are simultaneously present although typically one of the two is more prevalent (Dellarocas 2005b; Dellarocas 2006). Reputation systems are helpful in mitigating both types of trust issues. In contrast to the adverse selection setup, where feedback serves as a signaling device and thus helps traders to learn about initially unknown characteristics of their trading partners, feedback ratings can deter moral hazard by acting as a sanctioning device. Here, market participants punish traders for earlier opportunistic behavior by lowering their willingness to buy and also to pay for the offered products. In its most extreme case, traders immediately stop trading with someone who has received a negative feedback. Thereby, if the overall losses from missed trades outweigh the current gains from opportunism, feedback systems are able to promote honest and trustworthy trading behavior (Bar-Isaac and Tadelis 2008). Hence, with the help of such trigger strategies, feedback systems are able to establish a cooperative market environment for an infinite but also finite trading horizon (Kreps et al. 1982; Kreps and Wilson 1982; Milgrom and Roberts 1982). With regard to the comparison of short and long-lasting feedback systems, a trigger strategy can be realized with both of them. In contrast to the adverse selection case, it is not even necessary that traders use different feedback language between. In both systems, negative experiences can be passed on in ratings by e.g. simply reporting the received quality. Based on such a negative feedback following traders may implement the punishment and stop trading with these convicted black sheep. We run two exploratory sessions to test and compare the performance of short and long-memory feedback systems also in a market with moral hazard.

In order to study these two feedback systems in a moral hazard context, we use a similar design as for the adverse selection setup. We have matching groups of 8 subjects – 4 sellers and 4 buyers. Roles are randomly assigned at the beginning of the experiment and remain fixed for all periods. In each period, buyers and sellers are matched into pairs, with the restriction that the same pair does not interact in consecutive periods. Again, each seller offers a good of unknown quality $Q \in [0;100]$ at cost $c(Q) = Q$ to the buyer. Buyers have an individual valuation $v \sim U(100,300)$ for the offered product, which is randomly drawn and disclosed to them at the beginning of each period. Simultaneously, the buyer chooses his maximum bid and the seller determines the quality that is shipped to the buyer in case the product is sold. The product is sold if the bid at least matches the randomly drawn price *p* otherwise no trade occurs and both parties receive a profit of 0.[54] In case the product is sold, profits are $\pi_B = v \times Q / 100 - p$ for the buyer and $\pi_S = Q - p$ for the seller. After the trade, the buyer is asked to leave feedback on a continuous scale between 0 and 100. The difference between our two treatments is again the number of feedback ratings that are shown to the buyer before he takes his bidding decision. In treatment *'All MH'*, a long-memory feedback system is in place and buyers observe each individual rating whereas in treatment *'Last MH'* feedbacks are short-memory and buyers only get to know the most recent feedback rating. To elicit feedback ratings we again ask in both treatments to "Please rate the seller!" As additional information, buyers learn the number of periods in which the seller did not sell his product. Overall, the experiment lasted for 60 periods. Valuations, prices, and buyer-seller matching were drawn in advance to keep these factors constant across all matching groups. For each treatment we ran one session with four independent matching groups leading to 1920 seller-buyer observations.

Looking at sellers' quality decisions over time in Figure 3.6 we observe that the average level of quality is lower when a short-memory feedback system is in place (82 vs. 68). This suggests that at least some sellers expect to get away with lower quality when buyers only observe a short feedback history.[55] Regression results in Table 3.5 confirm this finding as the treatment variable for the short-memory feedback system has a negative influence on the overall quality level. This is also the case when we consider only the quality level of products, which are actually sold.

---

[54] In contrast to the adverse selection case the random price *p* was not drawn from a uniform distribution between 0 and 200. To guarantee that sellers cannot make losses the price ranges between 100 and 300. Values between 100 and 150 are drawn with 60% probability, values between 151 and 200 with 20%, values between 201 and 250 with 15%, and values between 251 and 300 with 5%. With this right-skewed distribution of prices, we increase the likelihood that the product is sold and thus the number of transactions and feedbacks.

[55] Because there is a significant endgame effect in the last five periods, all further analyses include only the first 55 periods.

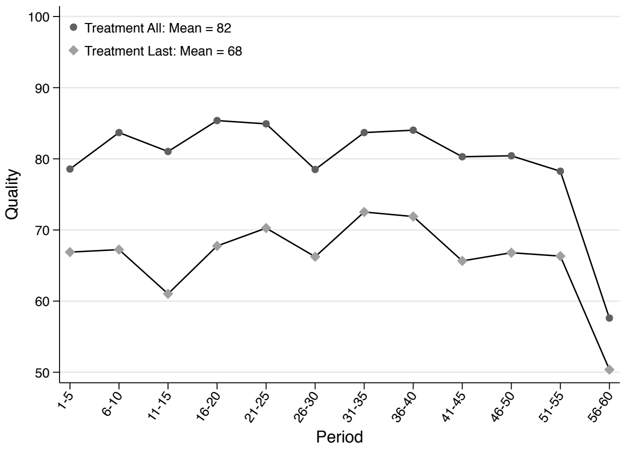


Figure 3.6: Sellers' quality choices across treatments.

Regarding buyers' behavior we observe that submitted bids are very similar across treatments (167.4 vs. 165.1). However, the share of buyers who decide not to place a bid increases from 14% to 24%. Regression results in Table 3.6 confirm that bidding behavior does not differ much between treatments. The treatment effect is weakly significant for the decision whether to submit a bid or not but not for the height of the bid conditional on bidding. In addition, we find no evidence that the most recent feedback rating is used differently in long and short-memory feedback systems, as the interaction effect is also not significant.

	Quality	**Quality sold**
T4 *Last*	-12.764**	-12.075**
	(-2.615)	(-2.736)
Most recent feedback	0.107**	0.167***
	(2.636)	(3.787)
# unsold products	-0.304	-0.399
	(-1.309)	(-1.229)
Period	0.165	0.239
	(1.217)	(1.527)
Constant	74.203***	68.699***
	(15.034)	(17.579)
R^2 within	0.004	0.011
R^2 between	0.454	0.605
R^2 overall	0.307	0.260
N	1669	761

Table 3.5: Random-effects regression with quality as dependent variable. Standard errors clustered on matching group level. *t* statistics in parentheses; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Also feedback behavior is rather similar across treatments. In both feedback systems buyers most often simply report the received quality. In Treatment *All* (*Last*) 77.4% (82.0%) of all feedbacks are equal to quality. Interestingly, if buyers deviate from reporting quality, they tend to give lower feedback ratings in both treatments (*All*: 23.3%; *Last*: 14.0%).

	Bid yes/no		Bid continuous	
T4 *Last*	-0.443^{+}	-0.169	-0.701	2.318
	(-1.714)	(-0.433)	(-0.104)	(0.189)
Valuation	0.012***	0.018***	0.533***	0.574***
	(14.828)	(14.890)	(11.760)	(12.749)
Most recent feedback		0.022***		0.412***
		(8.856)		(3.826)
Most recent feedback x T4		0.001		-0.034
		(0.417)		(-0.346)
# unsold products		-0.078***		-1.634***
		(-5.314)		(-4.688)
Period	-0.008**	0.031**	0.347**	0.967***
	(-3.099)	(3.238)	(3.148)	(4.349)
Constant	-0.741**	-3.005***	42.831***	10.729
	(-3.131)	(-8.186)	(5.317)	(0.655)
Log likelihood	-643.6	-425.4		
R^2 within			0.415	0.572
R^2 between			0.045	0.005
R^2 overall			0.318	0.445
N	1728	1669	1393	1338

Table 3.6: Random effects probit regressions with bid (0 = no; 1 = yes) as dependent variable in the first two models. Random effects regressions with continuous bid as dependent variable in the last two models. Standard errors clustered on matching group level in models with continuous bid. *t* statistics in parentheses; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Finally, we turn to the comparison of market performance measures. As a consequence of the changed bidding behavior, the share of sold products is lower in the short-memory feedback system (*All*: 52%; *Last*: 41%). This in turn has an effect on the overall level of efficiency.[56] According to the first regression in Table 3.7, efficiency decreases by almost 14% when only the most recent feedback system is displayed. This

[56] We measure efficiency as the ratio between the actual surplus and the maximum possible surplus. The maximum surplus is equal to the buyer's valuation minus the price. We leave out auctions where the valuation is smaller than the randomly drawn price because then trade cannot be efficient.

loss of welfare has to be covered mostly by buyers whose profits decrease from 27.4 ECU to 19.3 ECU, a reduction of 30%. Sellers' profits, in contrast, shrink only by 13% from 25.5 ECU to 22.1 ECU. According to models 2 and 3 in Table 3.7 only the profit reduction for buyers is significant on a 10% level.

	Efficiency	**Buyer profits**	**Seller profits**
T4 *Last*	-0.136**	-8.093+	-3.433
	(-2.734)	(-1.760)	(-1.303)
Period	0.002	0.151+	0.009
	(1.641)	(1.713)	(0.147)
Constant	0.518***	23.199***	25.266***
	(11.007)	(4.458)	(14.894)
R^2 within	0.004	0.002	0.001
R^2 between	0.240	0.200	0.058
R^2 overall	0.027	0.009	0.002
N	1272	1728	1728

Table 3.7: Random effects regression with efficiency, buyers' period profits, and sellers' period profits as dependent variable. Standard errors clustered on matching group level in models with continuous bid. *t* statistics in parentheses; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

To summarize, we note that long-memory feedback systems outperform feedback systems, where only the most recent rating is displayed. In the short-memory feedback system, sellers ship lower quality and buyers react by submitting bids less often. However, none of the buyers – neither in T3 or T4 – follows a harsh trigger strategy as they all buy from sellers who received feedback ratings of less than 100. Overall, the more cautious bidding behavior in the short-memory feedback system does not fully prevent buyers from sellers' opportunism, as they have to bear the major part of the welfare losses due to lower quality levels.

3.8 *APPENDIX B – INSTRUCTIONS*

Instructions (Treatment All)

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.

In this experiment there are buyers and sellers. All participants are in the role of buyers, sellers are played by the computer. There are as many computerized sellers as there are buyers. The experiment consists of 60 rounds. At the beginning of each round, each buyer is randomly matched with one computerized seller. It is guaranteed that a buyer will not interact with the same seller in two consecutive rounds.

In each round, each computerized seller offers a good with an unknown quality to the matched buyer. The quality of the good lies between 0 and 100% and is only revealed to the buyer if he purchases the good. In the following, quality between 0 and 49 is referred to as low quality and quality between 50 and 100 as high quality. Each buyer is randomly assigned a valuation for the good, which lies between 100 and 300 ECUs. The valuation represents the value of the good for the buyer if the quality is 100. Valuations are newly randomly drawn in each round and every integer value between 100 and 300 has the same probability to be selected.

There are two different types of computerized sellers: high sellers and low sellers. There are as many high sellers as there are low sellers. As a buyer, you are matched with sellers of both types over the course of the experiment.

High sellers are more likely to ship high quality than are low sellers. Specifically, high sellers ship quality between 0 and 49, each number with 0,5% probability and between 50 and 100, each number with 1,5% probability. Low sellers ship quality between 0 and 49, each number with 1,5% probability and between 50 and 100, each number with 0,5% probability. Overall, a high seller ships quality between 0 and 49 with 25% probability and between 50 and 100 with 75% probability. A low seller ships with 75% probability quality between 0 and 49 and with 25% quality between 50 and 100. This means that on average, a high seller ships quality of 62 and a low seller ships quality of 37. In the experiment, there are as many high sellers as there are low sellers.

Probabilities for each quality level conditional on seller type

			Seller type	
			Low	High
Quality	Low	0… 49	1,5%	0,5%
	High	50… 100	0,5%	1,5%

Each round consists of two stages, the **purchase stage** and the **transaction stage**.

In the **purchase stage**, buyers get to know their own valuation and see all feedback ratings which preceding buyers have left for this seller. Furthermore, there is a profit calculator where buyers can enter hypothetical prices and quality levels. The calculator displays the hypothetical profit for the entered values given the buyer's valuation in the current round. In this stage buyers submit their maximum price they are willing to pay for the offered good. The maximum price must be at least 0 but may not exceed 300. This maximum price is then compared to a random price, which is randomly drawn between 0 and 200 where each integer value has the same probability to be selected:

1. If the maximum price is smaller than the random price, the buyer does not purchase the good and pays nothing.
2. If the maximum price is equal or larger than the random price, the buyer purchases the good and pays the random price.

In case the buyer does not purchase the good, he pays nothing, earns a profit of 0 ECUs in this round and a new round starts immediately.

In case the buyer purchases the good, the buyer proceeds to the **transaction stage**. In this stage the buyer receives the good and learns the quality of the good. The actual value of the good for the buyer equals his/her valuation for the good times the received quality. Thus the actual value of the good for the buyer is 0 ECU if the received quality is 0, and equal to his/her valuation if the received quality is 100.

The period profit in ECU for a buyer who purchases the good is:

*Buyer's period profit = (Quality / 100) * Valuation – Random price*

Furthermore, in the transaction stage the buyer is asked to leave a feedback for the seller on a scale from 0 to 100. This feedback and all previous feedbacks are then shown to buyers that interact with this computerized seller in the following round.

After the transaction stage the round ends and a new round with newly matched buyer-seller pairs begins as described above.

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

3.9 APPENDIX C – SCREENSHOTS

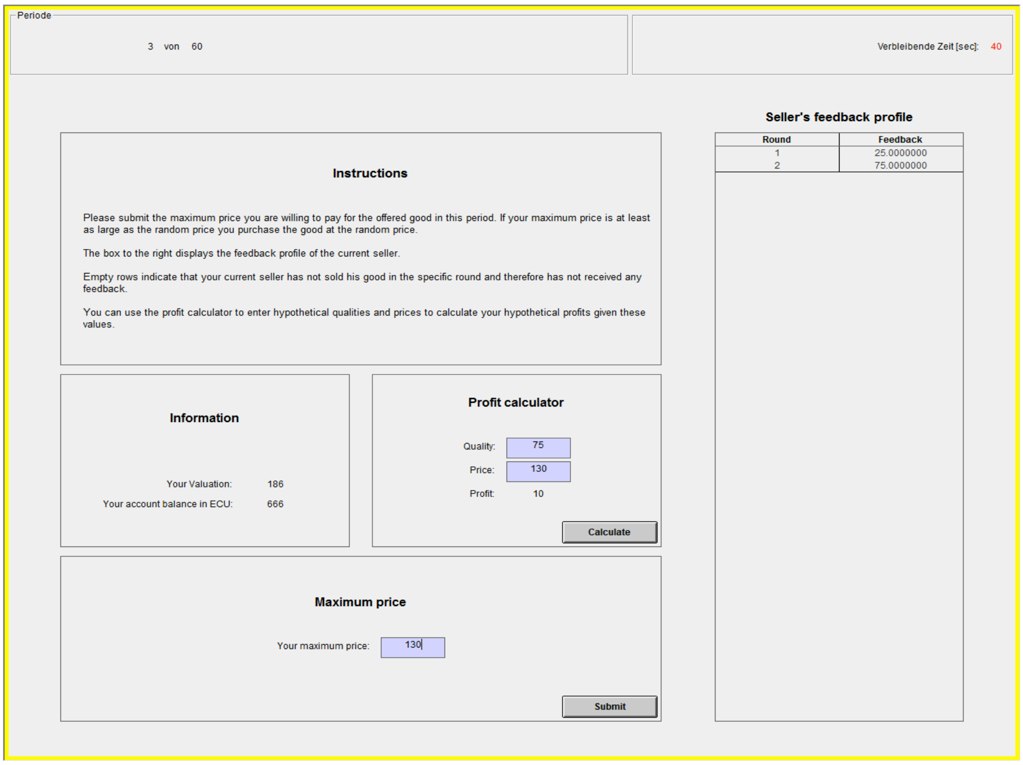


Figure 3.7: Purchase decision

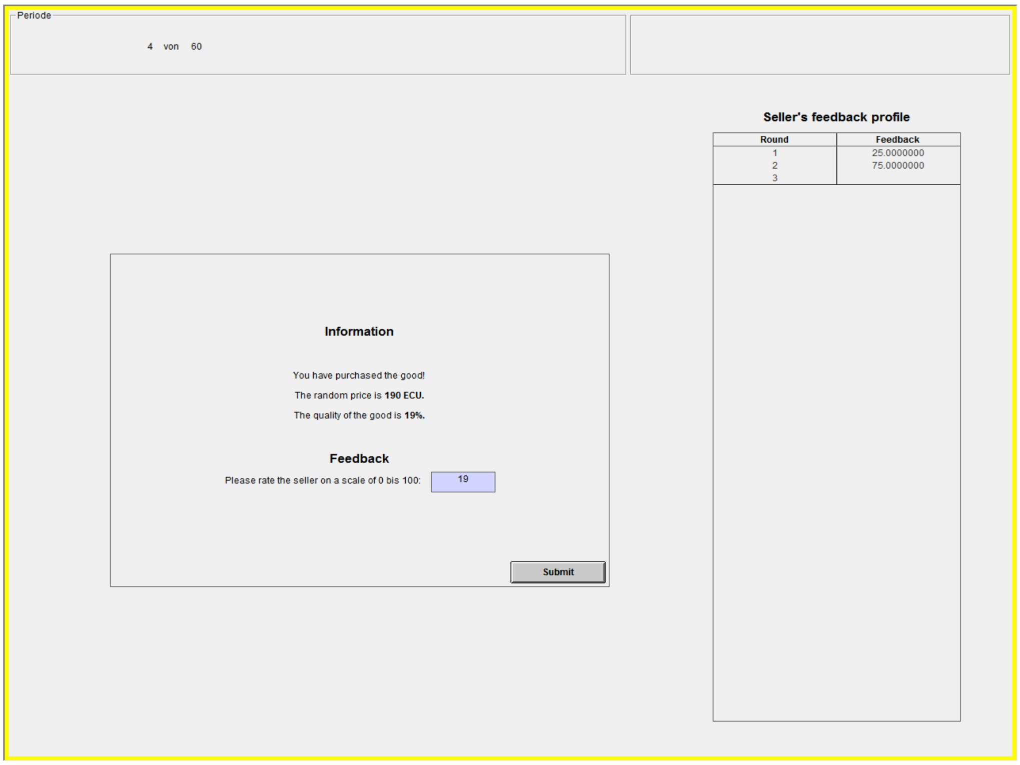


Figure 3.8: Feedback rating

Chapter 4

The Influence of Social Identity and Trading Frequency on the Provision of Feedback

4.1 Introduction

In the last two decades the rise of the Internet has helped to transform the world into a global marketplace.[57] Businesses and consumers from diverse markets around the world are brought together online to trade a huge variety of goods and services within and across borders. For example, on eBay – one of the leading online auction platforms for customer-to-customer and business-to-customer sales operating in more than 30 countries worldwide – already 20% of the trading volume takes place across borders (eBay Enterprise 2014).[58] Accordingly, recent market research by the electronic payment system PayPal reports that already 34.1 million private consumers from the US purchased goods online from foreign countries – spending more than $40 billion in

---

[57] This paper is single-authored. I thank Axel Ockenfels and Gary E. Bolton for their significant contributions to the initial idea, the experimental design and the hypotheses. Also, I would like to thank Gary E. Bolton for his hospitality while conducting the experiments at the University of Texas at Dallas. In addition, I received very helpful comments from Bettina Rockenbach, Ben Greiner, Mattia Nardotto, Peter Werner, Christoph Feldhaus, Florian Goessl, David Kusterer, Anne Schielke, Tobias Stangl. I am also thankful to Michael Cristescu, Owen Ma and Oliver Baker for their help with programming, testing and organizing the experiments. Financial support of the German Research Foundation (DFG) through the Research Unit "*Design & Behavior – Economic Engineering of Firms and Markets*" (FOR 1371) is gratefully acknowledged.

[58] To facilitate and expand international trade even further, eBay recently launched a global shipping program, which handles shipping, customs forms, and import charges for sellers in cross-border transactions. Similar to eBay, Amazon – one of the largest online retailers worldwide – and its third-party marketplace currently have websites in 13 countries and provide shipping to even more than 60 countries for international transactions. For B2B transactions, alibaba.com, offers a wholesale platform for suppliers – especially from traditional manufacturing countries such as China, India, and the USA – to find international companies for their manufactured goods.

2013 alone – and projects that this amount will double within the next five years (Paypal 2013).[59]

In line with this development, economic studies found that Internet penetration within a country is positively linked to the growth of import and export of goods and services (Freund and Weinhold 2002, 2004). Furthermore, geographic distance and national borders between trading partners matter much less in online than in comparable offline transactions (Hortaçsu, Martínez-Jerez, and Douglas 2009; Lendle et al. 2012). Besides reducing search and communication costs, the Internet also promotes international trade by eliminating trust barriers between geographically separated sellers and buyers. In this regard, reputation systems based on voluntary ratings are a key feature of eBay and other online trading platforms to establish trust and trustworthiness among their community members (Dellarocas 2003). A good reputation substitutes for a personal, repeated relationship and mitigates the distance effect between trading partners in national and international trades (Hortaçsu, Martínez-Jerez, and Douglas 2009; Lendle et al. 2012).

Voluntary, subjective feedback ratings, however, have the drawback that their provision as well as their content can be biased for strategic and psychological reasons (Dellarocas and Wood 2008; Bolton, Greiner, and Ockenfels 2013; Bolton, Kusterer, and Mans 2014). This, in turn, reduces the informativeness of a feedback system, sets lower incentives for trustworthy behavior, and decreases market efficiency, eventually. The geography of trade may provide additional reasons for selective feedback provision and thus may also lead to reduced and biased feedback information. Field data from eBay, including more than 600,000 completed listings, provides exploratory evidence that the provision of feedback is influenced by the geography of trade.[60] Despite the public good nature of feedback information, 72% of sellers and buyers leave a rating when trading within their own domain. In contrast, when trading partners come from different countries, feedback provision rates drop to 50% for sellers as well as for

[59] A similar development is expected for the European retail market (EMOTA 2014).

[60] This dataset was compiled by and already used in Bolton, Greiner, and Ockenfels (2013) to analyze feedback reciprocity. It consists of all successful transactions from six different categories on eight international domains in November and December 2006.

buyers.[61] Consequently, the share of auctions in which neither the seller nor the buyer leaves feedback is twice as large for international than for national trades (44% vs. 22%). Overall, this rating reluctance then leads to a lower level of informativeness within the feedback system.

In this study, we take a closer look at individual rating behavior and analyze through which channels the geography of trade affects the provision of feedback. More specifically, we investigate two prevalent and essential differences of national and international trades – social identity and individuals' trading frequency on a particular market (or market participation) – which both may provide an explanation for the observed variance in the provision of feedback.

First, a crucial difference between within- and cross-border transactions is that in the former traders typically have the same nationalities while nationalities presumably differ in the latter. Sharing the same nationality, traders may perceive their domestic trading partners as members of the same social group and thus assume a shared social identity (ingroup). In contrast, in international transactions foreign trading partners may be recognized as members of a different social group (outgroup). Following studies from social psychology and economics, which demonstrate that social identity influences reciprocity in a way that among ingroup members, rewarding good behavior is more likely while punishing bad behavior is less likely, we argue that social identities of seller and buyer also affect the provision of feedback conditional on the transaction experience. We expect that a shared social identity strengthens the reward but weakens the punishment motive in the provision of feedback. Hence, we expect that in transactions between ingroup members, satisfied traders are more likely to leave a feedback rating while unsatisfied buyers remain silent about their negative transaction experiences more often.

---

[61] Interestingly, other patterns of feedback giving can be found in national and international transactions. For example, reciprocity is an important determinant for the provision of feedback. The share of auctions with mutual feedback is 65% when trading partners come from the same domain and 44% when they come from different domains. If buyers and sellers gave feedback independently of each other, one would expect the share of mutual ratings to be 72% x 72% = 52% for national trades and 50% x 50% = 25% for international trades. Accordingly, one-sided feedback occurs in only 13% (12%) of all within-border (cross-border) trades. Another similarity of national or international trades is that, in case of mutual feedback, buyers are twice as likely as sellers to give feedback first (66% and 65%, respectively).

The second factor that influences feedback provision refers to a trader's frequency to participate in a specific national or international market. In online reputation systems, feedback information constitutes a public good and individuals are more willing to contribute if they participate more often and thus benefit more from this public good (e.g. Fisher et al. 1995). For example, on eBay more active members are also more likely to leave a feedback rating (Dellarocas and Wood 2008; Jian, Mackie-Mason, and Resnick 2010). Similarly, when trading on several – national and international – markets, individuals might provide feedback based on the frequency with which they trade on a particular market. Regarding E-commerce, although there are also professional international traders, for most sellers and buyers cross-border trades are rather the exception than the rule.[62] Accordingly, this public good rationale could also explain why feedback provision is more widespread in within-border trades.

With observational data, it is impossible to disentangle these two factors because they usually coincide in national transactions. In our eBay dataset, for example, even among those traders who have a history to trade nationally and internationally, domestic trades account for nearly two thirds (64%) of all observed transactions. Another crucial disadvantage of empirical eBay studies is that true transaction experience is unobservable. Even focusing on standardized goods cannot fully solve this omitted variables issue because additional factors, which also make for a satisfactory transaction experience (such as product description and presentation, communication with customers, etc.), are hard to measure objectively (Resnick et al. 2006). A similar problem arises for further possible differences between national and international transactions, which may contribute to the lower feedback rates in cross-border trades, e.g., different quality standards, longer delivery times, miscommunication due to language barriers, and self-selection into international trades.

For these reasons, we rely on an experimental design, where we are able to vary social identity and trading frequency exogenously while keeping all other factors constant or under control. Our workhorse is a repeated trust game which consists of 40 periods with two stages: (1) a trading stage where a buyer (trustor) decides whether or not to buy a good offered by a seller (trustee) and the seller chooses the quality level of the good

[62] See report on Internet use and skills by the European Commission (2013).

(amount returned) and (2) a feedback stage where the buyer has the option to leave a public feedback rating – positive or negative – about the seller at a small cost.[63] Trading is efficient and increases social welfare, as seller's costs of quality are smaller than buyer's valuation for quality. We run this trust game simultaneously in laboratories at two different universities in Germany and the US. To investigate the effects of social identity on trading behavior in general and buyers' feedback provision in particular, we match subjects within and across universities. Moreover, to analyze trading frequency as factor for feedback provision, we create two separate markets. While sellers always stay on one market, buyers switch between markets. Thereby, each buyer is assigned a home market, where he conducts most of his trades and an away market, which he visits less frequently. Overall, we obtain a 2x2 within-subject design (ingroup vs. outgroup trade and home vs. away market), which allows separating the effects of social identity and trading frequency on the provision of feedback.

Our experiment contributes to the emerging literature on social and psychological motives to share information in online markets and communities. With the help of survey data, Dholakia, Bagozzi, and Pearo (2004) show that identification with an online community and stronger norms within this group have a positive effect on contribution behavior. Several studies analyze the provision of ratings on MovieLens – an online movie recommendation platform. For example, Beenen et al. (2004) find that individuals can be motivated to submit reviews by setting specific individual or group goals and by making the uniqueness and value of a contribution more salient. Similarly, in artificially created discussion groups of MovieLens members, those being reminded of their unique taste in movies are more active in the discussion (Ludford et al. 2004). Chen et al. (2010) implement a large-scale field experiment to investigate how social comparison information affects users' contributions on MovieLens. Their findings demonstrate that social information about the median member's review behavior induces below median users to increase their contributions while those already above the median reduce rating effort. Analyzing customer restaurant reviews on different platforms, Wang (2010) concludes that prolific reviewers care more about their online

[63] The accumulated numbers of positive and negative ratings are stored in a sellers' feedback profile, which is shown to buyers before they make their purchase decision. A more detailed description of the experimental design is given in the following section.

social image and for this reason submit ratings to establish a good social image within the online community.

Social identity has received much attention since Akerlof and Kranton (2000) introduced this socio-psychological concept into the field of economics. However, there are only few previous studies focusing on the economic effects of social identity in market environments. In an experimental study, Li, Dogan, and Haruvy (2011) show that social identity affects the selection of trading partners and the determination of prices in repeated transactions in an oligopolistic market. Sellers are more likely to make offers to buyers who belong to the same social group. Analogously, buyers also prefer offers from ingroup sellers. As a consequence, outgroup sellers choose considerably lower prices than their ingroup counterparts. In a repeated trust game, Heap and Zizzo (2009) find evidence for outgroup discrimination in terms of lower trust and trustworthiness but no opposite effect of ingroup favoritism. Investment decisions as well as return rates are significantly lower in outgroup matches than in a baseline without induced group affiliations while there are no differences between ingroup matches and the baseline case. Accordingly, they find that inducing group identities rather has a negative effect on welfare. With our study, we attempt to further investigate the role of social identity in market settings. In particular, we are interested in how social identity affects individuals' willingness to leave feedback about other market participants and how this may translate into differences in trading behavior and in market performance in the long run.

The results of our experiment show that a common social identity increases the probability that extreme – positive or negative – transaction experiences are reported. For example, in ingroup transactions, buyers' reporting probabilities are around 25% higher when quality is either 10% or 90% of the maximum attainable quality level in comparison to seller-buyer pairs who are from different universities. In contrast, buyers' trading frequency on a particular market does not seem to affect their willingness to leave a feedback rating. In addition, none of our treatment variations influences the content of the feedback rating. In this respect, buyers have a consistent way of mapping quality into either positive or negative ratings. With regard to trading behavior, a shared social identity increases buyers' trust only when they deal with a seller who has a low percentage of positive ratings. Buyers show the same level of trust towards ingroup and outgroup sellers with medium or high feedback scores. This suggests that the reputation mechanism helps to level buyers' trust towards ingroup and outgroup trading partners. At the same time and in contrast to other studies, sellers also do not discriminate against outgroup buyers and on average provide similar levels of quality in ingroup as well as outgroup transactions.

In the next section we describe our experimental design in more detail and derive our hypotheses based on the related literature. Following, we present the analyses and our main results. The last section discusses these results in regard to our hypotheses and concludes with possible implications for real-world feedback systems.

4.2 *EXPERIMENTAL DESIGN AND HYPOTHESES*

In order to investigate the different effects of social identity and market participation on trading behavior and feedback provision, we use a multi-period trust game with buyer-seller framing and a within-subject design. Participants in our experiment are recruited from the University of Cologne (UoC) and from the University of Texas at Dallas (UTD). Recruitment from two different universities ensures natural groups as a social identity manipulation for sellers and buyers.[64] At the same time, trading frequency is only varied for buyers by assigning each of them a home market and an away market. The difference between these two markets is that buyers have a higher probability (80%) to trade on their home market and thus interact more frequently with sellers from this market.

Figure 4.1 gives a detailed overview of our matching procedure. We use matching groups of 16 subjects in which half of the subjects are sellers and the other half are in the role buyers. Of these eight sellers and buyers four are from UoC and from UTD, respectively. Roles remain constant throughout the entire experiment. Based on their university affiliation sellers are assigned to one of two different markets with four sellers each: a Cologne market and a Dallas market. In contrast to sellers, who are

[64] In order to have identical instructions for participants at the University of Cologne and at the University of Texas at Dallas we first wrote a German version, which was then translated into English and proofread by a native speaker. This version was then translated back into German by a third person who is a native speaker in English and German. A comparison with the original German version showed only minor differences, which were then aligned. An English version of the instructions can be found in Appendix B.

assigned to one particular market for the entire experiment, buyers switch regularly between their home and their away market: with a probability of 80% (20%) buyers trade with a seller from their home (away) market.

In addition, we have two types of buyers. Cologne is the home market for type C buyers while Dallas is the home market for type D buyers. In each matching group there are four of each buyer type and for each type two subjects come from UoC and from UTD, respectively. In detail, we have two buyers from UoC whose home market is the Cologne market, two buyers from UoC whose home market is the Dallas market, two buyers from UTD whose home market is the Cologne market and two buyers from UTD whose home market is the Dallas market. As a side effect of this within-design, social and market identity overlap for half of the buyers as they share a common social identity with sellers on their home market (type C from UoC and type D from UTD). For the other half (type D from UoC and type C from UTD) social identity and market participation do not correspond as sellers on their home market are from the other university. This between-subjects variation combines the expected effects of social identity and trading frequency.

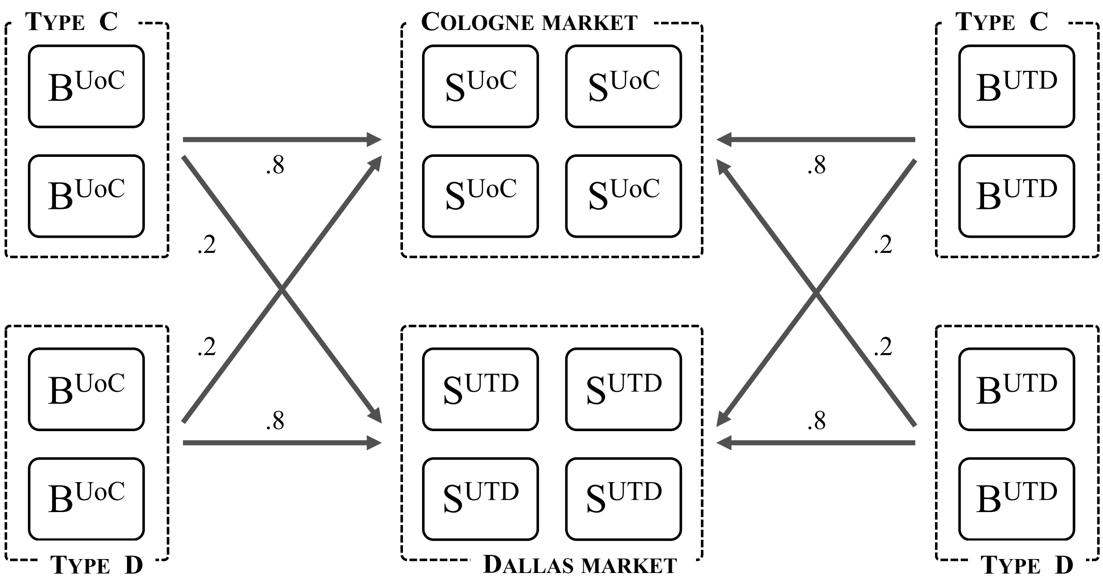


Figure 4.1: Matching procedure within each matching group of 16 subjects.

To ensure that there are an equal number of sellers and buyers on each market, we draw a random number between 0 and 1 in each period. If the value of this number is less than or equal to 0.8 all buyers trade on their home market and on the away market otherwise. Thereby, it is also guaranteed that there are two ingroup and two outgroup buyers on each market in every period. As a result, in following periods buyers from the in- and outgroup always profit in the same way from the provided feedback information. Hence, the motivation to provide more or better information on a particular market because there are more ingroup buyers trading on this market might drive behavior on eBay but not in our experiment. After buyers are assigned to one of the two markets, they are randomly matched to a seller on this market under the constraint that this is not the same seller-buyer pair as in the previous period. Each seller-buyer pair then plays a trust game with a one-sided feedback option for the buyer (see Figure 4.2).

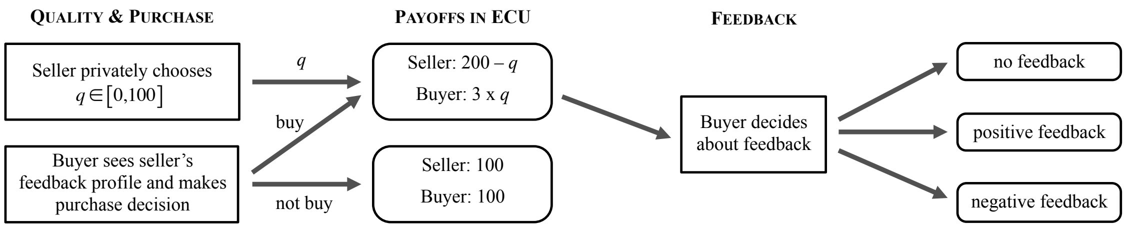


Figure 4.2: Trust game with one-sided feedback option

At the beginning of each period, buyers are informed whether they trade on their home market and thus also know whether the seller has the same social identity or not. Sellers, in contrast, only get to know whether the matched buyer is from the UoC or UTD. Thus sellers only know buyers' social identity but not their home market. In each period, both parties receive an endowment e = 100 ECU. The buyer then decides whether he wants to trust the seller and purchase his product at price $p = e = 100$. Simultaneously, the seller chooses a quality level q between 0 and 100 with linear costs $c = q$ for the product if it is sold. Hence, payoffs are 100 ECU for each player if the buyer does not purchase the good.[65] In case of a purchase, seller profit is given by his endowment e plus price p net of the costs for quality and the buyer receives a profit in height of the quality multiplied by three. Thus, there are gains from trade as welfare is maximally doubled when the seller ships full quality of 100. However, in this case the buyer receives three quarters of the total surplus. The equal split, where both parties receive a payoff of 150 ECU, is realized at a quality level of 50. After the buyer is informed about the received

[65] To accommodate for currency differences between Germany and the US, we use two different exchange rates in our experiment. 250 ECU = $1 for participants at the University of Texas and 350 ECU = 1€ for participants at the University of Cologne. The ratio of these two conversion rates correspond approximately to the exchange rate between $ and € of $1 = 0.74 € at the time of the experiment. Subjects were informed that we use different exchange rates that align earnings between participants from both universities. In addition to their period earnings from trading subjects received a show-up fee of $5 (3€).

quality, he has the option to leave a feedback rating (positive or negative) about the seller at the cost of 1 ECU. A seller's feedback information in the form of the number of positive and the number of negative ratings is presented to buyers who are paired with this seller in following periods.[66] Before a new period starts, the seller is informed about a newly received feedback rating. This stage game is repeated for 40 periods. Between periods, sellers and buyers are matched according to the matching procedure described above.

Hypotheses

With regard to traders' behavior in the first stage of the experiment, it is well known from social psychology that a shared group or social identity influences individual behavior in ways that favor ingroup members at the expense of outgroup members (Tajfel et al. 1971; Tajfel and Turner 1979). In recent years the concept of social identity has been incorporated into the field of economics and it has been shown that individuals' social preferences are group contingent and that people display more other-regarding behavior when matched with an ingroup member. Chen and Li (2009), for example, show in a wide class of sequential two-person games that subjects' charity concerns for their partners' earnings are larger in ingroup relationships, and that first and second movers are more likely to choose welfare maximizing options for ingroup matches. Similarly, in repeated trust games, trust and trustworthiness are larger between participants with the same social identity than with different identities (Heap and Zizzo 2009; Tsutsui and Zizzo 2014). Using ethnicity as natural identity, Fershtman and Gneezy (2001) and Haile, Sadrieh, and Verbon (2008) also find that trust is significantly lower towards members of a different ethnic group. Hence, although a feedback system should help to foster trust and trustworthiness among all subjects, we still expect to find similar identity-contingent behavioral patterns from buyers and sellers in our experiment.

Hypothesis 1a: Trust

Buyers are more likely to trust when paired with an ingroup than an outgroup seller and thus more purchases occur in ingroup matches.

[66] See APPENDIX C for screenshots of the experiment.

Hypothesis 1b: Quality

Sellers discriminate between ingroup and outgroup buyers and send higher levels of quality to ingroup trading partners.

In contrast to traders' social identities, trading frequencies are only privately known to the buyer in our experiment, and therefore cannot influence seller behavior. At the same time we do not expect that buyers' market participation alter their level of trust neither with a seller from their 'home' nor 'away' market. Hence, market participation should not affect first stage behavior of sellers and buyers but may alter the provision of feedback in the second stage.

The main focus of our paper lies on the provision of feedback ratings and how this is affected by social identity and trading frequency. In general, on electronic markets feedback systems are installed to mitigate trust issues, which arise due to asymmetric information between trading partners. Asymmetric information might be present in form of moral hazard, e.g. when sellers receive the payment first and then have the opportunity to deceive the buyers, or as adverse selection, where seller types are ex-ante unknown to buyers.[67] Therefore, feedback systems are installed as sanctioning and signaling device to establish trust and cooperation on markets with self-interested participants. Public ratings provide information about a transaction experience with a specific trader, thereby affect future customers' willingness to pay and thus reward (punish) him for his good (bad) behavior.[68] Accordingly, feedback information constitutes a public good because all traders can employ these publicly and costlessly available ratings to assess the risks associated with building new trade relationships. Despite its public good nature, on eBay around 50 to 70% of traders are willing to leave

---

[67] In some real-world settings both types of informational asymmetries are present meaning that traders behave opportunistically and also have different innate levels of ability. However, one of the two issues is usually more prevalent (Dellarocas 2005). For example, on eBay seller moral hazard is the main issue as sellers may be tempted to keep the sold product after receiving the payment. Customer reviews for services (angieslist.com), hotels and restaurants (yelp.com, Google reviews) or products (amazon.com) are examples where adverse selection is more prevalent and thus feedback systems rather have an informative function. In these cases relatively 'steady' but only privately known product attributes and qualities are being reviewed.

[68] With the help of a sophisticated experimental design Lafky (2014) is able to disentangle seller-centric and buyer-centric motives in a moral hazard setting. His results demonstrate that buyers are motivated to leave feedback by the urge to reciprocate and educate sellers but also by the concern for informing and protecting other buyers.

a feedback rating (Resnick and Zeckhauser 2002; Dellarocas and Wood 2008; Jian, Mackie-Mason, and Resnick 2010; Bolton, Greiner, and Ockenfels 2013). Further empirical evidence consolidates the assumption that users indeed rely on this public feedback information as, for instance, on eBay prices and sales probability are positively influenced by favorable feedback profiles (Ba and Pavlou 2002; Bajari and Hortacsu 2003; Houser and Wooders 2006; Livingston 2005; Melnik and Alm 2003; McDonald and Slawson 2002; Resnick and Zeckhauser 2002; Resnick et al. 2006). The same holds true for online customer reviews on products and services (Godes and Mayzlin 2004; Chevalier and Mayzlin 2006; Li and Hitt 2008; Zhu and Zhang 2010; Luca 2011).

Despite these positive effects, reputation systems based on voluntary feedback provision entail the problem that only specific transaction experiences are reported and that the available information only gives a biased view of the actual risks of trading with a particular seller.[69] Recent studies of eBay-like feedback systems and consumer product reviews suggest that raters follow a brag-and-moan approach and therefore frequently report extreme experiences but remain silent for mediocre outcomes (Anderson 1998; Hu, Pavlou, and Zhang 2006; 2009; Dellarocas and Wood 2008; Lafky 2014). Under the assumption that raters who submit a rating always report truthfully, Dellarocas and Wood (2008) estimate that on eBay satisfied buyers have the highest probability to give a corresponding feedback to their interaction partner followed by very dissatisfied customers leaving a negative rating. In contrast, mildly dissatisfying experiences are often not reported and thus do not show up in a seller's feedback profile. For our experimental setup, we expect a similar relationship between product quality and the likelihood to give a feedback rating.

Hypothesis 2: Feedback provision and transaction experience

Between transaction experience and the probability to provide feedback exists a u-shaped relationship meaning that buyers are more likely to report extreme (positive as well as negative) than mediocre experiences.

---

[69] In two-sided feedback systems for example, reciprocity plays an important role in the provision of ratings as the threat of retaliatory feedback inhibits many traders to post negative feedback after non-satisfying transactions (Dellarocas and Wood 2008; Jian, Mackie-Mason, and Resnick 2010; Bolton, Greiner, and Ockenfels 2013).

Looking at the data from eBay we observe that in cross-border transactions traders are less likely to submit a feedback rating, and thus twice as many auctions remain unrated. We presume that this decline in the number of ratings can be at least partially explained by differences in social identities between international trading partners as well as traders' lower frequency to interact on a foreign market.

As outlined above, encouraging sellers to behave honestly is one of the main functions of a reputation system. This can be achieved by rewarding good sellers with positive feedback as well as by punishing bad sellers with negative feedback as both result in monetary consequences for the seller in terms of increased or decreased future profit. By that means positive reciprocity reinforces good behavior while negative reciprocity discourages deceptive behavior. Moreover, reciprocal behavior is also affected by a shared social identity. Among members of the same group reciprocating or rewarding good behavior is more pronounced (Chen and Li 2009). At the same time, higher charity concerns also induce people to judge and punish ingroup members more leniently (Bernhard, Fischbacher, and Fehr 2006a; Chen and Li 2009; Mussweiler and Ockenfels 2013). Hence, we expect that a shared social identity between traders has a similar effect on the reciprocal nature of feedback provision.

Hypothesis 3: Feedback provision and social identity

Satisfied buyers are **more** likely to give feedback when paired with an ingroup than an outgroup seller, while unsatisfied buyers are **less** likely to give feedback when paired with an ingroup than an outgroup seller.

Besides the social relationship between sellers and buyers, we also investigate how a buyer's frequency to trade on a particular market – his market participation – and his willingness to provide a feedback rating afterwards are related. This market participation refers to the public good nature of ratings in feedback systems. Buyers who interact more frequently on a particular market benefit more from disciplined, trustworthy sellers and also from information provided on the market. Hence, those buyers benefit more from the feedback public good and therefore have a higher interest in establishing a norm of feedback provision among traders on their home market. In terms of public good experiments this can be compared to having a higher marginal per capita return (MPCR). This is the individual benefit a participant receives in return for each unit he or any other group member contributed to the public good. Higher MPCRs for all group members have a positive effect on their contributions (Isaac, McCue, and Plott 1985; Isaac and Walker 1988). Also in groups with heterogeneous MPCRs, those members who receive larger returns from the public good also contribute more and are less likely to free-ride (Fisher et al. 1995; Tan 2008; Reuben and Riedl 2009; Fischbacher, Schudy, and Teyssier 2014). On eBay, an analogous effect can be observed: those members who are more active are also more likely to leave a feedback rating (Dellarocas and Wood 2008; Jian, Mackie-Mason, and Resnick 2010). In line with these results, we expect that the likelihood to interact on a particular market affects the willingness to submit a rating.

Hypothesis 4: Feedback provision and trading frequency
Buyers are more likely to leave feedback for transactions on their more frequently visited market.

As we have discussed before, most traders on eBay conduct the majority of trades on their domestic market while trades on international markets are rather the exception than the rule. Accordingly, trading frequency might help to explain why in cross-border transactions sellers and buyers remain silent more often.

Procedure
All sessions took place in June 2014 at the Cologne Laboratory for Economic Research (CLER, UoC) and the Laboratory for Behavioral Operations and Economics (LBOE, UTD) including ten matching groups with 160 subjects in total (80 from each university).[70] To make sure that the stage game and the matching procedure were well understood, subjects had to answer several control questions before the start of the experiment.[71] At the end of the experiment, subjects answered a short questionnaire asking for some demographic characteristics, participant's affiliations with their own and the other university and how they made their decisions in the experiment.[72]

---

[70] Subjects in Cologne were recruited via ORSEE (Greiner 2004), for US participants a similar system called SONA (www.sona-systems.com) was used.

[71] Instructions were given to participants as handouts. Control questions and the actual experiment were computer-based using the experiment software SoPHIE – Software Platform for Human Interaction Experiments (Hendriks 2012).

[72] Average age of the participating subjects from the two different universities is similar (25.3 vs. 24.7) but the share of women is larger at Cologne (50% vs. 25%). Participants at both laboratories are students from a large variety of fields but most common are economics and business students. We asked for affiliations with the statement: "How much do you feel affiliated with the University of Cologne (Dallas

Sessions lasted approximately for 100 minutes and average earnings were 21€ for participants from UoC and $29 for UTD participants.[73]

4.3 RESULTS

Our stage game is repeated for 40 periods. In order to mitigate the influence of endgame considerations we do not include the final five periods in all analyses.[74] Although the main focus of our study is on the provision of feedback and how it is affected by a buyer's social identity and market participation, we first have a look at buyers' and sellers' trading behavior in our four treatment conditions. Buyers' purchase decisions and sellers' quality choices are made simultaneously in the first stage of the experiment.

Trading behavior

When buyers make their purchase decision they know whether they trade on their home or away market, from which university the seller comes from, and the number of positive and negative ratings the seller has received so far. Overall, the product is sold in 80% of all transactions. Descriptively, we do not observe large treatment differences regarding buyer behavior.[75] Buyers are slightly more likely to buy from an ingroup than from an outgroup seller (82% vs. 79%). Using a Wilcoxon sign rank test (WSR) based on aggregated matching groups, this difference is not significant (p = 0.386).[76] This means that buyers do not trust ingroup sellers more than outgroup sellers in general. However, in our experimental setup, the feedback system provides information about sellers' trustworthiness and thereby may have an additional effect on buyers' trust besides a shared social identity. Similar to eBay, we calculate the percentage of positive

---

at Texas)?", asking participants for a rating on a 9-point Likert scale ranging from 1 "not at all" to 9 "very much". Participants at Cologne state a slightly lower affiliation with their home university than participants at Dallas (6.3 vs. 7.2). Affiliation with the partner university is also lower among Cologne subjects (1.8 vs. 4).

[73] Recommended hourly compensations for subjects in the two laboratories are 10€ and $15, respectively. Hence, subjects from UoC and UTD received comparable payoffs on average (the recommended compensation for a 2-hour experiment).

[74] We observe an endgame effect for sellers' quality choice as well as for buyers' purchase and feedback decision. Nevertheless all results remain qualitatively the same if we run analyses including all periods.

[75] More descriptive statistics can be found in Table 4.5 in Appendix A.

[76] All Wilcoxon sign rank tests are two-tailed and are based on means aggregated on the matching group level.

feedbacks (percent positive; PP) received so far as a measure of sellers' trustworthiness.[77]

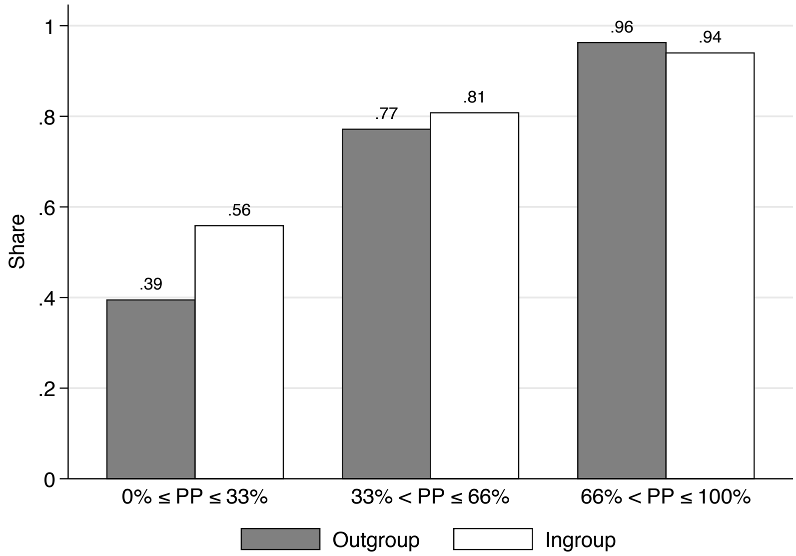


Figure 4.3: Buyers' purchase decision for ingroup and outgroup sellers with different levels of percent positive (PP).

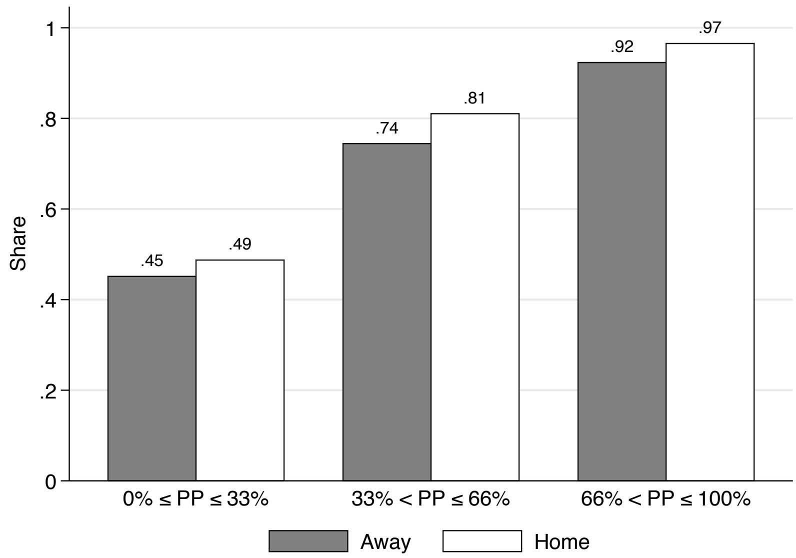


Figure 4.4: Buyers' purchase decision on home and away market and for sellers with different levels of percent positive (PP).

---

77 Looking at Table 4.6 in Appendix A shows that the first quality regression with percentage positive as independent variable has a higher model fit in terms of log likelihood than the second model where the number of positive and negative feedback ratings are included separately. Hence, percent positive seem to be a better predictor for seller trustworthiness.

Looking at different levels of percent positive, we observe that social identity between traders interacts with sellers' feedback profile. For sellers with a low percentage of positive feedbacks (PP ≤ 33%) a shared social identity increases the likelihood that the buyer makes a purchase by 17 percentage points (Figure 4.3). This difference is weakly significant on a 10% level (WSR; p = 0.083). In contrast, when buyers interact with sellers who have a medium (33% < PP ≤ 66%) or high (PP ≥ 66%) percentage of positive feedbacks social identity does not seem to make a difference.

With regard to market participation, the difference in buyers' trading behavior is a bit more pronounced – 82% decide to buy on their home market while this is only the case in 77% on the away market – and significant on a 5% level (WSR; p = 0.037). Percentages in Figure 4.4 suggest that this positive home bias in trust exists across all levels of feedback profiles, however, the difference between home and away purchases is only significant for medium feedback scores (WSR; p = 0.047).

Purchase (yes/no)	**Model 1**	**Model 2**	**Model 3**
Ingroup	0.129*	0.363**	0.130*
	(2.028)	(2.896)	(2.049)
Home	0.167*	0.162*	0.087
	(2.444)	(2.371)	(0.658)
Percent positive	0.024***	0.027***	0.023***
	(20.516)	(16.278)	(12.363)
Ingroup X percent positive		-0.005*	
		(-2.168)	
Home X percent positive			0.002
			(0.709)
UTD	0.276***	0.253***	0.274***
	(4.341)	(3.932)	(4.303)
Period	-0.016***	-0.015***	-0.015***
	(-4.560)	(-4.530)	(-4.537)
Intercept	-0.864***	-0.969***	-0.813***
	(-5.961)	(-6.314)	(-5.025)
N	2677	2677	2677
Log likelihood	-1010.3	-1007.9	-1010.0

Table 4.1: Random effects Probit regression with buyers' purchase decision as dependent variable. Dummy variables for matching groups included in all models. t statistics in parentheses; * p < 0.05, ** p < 0.01, *** p < 0.001.

To investigate these interaction effects on buyers' trust in more detail and to capture the dynamics over time in our experimental setup, we use Probit regressions with buyers' purchase decision as dependent variable.[78] Besides a period variable, we also include a dummy variable for buyers from UTD to control for cross-cultural and subject pool differences. Results in Table 4.1 show that subjects are significantly more likely to trust a seller with the same social identity. Also, a seller's feedback profile helps to establish trust and has a positive effect on buyers' willingness to purchase. The interaction between these two indicators of trustworthiness in Model 2 confirms that especially sellers with a low percentage of positive feedbacks profit from their ingroup status and are trusted more than their outgroup counterparts. According to the predicted probabilities (see left panel of Figure 4.11 in Appendix A), sellers who have more negative than positive ratings receive an ingroup trust premium while sellers with more positive than negative feedback are not treated differently. This partially corresponds to Hypothesis 1a stating that a shared social identity increases trust and promotes trade in ingroup matches. However, with a reputation system in place, buyers also rely on the information contained in feedback ratings of previous customers. When we control for a seller's feedback score, time and university affiliation, buyers trade significantly more often on their home market. This is no longer the case – probably due to multicollinearity – if we include an interaction with percent positive, which is also not significant (Model 3). Taken together, this suggests that the home bias persists across most levels of percent positives (see right panel of Figure 4.11 in Appendix A). Trust in sellers diminishes significantly over time. Besides, participants from UTD seem to be more trusting than participants from UoC.

When choosing the level of quality, sellers only know their trading partner's social identity but not whether this is their home or away market. Trading frequency, therefore, cannot play a role for sellers and – according to Hypothesis 1b – they should only condition their quality decision on whether they are paired with an ingroup or outgroup buyer.

Figure 4.5 shows that the distributions of quality choices made by sellers are very similar for ingroup and outgroup transactions. For both types, in more than 50% of all periods sellers choose the midpoint of the quality scale, which – neglecting feedback

[78] To account for the matching group structure in our experiment, we include dummy variables for each matching group. As a robustness check we also ran the same regressions using simple probit models with clustering on the matching group level. As can be seen in Table 4.7 and Figure 4.12 in Appendix A, the results remain qualitatively the same.

costs – leads to equal payoffs for both players. Overall, the average quality level is 46.23 in outgroup trades and 46.53 in ingroup trades, respectively, and this difference is not significant (WSR; $p = 0.799$).

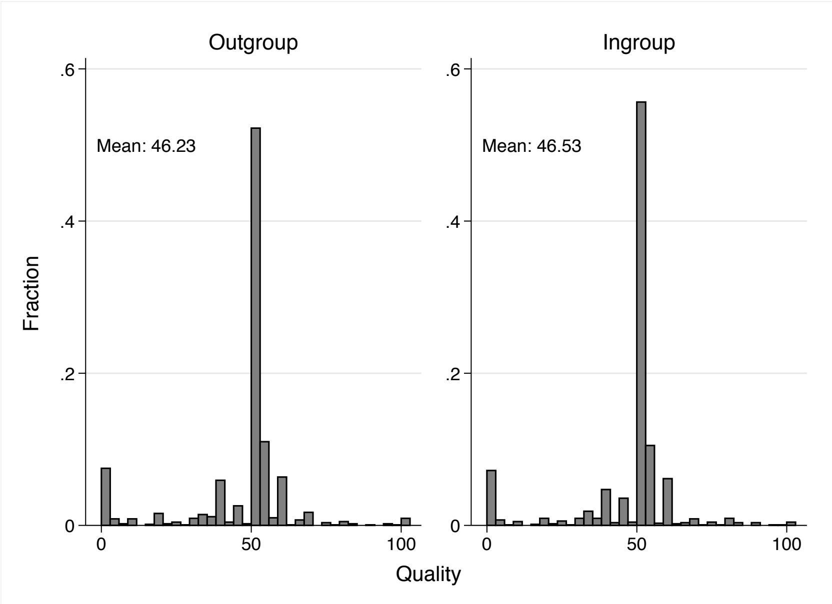


Figure 4.5: Histograms of all sellers' quality choices in ingroup and outgroup trades.

This also holds when we regress quality on ingroup and control for sellers' feedback profile, time and university affiliation (see Table 4.6 in Appendix A). Looking at sold qualities only does not lead to different results. This finding is in contrast to our hypothesis and earlier results in basic trust games by Heap and Zizzo (2009) who find that second movers discriminate based on social identity. According to their findings, this difference occurs because subjects – trustors and trustees – discriminate against outgroup members, but not because of ingroup favoritism. In our setup, however, the second stage might have an additional, opposing effect on trading behavior. Feedback systems and the threat of a bad reputation have been shown to curb moral hazard in trust games, and thereby are also able to enhance buyers' trust in sellers' trustworthiness (Bolton, Katok, and Ockenfels 2004). Hence, a feedback system might also give outgroup buyers leverage to react against discrimination so that sellers no longer – or, if so, only to a lower degree – differentiate between ingroup and outgroup members.

Feedback behavior

The main focus of our study lies on buyers' motivation to provide feedback. At first, we look at the relationship between received quality and the probability to give feedback. As can be seen in Figure 4.5, the distribution of quality is concentrated around the midpoint of the quality scale and the equal split quality level is chosen most often.

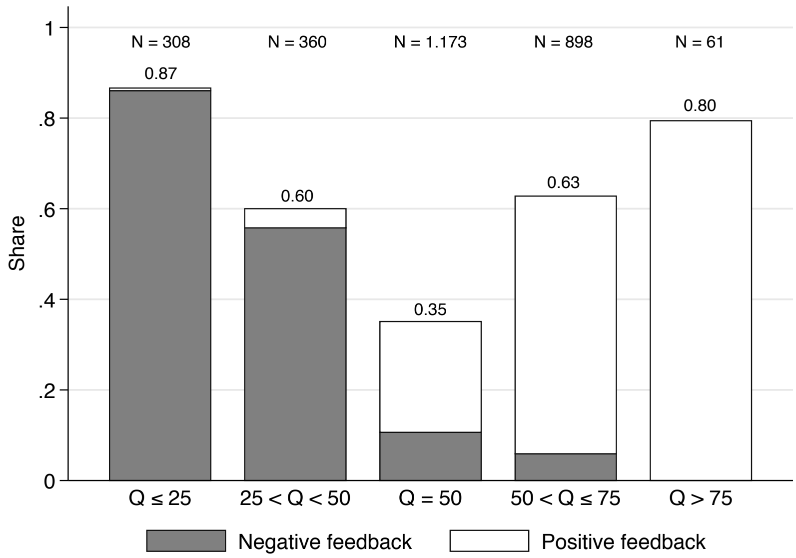


Figure 4.6: Buyers' provision of feedback over different levels of quality. Numbers on top of bars represent the overall percentage of provided feedback for this quality level irrespective of whether it is positive or negative.

We account for this centered distribution by aggregating quality into five categories: very low quality (≤ 25), low quality (> 25 & < 50), medium quality (= 50), high quality (> 50 & ≤ 75) and very high quality (> 75).

Looking at the descriptive results we clearly see that extreme – negative or positive – experiences are more likely to be reported (see Figure 4.6). Very low or very high levels of quality are reported in 87% and 80% of all transactions while only around a third of all sellers shipping medium quality receive a feedback rating. With regard to feedback content, there seems to be a consistent standard among buyers how to rate the received quality. In case quality is below 50, 96% of all ratings are negative, vice versa, when quality exceeds 50, 91% of all ratings are positive. Only when quality is exactly 50 there is some variance in ratings as 70% are positive. The u-shaped relationship between quality and feedback provision can also be seen in Model 1 of Table 4.2 where quality has a negative effect but its quadratic term is significantly positive. Overall, this provides strong support for Hypothesis 2, stating that buyers tend to 'brag and moan' and report extreme experiences more often. Looking at the control variables, buyers' stated expectations about quality do not seem to affect feedback provision, and also there are no differences in rating probability between buyers from UoC and UTD. In addition, we observe that buyers' willingness to leave feedback decreases over time and also the higher a seller's feedback score already is.

Feedback (yes/no)	**Model 1**	**Model 2**	**Model 3**
Ingroup	0.075	1.151^{***}	0.074
	(1.315)	(3.457)	(1.297)
Home	0.043	0.041	0.159
	(0.685)	(0.647)	(0.561)
Quality	-0.064^{***}	-0.044^{***}	-0.056^{***}
	(-10.406)	(-5.852)	(-6.107)
Quality^2	0.001^{***}	0.001^{***}	0.001^{***}
	(9.206)	(5.004)	(5.040)
Ingroup X quality		-0.059^{***}	
		(-4.084)	
Ingroup X quality^2		0.001^{***}	
		(3.946)	
Home X quality			-0.015
			(-1.257)
Home X quality^2			0.000
			(1.588)
Exp. quality	0.004^{*}	0.004	0.004
	(1.976)	(1.787)	(1.923)
UTD	-0.104	-0.103	-0.102
	(-1.712)	(-1.673)	(-1.668)
Percent positive	-0.003^{**}	-0.003^{**}	-0.003^{**}
	(-2.999)	(-2.987)	(-3.083)
Period	-0.017^{***}	-0.018^{***}	-0.017^{***}
	(-5.709)	(-5.901)	(-5.720)
Intercept	1.594^{***}	1.240^{***}	1.551^{***}
	(7.247)	(5.071)	(5.669)
N	2139	2139	2139
Log likelihood	-1326.4	-1316.9	-1325.0

Table 4.2: Random effects Probit regression with buyers' feedback provision as dependent variable (gave feedback = 1). Dummy variables for matching groups included in all models. *t* statistics in parentheses; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

We now turn to the differences in feedback provision due to our treatment variations: buyers' social identity and market participation. With regard to social identity between traders, Figure 4.7 shows that for low, medium and high levels of quality the probabilities to leave a feedback rating are more or less identical for in- and outgroup matches. However, for extreme quality levels buyers are more likely to report their experience if they belong to the same social group as the seller. Descriptively, for very low levels of quality, the rating probability increases by 10 percentage points and for very high levels of quality by even 26 percentage points. Regression results in the second model of Table 4.2 support these findings.[79] First, the base effect of ingroup is now significant. Furthermore, its interactions with quality and quality squared are highly significant and go into the same direction as the base effects of quality and its quadratic term. This indicates that a shared social identity further strengthens the u-shaped relationship between quality and feedback provision. Looking at the predicted probabilities of buyers to leave a feedback at different levels of quality, it can be seen that only for medium levels of quality and perfect quality predicted reporting probabilities are the same for in- and outgroup matches (see Figure 4.8).

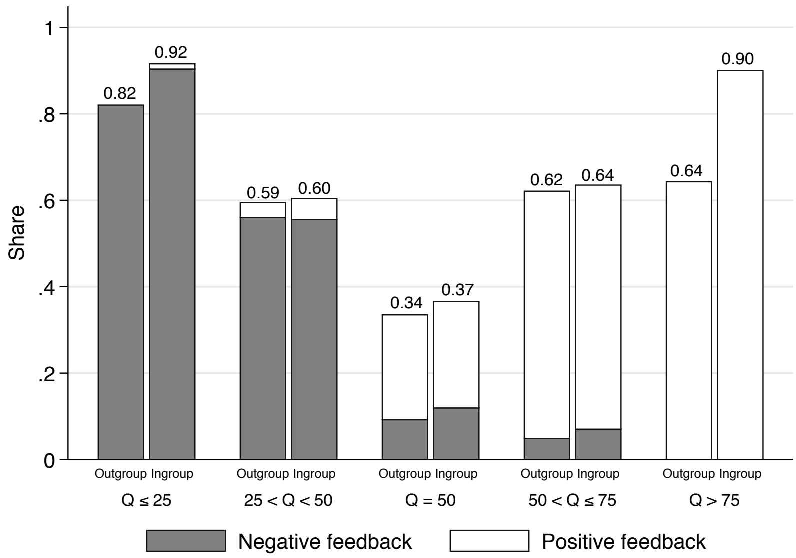


Figure 4.7: Buyers' provision of feedback for in- and outgroup sellers at different levels of quality. Numbers on top represent the overall percentage of provided feedback for this quality level irrespective of whether it is positive or negative.

---

[79] As a robustness check without 3-way interaction effects, we ran similar regressions with negative (for quality of 50 and below) and positive feedback (for quality of 50 and above) as dependent variable. The results are very similar to those presented here and can be found in Table 4.8 and Figure 4.13 in Appendix A. Additionally, Table 4.9 and Figure 4.14 in Appendix A show the results of simple probit models with clustering on the matching group level.

For example, an ingroup seller who ships 0 quality receives a feedback – and most likely a negative one – almost for sure (98%) whereas his outgroup counterpart gets away with no feedback nearly in one out of five transactions (82%). Similarly, for high quality levels between 60 and 90 reporting probabilities are 14 to 36% larger in ingroup than in outgroup matches. Taken together, descriptive as well as multivariate results clearly show that with regard to feedback giving, positive and negative reciprocity are more pronounced in ingroup than in outgroup matches.

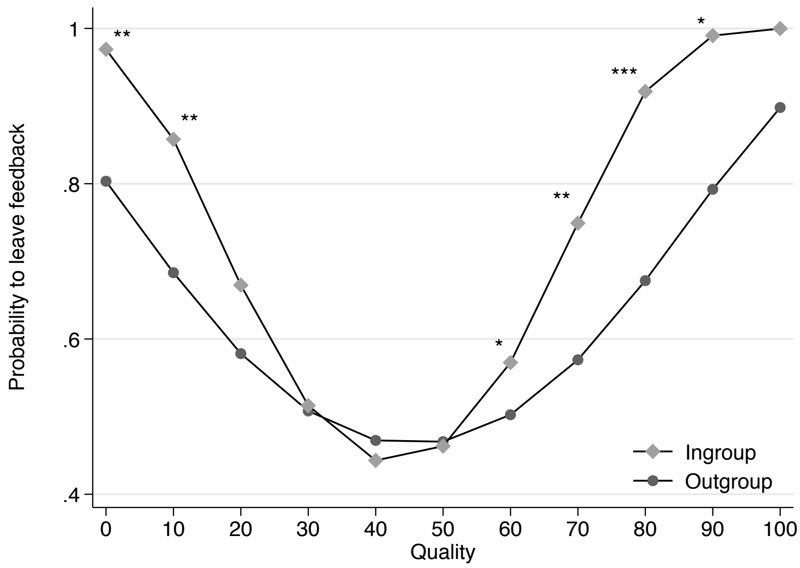


Figure 4.8: Predicted probability to leave feedback for in- and outgroup matches at different quality levels (based on Model 2 in Table 4.2). * ***p*** **<** **0.05,** ** ***p*** **< 0.01,** *** ***p*** **< 0.001.**

For our second identity manipulation, results are less informative. Buyers' likelihood to leave a feedback rating on their home or away markets is very similar across all quality levels except for the highest category where feedback is given on home markets in 86% and away markets in only 67% (Figure 4.9). Looking at the home market effect via regressions we observe no significant differences (Model 1 and 3 in Table 4.2).[80] Neither the baseline effect of home market nor its interactions with quality and quality squared are significant. Predicted probabilities of trading on the home market at different quality levels allow a more detailed analysis (see Figure 4.10). For quality levels of 50 and below the curves are almost identical. In contrast, for quality levels

[80] Although our hypothesis regarding the effect of trading frequency on feedback provision only assumes an upward shift independent of quality, in Model 3 we also test for a non-linear interaction between trading frequency and quality comparable to Model 2 for social identity.

above the equal quality buyers seem to be more willing to leave a rating when trading on their home market.

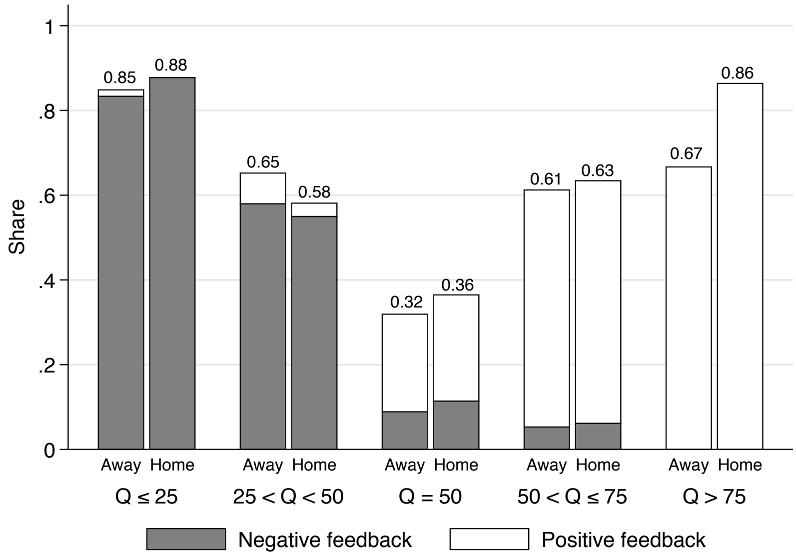


Figure 4.9: Buyers' provision of feedback on home and away market over different levels of quality. Numbers on top represent the overall percentage of provided feedback for this quality level irrespective of whether it is positive or negative.

However, according to the marginal effects, reporting probabilities at these particular levels are also not significantly different. Overall, we do not find evidence that trading frequency has a significant effect on the provision of feedback in our experiment.

As mentioned earlier, besides the two within-subject variations of social identity and market participation, our experimental design also creates a between-subject variation in the sense that exactly for half of the buyers social identity and market participation overlap – sellers on buyers' home market have the same university affiliation[81] – whereas for the remaining half of buyers social identity and market participation do not correspond.[82] Thus, for the former group of buyers our treatment variations coincide and the effects of social identity and trading frequency are difficult to disentangle. Hence, if our results were only driven by this subgroup it would not be possible to ascribe the effect to one of our manipulations in particular. For this reason, we run the

[81] This is the case for those buyers from UoC (UTD) who trade *more* frequently on the Cologne (Dallas) market.

[82] This is the case for those buyers from UoC (UTD) who trade *less* frequently on the Cologne (Dallas) market.

second model of Table 4.2 again for these two groups separately. The results are displayed in Table 4.3. For the ease of comparison the first model is identical to the original regression with all buyers. Model 2 includes only feedback decisions' of buyers who face ingroup sellers on their home market whereas the last model considers only buyers who face outgroup sellers on their home market.

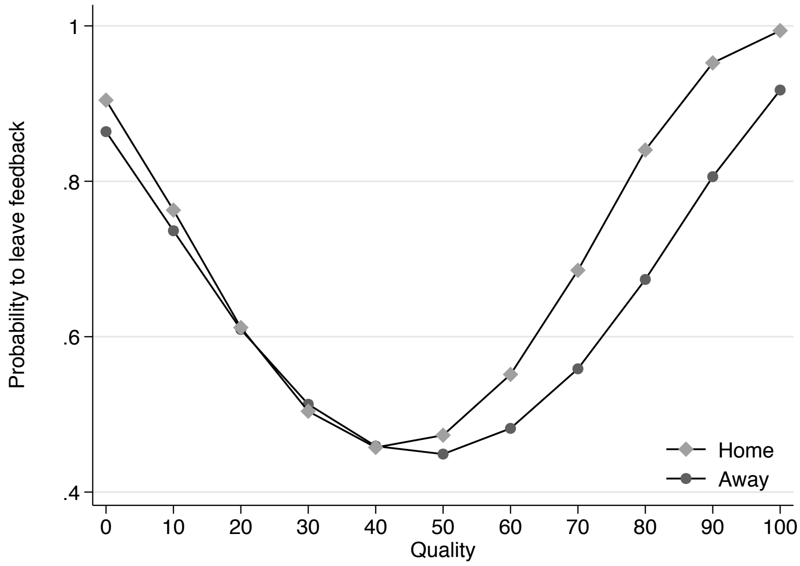


Figure 4.10: Predicted probability to leave feedback for home and away markets at different quality levels (based on Model 3 in Table 4.2). * $p <$ **0.05,** ** $p < 0.01$, *** $p < 0.001$.

Looking at the effects of our ingroup manipulation reveals that the general pattern is consistent for both subgroups. The base effect and the interaction with quality squared are significantly positive whereas the simple interaction between ingroup and quality shows a negative effect. However, size and significance levels of these effects are somewhat smaller for buyers whose social identity and market participation do not overlap.[83] Taken together, the results illustrate that positive and negative reciprocity towards ingroup members are more pronounced in both subgroups, but this treatment manipulation seems to be slightly more distinct for buyers who trade more often with sellers from their own university.

[83] This difference in size and significance can also be observed in the predicted probabilities based on the marginal effects in Figure 4.15 in Appendix A. Curves have similar shapes but differences for in- and outgroup matches are more pronounced for buyers who trade more frequently with sellers from the same university.

Feedback (yes/no)	Model 1	Model 2	Model 3
	all buyers	ingroup = home	ingroup ≠ home
Ingroup	1.147***	1.617**	1.003*
	(3.451)	(2.805)	(2.365)
Quality	-0.044***	-0.027*	-0.056***
	(-5.834)	(-2.220)	(-5.489)
Quality^2	0.000***	0.000	0.001***
	(4.984)	(1.655)	(4.995)
Ingroup X quality	-0.059***	-0.093***	-0.053**
	(-4.089)	(-3.812)	(-2.710)
Ingroup X quality^2	0.001***	0.001***	0.001*
	(3.957)	(3.879)	(2.177)
Exp. quality	0.004	0.009**	-0.004
	(1.797)	(3.033)	(-1.201)
UTD	-0.103	-0.399***	0.189*
	(-1.677)	(-4.323)	(2.154)
Percent positive	-0.003**	-0.002	-0.007***
	(-3.012)	(-1.405)	(-3.932)
Period	-0.018***	-0.013**	-0.024***
	(-6.030)	(-2.985)	(-5.465)
Intercept	1.271***	1.327***	1.455***
	(5.300)	(3.514)	(4.358)
N	2139	1063	1076
Log likelihood	-1317.1	-614.1	-635.2

Table 4.3: Random effects Probit regression with buyers' feedback provision as dependent variable (gave feedback = 1). Model 1 includes all buyers, Model 2 only those buyers where ingroup trades and home market overlap and Model 3 only those buyers where ingroup trades and home market do not overlap. Dummy variables for matching groups included in all models. *t* statistics in parentheses; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Feedback content

As we have seen earlier, rating standards with regard to feedback content seem to be very similar among buyers. For quality levels below 50, 96% of the ratings are negative while for quality levels above 50, 91% are positive. Hence, when sellers deviate from the equal split quality, neither social identity nor trading frequency can have a considerable influence on feedback content.[84] When buyers receive exactly half of the

[84] In fact, for quality below 50 only 5% (3%; 3%; 6%) of buyers give a positive feedback to ingroup (outgroup; home market; away market) sellers. The respective figures for quality above 50 are 90% (ingroup), 92% (outgroup), 91% (home market) and 92% (away market).

quality, there are also only minor descriptive treatment differences. In these cases, 67% of ingroup sellers receive a positive feedback, while with 73% probability positive feedbacks are slightly more likely in outgroup matches. Using a Wilcoxon sign rank test this difference is not significant (p = 0.508). The same holds true for trading frequency where buyers give a positive feedback on their home (away) market in 69% (72%) of all reported transactions (WSR; $p = 0.414$).[85]

Efficiency

In our experimental setup, the welfare-maximizing outcome is obtained if the seller ships full quality of 100. In this case the seller makes a profit of 100 ECU while the buyer ends up with 300 ECU, not taking feedback costs into consideration. Lowest possible surplus occurs if either the seller ships 0 quality or if there is no trade in the first place. In both cases the total surplus is 200 ECU either for the seller only or evenly split between the two. Hence, two factors may cause efficiency losses: buyers deciding not to purchase and sellers shipping less than maximum quality. According to Hypotheses 1a and 1b, we would expect that efficiency is higher in ingroup matches because buyers are more trusting and sellers ship higher quality. With regard to market participation we did not make any assumptions regarding its influence on buyer trust and seller trustworthiness.

	Trust Buy	**Trustworthiness** Quality	**Efficiency** % of maximum surplus
Ingroup	81.8% (38.61)	46.53 (16.49)	69.4% (11.23)
Outgroup	79.0% (40.75)	46.23 (17.38)	68.8% (11.77)
Home market	81.9% (38.49)	47.13 (16.14)	69.6% (11.18)
Away market	77.0% (42.08)	44.73 (18.46)	68.1% (12.13)

Table 4.4: Descriptive results for buyer trust, seller trustworthiness and transaction efficiency. Efficiency is measured as percentage of the maximum possible surplus. It ranges between 49.75% – if the seller ships zero and the buyer has feedback costs of 1 ECU – and 100%. Standard deviations are displayed in parentheses.

As we have discussed earlier, we observe only minor treatment differences with regard to buyers' purchase decisions and sellers' quality choices (see also Table 4.4). In ingroup matches trust (82% vs. 79%) and trustworthiness (46.5 vs. 46.2) are slightly higher than in outgroup transactions. However, these differences do not translate into

[85] Also several regressions including different ranges of quality levels do not show any differences in feedback content due to our treatment manipulations (see Table 4.10 in Appendix A).

significant efficiency gains (69.4% vs. 68.8%; WSR; p = 0.285). Comparing these figures for home and away markets, differences are a bit more pronounced. In home market transactions trust increases from 77% to 82%, and average quality also rises from 44.7 to 47.1. These differences are large enough to generate small, but significant efficiency gains of 1.5 percentage points (69.6% vs. 68.1%; WSR; p = 0.022). A tobit regression confirms these results (see Table 4.11 in Appendix A).

4.4 *Discussion and Conclusion*

Online marketplaces provide the opportunity to do business with transaction partners from around the world. Even non-professional traders or small companies now have access to global markets. Key features of these Internet markets are feedback systems, which can be used to publicly report transaction experiences and thereby promote trust and facilitate national and international trade. Via its effects on future profits, positive feedback promotes good behavior while negative ratings punish and discipline deceptive traders. So far, it has not been analyzed whether feedback ratings are used differently in national and international transactions. Empirical evidence from eBay suggests that the provision of feedback is lower for transactions between international traders. In this paper we took a first attempt to experimentally investigate which factors may cause traders to give ratings more often in within-border as compared to cross-border transactions. More specifically, we examined whether and how a shared social identity between seller and buyer and buyers' frequency to trade on a particular market affect rating and also trading behavior.

First, with regard to feedback giving, our results show that buyers are in general more likely to report extreme experiences. This is in line with earlier results by Dellarocas and Wood (2008) or Lafky (2014). This brag-and-moan way of reporting might be popular among buyers because they perceive remaining silent as some sort of neutral feedback rating. Interestingly, as a consequence, exactly those quality levels that occur most frequently are the least likely to be reported. Hence, buyers' choice not to report an experience conveys information about his level of satisfaction and therefore can be helpful to future traders to better assess their trading partner. However, as the number of unrated transactions is not included in feedback profiles – this is the case in our experiment but also on eBay and many other online feedback systems – these scores provide only a biased view on sellers' level of trustworthiness. Extending feedback profiles with a statistic for transactions without rating should improve informativeness of the feedback system to predict seller trustworthiness.

The 'overreporting' of extreme experiences is even more pronounced when buyer and seller belong to the same social group. Using university affiliations as natural group identity, we observe that positive (rewarding good quality with a positive rating) and negative reciprocity (punishing inferior quality with a negative rating) are more likely in ingroup matches. In these cases, reporting probabilities for very high and very low levels of quality increase by 14 to up to 37%. While the former effect is in line with our hypothesis, the latter goes into the opposite direction of our hypothesis and earlier findings from laboratory and field experiments. For example, Bernhard, Fischbacher, and Fehr (2006b), Chen and Li (2009), and Mussweiler and Ockenfels (2013) find that punishment towards members of the same social group is less severe. This leniency is attributed to individuals' larger concerns for other group members' payoffs. A possible explanation for our contrary results might come from an 'extension' of social identity theory: The so-called 'black sheep effect' assumes that deviant behavior by ingroup members is perceived as potential threat to the group's identity and therefore is judged more extremely than by outgroup subjects (Marques, Yzerbyt, and Leyens 1988; Marques and Yzerbyt 1988; Marques, Abrams, and Serôdio 2001; Pinto et al. 2010).[86] In this context it is assumed that harsher condemnation of ingroup members serves to maintain the group's positive identity. Empirical support for this derogation effect in economic settings was found by Shinada, Yamagishi, and Ohmura (2004) and McLeish and Oxoby (2007) who show that non-cooperative behavior by ingroup members is punished more extreme than by outgroup members.[87]

As a second treatment manipulation, we created two different markets and varied buyers' trading frequencies on these particular markets. Contrary to our hypothesis we

---

[86] According to the 'black sheep effect' ingroup members are also judged more positively for positive, cooperative behavior, which is also in line with our finding of higher positive reciprocity towards ingroup members. In a similar vein, Akerlof and Kranton (2000) argue that deviant behavior of other group members may be perceived as a threat to the group's identity and therefore may evoke negative responses to protect the positive image of the group.

[87] Additionally it might be important to note that in most other studies punishment has direct monetary consequences. In our experiment, however, a negative feedback has only an indirect effect on sellers' payoffs. This circumstance may also decrease the influence of charity concerns and thus reduces a barrier to ingroup punishment.

do not observe that feedback provision is affected by market participation. Descriptively, buyers receiving high levels of quality seem to be more likely to leave a rating on their home market, but this difference turns out to be insignificant. Hence, even on less frequently visited markets buyers are willing to leave a rating and contribute to the feedback public good. A possible explanation for this finding could be that punishment is used to express negative emotions after suffering from others' unfair or selfish behavior (Fehr and Gächter 2002; Xiao and Houser 2005). From this point of view, it might well be that receiving low levels of quality evokes the same negative emotions irrespective of whether a trade takes place on the home or away market.

The particular content of a feedback rating is not affected by our treatment variations. Regardless of the seller's social identity and the current market, buyers have uniform rating standards. In brief, in almost all cases quality above 50 is rewarded with a positive rating while below 50, buyers typically leave a negative rating. When sellers choose to split equally, buyers reciprocate with a positive rating in around two-thirds of all rated transactions regardless of whether it is an ingroup or outgroup trade or it takes place on the home or away market.

With regard to trading behavior – buyers' purchase and sellers' quality decision – it is interesting to note that both are rather unaffected by our social identity variation. Based on earlier studies, we would have expected that trust and trustworthiness are larger in ingroup than in outgroup matches. In this regard the implemented reputation system seems to be considered a reliable mechanism to assess the trading risk with a particular seller. Sellers who have a medium or high percentage of positive feedbacks have equal probabilities to sell their product. Only for sellers with a low feedback score we observe that buyers are more trusting towards other ingroup members and purchase more often. In addition, sellers also seem to be aware that, with a feedback system in place, they cannot discriminate against outgroup buyers in terms of quality without jeopardizing their good reputation. Hence, ingroup and outgroup buyers receive virtually the same quality on average. Buyers' individual frequency to trade on a particular market also has a positive effect on their level of trust. Independent of the seller's feedback score, buyers are more likely to buy on their home market than on the away market. As sellers do not get to know buyers' trading frequencies they cannot differentiate between home and away market buyers.[88]

Taken together, social identity increases buyers' willingness to leave feedback after receiving very high or very low quality while the frequency to trade on a market does not seem to affect feedback provision. Hence, social identity might be one reason why we observe more feedback ratings in national than in international transactions on eBay. However, from our experiment it is difficult to assess whether this stronger bragging and moaning has positive or negative implications for a marketplace. In fact, extreme experiences are more likely to be reported but occur rather infrequently while reporting probabilities for the most common transaction outcome – medium quality – remain unchanged. Thus, while in ingroup transactions overall feedback provision slightly increases this also may give a rather selective view on actually traded quality. A drawback of our within-subject design is that ingroup and outgroup buyers likewise trade on each market. Therefore, it is not possible to investigate the long-term effect of this ingroup rating bias on the effectiveness of the feedback system because members of both groups contribute to and profit from the feedback public good on each market. For future research, an experimental setup with unmixed markets would allow having a better look into the effects of this ingroup bias on the informativeness of feedback profiles, and on whether this eventually translates into differences in market efficiency.

[88] In fact, qualities buyers receive on their home and away market are 47 and 45, respectively.

		Periods			
		All	**1. - 12.**	**13. - 24.**	**25. - 35.**
Quality	$N = 2800$	46.38	48.35	47.89	42.58
Equal split	$N = 2800$	41.89%	49.27%	43.85%	31.70%
Sold products	$N = 2800$	80.39%	82.81%	80.83%	77.27%
Efficiency	$N = 2800$	69.12%	70.11%	69.58%	67.54%
Feedback given	$N = 2251$	52.15%	58.60%	51.29%	45.59%
Positive feedback	$N = 1174$	62.01%	63.52%	63.07%	58.39%
Negative feedback	$N = 1174$	37.99%	36.48%	36.93%	41.61%
Percent positive	$N = 2677$	60.92%	60.97%	60.85%	60.95%

Table 4.5: Descriptive statistics.

Quality	**Model 1**	**Model 2**
Ingroup	0.283	0.193
	(0.676)	(0.658)
Percent positive	0.085^{***}	
	(0.012)	
# positive feedbacks		0.300^{*}
		(0.121)
# negative feedbacks		-1.294^{***}
		(0.180)
UTD	-1.045	-0.649
	(0.684)	(0.666)
Period	-0.299^{***}	-0.152^{*}
	(0.035)	(0.059)
Intercept	43.266^{***}	46.604^{***}
	(1.550)	(1.296)
N	2677	2800
Log likelihood	-10904.1	-11404.2

Table 4.6: Random effects tobit regression with quality as dependent variable (lower limit = 0; upper limit = 100). Dummy variables for matching groups are included in both models. *t* statistics in parentheses; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

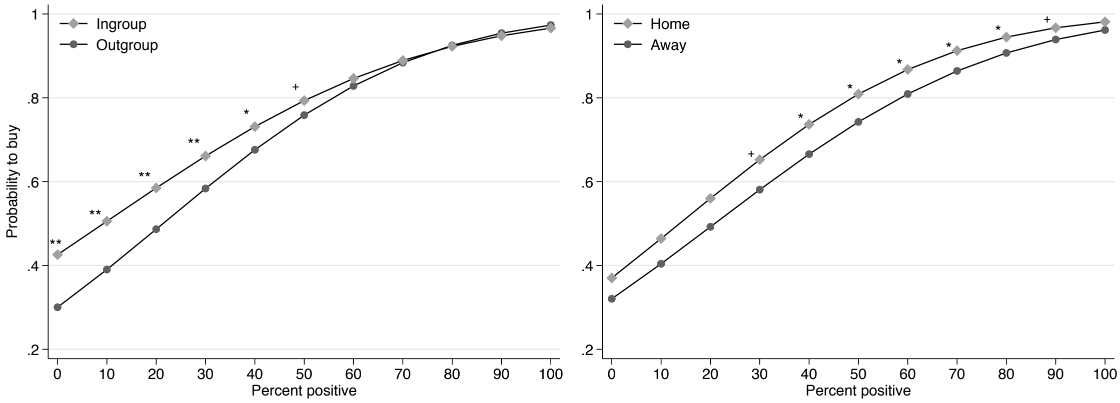


Figure 4.11: Probability to buy at different feedback scores. Left panel for in- and outgroup matches based on Model 2 in Table 4.1. Right panel for home and away markets based on Model 3 in Table 4.1. $^{+}p < 0.1$, $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

Purchase (yes/no)	Model 1	Model 2	Model 3
Ingroup	0.106	0.350**	0.108
	(0.908)	(2.973)	(0.932)
Home	0.163	0.158	0.071
	(1.576)	(1.536)	(0.468)
Percent positive	0.023***	0.025***	0.021***
	(6.139)	(6.398)	(4.663)
Ingroup X percent positive		-0.005***	
		(-4.696)	
Home X percent positive			0.002
			(0.545)
UTD	0.260	0.238	0.258
	(1.744)	(1.628)	(1.709)
Period	-0.013	-0.013	-0.013
	(-1.748)	(-1.754)	(-1.741)
Intercept	-0.383*	-0.492*	-0.324
	(-2.179)	(-2.509)	(-1.557)
Log likelihood	-1068.1	-1065.4	-1067.8
N	2677	2677	2677

Table 4.7 Simple probit models with buyers' purchase decision as dependent variable. Standard errors clustered on matching group level. *t* statistics in parentheses; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

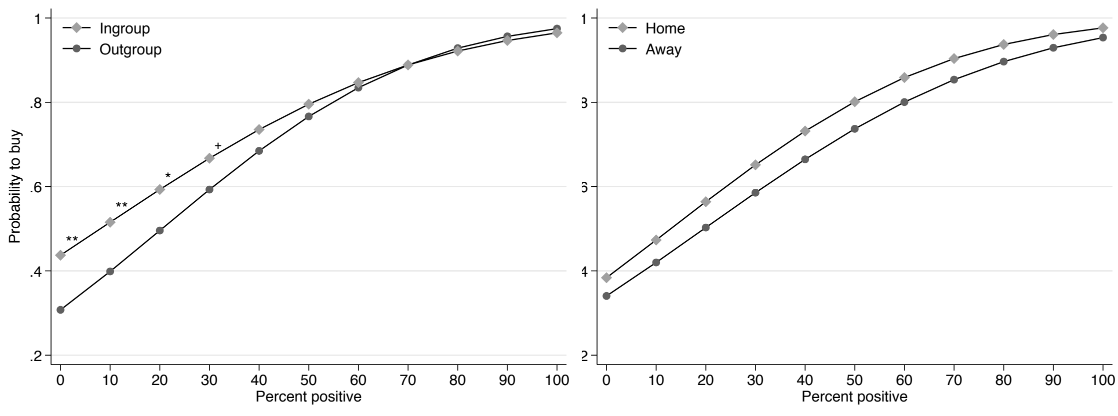


Figure 4.12 Probability to buy at different feedback scores. Left panel for in- and outgroup matches based on Model 2 in Table 4.7. Right panel for home and away markets based on Model 3 in Table 4.7. $p < 0.1$, * $p < 0.05$; ** $p < 0.01$.

	Prob(neg FB)	**Prob(pos FB)**
	$Q \leq 50$	$Q \geq 50$
Ingroup	0.372^{*}	-1.247^{**}
	(1.981)	(-2.873)
Home	0.159^{+}	0.020
	(1.817)	(0.307)
Quality	-0.039^{***}	0.025^{***}
	(-11.641)	(4.835)
Ingroup X quality	-0.005	0.023^{**}
	(-1.076)	(2.890)
Exp. quality	0.030^{***}	0.006^{***}
	(13.212)	(3.489)
UTD	-0.315^{***}	-0.085
	(-3.571)	(-1.347)
Percent positive	-0.003^{*}	0.006^{***}
	(-2.037)	(5.070)
Period	-0.014^{***}	-0.012^{***}
	(-3.337)	(-3.772)
N	1740	2040
Log likelihood	-630.6	-1196.7

Table 4.8: Random effects Probit regression with negative/ positive feedback as dependent variable. Model 1 is for observations with quality of 50 and below. Model 2 is for observations with quality of 50 and above. Dummy variables for matching groups are included in both models. *t* statistics in parentheses; $^{+}$ 0.1, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

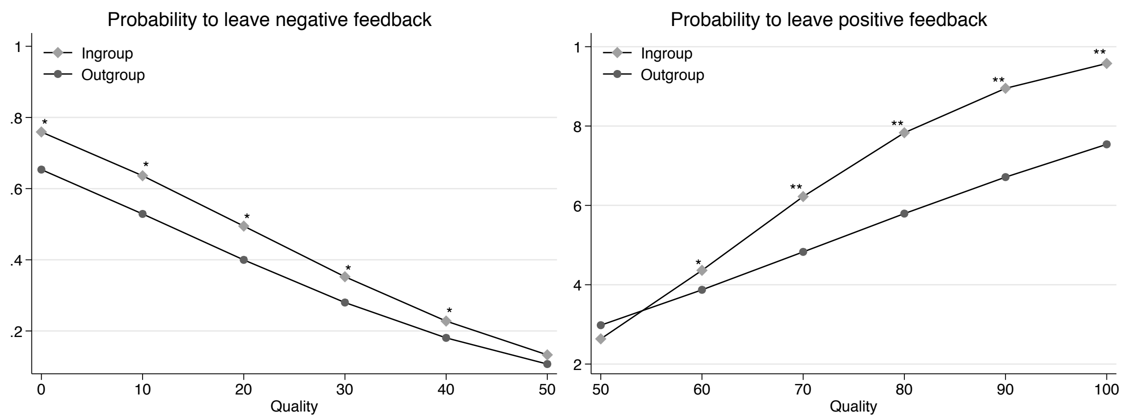


Figure 4.13: Probability to leave negative or positive feedback. Left panel based on Model 1 in Table 4.8. Right panel based on Model 2 in Table 4.8. * $p < 0.05$, ** $p < 0.01$.

Feedback (yes/no)	Model 1	Model 2	Model 3
Ingroup	0.058	1.049*	0.057
	(0.073)	(0.463)	(0.073)
Home	0.037	0.033	0.175
	(0.081)	(0.080)	(0.387)
Quality	-0.065***	-0.045***	-0.055***
	(0.010)	(0.010)	(0.014)
Quality^2	0.001***	0.000***	0.001**
	(0.000)	(0.000)	(0.000)
Ingroup X quality		-0.061**	
		(0.021)	
Ingroup X quality^2		0.001**	
		(0.000)	
Home X quality			-0.017
			(0.022)
Home X quality^2			0.000
			(0.000)
Exp. Quality	0.003	0.002	0.003
	(0.006)	(0.006)	(0.006)
UTD	-0.084	-0.079	-0.082
	(0.158)	(0.160)	(0.157)
Percent positive	-0.003	-0.003	-0.003
	(0.003)	(0.003)	(0.003)
Period	-0.017***	-0.018***	-0.017***
	(0.002)	(0.002)	(0.002)
Intercept	1.644***	1.357***	1.585***
	(0.166)	(0.250)	(0.231)
Log likelihood	-1391.6	-1380.6	-1389.9
N	2139	2139	2139

Table 4.9 Random effects linear probability models with buyers' feedback decision as dependent variable. Standard errors clustered on matching group level. *t* statistics in parentheses; $^{+}$ 0.1, * *p* < 0.05, ** *p* < 0.01, *** *p* < 0.001.

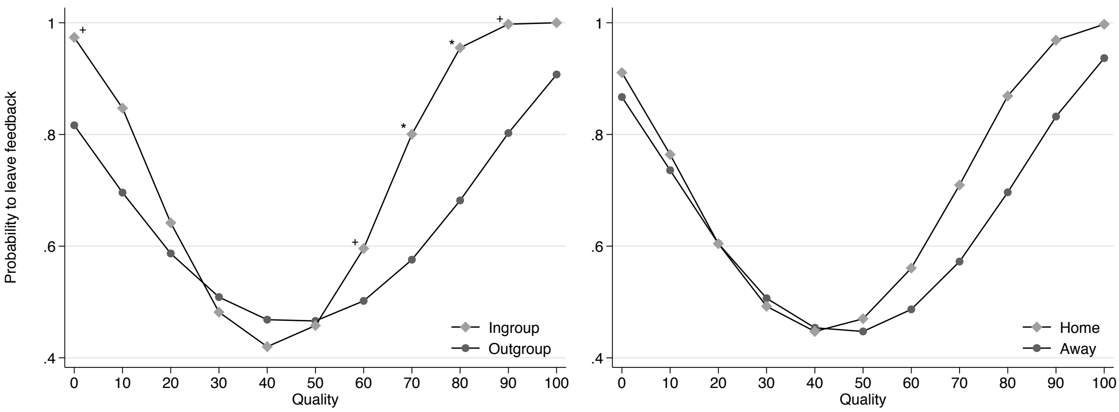


Figure 4.14 Predicted probability to leave feedback at different quality levels. Left panel for in- and outgroup matches based on Model 2 in Table 4.9. Right panel for home and away markets based on Model 3 in Table 4.9. $^{+}$ 0.1, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

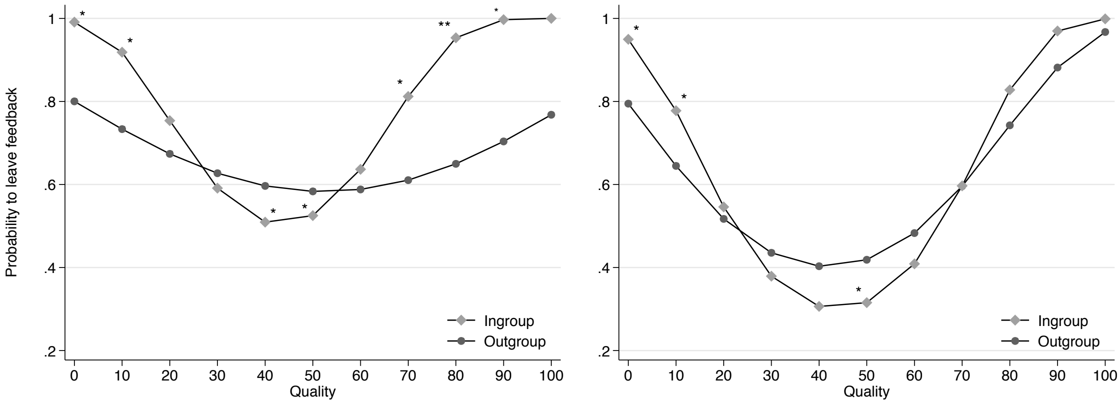


Figure 4.15: Predicted probability to leave feedback for in- and outgroup matches at different quality levels. Left panel for buyers where social identity and market participation overlap (based on Model 2 in Table 4.3). Right panel for buyers where social identity and market participation do not overlap (based on Model 3 in Table 4.3). $^{+}p < 0.1$, * $p < 0.05$, ** $p < 0.01$.

Feedback (pos/neg)	Model 1	Model 2	Model 3
	$25 \leq Q \leq 75$	$40 \leq Q \leq 60$	$Q = 50$
Ingroup	-0.127	-0.125	-0.331
	(-1.022)	(-0.971)	(-1.556)
Home	-0.182	-0.176	-0.157
	(-1.308)	(-1.224)	(-0.700)
Quality	0.219^{***}	0.243^{***}	
	(14.202)	(13.024)	
Exp. quality	-0.049^{***}	-0.049^{***}	-0.063^{***}
	(-9.715)	(-9.343)	(-7.371)
UTD	0.384^{**}	0.437^{**}	0.471^{*}
	(2.710)	(2.963)	(2.043)
Percent positive	0.008^{***}	0.009^{***}	0.006
	(3.598)	(3.655)	(1.515)
Period	-0.009	-0.009	-0.021
	(-1.399)	(-1.338)	(-1.824)
Intercept	-7.194^{***}	-8.402^{***}	5.192^{***}
	(-9.366)	(-8.868)	(7.078)
N	940	844	315
Log likelihood	-280.6	-265.2	-113.4

Table 4.10: Random effects Probit regression with feedback content as dependent variable (positive feedback = 1). Dummy variables for matching groups are included in all models. *t* statistics in parentheses; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Efficiency	Model 1	Model 2
Ingroup	0.865	2.281^{*}
	(1.572)	(2.311)
Home	1.536^{*}	2.570^{**}
	(2.548)	(3.025)
Ingroup X home		-2.054
		(-1.729)
UTD	2.021^{***}	2.023^{***}
	(3.673)	(3.680)
Period	-0.152^{***}	-0.152^{***}
	(-5.481)	(-5.482)
N	2800	2800
Log likelihood	-9430.1	-9428.6

Table 4.11: Random effects Tobit regression with efficiency as dependent variable (lower limit = 49.75; upper limit = 100). Dummy variables for matching groups are included in all models. *t* statistics in parentheses; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

4.6 Appendix B – Instructions

Instructions

Welcome and thank you for participating in this experiment. Please use the Login Code to log into the experiment. Please keep this Login Code until the end of the experiment. We need this Login Code to pay you your earnings.

You can earn money in this experiment. The specific amount depends on your decisions and the decisions of other participants. Take the time to read the instructions carefully and please make sure that you understand everything. If you have questions, please raise your hand and one of the supervisors will come to help you. From now on until the end of the experiment, please do not communicate with other participants. For the experiment to run smoothly, it is important that all participants focus solely on the experiment. For this reason, please put away your cellphones, magazines, and study materials. If you do not comply with these rules we may have to exclude you from the experiment and all payments.

This experiment will be conducted here and in a laboratory at the University of Cologne (Germany). This means that subjects from the ***University of Texas at Dallas*** *and from the* ***University of Cologne*** *participate in this experiment at the same time. During the experiment you interact with subjects from the University of Texas at Dallas (participants from Dallas) as well as with subjects from the University of Cologne (participants from Cologne). Whether you interact with a participant from Dallas or Cologne will be shown on your screen. All participants receive the same information. All decisions that you take are anonymous.*

In the experiment we use ECU (Experimental Currency Unit) as the monetary unit. At the end of the experiment your payoffs will be converted from ECUs into US Dollars (250 ECU = 1 US Dollar) and paid out in cash plus a show-up-fee of $5. The conversion rate for the participants at the University of Cologne is adjusted to the current exchange rate between US Dollars and Euro.

In this experiment there are **buyers** and **sellers**. Your role will be randomly determined at the beginning of the experiment and remains the same for the entire experiment. There are as many buyers as there are sellers. All participants have the same probability to be in the role of the buyer and seller, respectively. Before the first period starts, you will be informed whether you are in the role of a buyer or seller. In total, the experiment lasts for **40** periods.

Markets and buyer types

In this experiment there are two markets: a **Dallas market** and a **Cologne market**. Sellers always stay on their home market: **Sellers from Dallas** always trade on the **Dallas market** and **sellers from Cologne** always trade on the **Cologne market**.

Buyers do not stay on the same market, but switch between the two markets. There are two different types of buyers: **type D** buyers are more likely to trade in Dallas and **type C** buyers are more likely to trade in Cologne. Specifically:

- Buyers of **type D** in each round have a probability of 80% to trade on the Dallas market and a probability of 20% to trade on the Cologne market.
- Buyers of **type C** in each round have a probability of 20% to trade on the Dallas market and a probability of 80% to trade on the Cologne market.

There are as many buyers of type D as of type C. Each buyer, regardless of whether he/she is a participant from Dallas or Cologne, has the same probability to be chosen a type D or type C buyer. At the beginning of the experiment, each buyer will be randomly assigned a type (type D or type C), which remains the same for the entire experiment.

At the beginning of each round, each buyer is assigned to trade on the Dallas market or the Cologne market. The assignment is done according to the probabilities laid out above. The market chosen will be shown on the buyer's screen. Each buyer is then matched randomly to a seller on this market. It is guaranteed, that buyers are not matched to the same seller in two consecutive rounds. The buyer gets to know how many positive and negative feedbacks the matched seller received so far. The seller gets to know whether the matched buyer is a participant from Dallas or Cologne.

Trade

At the beginning of each round, sellers and buyers receive an endowment of 100 ECU. In each round the seller offers a good to the buyer and the buyer decides whether he/she wants to buy the good at a price of 100 ECU. If the buyer decides not to buy the good both, seller and buyer, keep their endowment of 100 ECU and the round ends. If the buyer decides to buy the good, he/she pays a price of 100 ECU to the seller. At the same time the seller chooses the quality of the good he/she wants to send to the buyer in case he/she buys the good. Quality ranges between 0 and 100. Each quality point costs the seller 1 ECU and increases the good's value to the buyer by 3 ECU, for example:

- If the quality is 0, the seller has costs of 0 ECU and the buyer receives a good with a value of 0 ECU.
- If the quality is 50, the seller has costs of 50 ECU and the buyer receives a good with a value of 150 ECU.
- If the quality is 100, the seller has costs of 100 ECU and the buyer receives a good with a value of 300 ECU.

Overview of period payoffs for sellers and buyers:

	Seller	**Buyer**
Buyer does not buy:	100 ECU	100 ECU
Buyer buys:	200 ECU – Quality	3 x Quality

Feedback

The buyer then gets to know the quality that the seller has chosen. Afterwards, the buyer may leave a feedback rating for the seller. Leaving a feedback rating costs 1 ECU. The feedback can either be “positive” or “negative”. In the following rounds buyers get to know how many positive and how many negative feedbacks the matched seller has received so far. The seller gets to know whether and what kind of feedback he/she received from the matched buyer and a new round starts.

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

4.7 *Appendix C – Screenshots*

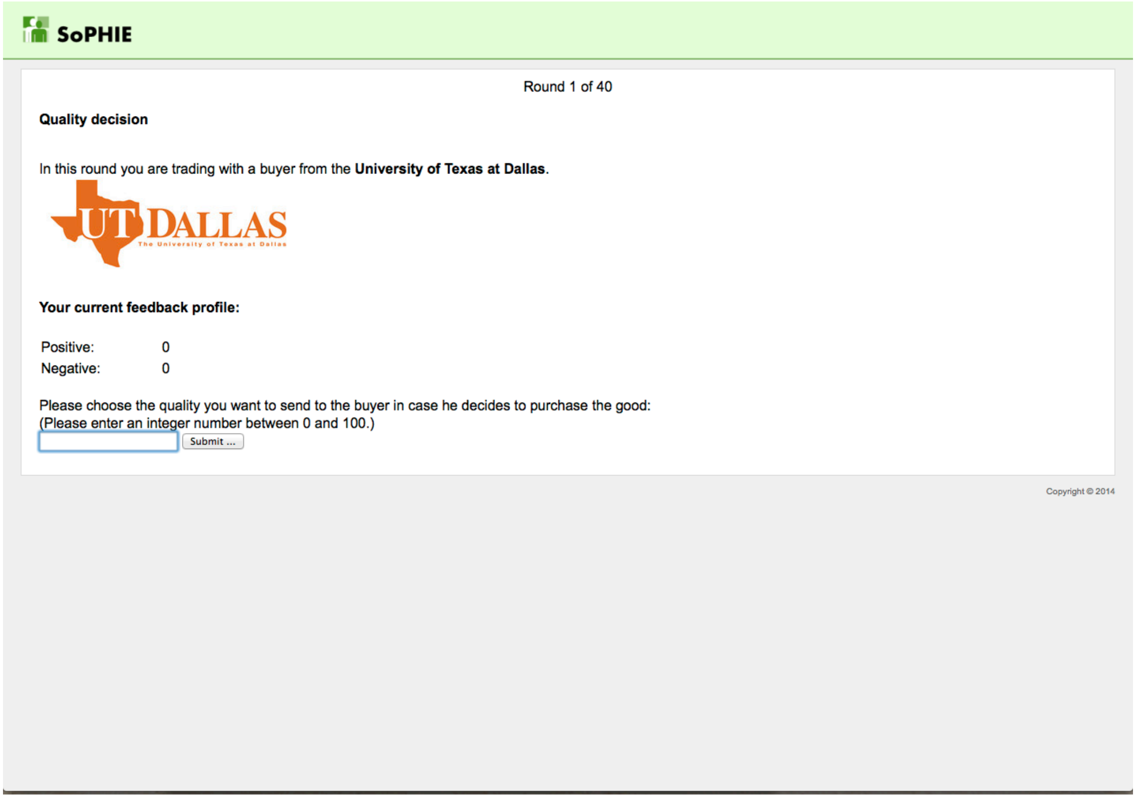


Figure 4.16: Quality decision

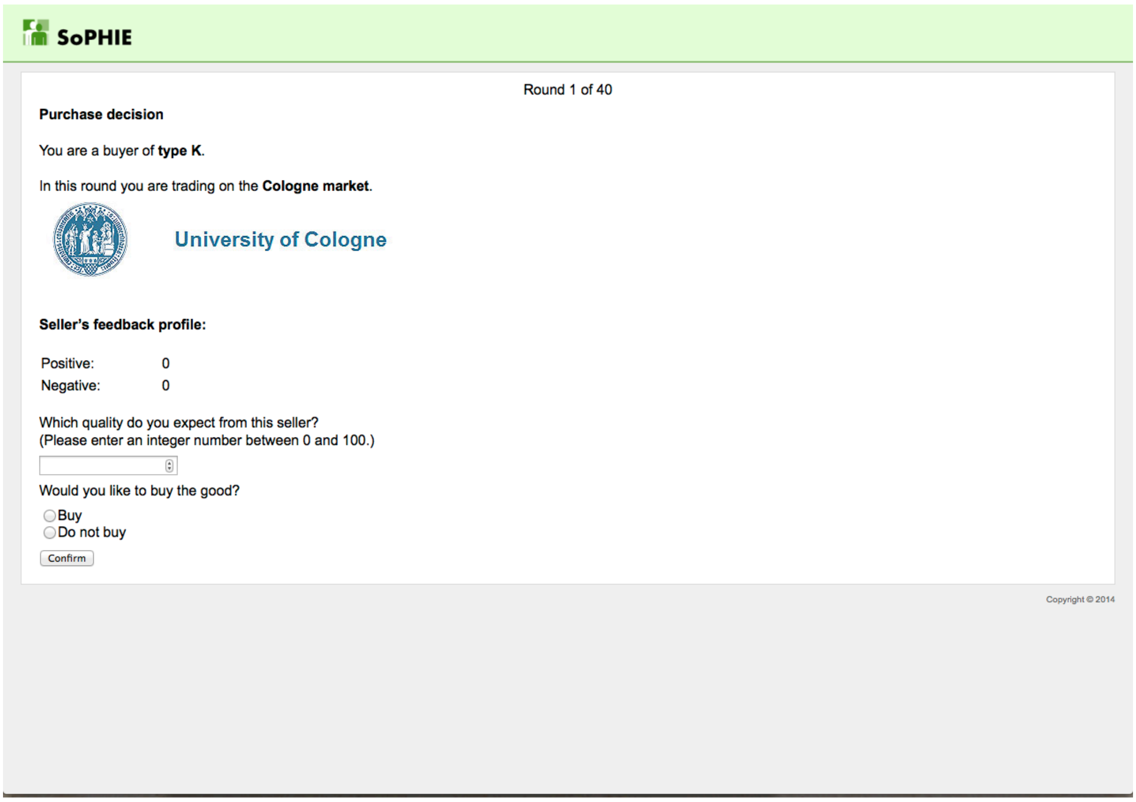


Figure 4.17: Purchase decision

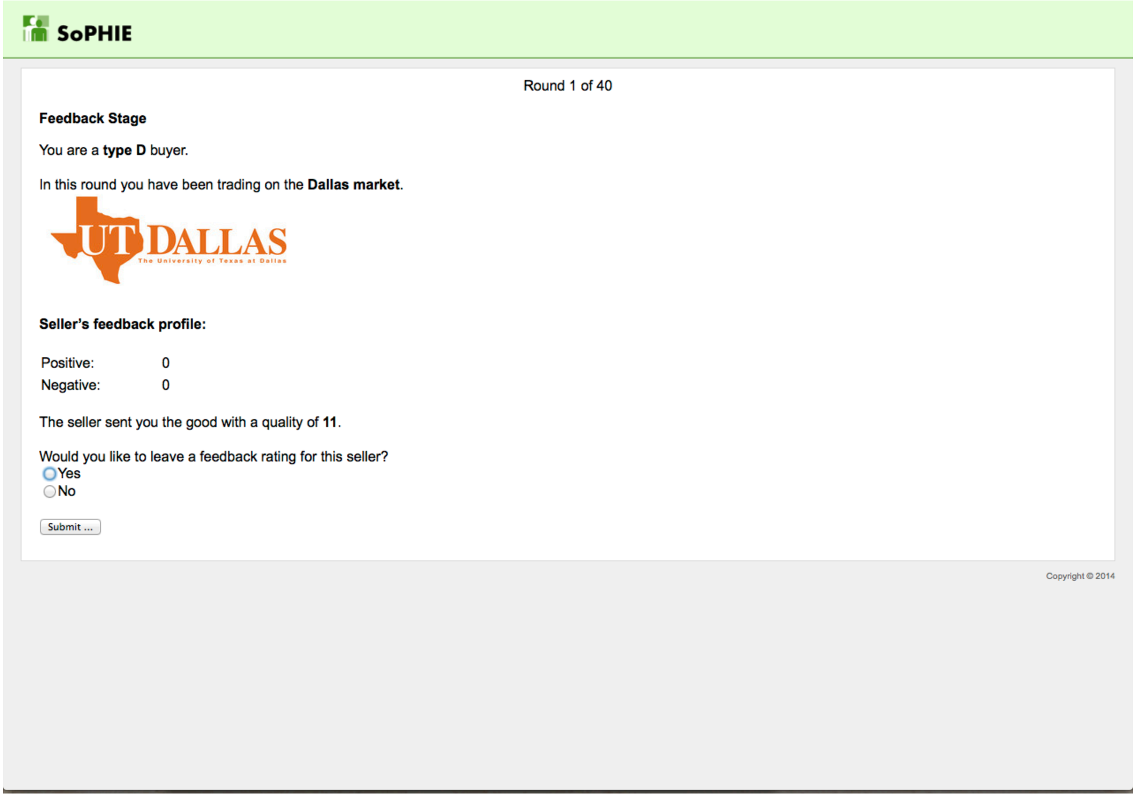


Figure 4.18: Feedback decision

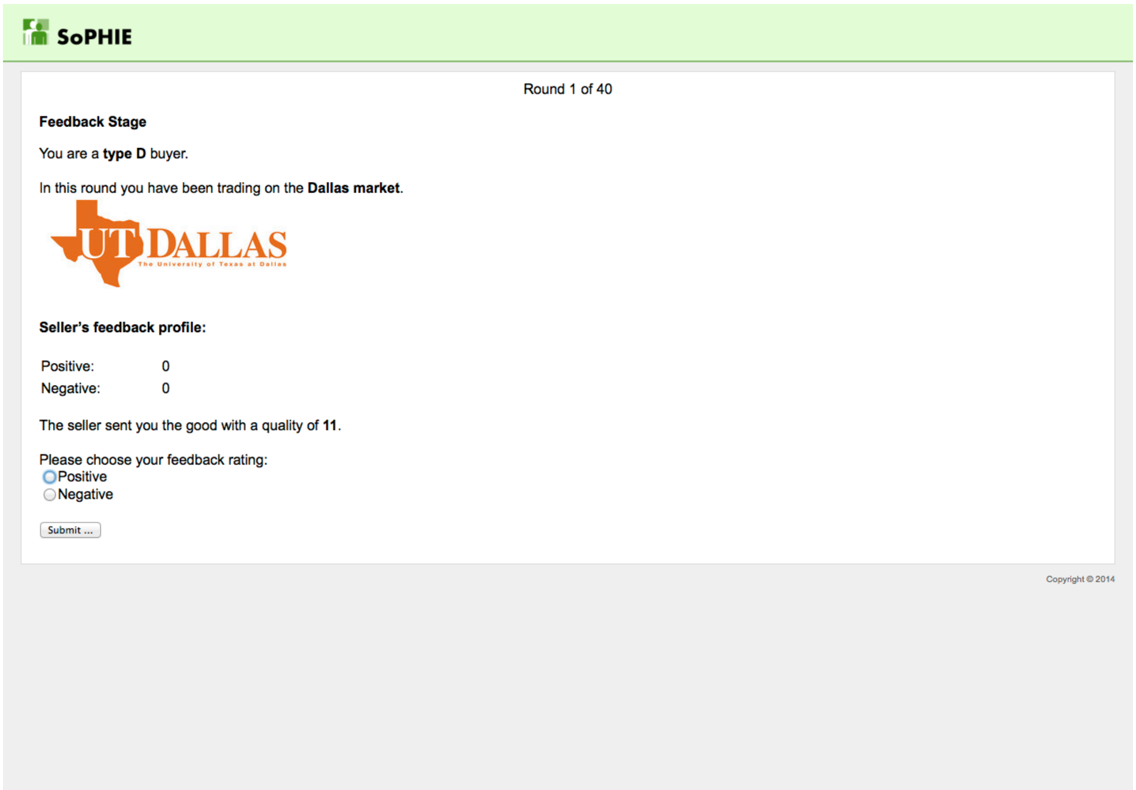


Figure 4.19: Feedback rating

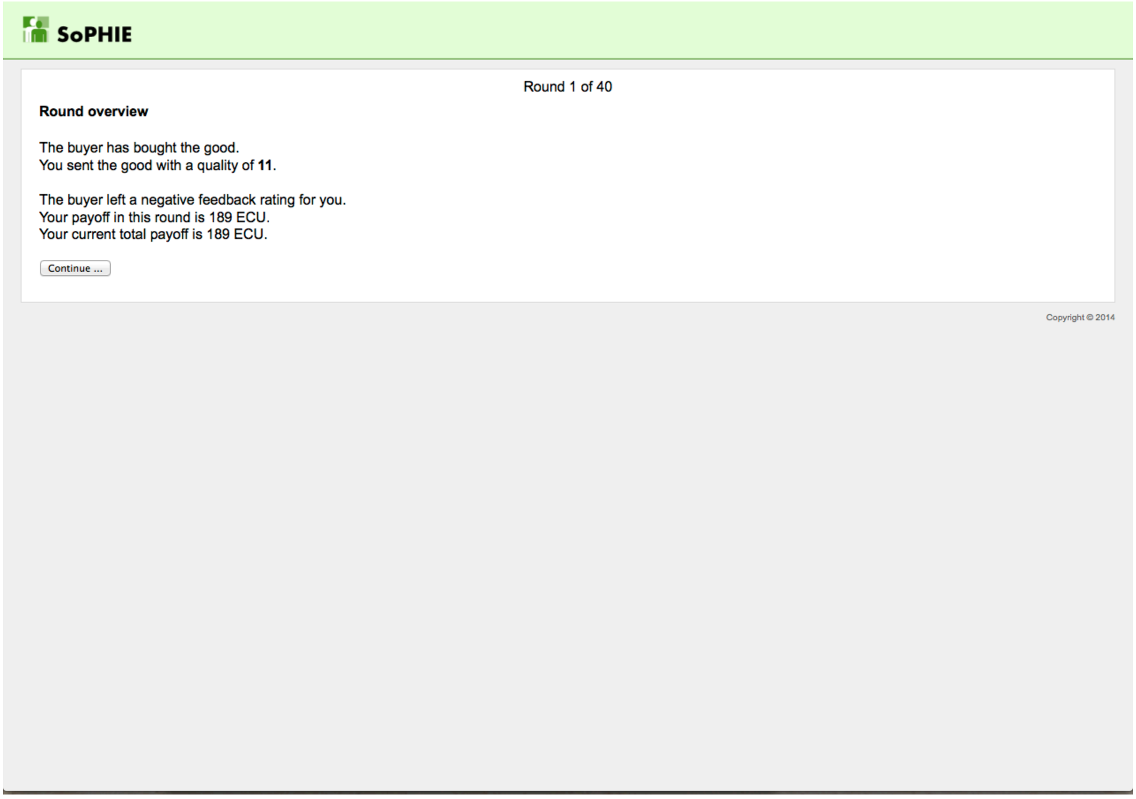


Figure 4.20: Period summary seller

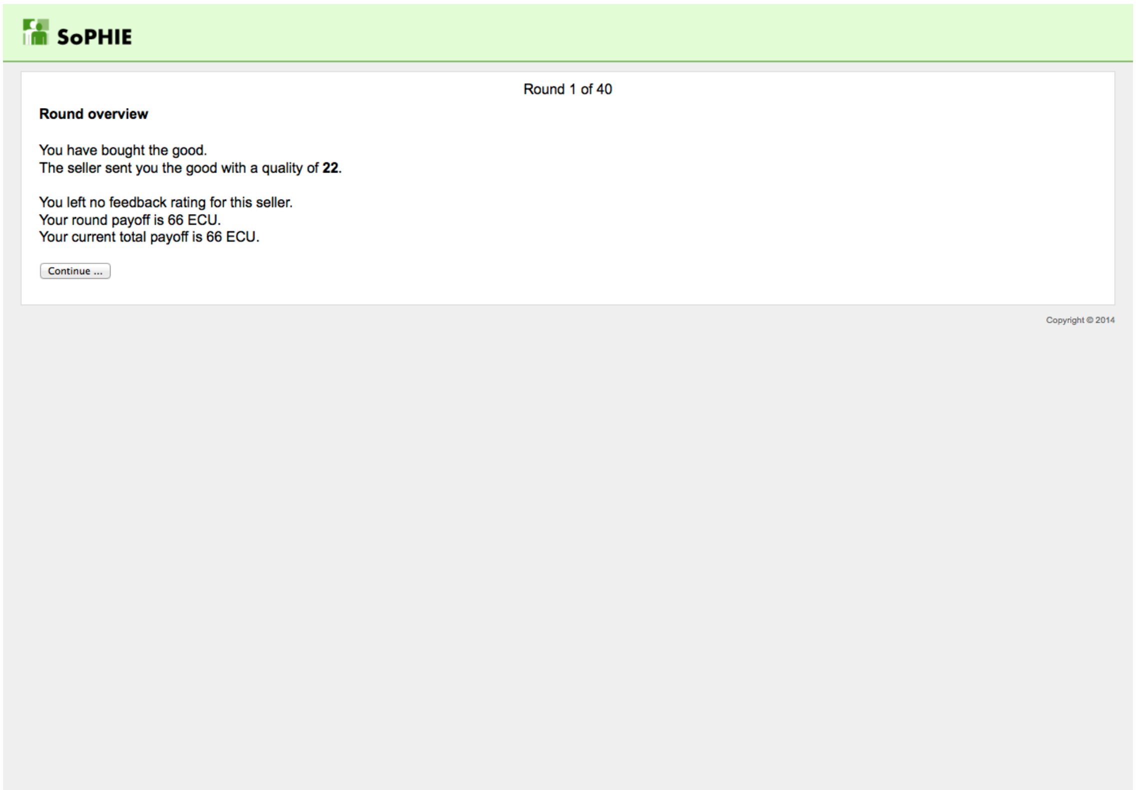


Figure 4.21: Period summary buyer

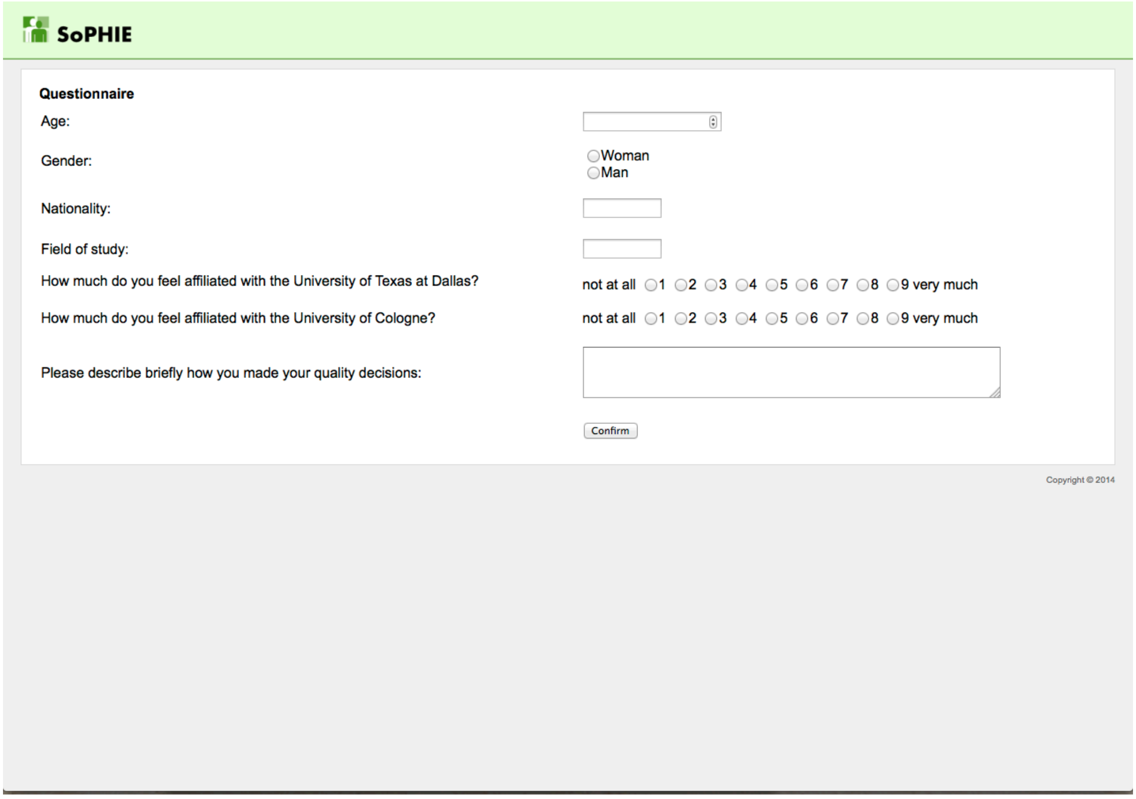


Figure 4.22: Questionnaire seller

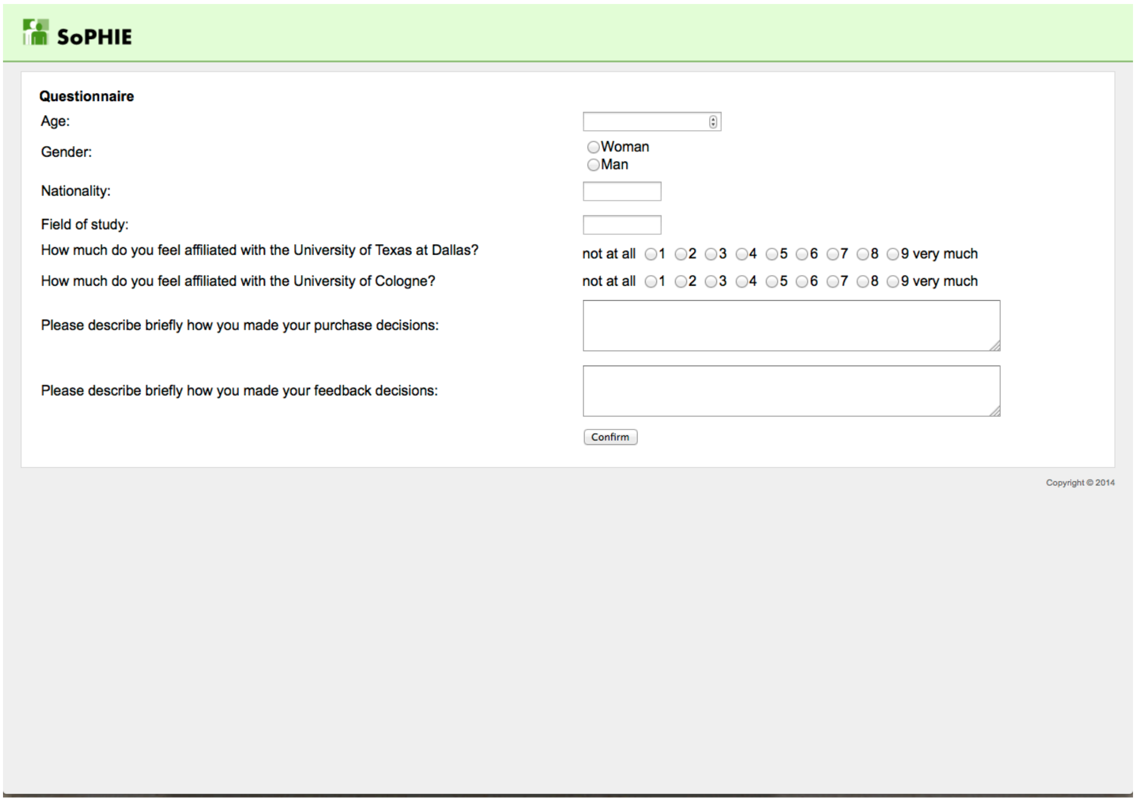


Figure 4.23: Questionnaire buyer

Chapter 5

Who Do You Lie to? Social Identity and the Costs of Lying

5.1 INTRODUCTION

Communication is a central element in almost every economic interaction.[89] However, when people communicate they may not always tell the truth. In fact, in situations where one party possesses better information than the other, standard economic theory would predict that fully rational people do not tell the truth whenever the monetary benefits exceed the expected fines from detection (e.g. Becker 1968). In contrast to this rational perspective, some recent experimental studies have shown that in situations with asymmetric information individuals are not always willing to capitalize on their informational advantage, and do not tell lies despite monetary incentives to do so (Gneezy 2005; Gibson, Tanner, and Wagner 2013; Hurkens and Kartik 2009; Sutter 2009; Fischbacher and Föllmi-Heusi 2013). Thus, not all people follow a solely consequentialist approach and assess the monetary outcome only but incur costs of lying and also take the means to the end into account.

In this paper we investigate whether and how an individual's propensity to lie is influenced by characteristics of the receiver. In particular, we study in a modified sender-receiver game whether a sender's costs of lying are larger when sender and receiver share a common social identity. We argue that social identity can affect lying

[89] This chapter is joint work with Christoph Feldhaus. Both authors contributed equally to the idea of this project, the design and organization of the classroom experiment, statistical analyses, and the writing of this draft. We thank Axel Ockenfels, Matthias Sutter, Peter Werner, David Kusterer, and Christopher Zeppenfeld for helpful comments and suggestions. Financial support of the German Research Foundation (DFG) through the Research Unit "*Design & Behavior – Economic Engineering of Firms and Markets*" (FOR 1371) is gratefully acknowledged.

behavior in two different ways. First, senders take receivers' payoffs more into consideration if they share an identity (*allocative* costs of lying). Second, in addition to these monetary consequences, senders might also be more reluctant to tell a lie to someone who is a member of the same social group, because they feel more obliged to live up to social norms (*social* costs of lying). A better knowledge of whether and through which channel(s) social identity affects deceptive behavior helps to understand what drives individuals' costs of lying.

The subsequent section discusses the related literature and introduces our hypotheses. In the following, we describe the experimental design including our treatment variations. We then present our results, which indicate that social identities do not affect lying behavior and that neither allocative nor social costs play an important role in our setting. Finally, the conclusion discusses potential explanations for these results.

5.2 RELATED LITERATURE AND HYPOTHESES

Recent experimental evidence clearly demonstrates that many decision makers have an aversion to lie and include more truthful information in their messages than predicted by economic theory (Dickhaut et al. 1995; Cai and Wang 2006; Sánchez-Pagés and Vorsatz 2007).[90] When deciding whether to tell the truth or not, individuals consider their own monetary gains as well as the losses their lie may cause to others (Gneezy 2005; Sutter 2009; Erat and Gneezy 2012; Gneezy et al. 2013; Conrads et al. 2014). For a considerable fraction of people the aversion to lie is so strong that they prefer to tell the truth even if lying may lead to a Pareto-improvement (Erat and Gneezy 2012; López-Pérez and Spiegelman 2012). Recent economic models therefore incorporate this intrinsic lying aversion and assume that individuals experience psychological disutility from misrepresenting private information (e.g. Kartik 2009).

In order to gain a better understanding of these costs of lying, various factors that affect an individual's decision to tell the truth have been investigated. Besides monetary

[90] The aversion to lie seems to be driven by individuals' urge not to act or appear dishonest and not by the preference to tell the truth. For instance, Sánchez-Pagés and Vorsatz (2007) show that otherwise honest subjects are likely to remain silent and send neither false nor correct information when given the opportunity. Similarly, people often try not to appear dishonest and disguise their lies by telling just 'incomplete lies' which do not maximize their profits (Conrads et al. 2014, 2013; Fischbacher and Föllmi-Heusi 2013).

consequences, also situational factors such as diffusion of responsibility and the number of people who benefit from the lie have been found to play an important role (Gino et al. 2009; Wiltermuth 2011; Conrads et al. 2013; Erat 2013; Gino et al. 2013). Furthermore, honesty is also affected by contextual factors – characteristics of the lie itself – for example when the content is personal or the more personal information needs to be exaggerated (Lundquist et al. 2009; Cappelen et al. 2013). Additionally, individual characteristics of the potential liar, e.g., gender, self-control, ego depletion and certain personality traits, correlate with lying behavior (Trevino 1986; Ross and Robertson 2000; Tyler and Feldman 2004; Tyler et al. 2006; Dreber and Johannesson 2008; Mead et al. 2009; Childs 2012; Cappelen et al. 2013). This demonstrates that the aversion to lie is heterogeneous between and within individuals and not an either-or decision between always or never telling the truth (Gibson et al. 2013).

In recent years the psychological concept of social or group identity developed by Tajfel and Turner (1979) has been incorporated into the field of economics. Social identity is usually defined as a person's sense of self which is derived from actual or perceived membership in social groups (Akerlof and Kranton 2000; Chen and Li 2009). Affiliation and identification with a social group induce people to interact differently with members from this particular group (ingroup) as compared to members from another group (outgroup). Experimental studies from psychology and economics show that a shared social identity affects people's social preferences and thereby leads to ingroup favoritism, meaning that decisions are taken to the advantage of the ingroup and often at the expense of the outgroup. In this sense, an individual's social preferences and the resulting actions can be group contingent. For example, Chen and Li (2009) find that in dictator and response games, subjects display more altruism and less envy in allocation decisions when paired with another ingroup subject than with someone from an outgroup. Besides altruistic behavior also reciprocal behavior is affected by group identity: Reciprocating good behavior is more pronounced among ingroup members while punishing bad behavior is stronger with an outgroup match (Bernhard et al. 2006; Goette et al. 2012; Mussweiler and Ockenfels 2013). Moreover, it has been shown that a shared social identity increases cooperation in dilemma games (Eckel and Grossman 2005; Goette et al. 2006; Charness et al. 2007; Goette et al. 2012), has a positive effect on trust and reciprocity (Heap and Zizzo 2009) and improves efficiency in coordination games (Bornstein et al. 2002; Chen and Chen 2011).

However, to the best of our knowledge, it has not been investigated yet how the relationship between sender and receiver influences the sender's propensity to lie in an economic setting. In the context of misrepresenting private information we argue that social identity may affect lying behavior in two different ways. First, in economic settings, lies usually result in different monetary allocations for the liar and the receiver of a lie. As outlined above, people have stronger charity concerns for members of the same social group and therefore put a higher weight on their interaction partner's profit in ingroup relationships (Chen and Li 2009). Thus, senders should be more likely to refrain from telling a harmful lie if an ingroup receiver is affected. Second, in addition to these outcome-based considerations, individuals may also have a stronger intrinsic motivation to behave honestly towards a member from the same group per se. According to Jones (1991), lying itself, as an unethical and immoral act, is also affected by the proximity between the potential liar and the potential victim irrespective of the monetary consequences. He argues that the moral imperative becomes stricter with the feeling of closeness on a social, cultural, psychological or even physical level between two parties. Similarly, DePaulo and Kashy (1998) argue that telling a lie violates important ideals of close relationships such as relatedness, openness and trustworthiness and that for these reasons, people feel more uncomfortable and distressed when they lie to people to whom they are closer. Therefore, lying to someone from your ingroup might also incur higher costs because you feel closer to this interaction partner. Survey data on everyday lies find that people are less likely to tell a harmful lie the closer the relationship with the recipient of the lie is (DePaulo and Kashy 1998; Ennis et al. 2008). In a similar vein, using survey data, Ross and Robertson (2000) demonstrate that in an organizational context the probability to lie depends on the closeness of the relationship between liar and victim as salespeople are less likely to represent false information to someone from their own firm than from a competitor. In addition, Chakravarty et al. (2012) show that subjects are less likely to deceive a friend than an anonymous stranger in an experimental setup where lying may lead to an increase of their own profit but decreases the other person's profit.

Following these arguments, we assume that there are two distinct preference-based costs of lying: *allocative costs*, related to the monetary outcomes of the lie and *social costs* associated with the closeness of the relationship. We hypothesize that these preference-based types of costs are more important when the sender is matched with an individual from the same social group in comparison to an outgroup match. Therefore, lies should be less likely between ingroup members.

5.3 EXPERIMENTAL DESIGN

Our experimental design is similar to the two-player sender-receiver game used by Erat and Gneezy (2012) except that we introduce a third inactive person (dummy). As in Erat and Gneezy (2012) a sender has private information about the true state of the world and sends a (dis)honest message about the state of the world to the receiver. The receiver then takes an action, which, in combination with the state of the world, determines the payoffs for all group members. The new, third player receives the same payoff as the receiver but neither gets the message of the sender nor takes a decision. Thereby, the sender lies only to the receiver but not to the dummy player and thus this third player should only affect the sender's *allocative* but not her *social* costs of lying. By varying the identities of the receiver and the dummy separately we are able to test how allocative and social costs are affected by the relationship between the liar and the potential victims.

The three-player deception game

For each group we independently rolled a six-sided dice and communicated the outcome only to the sender in the instructions.[91] After being informed about the outcome of the dice roll, we asked the sender to choose one of six possible messages – one for each possible outcome – to be sent to the matched receiver. This set of possible messages consisted of: "The outcome of the roll of dice was *i*", where $i \in \{1, 2, 3, 4, 5, 6\}$. We made clear, that the chosen message would be transmitted only to the receiver and that this message would be the only information given to the receiver regarding the roll of the dice. The sender also knew that the dummy would not receive any information and would not take any decision in the experiment. The sender was told that after receiving her message we would ask the receiver to also choose a number from the set {1, 2, 3, 4, 5, 6} and that the receiver's choice would determine payoffs for all participants in the following way: If the receiver chooses the actual outcome of the dice, all participants receive a payment of 7 € and otherwise – if the receiver chooses a

[91] In the instructions of the game we used a neutral framing: PARTICIPANT A, B and C.

different outcome – the sender gets 11 € while receiver and dummy get 5 €. Furthermore, we told the sender that she is the only one who knows these two payment schemes and that receiver and dummy would only be informed about their own individual payoffs of the implemented option. Erat and Gneezy (2012) argue that this expanded message space with six instead of two options makes it less likely that senders engage in sophisticated deception – telling the truth just because one believes that the receiver will not follow (Sutter 2009). In order to make an honest message more profitable than a deceptive message, a sender would need to believe that more than 83% of the receivers will not follow the sent message. We assume that most senders interpret their message rather as an allocative and not as a strategic decision and send a truthful message only when they really want to be honest.[92]

	Sender	**Receiver**	**Dummy**
Option 1 Receiver chooses **actual** number	7 €	7 €	7 €
Option 2 Receiver chooses **different** number	11 €	5 €	5 €

Table 5.1 Payoff schemes depending on receiver's choice

The receiver and the dummy were informed that we rolled a dice for the sender, revealed the outcome only to the sender and asked her to send a message only to the receiver.[93] After receiving this message the receiver would be asked to choose a number from the set of possible outcomes {1, 2, 3, 4, 5, 6}. They were told that this choice would determine the payoffs for all three participants but that they would only get to know their own payoff at the end of the experiment and neither the other players' payoffs nor the payoff scheme of the alternative option. They only knew that if the receiver chooses the number that corresponds to the actual outcome of the dice one option would be implemented and for any other choice a second option would be implemented. It was common knowledge that even after the experiment the actual outcome of the dice would not be revealed to the receiver and the dummy. Hence, a lie by the Sender could not be detected by either of the two interaction partners.

[92] Additionally, Gneezy (2005) shows that in his experimental design with only a binary message space, 82% of the senders expect the receiver to follow their recommended option.

[93] We made very clear in the instructions that the dummy player would neither receive the sender's message nor take a decision in this experiment.

Treatment manipulations

We use a between-subjects design to investigate how social identity affects the allocative and social costs of lying. Senders are matched with receivers and dummies either from their in- or outgroup or a combination of both identities. We employ natural groups to induce social identities, as our participants were students from two different universities: the University of Cologne (UoC) or the University of Duesseldorf (UoD). Subjects in the role of the sender were always from UoC while receiver and dummy participants were enrolled either at UoC (ingroup match) or at UoD (outgroup match).[94] Thereby we have a full factorial 2x2 design with the university affiliations of receiver and dummy players as treatment variations: a sender either shares the social identity with both receiver and dummy (T1 CC), with only one of the two – either with the receiver (T2 CD) or the dummy (T3 DC) – or with none (T4 DD). According to our hypothesis, there are allocative and social costs of lying and a shared identity increases both these preference-based costs. Therefore, lying behavior should differ between treatments and result in the following order of treatments regarding the share of lies:

T4 DD Outgroup/Outgroup	>	**T3 DC** Outgroup/Ingroup	>	**T2 CD** Ingroup/Outgroup	>	**T1 CC** Ingroup/Ingroup

Costs of lying should be lowest when a sender is matched with two outgroup members (T4). In comparison, being matched with an ingroup dummy (T3) increases a seller's allocative costs while being matched with an ingroup receiver (T2) increases allocative and social costs. Finally, when both partners come from the same social group (T1) allocative costs rise again and overall costs should be highest.

These preference-based explanations assume that senders care about group identity per se. Hence, for these allocative and social costs to have an impact on senders' lying behavior it is sufficient that only the sender knows about the identities of her interaction partners.[95] Therefore, we inform only senders whether the matched receiver and dummy

[94] These two natural groups have already been used successfully in earlier experiments on social identities and lead to significant differences in, e.g., punishment behavior (Mussweiler and Ockenfels 2013) and dictator giving (Ockenfels and Werner 2014).

[95] In their seminal paper, Chen and Li (2009) also do not inform the recipients whether the received money comes from an ingroup or outgroup decision maker. However, this is not explicitly mentioned in the instructions. Moreover, due to their specific design, each participant takes the role of the recipient and the role of the decision maker at the same time. Therefore, all participants are aware of the fact that members of two different groups are involved in the experiment.

are students from UoC or UoD.[96] Conversely, receiver and dummy participants did not get to know that senders were from the University of Cologne and senders were informed about this.[97]

Procedure

We collected the data for all three different roles sequentially. Sender behavior was collected first in two classroom sessions, while receiver and dummy subjects were recruited via Orsee (Greiner 2004) from the subject pool of either the laboratory at the University of Cologne or at the University of Duessseldorf. Sessions were run in the respective laboratory using pen and paper. This sequential procedure made it necessary that decisions of senders and payments to senders take place in separate sessions. Therefore, we revisited both classroom sessions in the following week handing out the payments to the senders. For this purpose all senders received a card with a unique identification number along with the instructions.[98] This identification number was also printed on the decision sheet on which senders recorded their message to the receivers. Subjects had to hand over their ID card to us in order to receive their payment. We randomly selected 100 of all participating senders for payments. We draw these 100 senders after the first two classroom sessions and published these ID numbers via an online learning tool to all course members and also at the beginning of each payment session in the classroom.

To collect sender behavior we visited two consecutive sessions of the undergraduate course "Introduction into Microeconomics". These two sessions are identical courses at different times and are offered to students as alternatives due to a limited number of seats and possible overlaps with other courses. Since the teaching content in both

[96] To increase salience, university affiliation of the receiver and the dummy were mentioned throughout the senders' instructions.

[97] Using this information structure, we focus at first on preference-based explanations for the influence of social identity on lying behavior. In a second experiment we test an alternative, belief-based explanation which is discussed and presented at the end of the results part.

[98] Instructions for sender's were handed out in large envelopes and consisted of 4 items: ID card, general instructions, game instructions with decision sheet and questionnaire. Instructions for receivers and dummies consisted only of game instructions with decision sheet and questionnaire. A full set of instructions for all three roles can be found in the Appendix.

sessions is identical students usually attend only one of the two.[99] In general, students from different fields are attending, while most of them study business administration or economics and are usually in their first year at university. In both sessions we randomly distributed envelopes of all four treatments. Participation was voluntary but almost all students participated.[100] Overall 545 students participated as senders in the two classroom sessions, 284 in the first and 261 in the second. We have a similar number of sender participants in all four treatments, ranging between 131 and 140. Additionally, 100 receivers (53 from UoC) and 100 dummies (53 from UoC) took part in the laboratory sessions.[101]

After subjects made their decisions they were asked to complete a short questionnaire including demographic information such as age, gender, course of study, year of study, zip code and mother tongue. To elicit senders' affiliation with the University of Cologne we also asked them to indicate whether studying at the University of Cologne is an important part of their self-image on a nine-point Likert scale (1-9).[102]

5.4 *RESULTS*

Across all treatments, the average probability to lie is 49% (see Table 5.2).[103] We thus find clear evidence for an aversion to lie, as half of the participants are not willing to send a false number to increase their own payoffs. In our setup the share of lies is very similar to the results in comparable treatments in Gneezy (2005) where lying also has

---

[99] We cannot fully rule out that some students attended both classroom sessions. However, we made very clear that it is not allowed to participate more than once in this experiment. Based on answers given in the questionnaire we suspect in one case that a subject participated twice. We excluded both observations.

[100] Only five participants decided not to write down a message on the decision sheet. We excluded these observations.

[101] When recruiting receivers and dummy participants for the laboratory sessions we made sure that they are students at UoC or UoD and do not attend the course where we conducted the sender sessions.

[102] Table 5.4 in Appendix A lists the demographic characteristics from the post-experimental questionnaire of the participants in the role of senders. Nearly half of the subjects are female, the average age is 20 and most of them speak German as their mother tongue (83%). Nearly two thirds are business students, one third comes from economics while the rest has another background. More than 90% are in their first year at university but already show an average affiliation of 6 on Likert-scale from 1 to 9. Randomization into the four different treatments worked very well as there are no major treatment differences for all characteristics with the exception that in T1 there are less business and more economics students ($p < 0.05$, chi^2-test) and subjects in T2 are slightly older than in all other three treatments ($p < 0.1$, t-test).

[103] The variable lie is a dummy variable equal to 1 if the message sent by the sender is a different number than the actual outcome of the roll of the dice.

no welfare effects as the monetary gain of the sender equals the loss of the receiver.[104] However, between our four treatments there are no significant differences in the probability to lie. In all treatments roughly half of the subjects send a deceptive message to the receiver (between 47% and 51%).[105]

As a robustness check, we look at lying behavior in different subpopulations for which the social identity of the interaction partner should be even more important or who might have been better able to understand the instructions. These subgroups are 1) participants who report a high (> 5) affiliation with the University of Cologne, 2) subjects who currently live in Cologne (based on zip code) and 3) those who speak German as a first language.[106] However, we do not find any differences in lying behavior between treatments in one of these subgroups (Table 5.2).

Share of lies	**T1 CC**	**T2 CD**	**T3 DC**	**T4 DD**	**Overall**	**N**
All subjects	51%	48%	47%	51%	49%	545
High affiliation	53%	47%	50%	48%	49%	371
From Cologne	53%	48%	47%	54%	50%	398
Mother tongue German	49%	50%	47%	50%	49%	453

Table 5.2: Share of lies across treatments for different groups of subjects.

As a further test we ran Probit regressions for the same groups with lie as dependent variable and controlling for demographic characteristics of our subjects (see Table 5.3).[107] In all four models, none of the treatment dummies have a significant effect on the probability to lie. Additionally, pairwise comparisons between all treatment dummies using Wald tests do not yield any significant difference. Overall, we find no evidence that social identity affects individual lying behavior neither through higher allocative nor through higher social costs.

[104] Receiver behavior is less trusting than in Gneezy (2005) as in our experiment only 65% of the receiver choose to follow the message by the sender. However, this figure is in line with trust rates from other lying experiments such as Hurkens and Kartik (2009).

[105] A pairwise comparison of treatments using *chi*^{2}-tests yields a p value of 0.24 for T2 versus T4, all other comparisons result in larger p values. For more details see Tables 5-8 in Appendix A.

[106] We differentiate between native and non-native German-speaker because it could be that foreign students may misunderstand the instructions especially with regard to the difference between the role of the receiver and the dummy. In addition, we also hypothesize that our social identity manipulation is more effective with German participants.

[107] We exclude seven additional observations in the regressions because these subjects did not fill in all information in the post-experimental questionnaire. One participant did not fill in any information in the questionnaire; six participants answered all questions except for zip code.

Regarding our control variables, we find that older subjects are less likely to send a deceptive message. Furthermore, looking at participants from Cologne, we also find evidence for a gender effect, indicating that women are less likely to lie. However, for the entire group of participants this effect is only significant on a 10% level. These results are in line with findings in other studies from e.g. Ross and Robertson (2000), Dreber and Johannesson (2008) and Conrads et al. (2013) who also find that age has a negative effect on lying and that women are more honest than men.[108]

Lie (yes/no)	**All subjects**	**Affiliation > 5**	**Cologne**	**German**
T2 CD	-0.040	-0.113	-0.076	0.062
	(0.156)	(0.190)	(0.178)	(0.171)
T3 DC	-0.093	-0.105	-0.102	-0.072
	(0.152)	(0.186)	(0.178)	(0.168)
T4 DD	0.006	-0.147	0.047	-0.007
	(0.152)	(0.184)	(0.177)	(0.166)
Age	-0.089***	-0.075**	-0.082**	-0.101***
	(0.025)	(0.029)	(0.029)	(0.028)
Female	-0.197	-0.159	-0.306*	-0.164
	(0.111)	(0.134)	(0.130)	(0.123)
Economics	-0.003	-0.031	0.066	-0.026
	(0.119)	(0.144)	(0.142)	(0.132)
Other field of study	0.224	0.331	0.126	0.318
	(0.230)	(0.341)	(0.240)	(0.245)
Constant	1.904***	1.673**	1.839**	2.086***
	(0.528)	(0.617)	(0.611)	(0.596)
Log-likelihood	-368,76	-251,71	-268,17	-305,58
Observations	544	370	397	452

Table 5.3: Probit regressions on lying for different groups of subjects. Standard errors in parentheses; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Do senders' beliefs matter for the effect of social identity on lying behavior?

Up to now, we focused on preference-based explanations for the influence of social identity on the costs of lying. Therefore, only senders were informed about other players' social identities. However, in general, social identity may also have an effect on lying behavior by inducing senders to have different beliefs about the payoffs receivers and dummy players expect from an ingroup an outgroup, respectively. Not

[108] There is an ongoing debate on which gender is more likely to lie. Besides the aforementioned studies which show that women are lying less often, there are also other studies finding no differences between the sexes (DePaulo et al. 1996; Childs 2012; Cappelen et al. 2013) or even more deceptive behavior by women (Tyler and Feldman 2004; Tyler et al. 2006).

living up to these perceived expectations may then lead to a feeling of guilt on the part of the sender. Recent studies have shown that such second-order beliefs indeed also motivate ingroup favoritism as it is more pronounced when participants' identities are public knowledge instead of dictators' private knowledge (Güth et al. 2009; Ockenfels and Werner 2014). Yamagishi and Mifune (2008) even find no evidence for an ingroup favoritism in a dictator game when the dictator is privately informed about the group membership of the receiver. In a sender-receiver setup like ours, however, the potential role of sender's beliefs about her partners' expectations is more complex because only the sender has full information on the structure and payoffs of the game. In contrast, receivers and dummy players do not know the payoffs behind the two possible options and thus are also unaware that the sender has a financial incentive to send a dishonest message. Still, Battigalli et al. (2013) demonstrate that the theory of guilt aversion by Battigalli and Dufwenberg (2007) can explain lying behavior as observed in the sender-receiver game by Gneezy (2005). They argue that a sender still anticipates bigger disappointment on the part of the receiver (dummy player) even when the receiver does not know the alternative payoff(s).[109] Conditional on a given payoff, a sender might perceive the receiver's (dummy player's) disappointment to be even greater on average when resulting from an ingroup interaction.

In this regard, our current design with private information about identities may constitute a (too) conservative test on how social identities influence lying behavior, as it does not allow for senders' beliefs about partners' expectations to vary between treatments. In order to incorporate such a belief-based effect of social identity on senders' lying behavior, we run an additional set of treatments with the only difference that at the beginning of the experiment each player – irrespective of his role – gets to know the identities of the other two players he/she is matched with. Overall, 341 students from two first year economics courses participated in this second

[109] Battigalli et al. (2013) emphasize that this simple form of guilt cannot be the only driver of people's aversion to lie because many people still refrain from lying even if this leads to Pareto improvements as seen in Erat and Gneezy (2012). In a recent experimental study, Peeters et al. (2015) investigate the role of second-order beliefs for truth telling. According to their results, senders' messaging behavior is uncorrelated with their second-order beliefs. Similarly, with regard to promises, Vanberg (2008) finds that people mainly seek to keep their own but not other dictators' promises although their second-order beliefs about receivers' expectations are very similar in both cases. Both studies suggest that people have a preference for honesty or promise-keeping per se rather than fulfilling perceived expectations of others.

experiment.[110] The average lying rate over all treatments is 47% and is almost identical to the first experiment (49%). With regard to treatments, lying rates are 45% in T1 CC, 43% in T2 CD, 58% in T3 DC, and 42% in T4 DD.[111] Hence, three out of four treatments show very similar lying rates which are not significantly different from each other (pairwise comparisons using chi^2-tests result in p values of at least 0.54). Only the third treatment stands out, as there are more dishonest than honest messages. This third treatment is not significantly different from T1 (chi^2; $p = 0.16$), but significantly different from T2 and T4 (chi^2; $p = 0.05$ and $p = 0.04$, respectively).[112] While the difference between T2 and T3 is in line with the hypothesized order of treatments, the latter is not, as we expected an increase in lying behavior from T3 to T4.

Hence, also when identities are revealed to all players and senders' beliefs can play a role, we do not find compelling evidence that social identities influence lying behavior in line with our predictions.

5.5 DISCUSSION AND CONCLUSION

In this paper we investigated whether the social relationship between the potential liar and the possible victim alters people's propensity to lie. More specifically, we argued that a shared social identity of sender and receiver might increase senders' aversion to lie by raising the allocative and social costs of lying. We varied social identities by matching participants from different universities. The same natural identities have already been used successfully in recent laboratory and classroom experiments to study effects of social identity (Mussweiler and Ockenfels 2013; Ockenfels and Werner 2014).[113] In contrast to our hypothesis, however, our experimental results from a modified three-person sender-receiver game do not provide evidence that social

---

[110] The second part of the experiment took part in June 2015 in the courses "Introduction into Microeconomics" and "Introduction into Macroeconomics". To make sure that no one participated twice, we asked all participants of the second part in the post-experimental questionnaire which of the three courses they currently are or have been attending (Microeconomics in winter 2013/14, Microeconomics in summer 2015, Macroeconomics in summer 2015). Based on this, we excluded 31 participants who either stated to have attended more than one of these courses or did not answer all three questions.

[111] Looking at the same subpopulations (affiliation > 5; from Cologne; German) as before does not change the lying rates across treatments substantially (see Table 5.9 in the Appendix).

[112] Using a probit regression to control for additional factors such as age, gender, and course of study shows similar treatment differences but with lower significance levels (see Table 5.10 in the Appendix).

[113] In addition, Gino, Ayal, and Ariely (2009) successfully used different university affiliations even within the same city to induce different social identities.

identities play a role for lying behavior. Despite the varying social identity of the two victims of the lie, consistently across all four treatments roughly half of our participants decide to send a message containing false information in order to increase their own payoffs at the expense of the receiver and the dummy player. This overall share of lies is in line with earlier results from Gneezy (2005) and confirms that for a substantial fraction of participants the costs of lying are too high to exploit an informational asymmetry for their financial advantage. However, these lying costs do not seem to vary between in- and outgroup relationships, neither in terms of allocative nor of social costs. Additionally, also when identities are revealed to all players and senders' beliefs about others' expectations may be different for in- and outgroup interactions, we do not observe an effect of social identities on lying behavior. Overall, this suggests that individual lying behavior cannot be easily manipulated by the composition of social identity among the potential victims of the lie.[114]

Regarding allocative costs this result suggests that a subject's propensity to lie is a robust individual characteristic that cannot be easily manipulated by a variation of social identities. Consequently, in a lying context subjects do not discriminate between in- and outgroup members based on differences in social preferences as in e.g. Chen and Li (2009). In this respect, our results indicate that discrimination based on social groups does not extend to richer moral contexts such as deception and lying.[115] An alternative explanation for our results could be that those subjects who have different social preferences for in- and outgroup members are the same who have a higher aversion to lying. Cappelen et al. (2013) and Chakravarty et al. (2012) for example find that subjects who have higher social preferences in general are also less likely to lie. Following this argument, it could be interesting to investigate the relationship between ingroup favoritism and lying aversion on an individual level by combining a simple allocation game with a deception game.

Regarding the social costs of lying, based on evidence from social psychology we argued that lying is perceived as more unethical the closer individuals are to each other.

---

[114] In another project, we investigated whether individuals adjust their own lying behavior to the observed level of dis(honesty) of other group members. However, we found no evidence that lying is conditional on the lying behavior of other subjects. A brief summary of this study can be found in Appendix C.

[115] Similarly, we found in another experimental study that social identity also does not affect individuals' behavior in tournaments regarding constructive and destructive effort. A short summary of this project can be found in Appendix D.

A possible explanation why we do not find evidence for group-contingent social costs of lying might be that an important motivational factor to tell the truth in close relationships are not only moral reasons but also the fear of being caught as a liar. People suspect that close relationship partners are more likely to spot a lie immediately due to verbal and non-verbal cues and also discover the truth more often in the future because of more frequent interactions (DePaulo and Kashy 1998; Anderson et al. 1999). In addition, being caught as a liar might have more severe consequences for a close relationship as it violates important ideals of the relationship such as openness and trustworthiness. In our anonymous, one-shot interaction, however, senders do not have to worry about being caught lying and the corresponding consequences since neither the receiver nor the dummy ever get to know the actual outcome of the role of the dice, the sender's payoff or the alternative payment option. Overall, the mere feeling of closeness due to a shared social identity does not seem to affect individual lying behavior in our experimental setup.

Taken together, in contrast to other experimental settings regarding altruism, trust, reciprocity, or cooperation, the social relationship between individuals does not affect behavior in our deception game. Hence, more research is needed to identify factors that distinguish lying behavior in our setting from other allocation decisions as in Chen and Li (2009) and make deception robust to social identity manipulations.

5.6 *Appendix A – Supplementary tables and figures*

	T1 CC	T2 CD	T3 DC	T4 DD	Overall
Female	50%	44%	51%	49%	49%
Age	20.2	20.7	20.3	20.2	20.3
German	85%	83%	80%	84%	83%
Business	53%	67%	64%	62%	61%
Economics	42%	28%	29%	31%	33%
Other	5%	5%	7%	7%	6%
Term	1.3	1.3	1.2	1.2	1.2
Affiliation	6.0	6.1	6.0	6.2	6.1
N	140	131	140	134	545

Table 5.4: Demographics of participants.

All subjects	T1 CC	T2 CD	T3 DC
T1 CC	-	-	-
T2 CD	0.67	-	-
T3 DC	0.55	0.88	-
T4 DD	0.90	0.58	0.47

Table 5.5: *P* values of pairwise *chi*^{2} tests of shares of lies for all subjects.

High affiliation	T1 CC	T2 CD	T3 DC
T1 CC	-	-	-
T2 CD	0.46	-	-
T3 DC	0.66	0.76	-
T4 DD	0.51	0.93	0.83

Table 5.6: *P* values of pairwise *chi*^{2} tests of shares of lies for subjects with a high affiliation to the UoC.

From Cologne	T1 CC	T2 CD	T3 DC
T1 CC	-	-	-
T2 CD	0.49	-	-
T3 DC	0.44	0.94	-
T4 DD	0.89	0.43	0.38

Table 5.7: *P* values of pairwise *chi*^{2} tests of shares of lies for subjects living in Cologne.

Mother tongue German	T1 CC	T2 CD	T3 DC
T1 CC	-	-	-
T2 CD	0.80	-	-
T3 DC	0.83	0.64	-
T4 DD	0.90	0.89	0.74

Table 5.8: *P* values of pairwise *chi²* tests of shares of lies for subjects with German as first language.

Share of lies	T1 CC	T2 CD	T3 DC	T4 DD	Overall	N
All subjects	45%	43%	58%	42%	47%	341
High affiliation	48%	42%	58%	42%	47%	228
From Cologne	41%	46%	58%	38%	46%	263
Mother tongue German	47%	41%	58%	39%	46%	285

Table 5.9: Share of lies across treatments for different groups of subjects in second experiment.

Lie (yes(no)	All subjects	Affiliation > 5	Cologne	German
T2 CD	-0.013	-0.111	0.169	-0.107
	(0.204)	(0.248)	(0.235)	(0.225)
T3 DC	0.281	0.215	0.420	0.250
	(0.203)	(0.254)	(0.237)	(0.217)
T4 DD	-0.044	-0.136	-0.065	-0.187
	(0.199)	(0.243)	(0.229)	(0.218)
Macroeconomics	0.288	0.310	0.409*	0.419*
	(0.160)	(0.198)	(0.183)	(0.176)
Age	-0.062*	-0.072*	-0.066	-0.055
	(0.026)	(0.032)	(0.034)	(0.029)
Female	-0.102	-0.204	-0.116	-0.232
	(0.144)	(0.176)	(0.165)	(0.158)
Economics	0.091	0.150	0.136	-0.013
	(0.156)	(0.198)	(0.177)	(0.174)
Other	-0.336	-0.228	-0.141	-0.263
	(0.234)	(0.306)	(0.267)	(0.244)
Constant	0.844	1.129	0.633	0.662
	(0.647)	(0.790)	(0.801)	(0.703)
Log-likelihood	-223,62	-148,4	-171,23	-185,11
Observations	339	227	263	284

Table 5.10: Probit regression on lying for different groups of subjects in second experiment. Standard errors in parentheses; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

5.7 *APPENDIX B – INSTRUCTIONS*

Instructions for senders

Welcome and thank you for participating in this experiment. Please read the instructions carefully. You can earn money based on your decisions. We will randomly choose 100 participants and will pay them the payoff of this experiment. Which participants were chosen will be disclosed in the lecture on 20. November 2013. Following the lecture on 20 November 2013 these 100 participants will receive their payoff in cash.

In addition to the instructions for this experiment you find enclosed:

- **Payoff-ID:** Your personal identification number is printed on the ID. Based on this identification number we will randomly choose 100 participants. In case you are chosen, your **payoff can only be paid to you if you hold your identification number**. Thus, please, take your identification number with you immediately and do not forget to take it with you to the lecture next week. It is not sufficient to remember your identification number. Without your payoff-ID we cannot hand out your payoff.
- **Blue envelope:** Please do not open this envelope until asked to by the experimenter.
- **Red envelope:** Please do not open this envelope until asked to by the experimenter.

Please decide **on your own** and do **not** talk to your classmates. If you have any questions, please raise your hand, and an experimenter will come to you. If you do not follow these rules we will have to exclude you from this experiment and all payoffs.

Game instructions

If your identification number is chosen, you will be matched to two randomly chosen persons with whom you will interact. In the following parts of the instructions you will be referred to as Person A. The other two participants will be referred to as Person B and Person C. Neither you nor the other two persons will learn the identity of the others. Thus, your decisions in this experiment are completely anonymous. The only information you receive on participant B and C is the University they attend:

- **Participant B** is a student at the **University of Cologne**.
- **Participant C** is a student at the **University of Cologne**.

Participants B and C do **not** know that you are a student at the University of Cologne.

Prior to the experiment we threw a six-sided dice for every participant A.

Rolling the dice has resulted in "**1, 2, 3, 4, 5 or 6**" for you.

Participants B and C are **not** informed of the result of your dice roll. However, they will be told, that you as Participant A know the result. We would like to ask you to send a message to **participant B (Cologne)**. **Participant C (Cologne)** does **not** receive a message. You have the choice between six different messages:

- "The result of the dice roll was 1."
- "The result of the dice roll was 2."
- "The result of the dice roll was 3."
- "The result of the dice roll was 4."
- "The result of the dice roll was 5."
- "The result of the dice roll was 6."

Which message you send is your decision. The message will **only** be shown to **participant B (Cologne)**. We will then ask participant B to choose a number between 1 and 6. The number participant B chooses will determine your payoff as well as the payoff of participant B and C:

- If participant B chooses the number that has been thrown, you receive **7€**, participant B (Cologne) receives **7€** and participant C (Cologne) receives **7€**, as well.
- If participant B does not choose the number that has been thrown, you receive **11€**, participant B (Cologne) receives **5€** and participant C (Cologne) receives **5€**, as well.

The following table illustrates the payoffs for you, participant B and participant C depending on the number participant B chooses:

	Your payoff	**Payoff participant B** (University of Cologne)	**Payoff participant C** (University of Cologne)
Participant B chooses the **actual** number	7€	7€	7€
Participant B chooses a **different** number	11€	5€	5€

You are the only one who knows these possible payoffs. Participant B and participant C do not know these possible payoffs and, also do not know how the decision of participant B affects the payoffs. Both participants only know that you know all possible payoffs.

After participant B has chosen a number, participant B and participant C are informed about their payoffs. Participant B and participant C are informed neither about the actual dice roll nor about the payoff you receive. Also, they do not know the payoffs you would have received if participant B had chosen a different number.

Please now fill out the decision sheet, stating the message you want to send to participant B.

Instructions for receivers

Welcome and thank you for participating in this experiment. Please read the instructions thoroughly. You can earn money based on your decisions. Independent of this you will receive 2,50 € as a show-up-fee.

Two other participants have been randomly assigned to you. Through the following parts of the instructions you are referred to as participant B and the other two participants as participant A and C. Participant C does not make any decisions in this experiment. She/he receives a payoff depending on your decision. Neither you nor the other participants are informed about the identity of the other participants. Your decisions are completely anonymous in this experiment.

Prior to the experiment we rolled a six-sided dice for participant A. You will not be informed about the result; however, participant A knows the result. After informing participant A of the result, she/he has the possibility to send you a message. You are the only one who receives this message. Participant C does not receive any message. Participant A can send one of six different messages:

1. "The result of the dice roll was 1."
2. "The result of the dice roll was 2."
3. "The result of the dice roll was 3."
4. "The result of the dice roll was 4."
5. "The result of the dice roll was 5."
6. "The result of the dice roll was 6."

Participant A has sent the following message:

"The result of the tossing of the dice was "(1, 2, 3, 4, 5, or 6)"."

This message is the only information you receive regarding the role of the dice. Now, we would like to ask you to choose a number between 1 and 6. The number you choose determines your payoff as well as the payoffs of participant A and C:

- If you choose the actual number, you, participant A and participant C will be paid **according to option 1**.
- If you choose a different number, you, participant A and participant C will be paid **according to option 2**.

Only participant A knows the exact payoffs of option 1 and 2 for all participants.

Your choice:

Please enter your choice of number (1, 2, 3, 4, 5 or 6) here:

I choose number ____.

After deciding on a number, please hand the sheet to the experimenter. Next, you will receive a short questionnaire.

Instructions for dummy players

Welcome and thank you for participating in this experiment. Please read the instructions thoroughly. You can earn money in this experiment. How much depends on the decisions of the other players. Independent of this you receive 2,50 € as a show-up-fee.

For this experiment two other players are randomly assigned to you. Through the following parts of the instructions you will be referred to as participant C and the other two participants will be referred to as participant A and participant B. As participant C, you do not make any decisions in this experiment. Your payoff will be determined depending on the decision of participant B. Neither you nor the two other participants are informed about the identity of the other participants.

Prior to the experiment we rolled a six-sided dice for participant A. You will not be informed of the result; however, participant A knows the result. After informing participant A of the result, she/he has the possibility to send a message to participant B. Only participant B receives this message. You will not receive any message. Participant A had to choose between six different messages:

1. "The result of the dice roll was 1."
2. "The result of the dice roll was 2."
3. "The result of the dice roll was 3."
4. "The result of the dice roll was 4."
5. "The result of the dice roll was 5."
6. "The result of the dice roll was 6."

This message is the only information participant B has concerning the result of the dice roll. After participant B received the message, she/he chooses a number between 1 and 6. The number participant B chooses determines your payoff as well as the payoff of participants A and B.

If participant B chooses the actual number, you will be paid corresponding to option 1.

If participant B chooses a different number than the actual number, you will be paid corresponding to option 2.

Only participant A knows the exact payoffs of option 1 and 2 for all participants.

You will not make decision in this experiment. The decisions of participants A and B determine your payoff. We would lie to ask you to answer the following questionnaire.

5.8 *Appendix C – Summary conditional lying*

In this project, we experimentally investigate whether social information on others' behavior affects the costs of lying and thereby influences individuals' willingness to lie.[116] More specifically, we examine whether observing (dis)honest behavior of peers induces individuals to adapt their own lying behavior to the observed group norm. Previous studies have already shown that social norms and social information influence individuals' economic behavior. In the context of dictator giving, for example, showing dictators varying information about the share of fair versus selfish decisions made by previous participants influences their propensity to follow and to give a fair amount themselves (Bicchieri and Xiao 2009; Cason and Mui 1998; Krupka and Weber 2009). With regard to dishonesty this contagion effect among individuals has been investigated by Gino, Ayal, and Ariely (2009) and Innes and Mitra (2013). Gino and colleagues show that cheating in an experiment increases if a confederate leads by negative example. Mimicking Gneezy's seminal experimental design (Gneezy 2005), Innes and Mitra present individuals the behavior of several previous subjects sending either a very high or very low share of truthful messages. Their results generally suggest that also honesty and dishonesty can be contagious.

In this project we want to further elaborate on the effect of social information on lying behavior. Using a within-subject design inspired by Falk, Fischbacher, and Gächter (2010), we test whether subjects adjust their individual level of (dis)honesty over time to the observed lying behavior of other group members. In order to do so, we use an experimental design in which each subject has to make the same decision twice in two different groups to which they belong. Thereby, we are able to investigate whether within-subject behavioral differences in the two groups can be tracked back to differences in observed behavior of other group members. For example, a subject who is a member of a group where most other members send honest messages and at the same time also belongs to a second group where dishonest messaging is the prevalent norm

[116] This project was joint work with Roman Inderst and Axel Ockenfels. Financial support of the German Research Foundation (DFG) through the Research Unit "*Design & Behavior – Economic Engineering of Firms and Markets*" (FOR 1371) is gratefully acknowledged.

might adjust his/her individual lying behavior and send an honest message in the former and a dishonest message in the latter group.

The basic structure of our experiment is a standard information transmission game. Subjects are either in the role of the sender or receiver and remain in this role throughout all periods of the experiment. In total we have 16 senders and 16 receivers. Each sender belongs to two groups – "row group" and "column group" – with four members each. Figure 5.1 illustrates how groups are formed.

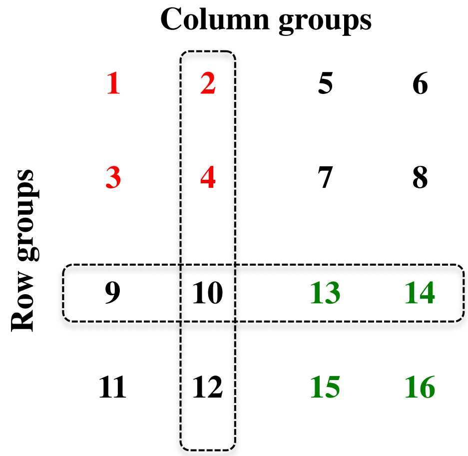


Figure 5.1: Matrix for the group formation of senders. Numbers represent different senders. Colors indicate different levels of individual dishonesty in the first period of the experiment. Red are the most dishonest senders, green are the most honest senders, black are in between.

For example, participant 10 forms a group with participants 2, 4, and 12 and another group with participants 9, 13, and 14. There were 8 different groups in total, 4 row groups and 4 column groups. In each combination of one row and one column group there is only one subject that belongs to both groups – the subject where row and column intersect. Group composition remains unchanged throughout the entire experiment.

At the beginning of each period, each of the 16 senders is randomly paired with two of the 16 receivers. The sender simultaneously plays identical but independent stage games with each of the receivers. For each sender-receiver pair nature draws a random state of the world $s \in \{7,8,9,10,11,12\}$. This state of the world is only revealed to the sender who then sends a message $m \in \{7,8,9,10,11,12\}$ to the paired receiver. The receiver has to guess the true state of the world by choosing his/her estimate $e \in \{7,8,9,10,11,12\}$, which – together with the true state s – determines his/her payoff.

If he guesses correctly ($e = s$) the payoff is equal to s (or e). If the estimate is lower than the true state ($e < s$) the payoff is equal to e. However, if the estimate exceeds the true state ($e > s$) the receiver gets $2s - e$. Hence, the receiver gets the highest payoff if he/she guesses correctly but is punished symmetrically for deviations from the true state. In contrast sender's payoff is always equal to the chosen action a and thus independent of the true state of the world. The following Table 5.11 illustrates the payoffs for senders and receivers for each state-action combination.

		Receiver's payoff						**Sender's payoff**
		State *s*						
		7	8	9	10	11	12	
Estimate *e*	7	7	7	7	7	7	7	7
	8	6	8	8	8	8	8	8
	9	5	7	9	9	9	9	9
	10	4	6	8	10	10	10	10
	11	3	5	7	9	11	11	11
	12	2	4	6	8	10	12	12

Table 5.11: Payoff matrix for senders and receivers conditional on state *s* and estimate *e*.

We use the strategy method to elicit senders' message decisions. Hence, before the true state of the world is revealed to the sender he/she chooses a message to be sent to the receiver for each possible state. The sender states his messages conditional on the true state separately for each of the two receivers. Thereby, it is possible for a sender to submit different message profiles for the two matched receivers. This double sender-receiver game is repeated for 15 periods and receivers do not get any feedback between periods.

To allow for an influence of social information, senders observe separately the messaging behavior of their peers in the row and column groups from the previous period. Therefore, the decision screen is divided vertically. The left hand side displays the average message for each possible state of the other three *row* group members while the right hand side displays the average message for each possible state of the other three *column* group members. On each side at the lower part of the screen, senders are then asked to submit their conditional messages: on the left hand side for the receiver in the row group and on the right hand side for the receiver in the column group. We expect that displaying the average messages of the other group members provides senders with information about the (different) prevalent norm within the two groups and thereby has an influence on their own messaging behavior. To ensure that some senders belong to two groups who show different levels of honesty, we covertly match senders into row and column groups based on a specific scheme. Depending on the behavior in the first period, all senders are ranked based on how much their conditional messages deviate from the true state. The four least honest senders (high liars) are placed on positions 1, 2, 3, and 4 in the group formation matrix (see Figure 5.1). Those senders whose messages are closest to the true state are placed on positions 13, 14, 15, and 16 (low liars). All other senders are placed on positions 5-12 (medium liars). With this matching procedure it is guaranteed that each medium liar belongs to a group with two high liars and to another group with two low liars and thus is confronted with competing norms. We expect that medium liars adjust their behavior to these differing norms and send more dishonest messages in the group with the two high liars and more honest messages in the group with the two low liars.

Average message	Medium liars		High liars	Low liars
	in groups with high liars	in groups with low liars		
State = 7	9.62	9.73	11.06	9.00
State = 8	10.00	10.00	11.11	9.52
State = 9	10.36	10.43	11.29	10.16
State = 10	10.86	10.77	11.41	10.78
State = 11	11.16	11.16	11.54	11.40
State = 12	11.41	11.30	11.66	11.82

Table 5.12: Average messages for each state of high, medium and low liars.

However, our results in Table 5.12 show that medium liars sent very similar messages on average in both groups. The largest difference is 0.11 and the average message in the group with high liars is not always larger than the average message in the group with low liars. Additionally, conditional on the true state, only in 15% of all cases messages from the same sender differ between groups. This high similarity between messages of medium liars across groups cannot be explained by low differences between high and low liars. Especially for low states, average messages of high liars are substantially larger than those of low liars (1.94 for state 7 and 1.59 for state 8). Regression results confirm that the observed average message of the other group members only has a small influence on individual messaging. In contrast, there is an almost perfect correlation between messages of the same sender for row and column receivers.

Thus, we conclude that in our experimental design social information has no effect on individuals' lying behavior because subjects do not adjust their level of honesty to the norm prevalent in the respective groups. Overall, we again observe that individuals' lying behavior is rather stable and cannot be easily manipulated.

5.9 *APPENDIX D – SUMMARY SOCIAL IDENTITY IN TOURNAMENTS*

In this project we investigate whether social identity considerations affect competitors' behavior in tournaments.[117] More specifically, we look at whether individuals adjust their level of productive and destructive effort provision (effort and sabotage) when they compete against an ingroup or an outgroup member. Group identity has been shown to affect individuals' social preferences leading to favoritism of ingroup members and discrimination of outgroup members (Chen and Li 2009; Filippin and Guala 2013; Goette et al. 2006, 2012). Extending a theoretical model by Harbring and Irlenbusch (2008, 2011) with group-related altruism and envy, we hypothesize that, in equilibrium, effort and sabotage decrease in tournaments within groups but increase when agents from different groups compete. However, with our experimental design we find no evidence that group identity affects productive or destructive effort. Also, the price spread has no influence on agents' effort or sabotage choices.

The model of Harbring and Irlenbusch (2008, 2011) analyzes a simple two-stage game with n agents and one principal. While in the first stage the principal chooses the price sum and spread, agents decide on their effort provision and level of sabotage in the second stage. As we were mainly interested in intra- and intergroup tournaments, we focus solely on the second stage and treat the price sum and spread as exogenous parameters. Additionally, to have clear ingroup or outgroup matches, we only consider the case with two agents i and j. Agents can exert two different activities: productive effort e and sabotage s. The first increases an agent i's own output whereas the latter decreases the output of the rival agent j. Both activities are costly for each agent i. For

[117] This project was joint work with Peter Werner. Financial support of the German Research Foundation (DFG) through the Research Unit "*Design & Behavior – Economic Engineering of Firms and Markets*" (FOR 1371) is gratefully acknowledged.

simplicity, Harbring and Irlenbusch (2011) assume symmetric and quadratic cost functions with $C_e(e_i) = e_i^2 / c_e$ and $C_s(s_i) = s_i^2 / c_s$, respectively. To incorporate production luck or measurement errors, a random component ε_i – uniformly distributed over the interval $\left[-\bar{\varepsilon}/2, +\bar{\varepsilon}/2\right]$ – also affects the output y_i of each agent: $y_i = e_i + \varepsilon_i - s_j$. In the tournament, both agents compete with their output y_i for a winning price M. The losing agent receives a smaller price of m with $0 < m < M$, and where Δ denotes the price spread $(M - m)$. Overall the expected payoff for agent *i* is given by

$$E\left[\pi(e_i, e_j, s_i, s_j)\right] = m + f^w(e_i, e_j, s_i, s_j)\Delta - \frac{e_i^2}{c_e} - \frac{s_i^2}{c_s} \qquad (1)$$

where $f^w(e_i, e_j, s_i, s_j)$ is agent *i*'s probability to win the tournament. Deriving the first-order conditions and solving for e_i and s_i yields the following symmetric equilibrium effort and sabotage choices of rational, risk-neutral, money-maximizing agents in the one-shot setting[118]:

$$e^* = \frac{c_e \Delta}{2\bar{\varepsilon}} \qquad \text{and} \qquad s^* = \frac{c_s \Delta}{2\bar{\varepsilon}}.$$

Based on these considerations, we introduce parameters for charity C $(0 < C < 1)$ towards ingroup members and envy E $(0 < E < 1)$ towards outgroup members in the following way

$$E\left[\pi(e_i, e_j, s_i, s_j)\right] = m + f^w(\cdot)\Delta - \frac{e_i^2}{c_e} - \frac{s_i^2}{c_s} - f^w(\cdot)CI\Delta - \left(1 - f^w(\cdot)\right)EO\Delta$$

$$= m - EO\Delta + f^w(\cdot)\Delta(1 - CI + EO) - \frac{e_i^2}{c_e} - \frac{s_i^2}{c_s} \qquad (2).$$

The two additional terms capture agent *i*'s additional charity concern in ingroup matches when he wins the tournament and envy in outgroup matches when he looses.[119]

---

[118] For a detailed exposition, see, e.g. (Harbring and Irlenbusch 2008, 2011; Orrison, Schotter, and Weigelt 2004).

[119] *I* and *O* are identity parameters indicating whether it is an intra- or intergroup tournament. *I* = 1 for ingroup matches, and 0 otherwise. *O* = 1 for outgroup matches, and 0 otherwise.

Winning the tournament against a member of the same group reduces the expected profit because agents also take the payoff of their peers more into consideration. At the same time, loosing a tournament against a member of another group causes disutility because subjects are more envious towards outgroup members (Chen and Li 2009). The equilibrium effort and sabotage choices in an intragroup tournament then are

$$e_I^* = \frac{c_e\Delta(1-C)}{2\overline{\varepsilon}} \qquad \text{and} \qquad s_I^* = \frac{c_s\Delta(1-C)}{2\overline{\varepsilon}} \qquad (3).$$

Thus, higher charity concerns for ingroup members decrease the marginal utility from productive and destructive effort. At the same time, the opposite is true for outgroup matches, where envy increases the marginal utility of both activities

$$e_O^* = \frac{c_e\Delta(1+E)}{2\overline{\varepsilon}} \qquad \text{and} \qquad s_O^* = \frac{c_s\Delta(1+E)}{2\overline{\varepsilon}} \qquad (4).$$

Comparing the equilibrium choices for effort and sabotage in ingroup, outgroup and baseline case without group affiliation we obtain

$$e_I^* < e^* < e_O^* \qquad \text{and} \qquad s_I^* < s^* < s_O^* .$$

Thus, effort and sabotage should be lowest in ingroup matches and highest in outgroup matches. At the same time, according to (3) and (4), the price spread Δ interacts with charity concerns and envy and therefore should reinforce their effects on effort and sabotage.

In order to test these effects, we design an experiment where we, at first, induce an artificial group identity among agents who then compete in two-person tournaments for 20 periods. Between periods, we vary the group composition (ingroup or outgroup match) as well as the price spread (high or low) leading to a 2x2 within-subject design. At the beginning of the experiment roles are randomly assigned to the participants. Each matching group consists of 12 participants with 8 of them being in the role of agents and 4 of them being in the role of principals. In contrast to the study of Harbring and Irlenbusch (2011), principals do not have to take any decision during the experiment as the price sum is fixed and the price spread is varied exogenously. They are compensated based on the outputs generated by their assigned agents. A further difference to Harbring and Irlenbusch (2011) is that we have a random instead of a fixed matching of agents and principals.

We use a simple estimation task to divide agents into two different groups (Böhm, Rockenbach, and Weiss 2013). On several sequential screens a random number of 'X' are shown only for a second. Each agent then has to guess the number of 'X'. Based on the median estimation, agents are divided into a group of four low estimators (called 'yellow group') and into a group of four high estimators (called 'blue group'). Following, in order to reinforce group identity, these two groups of agents compete in a short real-effort task.[120] Each agent contributes with his effort to the output of his group and the group with the highest overall output receives an additional payment of 2.5 € at the end of the experiment. Which team performed better in this real effort task is only revealed at the end of the experiment. After the real effort task, agents compete in tournaments. Therefore, in each period of the tournament stage, two agents are matched with one principal. Parameters are chosen similarly to Harbring and Irlenbusch (2011) with costs $c_e = 70$ and $c_s = 20$ and the random factor chosen from the interval $\varepsilon_i \in [-60;+60]$. The price spread is either low $\Delta_{Low} = 96$ (68 vs. 164) or high $\Delta_{High} = 144$ (44 vs. 188). Given these parameters, equilibrium effort and sabotage choices without group identity are $e^*_{Low} = 28$ and $s^*_{Low} = 8$ for the low price spread and $e^*_{High} = 42$ and $s^*_{High} = 12$ in case of a high price spread. Treatment variations over all periods are balanced so that each combination of our 2x2 design is played five times. At the beginning of each period, agents are informed about the price spread and whether they compete with an ingroup or outgroup member but not with which group member exactly. To make group identity more salient, we use the colors of the agent's own and the other agent's group on the screen where he makes his effort and sabotage decisions. For both activities, agents can only choose integer values: $e_i \in [0,...,100]$ and $s_i \in [0,...,50]$. Overall, we run two sessions with two matching groups each, leading to 640 agent decisions in total.[121]

[120] In the real effort task, subjects received a text printed on several pages and then were asked to identify several letters, e.g. the fifth letter of the second word on page three. The group with more correctly identified letters won the group competition.

[121] Sessions took place at the Cologne Laboratory for Economic Research (CLER). Participants were recruited from the CLER's subject pool using ORSEE (Greiner 2004) and the computerized experiment was coded in z-Tree (Fischbacher 2007).

	Effort		Sabotage	
Low spread & outgroup	47.11	(24.44)	14.30	(9.22)
High spread & outgroup	47.78	(23.54)	15.10	(9.45)
Low spread & ingroup	47.09	(21.56)	15.30	(9.35)
High spread & ingroup	47.66	(22.84)	15.61	(9.61)

Table 5.13: Effort and sabotage choices across treatments, standard deviations in parentheses.

Our descriptive results in Table 5.13, however, show that there were no large differences between treatment variations neither with regard to effort nor sabotage levels. Tobit regressions in Table 5.14 confirm these findings as none of our treatment variables has a significant influence on effort or sabotage. Also for low and high wage spreads we observe no differences for both activities.

	Effort I	Effort II	Sabotage I	Sabotage II
Ingroup	-0.071	-0.093	0.743	1.029
	(-0.051)	(-0.048)	(1.222)	(1.196)
High spread	2.290	2.268	0.846	1.129
	(1.553)	(1.125)	(1.304)	(1.274)
Ingroup x high spread		0.045		-0.571
		(0.016)		(-0.470)
Period	-0.422***	-0.422***	-0.061	-0.060
	(-3.301)	(-3.301)	(-1.076)	(-1.067)
N	640	640	640	640
Log likelihood	-2706.4	-2706.4	-2156.9	-2156.8

Table 5.14: Random effect Tobit regressions with effort (lower limit 0; upper limit 100) and sabotage (lower limit 0; upper limit 50) as dependent variable. *t* statistics in parentheses; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Overall, we suspect that due to the complexity of the tournament task agents did not notice – and thus also did not react – to the social identity variations. In addition, the history of play due to the repeated interactions may also bring about that – in contrast to other studies – social identity does not influence behavior.

References

Akerlof, George A. 1970. “The Market for Lemons: Quality Uncertainty and the Market Mechanism.” *Quarterly Journal of Economics* 84 (3): 488–500.

Akerlof, George A., and Rachel E. Kranton. 2000. “Economics and Identity.” *Quarterly Journal of Economics* 115 (3): 715–753.

Albinsson, Pia A, and B Yasanthi Perera. 2012. “Alternative Marketplaces in the 21st Century: Building Community through Sharing Events.” *Journal of Consumer Behavior* 11: 303–315.

Anderson, By Lisa R, and Charles A Holt. 1997. “Information Cascades in the Laboratory.” *The American Economic Review* 87 (5): 847–862.

Anderson, D. Eric, Matthew E. Ansfield, and Bella M. DePaulo. 1999. “Love’s Best Habit: Deception in the Context of Relationships.” In *The Social Context of Nonverbal Behvaior*, edited by Pierre Philippot, Robert S. Feldman, and Erik J. Coats, 372–409. Cambridge: Cambridge University Press.

Anderson, E. W. 1998. “Customer Satisfaction and Word of Mouth.” *Journal of Service Research* 1 (1) (August 1): 5–17.

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.

Arrow, Kenneth J. 1972. “Gifts and Exchanges.” *Philosophy & Public Affairs* 1 (4): 343–362.

———. 1974. *The Limits of Organization*. New York: W.W. Norton.

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.

Banerjee, Abhijit V. 1992. “A Simple Model of Herd Behavior.” *The Quarterly Journal of Economics* 107 (3): 797–817.

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

Battigalli, Pierpaolo, Gary Charness, and Martin Dufwenberg. 2013. "Deception: The Role of Guilt." *Journal of Economic Behavior & Organization* 93 (September): 227–232.

Battigalli, Pierpaolo, and Martin Dufwenberg. 2007. "Guilt in Games." *American Economic Review* 97 (2): 170–176.

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

Becker, Gary S. 1968. "Crime and Punishment: An Economic Approach." *The Journal of Political Economy* 76 (2): 169–217.

Becker, Gordon M., Morris H. DeGroot, and Jacob Marschak. 1964. "Measuring Utility by a Single-Response Sequential Method." *Behavioral Science* 9: 226–232.

Beenen, Gerard, Kimberly Ling, Xiaoqing Wang, Klarissa Chang, Dan Frankowski, Paul Resnick, and Robert E. Kraut. 2004. "Using Social Psychology to Motivate Contributions to Online Communities." In *Proceedings of 2004 ACM Conference on Computer Supported Cooperative Work*, 212–221. New York.

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.

Berger, Johannes, Christine Harbring, and Dirk Sliwka. 2013. "Performance Appraisals and the Impact of Forced Distribution: An Experimental Investigation." *Management Science* 59 (1): 54–68.

Bernhard, Helen, Urs Fischbacher, and Ernst Fehr. 2006a. "Group Affiliation and Altruistic Norm Enforcement." *American Economic Review* 96 (2): 217–221.

———. 2006b. "Parochial Altruism in Humans." *Nature* 442 (August 24): 912–915.

Bicchieri, Cristina, and Erte Xiao. 2009. "Do the Right Thing: But Only If Others Do So." *Journal of Behavioral Decision Making* 22: 191–208.

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

Bikhchandani, Sushil, David Hirshleifer, and Ivo Welch. 1992. "A Theory of Fads, Fashion, Custom, and Cultural Change as Informational Cascades." *Journal of Political Economy* 100 (5): 992–1026.

———. 1998. “Learning from the Behavior of Others: Conformity, Fads, and Informational Cascades.” *The Journal of Economic Perspectives* 12 (3): 151–170.

Böhm, Robert, Bettina Rockenbach, and Arne Weiss. 2013. “Experimental Evidence on Identity Biases in Voting Choices.” Working Paper.

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.

Bolton, Gary E, David J Kusterer, and Johannes Mans. 2014. “Inflated Reputations - Leniency and Moral Wiggle Room in Trader Feedback Systems.” Working Paper.

Bolton, Gary E., Elena Katok, and Axel Ockenfels. 2004. “How Effective Are Electronic Reputation Mechanisms? An Experimental Investigation.” *Management Science* 50 (11) (November): 1587–1602.

Bornstein, Gary, Uri Gneezy, and Rosmarie Nagel. 2002. “The Effect of Intergroup Competition on Group Coordination: An Experimental Study.” *Games and Economic Behavior* 41 (1) (October): 1–25.

Botsman, Rachel, and Roo Rogers. 2010. *What’s Mine Is Yours - The Rise of Collaborative Consumption*. New York, NY: Harper Business.

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. 2012. “Reputation on the Internet.” In *Oxford Handbook of the Digital Economy*, edited by Martin Peitz and Joel Waldfogel. Oxford University Press.

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

Cabral, Luis, and Lingfang (Ivy) Li. 2012. “A Dollar for Your Thoughts: Feedback-Conditional Rebates on eBay.”

Cai, Hongbin, and Joseph Tao-Yi Wang. 2006. “Overcommunication in Strategic Information Transmission Games.” *Games and Economic Behavior* 56 (1): 7–36.

Cannon, Sarah, and Lawrence H Summers. 2014. “How Uber and the Sharing Economy Can Win Over Regulators.” *Harvard Business Review*.

Cappelen, Alexander W., Erik Ø. Sørensen, and Bertil Tungodden. 2013. “When Do We Lie?” *Journal of Economic Behavior & Organization* 93 (September): 258–265.

Cason, Timothy N., and Vai-Lam Mui. 1998. “Social Influence in the Sequential Dictator Game.” *Journal of Mathematical Psychology* 42 (2-3) (June): 248–265.

Celen, Bogachan, and Shachar Kariv. 2004. “Distinguishing Informational Cascades from Herd Behavior in the Laboratory.” *American Economic Review* 94 (3): 484–498.

Chakravarty, Sugato, Yongjin Ma, and Sandra Maximiano. 2012. “Lying and Friendship.” Working Paper.

Charness, Gary, Luca Rigotti, and Aldo Rustichini. 2007. “Individual Behavior and Group Membership.” *American Economic Review* 97 (4): 1340–1352.

Chen, Roy, and Yan Chen. 2011. “The Potential of Social Identity for Equilibrium Selection.” *American Economic Review* 101 (6): 2562–2589.

Chen, Yan, F Maxwell Harper, Joseph Konstan, and Sherry Xin Li. 2010. “Social Comparisons and Contributions to Online Communities: A Field Experiment on MovieLens.” *American Economic Review* 100 (4): 1358–1398.

Chen, Yan, and Sherry Xin Li. 2009. “Group Identity and Social Preferences.” *The American Economic Review* 99 (1): 431–457.

Chevalier, Judith A, and Dina Mayzlin. 2006. “The Effect of Word of Mouth on Sales: Online Book Reviews.” *Journal of Marketing Research* 43 (3): 345–354.

Childs, Jason. 2012. “Gender Differences in Lying.” *Economics Letters* 114 (2): 147–149.

Coleman, James S. 1990. *Foundations of Social Theory*. Cambridge, Massachusetts: Belknap Press.

Conrads, Julian, Bernd Irlenbusch, Rainer Michael Rilke, Anne Schielke, and Gari Walkowitz. 2014. “Honesty in Tournaments.” *Economics Letters* 123 (1) (April): 90–93.

Conrads, Julian, Bernd Irlenbusch, Rainer Michael Rilke, and Gari Walkowitz. 2013. “Lying and Team Incentives.” *Journal of Economic Psychology* 34: 1–7.

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.

———. 2005a. “Reputation Mechanisms.” Working Paper.

———. 2005b. “Reputation Mechanism Design in Online Trading Environments with Pure Moral Hazard.” *Information Systems Research* 16 (2) (June): 209–230.

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.

DePaulo, Bella M., and Deborah A. Kashy. 1998. “Everyday Lies in Close and Casual Relationships.” *Journal of Personality and Social Psychology* 74 (1): 63–79.

DePaulo, Bella M., Deborah A. Kashy, Susan E. Kirkendol, Melissa M. Wyer, and Jennifer A. Epstein. 1996. “Lying in Everyday Life.” *Journal of Personality and Social Psychology* 70 (5) (May): 979–95.

Dholakia, Utpal M., Richard P. Bagozzi, and Lisa Klein Pearo. 2004. “A Social Influence Model of Consumer Participation in Network- and Small-Group-Based Virtual Communities.” *International Journal of Research in Marketing* 21 (3) (September): 241–263.

Dickhaut, John W., Kevin A. McCabe, and Arijit Mukherji. 1995. “An Experimental Study of Strategic Information Transmission.” *Economic Theory* 6 (3): 389–403.

Dreber, Anna, and Magnus Johannesson. 2008. “Gender Differences in Deception.” *Economics Letters* 99 (1): 197–199.

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).

eBay Enterprise. 2014. “Cross-Border Ecommerce: Catching the Trade Winds.” *Last Accessed: 19.08.2014*. http://www.ebayenterprise.com/commersations/blog/cross-border_ecommerce_catching_the_trade_winds.

Eckel, Catherine C., and Philip J. Grossman. 2005. “Managing Diversity by Creating Team Identity.” *Journal of Economic Behavior & Organization* 58 (3) (November): 371–392.

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.

EMOTA. 2014. “E-Commerce in Europe.” *Last Accessed: 19.08.2014*. http://media.wix.com/ugd/b18286_390bb25f5c1340fbbc9df4945b56ad16.pdf.

Ennis, E., A. Vrij, and C. Chance. 2008. “Individual Differences and Lying in Everyday Life.” *Journal of Social and Personal Relationships* 25 (1) (February 1): 105–118.

Erat, Sanjiv. 2013. “Avoiding Lying: The Case of Delegated Deception.” *Journal of Economic Behavior & Organization* 93 (September): 273–278.

Erat, Sanjiv, and Uri Gneezy. 2012. “White Lies.” *Management Science* 58 (4): 723–733.

European Commission. 2013. “Digital Agenda Scoreboard 2013.”

Falk, Armin, Urs Fischbacher, and Simon Gächter. 2010. “Living in Two Neighborhoods - Social Interaction Effects in the Laboratory.” *Economic Inquiry* 51 (1) (January 8): 563–578.

Falk, Armin, and James J Heckman. 2009. “Lab Experiments Are a Major Source of Knowledge in the Social Sciences.” *Science* 326 (October): 535–538.

Fehr, Ernst. 2009. “On the Economics and Biology of Trust.” *Journal of the European Economic Association* 7 (May): 235–266.

Fehr, Ernst, and Simon Gächter. 2002. “Altruistic Punishment in Humans.” *Nature* 415 (10): 137–140.

Feldhaus, Christoph, and Johannes Mans. 2014. “Who Do You Lie to? Social Identity and the Costs of Lying.” Working Paper.

Fershtman, Chaim, and Uri Gneezy. 2001. “Discrimination in a Segmented Society: An Experimental Approach.” *The Quarterly Journal of Economics* 116 (1): 351–377.

Filippin, Antonio, and Francesco Guala. 2013. “Costless Discrimination and Unequal Achievements in an Experimental Tournament.” *Experimental Economics* 16: 285–305.

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

Fischbacher, Urs, and Franziska Föllmi-Heusi. 2013. “Lies in Disguise - An Experimental Study on Cheating.” *Journal of the European Economic Association* 11 (3): 525–547.

Fischbacher, Urs, Simeon Schudy, and Sabrina Teyssier. 2014. “Heterogeneous Reactions to Heterogeneity in Returns from Public Goods.” *Social Choice and Welfare* 43 (1) (August 15): 195–217.

Fisher, Joseph, R. Mark Isaac, Jeffrey W. Schatzberg, and James M. Walker. 1995. “Heterogenous Demand for Public Goods: Behavior in the Voluntary Contributions Mechanism.” *Public Choice* 85 (3-4) (December): 249–266.

Fradkin, Andrey, Elena Grewal, David Holtz, and Matthew Pearson. 2014. “Bias and Reciprocity in Online Reviews: Evidence From Field Experiments on AirBnb.”

Freund, Caroline L, and Diana Weinhold. 2002. “The Internet and International Trade in Services.” *American Economic Review* 92 (2): 236–240.

———. 2004. “The Effect of the Internet on International Trade.” *Journal of International Economics* 62 (1) (January): 171–189.

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.

Geron, Tomio. 2013. “Airbnb And The Unstoppable Rise Of The Share Economy.” *Forbes Magazine*.

Gibson, Rajna, Carmen Tanner, and Alexander F. Wagner. 2013. “Preferences for Truthfulness: Heterogeneity among and within Individuals.” *American Economic Review* 103 (1): 532–548.

Gino, Francesca, Shahar Ayal, and Dan Ariely. 2009. “Contagion and Differentiation in Unethical Behavior: The Effect of One Bad Apple on the Barrel.” *Psychological Science* 20 (3) (March): 393–398.

———. 2013. “Self-Serving Altruism? The Lure of Unethical Actions That Benefit Others.” *Journal of Economic Behavior & Organization* 93 (September 1): 285–292.

Gneezy, Uri. 2005. “Deception: The Role of Consequences.” *The American Economic Review* 95 (1): 384–394.

Gneezy, Uri, Bettina Rockenbach, and Marta Serra-Garcia. 2013. “Measuring Lying Aversion.” *Journal of Economic Behavior & Organization* 93 (September): 293–300.

Godes, David, and Dina Mayzlin. 2004. “Using Online Conversations to Study Word-of-Mouth Communication.” *Marketing Science* 23 (4) (November): 545–560.

Goette, Lorenz, David Huffman, and Stephan Meier. 2006. “The Impact of Group Membership on Cooperation and Norm Enforcement: Evidence Using Random Assignment to Real Social Groups.” *The American Economic Review* 96 (2): 212–216.

Goette, Lorenz, David Huffman, Stephan Meier, and Matthias Sutter. 2012. “Competition Between Organizational Groups: Its Impact on Altruistic and Antisocial Motivations.” *Management Science* 58 (5): 948–960.

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.

Griffin, Dale, and Amos Tversky. 1992. “The Weighing of Evidence and the Determinants of Confidence.” *Cognitive Psychology* 24: 411–435.

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

Guala, Francesco, Luigi Mittone, and Matteo Ploner. 2013. “Group Membership, Team Preferences, and Expectations.” *Journal of Economic Behavior & Organization* 86 (February): 183–190.

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.

Güth, Werner, Matteo Ploner, and Tobias Regner. 2009. “Determinants of in-Group Bias: Is Group Affiliation Mediated by Guilt-Aversion?” *Journal of Economic Psychology* 30 (5) (October): 814–827.

Haile, Daniel, Abdolkarim Sadrieh, and Harrie a a Verbon. 2008. “Cross-Racial Envy and Underinvestment in South African Partnerships.” *Cambridge Journal of Economics* 32 (2001): 703–724.

Harbring, Christine, and Bernd Irlenbusch. 2008. “How Many Winners Are Good to Have?: On Tournaments with Sabotage.” *Journal of Economic Behavior & Organization* 65 (3--4): 682–702.

———. 2011. “Sabotage in Tournaments: Evidence from a Laboratory Experiment.” *Management Science* 57 (4): 611–627.

Heap, Shaun P Hargreaves, and Daniel John Zizzo. 2009. “The Value of Groups.” *The American Economic Review* 99 (1): 295–323.

Hendriks, Achim. 2012. “SoPHIE - Software Platform for Human Interaction Experiments.” *Working Paper*. Working Paper.

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.

Hortaçsu, Ali, F Asís Martínez-Jerez, and Jason Douglas. 2009. "The Geography of Trade in Online Transactions: Evidence from eBay and MercadoLibre." *American Economic Journal: Microeconomics* 1 (1): 53–74.

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.

Huck, Steffen, and Jörg Oechssler. 2000. "Informational Cascades in the Laboratory: Do They Occur for the Right Reasons?" *Journal of Economic Psychology* 21 (6): 661–671.

Hurkens, Sjaak, and Navin Kartik. 2009. "Would I Lie to You? On Social Preferences and Lying Aversion." *Experimental Economics* 12 (2): 180–192.

Innes, Robert, and Arnab Mitra. 2013. "Is Dishonesty Contagious?" *Economic Inquiry* 51 (1) (January 31): 722–734. doi:10.1111/j.1465-7295.2012.00470.x. http://doi.wiley.com/10.1111/j.1465-7295.2012.00470.x.

Isaac, R. Mark, Kenneth F. McCue, and Charles R. Plott. 1985. "Public Goods Provision in an Experimental Environment." *Journal of Public Economics* 26: 51–74.

Isaac, R. Mark, and James M. Walker. 1988. "Group Size Effects in Public Goods Provision: The Voluntary Contributions Mechanism." *The Quarterly Journal of Economics* 103 (1): 179–199.

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.

Jian, Lian, Jeffrey K Mackie-Mason, and Paul Resnick. 2010. "I Scratched Yours : The Prevalence of Reciprocation in Feedback Provision on eBay." *The B . E . Journal of Economic Analysis & Policy* 10 (1): 1–38.

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.

Jones, Thomas M. 1991. “Ethical Decision Making by Individuals in Organizations: An Issue-Contingent Model.” *The Academy of Management Review* 16 (2): 366–395.

Kahneman, Daniel F, Paul Slovic, and Amos Tversky. 1982. *Judgement under Uncertainty: Heuristics and Biases*. Cambridge: Cambridge University Press.

Kartik, Navin. 2009. “Strategic Communication with Lying Costs.” *Review of Economic Studies* 76 (4): 1359–1395.

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. 2006. “Last Minute Feedback.” Working Paper.

———. 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.

Klein, Tobias J., Christian Lambertz, and Konrad O. Stahl. 2013. “Market Transparency, Adverse Selection, and Moral Hazard.”

Kreps, David M, Paul Milgrom, John Roberts, and Robert Wilson. 1982. “Rational Cooperation in the Finitely Repeated Prisoners’ Dilemma.” *Journal of Economic Theory* 27: 245–252. doi:10.1016/0022-0531(82)90029-1.

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

Krupka, Erin, and Roberto a. Weber. 2009. “The Focusing and Informational Effects of Norms on pro-Social Behavior.” *Journal of Economic Psychology* 30 (3) (June): 307–320.

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.

Lendle, Andreas, Marcelo Olarreaga, Simon Schropp, and Pierre-Louis Vézina. 2012. "There Goes Gravity: How eBay Reduces Trade Costs." Working Paper.

Li, Lingfang (Ivy), and Erte Xiao. 2014. "Money Talks: Rebate Mechanisms in Reputation System Design." *Management Science* 60 (8) (August): 2054–2072.

Li, Sherry Xin, Kutsal Dogan, and Ernan Haruvy. 2011. "Group Identity in Markets." *International Journal of Industrial Organization* 29 (1) (January): 104–115.

Li, Xinxin, and Lorin M. Hitt. 2008. "Self-Selection and Information Role of Online Product Reviews." *Information Systems Research* 19 (4) (December): 456–474.

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

López-Pérez, Raúl, and Eli Spiegelman. 2012. "Why Do People Tell the Truth? Experimental Evidence for Pure Lie Aversion." *Experimental Economics*: 1–15.

Luca, Michael. 2011. "Reviews, Reputation, and Revenue: The Case of Yelp.com." Working Paper.

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.

Ludford, Pamela J., Dan Cosley, Dan Frankowski, and Loren Terveen. 2004. "Think Different: Increasing Online Community Participation Using Uniqueness and Group Dissimilarity." In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 6:631–638. New York.

Lundquist, Tobias, Tore Ellingsen, Erik Gribbe, and Magnus Johannesson. 2009. "The Aversion to Lying." *Journal of Economic Behavior & Organization* 70 (1-2): 81–92.

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.

Marques, Jose M, Dominic Abrams, and Rui G Serôdio. 2001. "Being Better by Being Right: Subjective Group Dynamics and Derogation of in-Group Deviants When Generic Norms Are Undermined." *Journal of Personality and Social Psychology* 81 (3): 436–447.

Marques, Jose M, and Vincent Y Yzerbyt. 1988. "Short Note The Black Sheep Effect : Judgmental Extremity towards Ingroup Members in Inter- and Intra-Group Situations." *European Journal of Social Psychology* 18: 287–292.

Marques, Jose M, Vincent Y Yzerbyt, and Jacques-Philippe Leyens. 1988. "The 'Black Sheep Effect': Extremity of Judgments towards Ingroup Members as a Function of Group Identification." *European Journal of Social Psychology* 18: 1–16.

Masclet, David, and Thierry Pénard. 2012. "Do Reputation Feedback Systems Really Improve Trust among Anonymous Traders? An Experimental Study." *Applied Economics* 44 (35) (December): 4553–4573.

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

McKnight, D. Harrison, and Norman L Chervany. 2001. "Trust and Distrust Definitions: One Bite at a Time." In *Trust in Cyber-Societies*, edited by Rino Falcone, Munindar Singh, and Yao-Hua Tan, 27–54. Berlin: Springer-Verlag.

McLeish, Kendra N., and Robert J. Oxoby. 2007. "Identity, Cooperation and Punishment." Working Paper.

Mead, Nicole L., Roy F. Baumeister, Francesca Gino, Maurice E. Schweitzer, and Dan Ariely. 2009. "Too Tired to Tell the Truth: Self-Control Resource Depletion and Dishonesty." *Journal of Experimental Social Psychology* 45 (3) (January): 594–597.

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.

Milgrom, Paul R., and John Roberts. 1982. "Predation, Reputation, and Entry Deterrence." *Journal of Economic Theory* 27 (2): 280–312.

Mill, John Stuart. 1848. *Principles of Political Economy with Some of Their Applications to Social Philosophy*. London: Longmans, Green and Co.

Miller, Nolan, Paul Resnick, and Richard Zeckhauser. 2005. "Eliciting Informative Feedback: The Peer-Prediction Method." *Management Science* 51 (9): 1359–1373.

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

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

Mussweiler, Thomas, and Axel Ockenfels. 2013. "Similarity Increases Altruistic Punishment in Humans." *Proceedings of the National Academy of Sciences of the United States of America* 110 (48) (November 26): 19318–23.

Nöth, Markus, and Martin Weber. 2003. "Information Aggregation with Random Ordering: Cascades and Overconfidence." *The Economic Journal* 113: 166–189.

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

Ockenfels, Axel, and Peter Werner. 2012. "'Hiding behind a Small Cake' in a Newspaper Dictator Game." *Journal of Economic Behavior & Organization* 82 (1): 82–85.

———. 2014. "Beliefs and Ingroup Favoritism." *Journal of Economic Behavior & Organization* (December): 1–10.

Paypal. 2013. "Modern Spice Routes: The Cultural Impact and Economic Opportunity of Cross-Border Shopping." *Last Accessed: 19.08.2014*. https://www.paypal-media.com/assets/pdf/fact_sheet/PayPal_ModernSpiceRoutes_Report_Final.pdf.

Peeters, Ronald, Marc Vorsatz, and Markus Walzl. 2015. "Beliefs and Truth-Telling: A Laboratory Experiment." *Journal of Economic Behavior & Organization* 113: 1–12.

Pinto, Isabel R, José M Marques, John M Levine, and Dominic Abrams. 2010. "Membership Status and Subjective Group Dynamics: Who Triggers the Black Sheep Effect?" *Journal of Personality and Social Psychology* 99 (1) (July): 107–19.

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.

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.

Reuben, E., and A. Riedl. 2009. "Public Goods Provision and Sanctioning in Privileged Groups." *Journal of Conflict Resolution* 53 (1) (November 12): 72–93.

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

Ross, William T., and Diana C. Robertson. 2000. "Lying: The Impact of Decision Context." *Business Ethics Quarterly* 10 (2): 409–440.

Rousseau, Denise M., Sim B. Sitkin, Ronald S. Burt, and Colin Camerer. 1998. "Not so Different after All: A Cross-Discipline View of Trust." *Academy of Management Review* 23 (3): 393–404.

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.

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

Sánchez-Pagés, Santiago, and Marc Vorsatz. 2007. "An Experimental Study of Truth-Telling in a Sender–receiver Game." *Games and Economic Behavior* 61 (1) (October): 86–112.

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.

Shinada, Mizuho, Toshio Yamagishi, and Yu Ohmura. 2004. "False Friends Are Worse than Bitter Enemies." *Evolution and Human Behavior* 25 (6) (November): 379–393.

Spiwoks, Markus, Kilian Bizer, and Oliver Hein. 2008. "Informational Cascades: A Mirage?" *Journal of Economic Behavior & Organization* 67 (1): 193–199.

Sutter, Matthias. 2009. "Deception Through Telling the Truth?! Experimental Evidence From Individuals and Teams." *Economic Journal* 119 (534): 47–60.

Tajfel, Henri, M. G. Billig, R. P. Bundy, and C. Flament. 1971. "Social Categorization and Intergroup Behavior." *European Journal of Social Psychology* 1 (2): 149–178.

Tajfel, Henri, and John Turner. 1979. "An Integrative Theory of Intergroup Conflict." In *The Social Psychology of Intergroup Relations*, edited by Stephen Worchel and William Austin, 33–47. Monterey, CA: Brooks/Cole.

Tan, Fangfang. 2008. "Punishment in a Linear Public Good Game with Productivity Heterogeneity." *De Economist* 156 (3) (September 23): 269–293.

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

Trevino, Linda Klebe. 1986. "Ethical Decision Making in Organizations: A Person-Situation Interactionist Model." *The Academy of Management Review* 11 (3) (July): 601–617.

Tsutsui, Kei, and Daniel John Zizzo. 2014. "Group Status, Minorities and Trust." *Experimental Economics* 17 (2) (May 11): 215–244.

Tyler, James M., and Robert S. Feldman. 2004. “Truth, Lies, and Self-Presentation: How Gender and Anticipated Future Interaction Relate to Deceptive Behavior.” *Journal of Applied Social Psychology* 34 (12): 2602–2615.

Tyler, James M., Robert S. Feldman, and Andreas Reichert. 2006. “The Price of Deceptive Behavior: Disliking and Lying to People Who Lie to Us.” *Journal of Experimental Social Psychology* 42 (1): 69–77.

Vanberg, Christoph. 2008. “Why Do People Keep Their Promises? An Experimental Test of Two Explanations.” *Econometrica* 76 (6): 1467–1480.

Wang, Zhongmin. 2010. “Anonymity, Social Image, and the Competition for Volunteers: A Case Study of the Online Market for Reviews.” *The B.E. Journal of Economic Analysis & Policy* 10 (1) (January 21): 1–34.

Welch, Ivo. 1992. “Sequential Sales, Learning, and Cascades.” *The Journal of Finance* 47 (2): 695–732.

Williamson, Oliver E. 1993. “Calculativeness, Trust, and Economic Organization.” *Journal of Law and Economics* 36: 453–486.

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

Wiltermuth, Scott S. 2011. “Cheating More When the Spoils Are Split.” *Organizational Behavior and Human Decision Processes* 115 (2) (July): 157–168.

Wolf, James R., and Waleed a. Muhanna. 2011. “Feedback Mechanisms, Judgment Bias, and Trust Formation in Online Auctions*.” *Decision Sciences* 42 (1): 43–68.

Xiao, Erte, and Daniel Houser. 2005. “Emotion Expression in Human Punishment Behavior.” *Proceedings of the National Academy of Sciences of the United States of America* 102 (20): 7398–7401.

Yamagishi, T., and N. Mifune. 2008. “Does Shared Group Membership Promote Altruism?: Fear, Greed, and Reputation.” *Rationality and Society* 20 (1) (February 1): 5–30.

Ye, Shun, Guodong Gao, and Siva Viswanathan. 2010. “The Good, the Bad, and the Ugly? An Empirical Investigation of Revoking Behavior on Ebay.”

Zhu, Feng, and Xiaoquan Zhang. 2010. “Impact of Online Consumer Reviews on Sales: The Moderating Role of Product and Consumer Characteristics.” *Journal of Marketing* 74 (2): 133–148.

Curriculum Vitae

Johannes Mans

Personal details

Place of birth	Lahr im Schwarzwald
Date of birth	04.10.1981
Nationality	German

Education

Since 10/2011	Research Assistant in DFG Research Unit „Design & Behavior – Economic Engineering of Firms and Markets“, University of Cologne
Since 10/2009	PhD student at the University of Cologne, Cologne Graduate School in Management, Economics and Social Sciences Supervisor: Prof. Dr. Axel Ockenfels
05/2014 - 06/2014	Visiting scholar, University of Texas at Dallas, Prof. Gary E. Bolton, PhD
01/2012 - 07/2012	Visiting scholar, Pennsylvania State University, Prof. Gary E. Bolton, PhD
10/2003 - 11/2009	Diploma in Economics (with special references to Social Sciences), University of Cologne
09/1992 - 06/2001	Abitur, Schiller Gymnasium Offenburg

Johannes Mans
22 June 2015

Erklärung

nach § 6 der Promotionsordnung vom 16. Januar 2008

Ich erkläre hiermit, dass ich die vorgelegte Arbeit ohne Hilfe Dritter und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe. Die aus anderen Quellen direkt oder indirekt übernommenen Aussagen, Daten und Konzepte sind unter Angabe der Quelle gekennzeichnet. Bei der Auswahl und Auswertung folgenden Materials haben mir die nachstehend aufgeführten Personen in der jeweils beschriebenen Weise entgeltlich/ unentgeltlich geholfen:

Weitere Personen – neben den in der Einleitung der Arbeit aufgeführten Koautorinnen und Koautoren – waren an der inhaltlich-materiellen Erstellung der vorliegenden Arbeit nicht beteiligt. Insbesondere habe ich hierfür nicht die entgeltliche Hilfe von Vermittlungs- bzw. Beratungsdiensten in Anspruch genommen. Niemand hat von mir unmittelbar oder mittelbar geldwerte Leistungen für Arbeiten erhalten, die im Zusammenhang mit dem Inhalt der vorgelegten Dissertation stehen. Die Arbeit wurde bisher weder im In- noch im Ausland in gleicher oder ähnlicher Form einer anderen Prüfungsbehörde vorgelegt. Ich versichere, dass ich nach bestem Wissen die reine Wahrheit gesagt und nichts verschwiegen habe.

Unterschrift: