DISPUTE RESOLUTION OR ESCALATION?

THE STRATEGIC GAMING OF FEEDBACK WITHDRAWAL OPTIONS IN ONLINE MARKETS

GARY BOLTON, BEN GREINER, AND AXEL OCKENFELS

22 November 2016

Abstract. Many online markets encourage traders to make good after an unsatisfactory transaction by offering the opportunity to withdraw negative reputational feedback in a dispute resolution phase. Motivated by field evidence and guided by theoretical considerations, we use laboratory markets with two-sided moral hazard to show that this option, contrary to the intended purpose, produces an escalation of dispute. The mutual feedback withdrawal option creates an incentive to leave negative feedback, independent of the opponent's behavior, to improve one's bargaining position in the dispute resolution phase. This leads to distorted reputation information and less trust and trustworthiness in the trading phase. Buyers who refuse to give feedback strategically, even when it comes at a personal cost, mitigate the detrimental impact. It is also mitigated in markets with one-sided moral hazard and a unilateral feedback withdrawal option.

Keywords: Dispute resolution system, market design, reputation, trust, reciprocity

JEL classification: C73, C9, D02, L14

Financial support from the German Science Foundation (DFG) through the research unit "Design & Behavior" (FOR 1371), the Australian Research Council (ARC) through Discovery Project DP130104557, and the U.S. National Science Foundation (NSF) is gratefully acknowledged. Ockenfels thanks the Stanford Economics Department for its hospitality. We thank seminar participants at Humboldt University, Simon Fraser University, Stanford University, University of Melbourne, University of Munich, University of New South Wales, New York University at Abu Dhabi, University of Sydney, University of Technology Sydney, Vienna University of Economics and Business, and at the eBay headquarters, as well as participants at the 2013 workshop of the DFG research unit "Design & Behavior", the 2013 Economic Science Association Meeting in Zurich, and the 2013 Australia-New Zealand Workshop in Experimental Economics for helpful comments. We are indebted to Markus Baumann, Paul Beckmann, and Kevin Breuer for excellent research assistance with the experiments. We thank eBay for the field data shown in Section II of this study, which were obtained under a cooperation agreement between the authors and eBay. In previous research, we used theory, laboratory and field experiments to help eBay (re-)design certain aspects of their auction rules and feedback forum, partly in order to mitigate various forms of social dispute; this study was motivated by, but independent from, our collaboration with eBay.

Bolton. University of Texas in Dallas, Naveen Jindal School of Management, Department of Managerial Economics, 800 W Campbell Rd 3.612, Richardson, TX 75080, USA, Tel: 1 (972) 883 5844, e-mail: gbolton at utdallas.edu.
Greiner. Wirtschaftsuniversität Wien, Institute for Markets and Strategy, Welthandelsplatz 1, 1020 Vienna, Austria, Tel: +43 1 31336 6021, e-mail: bgreiner at wu.edu.at.
Ockenfels. University of Cologne, Department of Economics, Albertus-Magnus-Platz, D-50923 Köln, Germany, Tel: +49/221/470-5761, e-mail: ockenfels at uni-koeln.de.

I. Introduction

The asynchronous timing and geographically dispersed nature of online markets requires traders to trust one another, and it can happen that disputes arise over whether the terms of trade were fulfilled. An important part of being trustworthy is 'making good'; that is, correcting errors or offering compensation when the terms of a deal cannot be met. To facilitate make good behavior, many online markets offer the incentive of getting a negative feedback rating withdrawn. A make good trader can preserve his reputation by turning an unhappy trade into a happy one, increasing the welfare of all involved. But as we will see there are also reasons to think that feedback withdrawal options can create opportunities for strategic gaming that weaken, rather than strengthen, the incentives to be trustworthy. In this paper, we examine evidence for this hypothesis, and for the consequences any gaming has on the marketplace.

TABLE 1: FEEDBACK WITHDRAWAL (FBW) RULES IN ONE- AND TWO-SIDED TRADING SYSTEMS

		FB-System	
		Two-sided	One-sided
FBW	Yes, unilateral	taobao.com ebid.net	amazon.com ioffer.com eBay post-2008 (limited to 0.5% of received feedback)
	Yes, mutual	eBay pre-2008 etsy.com discogs.com tradingpost.com.au ricardo.ch	
	No	trademe.co.nz mercadolibre.com listia.com	eBay mid-2008 marktplaats.nl eCrater.com

Table 1 describes the use of feedback withdrawal options by some of the better-known online markets. In theory, the choice of a one- or two-sided reputation system depends on whether there is scope for moral hazard or adverse selection problems on both buyer and seller sides of the market or on just one side. Some on the markets implement a one-sided feedback system (usually the buyer can rate the seller) while others implement a two-sided feedback system (buyer and seller rate one another).[1] Of the markets that allow feedback withdrawal, some allow a trader to withdraw her feedback unilaterally, while for others feedback

[1] The potential for moral hazard and adverse selection is transparent on platforms like Airbnb, which has a two-sided system. Other platforms are harder to judge. For example, while seller moral hazard is a known and accepted issue on eBay, the platform moved in 2008 from a two-sided to a one-sided (buyer rates seller) system. The extent of buyer moral hazard is debatable. On the one hand, there is little quantitative evidence identifying buyer moral hazard. On the other hand, anecdotal evidence from surveys, discussion forums, and eBay seller conferences speaks of problems such as the buyer purchases, but does not pay; the buyer raises unsubstantiated complaints; and the buyer blackmails the seller for over-fulfillment by threatening negative feedback.

withdrawal requires mutual trader agreement. Observe that there is a correlation between the sidedness of the feedback system and the choice of unilateral or mutual feedback withdrawal.

There are reasons to suspect feedback withdrawal options might be gamed, particularly *mutual* feedback options. At a broad level, feedback withdrawal is effectively a dispute resolution technique (the market platform records the feedback and sets the rule for feedback withdrawal). and it is well established that other dispute resolution methods, such as binding arbitration, in which a third party imposes a settlement, can invite unintended strategic behavior. Binding arbitration saves the negotiating parties the costs of an impasse (the intended purpose), yet expectations of what the arbitrator will decide can change negotiation strategies and, in turn, both the nature and probability of a non-arbitrated, voluntary settlement (Iyengar and Ashenfelter 2009, Bolton and Katok 1998, Deck and Farmer 2007). The distortions can be traced to the need to mandate the use of arbitration by agreement prior to a dispute arising, since the party accused of doing harm typically has little incentive to reach an agreement after an impasse is reached (Shavelle 1995). Feedback withdrawal options similarly aim to reduce impasse costs. But they take a novel and softer, incentive-based approach to encouraging participation, exploiting the traders' incentive to maintain a good reputation. The options and strategies available are also different than in traditional dispute resolution mechanisms. While the effectiveness of various online feedback system designs are studied by a rich and growing body of literature (e.g. Bolton et al. 2004, Bolton et al. 2013, Hui et al. 2016, Li and Xiao 2014, and Luca and Zervas forthcoming), we are not aware of another study that deals with the effectiveness of dispute resolution mechanisms.

More direct but anecdotal evidence for unintended behavior comes from the advice traders receive about how to navigate mutual feedback withdrawal systems. The advice is to play strategically. As Frank Fortunato (2007) explains to ecommerce-guide.com readers who trade in markets with two-sided feedback systems,

> 'Mutual Feedback Withdrawal' is the easiest and surest way to remove a negative from your rating. After receiving a negative feedback it is a good idea to contact the buyer and try to reason with the person. […] However, *I recommend calling them after leaving the other party a negative feedback in response*. It gives you *leverage in further negotiations*, and may be your only chance to do so because once you enter the Mutual Feedback Withdrawal process, eBay will not allow you to leave feedback for the transaction" [*emphasis added*].

If the recommended strategic feedback reply is successful at getting the other trader to remove their negative feedback even when the trader did not make good in the form of compensation, then this distorts feedback in a way that hurts the ability of future traders to accurately forecast who they should trust, and thus may eventually hamper trade efficiency. This line of argument seems to have contributed to eBay's decision in 2008 to abolish their mutual feedback withdrawal policy when moving to a one-sided system. EBay writes, Why is mutual feedback withdrawal being removed? eBay wants sellers to focus on buyer satisfaction up front, not after an issue arises. [...] The possibility of Feedback withdrawal leaves buyers open to potentially unwelcome contact from sellers attempting to have the buyer change the Feedback.[2]

Should strategic manipulation be suspected equally in both one and two-sided systems? On the one hand, feedback withdrawal changes the incentives for dispute. Perhaps this is enough to cause problems. On the other hand, Frank Fortunato's advice does not apply to one-sided systems since these systems are unilateral by design; replying to negative feedback with negative feedback is not possible.

In the next section, we review field evidence from eBay in the period when mutual feedback withdrawal option was an option. The data suggest that feedback giving is in fact manipulated. However, the implied leverage of strategic feedback giving appears lower than expected—receiving negative feedback does not induce buyers to concede as much as our theory would suggest. Based on the behavioral economics literature on *altruistic punishment* (Fehr and Gächter 2000, Balafoutas et al. 2014), we hypothesize that the reduced leverage is due to the unwillingness of some buyers to accept a seller's aggressive trade and feedback behavior. These buyers would thus not be willing to concede without the seller making good—even when this comes at a cost to themselves. Unfortunately, the field record does not allow an investigation of our hypotheses in more detail, because it does not allow us to relate dispute resolution behavior to either initial trading or the make-good stage. There are also a number of endogeneity issues associated with judging the field data, which we will describe later. Thus, to further examine the influence of online feedback withdrawal on trader actions and market performance, we construct a model that we then examine both theoretically and behaviorally. Our laboratory observations complement the theory and the field data by providing strong evidence for strategic gaming and, at the same time, a behavioral explanation for the disconnect between the strategic argument and the field data.

Taking the evidence together, we find that the mutual feedback withdrawal option creates an incentive to leave negative feedback, independent of the opponent's behavior, to improve one's bargaining position in the dispute resolution phase. The escalation can distort reputation information and hamper efficiency in the trading phase of the market. The detrimental impact is mitigated by buyers who choose to not allow the seller to get away with strategically aggressive behavior, even when it comes at a cost to themselves. We also test and find support for the prediction from our model that the gaming incentives due to a feedback withdrawal option largely vanish in markets with one-sided moral hazard.

[2] http://pages.ebay.com.au/help/feedback-changes.html, last accessed July 22, 2014 (not directly available as of April 2016, but available via archive.org or as PDF on request). The "mutual feedback withdrawal" procedure on eBay was introduced in 2004 and abolished in 2008. Soon after, eBay introduced a withdrawal policy allowing, with some restrictions, sellers to submit "feedback revision requests" to buyers, where buyers could change their feedback.

II. Field data

To get a closer look at how feedback withdrawal options influence feedback giving, we turn to field data from eBay. The dataset includes 573,567 transactions in June 2007 from seven countries and six categories.[3] Most importantly, the data include all feedback given on these transactions until about 4 months later (when the data was retrieved), including information on whether the mutual feedback withdrawal process was initiated and whether it led to a withdrawal of the respective feedback.

TABLE 2: FEEDBACK AND WITHDRAWAL FREQUENCIES IN EBAY FIELD DATASET

Feedback given by		Frequency of feedback given	Withdrawn	Unresolved	Frequency of eventual feedback
Buyer -> Seller	Positive	96.8%	0.5%	0.0%	97.2%
	Neutral	1.2%	8.3%	4.7%	1.1%
	Negative	1.9%	16.3%	9.5%	1.6%
Seller -> Buyer	Positive	97.7%	0.4%	0.0%	98.3%
	Neutral	0.3%	3.7%	3.3%	0.3%
	Negative	2.0%	12.7%	7.7%	1.8%

Notes: Columns 1 and 4 list the frequency of transaction feedback given (before any withdrawal), and four months after the transaction (after withdrawals up to the time of data retrieval), respectively. Column 2 shows the percentage of feedback in each feedback category (i.e. positive, neutral, and negative) that had been withdrawn up to the time of data retrieval, and Column 3 shows the percentage of feedback in each feedback category for which mutual feedback withdrawal was initiated but has not yet found a resolution up to the time of data retrieval. Approximately 45% of the removed feedback reported in Column 2 was withdrawn through the "mutual feedback withdrawal" process. Other reasons mainly include "unpaid item claims" and the "90 day rule".[4]

Table 2 shows the distribution of the initially submitted feedback on eBay, the percentage of feedback that is withdrawn, and the eventual distribution of feedback. Approximately 97% of all submitted feedback is positive (see Bolton et al. 2013, and Dellarocas and Wood 2008 for a discussion on the high frequency of positive feedback). Approximately 16% of the negative feedback submitted by buyers is eventually withdrawn; for another 10%, withdrawal was initiated but the process had not concluded at the time of data retrieval. The withdrawal numbers for neutral feedback and seller-to-buyer neutral and negative feedback are somewhat lower. The table also demonstrates that while initial feedback is already very positively skewed (96.8% positive), the eventual feedback distribution after withdrawals is even more so (97.7%). As a

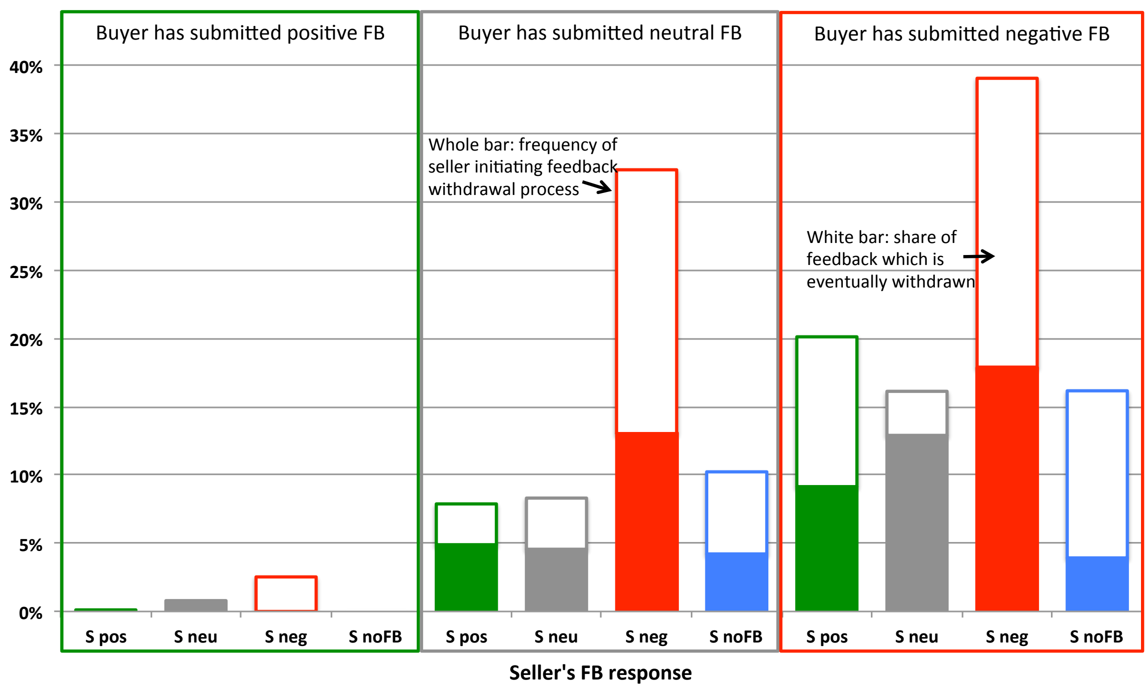
[3] Our analysis here is based on "*Dataset 2*" that is described in detail in Bolton et al. (2013), although they did not analyze data related to the feedback withdrawal option that we study here.

[4] Beside mutual feedback withdrawal, there are other processes that may have led eBay to delete feedback at that time. One of these processes is the "unpaid item claim", where a seller might contact eBay and complain that the buyer has not paid for an item. If the buyer does not respond to this complaint within a certain timeframe, then eBay will delete any feedback the buyer may have left. A second rule is that eBay deletes feedback from users who are excluded from the platform (based on negative feedback scores) within 90 days after their registration. These two processes may overlap with the mutual feedback withdrawal procedure, such that the data are not able to reveal how much feedback would have been withdrawn if mutual feedback withdrawal were the only withdrawal possibility. For example, a seller might initiate a mutual feedback withdrawal process and at the same time file an unpaid item claim, but before any of these processes are resolved, the buyer is excluded based on her negative feedback score.

consequence, it is the most valuable information (the rare information about misbehavior) that is potentially lost from the gaming of the mutual feedback withdrawal system.

Figure 1 shows the probability that a seller initiates the feedback withdrawal procedure and whether this endeavor was successful conditional on the buyer's first feedback and the seller's feedback response (a table with frequencies for all other cases of feedback timing can be found in Appendix B). Consistent with the first half of Fortunato's hypothesis, we observe a higher share of mutual feedback withdrawal process initiations by the second mover when she has responded with negative feedback to an observed non-positive feedback, rather than responding with no feedback at all. After having retaliated negative (neutral) feedback with negative feedback, the seller challenges the feedback in 39% (32%) of the cases, whereas only 16.2% (10.2%) of feedback are challenged when the seller did not respond with feedback; Fisher's Exact Tests yield $p<0.0001$ in both comparisons.[5]

FIGURE 1: FREQUENCY OF SELLER'S WITHDRAWAL PROCESS INITIATION AND SUCCESS CONDITIONAL ON BUYER'S FEEDBACK AND SELLER'S RESPONSE



The white part of the bars in Figure 1 represents the success rates of mutual feedback withdrawal initiations, with success defined as the eventual withdrawal of the other's feedback about oneself (up to the time of our data retrieval). Contrary to the hypothesis that strategic negative feedback leads to more negotiation power in the feedback withdrawal process, we find that the success rate after retaliating with

[5] Since we observe very little negative feedback in general, the total number of observations of initiated feedback withdrawal processes in this category is small, in particular for buyers, even though we began with a large dataset of more than half a million transaction observations. For the reported Fisher's Exact tests, the number of observations are: negative after negative N = 2,821, no feedback after negative N = 3,006, negative after neutral N = 445, no feedback after neutral N = 2,561. Table B.1 in Appendix B includes all observation counts.

negative feedback is either the same or even less than not responding with feedback (Fisher's Exact Tests, $p = 0.916$, $N = 144$ and $N = 262$, and $p < 0.0001$, $N = 262$ and $N = 486$ for buyer neutral first and buyer negative first, respectively). Thus, our data fail to verify the second half of Fortunato's hypothesis that retaliation toward negative feedback in-kind increases the chances of getting rid of it.

There are several hypotheses for why strategically retaliating with negative feedback may not pay as expected. One hypothesis is self-selection, that sellers who retaliate against negative feedback in the trading phase are traders with below average communication or social skills, which in turn would make it more likely that negotiations in the dispute resolution phase would fail. An alternative hypothesis, the one we will focus on in our laboratory study, is that leaving negative feedback on the seller reflects the buyer's impression that she has been treated unfairly in the transaction. For these already annoyed buyers, receiving an unjustified (strategic) negative feedback plus a request to withdraw one's own (truthful) negative feedback might actually increase their willingness to punish the seller, even if it comes at a cost. Indeed, in a related context, Nikiforakis (2008) found that victims of unjustified punishment in a public-good game (where unjustified was defined as being punished despite contributing more than the punisher) were more than twice as likely to retaliate punishment.

Due to limitations in the available field data, we cannot separate the different mechanisms that may be involved in the field. In particular, we are unable to investigate the role of behavior during dispute resolution, such as pecuniary make-good or apologies or threats. Moreover, we hypothesize that the reputation system design of a market, one-sided or two sided, interacts with the dispute resolution system to promote trade efficiency. A feedback withdrawal option, which creates a detrimental clash of negotiation threats in two-sided systems, may work well in one-sided systems that prevent such escalations. This also cannot be investigated with our field data, which is confined to a two-sided system. Thus, in order to investigate the interplay between the dispute resolution phase and the feedback system, we conducted laboratory experiments. The experiments allow us to control all terms of the economic transaction as well as the message space in the communication that takes place between sellers and buyers. It also allows us to investigate the impact of strategic and non-strategic behaviors on the (in)effectiveness of the feedback withdrawal option. Finally, the experiments level the playing field when studying the impact of dispute resolution systems under the two typical market types described in Table 1—markets with one-sided feedback systems (assuming that only one market side is subject to moral hazard), and markets with two-sided feedback systems (in markets with two-sided moral hazard). This will help us to understand the determinants of the benefits and costs of feedback withdrawal systems used in various online markets.

III. Experiment design

We study four different treatments along two dimensions: two-sided feedback systems with (2s-FBW) and without (2s-noFBW) feedback withdrawal options, and one-sided feedback systems with (1s-FBW) and without (1s-noFBW) feedback withdrawal options.

At the beginning of a market interaction in a two-sided feedback system treatment, both the buyer and the seller receive an endowment of 100 ECU (Experimental Currency Units), and are informed about any feedback their transaction partner received in previous interactions

In the first stage of the interaction, both the buyer and the seller are asked whether they want to trade. If both agree to trade, then the interaction continues; otherwise both earn their endowment. Like in the field, a decision not to trade allows the trader to avoid exposure to feedback from somebody that they do not want to interact with.

Next, the buyer and the seller engage in a transaction. There is potential moral hazard on both sides of the market. The buyer decides whether to pay his endowment of 100 ECU to the seller. At the same time, the seller decides on the quality of the product to ship, which can be between 0% and 100%. Each quality percentage point costs the seller 1 ECU and benefits the buyer 3 ECU. Thus, the efficiency gains from trade depend on the quality choice, while the buyer's payment has only distributional effects. Also, a quality choice of 100% maximizes efficiency, while a quality choice of 50% leads to an equal split of payoffs when the buyer has paid. This may create a normative conflict about what is sufficient quality to justify positive feedback. Normative conflict has been shown to be one important source of conflict escalation in Nikiforakis et al. (2012), and is thus a useful feature in a study of dispute resolution.

After the transaction, both the buyer and the seller are informed about each other's transaction choices and are asked to simultaneously submit either positive or negative feedback on each other. After feedback has been submitted, both parties are informed about the feedback they received, and simultaneously have the opportunity to make good by improving upon their payment/quality choice. In particular, if the buyer has not paid in the trading phase, he may do so now, while the seller may increase the quality. Subsequently, both parties are informed about the make-good behavior of their counterpart.

For treatment 2s-nFBW, the interaction ends here. In treatment 2s-FBW and if at least one feedback was negative, both the seller and the buyer can vote for or against a withdrawal of feedback. If both traders agree to withdraw, then the feedback for both parties would be made positive. Otherwise, feedback stays as given.[6]

In the treatments with a one-sided feedback system, only the buyer has to agree whether to trade. If he agrees, then the 100 ECU is automatically sent to the seller (so there is no scope for buyer moral hazard), and

[6] Research on the effects of feedback on eBay suggests that it is mainly negative feedback that matters for future payoffs of traders (Livingston 2004, Houser and Wooders 2005). Neutral feedback is often treated like negative feedback (as Figure 1 also suggests). Thus, to keep the experimental design simple, we decided to subsume neutral and negative feedback into one category, and positive and no feedback into another category.

the seller makes the choice in product quality. After the buyer is informed about that choice, she submits either positive or negative feedback. Then, after being informed about the feedback, the seller may make good by increasing the quality. In treatment 1s-noFBW, this ended the interaction. In contrast, in treatment 1s-FBW the buyer would be informed about the make-good choice of the seller and may choose whether to withdraw her negative feedback, so that it is turned into positive feedback.

Table 3 summarizes the stages of the buyer-seller interaction and the differences between treatments. It will be useful to refer to the first three stages as the *trading phase* and the last two stages as the *dispute resolution phase*.

TABLE 3: SUMMARY OF THE LABORATORY STAGE GAME

	Feedback System	
Stage	2-sided *(2s-noFBW + 2s-FBW)*	1-sided *(1s-noFBW + 1s-FBW)*
Feedback displayed	Both transaction partners' number of positive and negative feedback received in previous rounds.	Seller's number of positive and negative feedback received in previous rounds.
Trading phase		
1. Elect to trade	Buyer and seller simultaneously decide whether to trade. If one decides not to trade, then the round ends with round payoffs π_B=100, π_S=100.	Buyer decides whether to trade. If buyer decides not to trade, then the round ends with round payoffs π_B=100, π_S=100.
2. Transaction	Buyer decides whether to pay 100 ECU. Seller simultaneously decides on Quality Q_1 with $0 \leq Q_1 \leq 100\%$.	100 ECU are sent automatically. Seller decides on Quality Q_1 with $0 \leq Q_1 \leq 100\%$.
3. Feedback	Buyer and seller decide simultaneously whether they give positive or negative feedback.	Buyer decides whether he gives positive or negative feedback.
Dispute resolution phase		
4. Make good	If buyer has not made the payment in Stage 2, then he can now decide again whether to pay. Seller simultaneously decides on Quality Q_2 with $Q_1 \leq Q_2 \leq 100\%$.	Seller decides on Quality Q_2 with $Q_1 \leq Q_2 \leq 100\%$.
5. Feedback withdrawal *(2s-FBW and 1s-FBW only)*	If there was negative feedback for at least one side in Stage 3, both buyer and seller decide whether to vote for feedback withdrawal. If both vote for withdrawal, both feedback will be made positive.	If buyer has given negative feedback in Stage 3, then he decides whether to withdraw feedback. If buyer withdraws, then feedback will be made positive.
Payoffs	π_B= 100 – PricePaid + Q_2 * 3 π_S= 100 + PricePaid – Q_2	π_B= Q_2 * 3 π_S= 200 – Q_2

A note on the design: We include a make good phase in the baseline *noFBW* games because it is always possible to do in the field even without feedback withdrawal, and because it controls for level effects to better isolate the influence of having the *FBW* option when comparing treatment behavior. While not necessarily cheaper (market platform like eBay also offered feedback withdrawal free of charge), the *FBW* option may be more salient in the simplified environment of the laboratory than it is in the field. But the important point, given the level effect control, is that it is not automatically used either in the field (Figure 1) or, as we will see, in the laboratory (Table 6 below) environment.

The experiment was conducted at the Laboratory for Experimental Research at the University of Cologne. Participants were students mostly from economics and business administration, recruited via ORSEE (Greiner 2015). We conducted eight sessions (two sessions per treatment) with 248 participants (64 traders in each of 2s-FBW, 2s-noFBW, and 1s-FBW, and 56 traders in 1s-noFBW due to some no-shows in one session). Upon arrival, participants read the instructions (see Appendix C) and could privately ask questions. Once all questions were answered, the experiment began. The market was repeated for 60 rounds, with fixed buyer/seller role assignment but random trader matching in each round. Matching was restricted to groups of 8 participants (4 buyers and 4 sellers), yielding 8 (in one cell 7) matching groups per treatment. Sessions lasted approximately two hours. Participants were paid from their cumulative earnings over all 60 rounds in cash at the end of the experiment, with an exchange rate of 400 ECU = 1 EUR. The average payoff was EUR 20.03 (SD = 2.14) including a show-up fee of EUR 2.50.

IV. Hypotheses

The fundamental intuition behind our hypotheses is that permitting mutual feedback withdrawal weakens the incentive to be trustworthy in the first place. We would expect, however, that this change in incentive leads to different strategic behavior depending on the nature of moral hazard problems, and the associated difference in feedback systems (one-sided or two-sided).

The one-shot version of all of the stage games described in the previous section has many subgame-perfect Nash equilibria (SPNE), although all of them stipulate no buyer payment, zero quality, and no make-good. Hence, all yield payoffs are equal to the initial endowments. The same statements hold for the SPNEs for the finitely repeated version of these one-shot games under the usual assumptions. That said, there are two reasons to be dissatisfied with such an analysis. First, experiments on social dilemma games like the one we study here (Bohnet and Huck 2004, Bolton et al. 2003, Bolton et al. 2013, Nowak and Sigmund 1998) find that participants do cooperate and make higher profits than what the analysis anticipates. Second, while the SPNE solutions all imply zero market activity, they differ with respect to feedback giving and feedback withdrawal behavior; in fact, feedback behavior is, with regard to equilibrium, arbitrary.[7] Yet this prediction is clearly at odds with both experimental and field observations of these markets.

In principle, a finitely repeated, incomplete information model in which there is a small probability that a trader is truly trustworthy (e.g., Kreps et al. 1982) could make more plausible predictions of trade, payments, and positive quality. A full model of this kind is beyond the scope of the present paper. However, the SPNE analysis associated with the stage game, given some straightforward assumptions about the honesty of feedback and how it relates to the continuation payoffs, captures the intuition behind our hypotheses. The full analysis, including the complete game tree, is in Appendix A. Here, we provide an overview.

Our model is designed to illustrate why and where mutual feedback withdrawal creates an incentive problem. For this reason, our assumptions are chosen to allow for straightforward strategies that lead to robust cooperation without mutual feedback withdrawal. We first assume that the feedback system, in principle, allows simple 'tit for tat'-like behavior to promote trust and trustworthiness. That is, defection is followed by costly negative feedback and cooperation is followed by beneficial positive feedback. For this, we assume that traders who are monetarily indifferent will submit informative (honest) feedback, conditional on whether the opponent cooperated or not (otherwise, they submit feedback strategically),[8] and that receiving negative feedback imposes a reduction in continuation profits that outweighs profits from shirking in the single-stage game (see e.g. Bolton et al. 2013).[9] For simplicity in our exposition, we additionally assume that what the buyer considers to be a minimum 'satisfactory quality' level is known by the seller, so that negative feedback cannot result from the seller underestimating the quality level that would gain him positive feedback (see Appendix A for further specifications).

Given these assumptions we can reanalyze the SPNE for each stage game for each market. In markets without the feedback withdrawal stage (1s-noFBW and 2s-noFBW), traders do not make good in the dispute resolution phase, because there is no feedback withdrawal stage where this could have any effect. In the trading phase, each trader gives negative feedback if and only if the opponent defected (did not make the payment or did not provide at least the minimum satisfactory quality level). This in turn incentivizes them to

[7] The arbitrariness is driven by the fact that, in equilibrium, feedback does not affect economic outcomes. But even if it would, truthful feedback giving in large communities involving interactions with strangers is often characterized as a public good that, by the usual assumptions, contributes little value to the giver (e.g., Bolton et al. 2011).

[8] While we assume 'truth-telling' as a mere tie-breaker, Gneezy (2005) and Gneezy et al. (2013), among others, provide considerable evidence that some people are willing to forgo profits in order to avoid lying. Our experimental evidence provides support for this assumption; we observe mostly honest feedback in the system without FBW, when there are no strategic incentives to submit dishonest feedback.

[9] There is robust and strong evidence for a cost of negative feedback both within and beyond our experiment, which empirically justifies its critical role in our model. Empirical studies of eBay's feedback system (as summarized, for example, in in Dellarocas 2004 and Cabral 2012) generally find a negative impact of negative feedback on subsequent sales. Previous laboratory research, including Bohnet and Huck (2004), Bolton, Greiner, and Ockenfels (2013), and Bolton, Ockenfels, and Ebeling (2011) cited above, as well as other studies surveyed in Bolton, Greiner, and Ockenfels (2013), shows a negative correlation between received negative feedback and future profits. For the data collected in this study, we compared profits in rounds 2-60 after having received a negative feedback in round 1 to those after having received positive feedback. We indeed find costs of negative feedback, in particular for the seller side.

cooperate. That is, to initiate trade, make the payment, and deliver satisfactory quality. Hence, under our assumptions, both two-sided and one-sided markets with no feedback withdrawal option should robustly promote trust and trustworthiness.

Turning to the SPNE analysis for the market with a one-sided feedback system and a feedback withdrawal option: In the dispute resolution phase, the buyer withdraws negative feedback when the seller delivers satisfactory quality. The seller therefore makes good if she has received negative feedback (because of unsatisfactory quality) in the trading phase. In the trading phase, the buyer submits negative feedback whenever the seller does not deliver sufficient quality. Thus, in equilibrium, the seller has the choice to deliver satisfactory quality right away or to deliver lower quality first and then improve quality to a satisfactory level later. In any case, the seller will deliver sufficient quality by the end of the round, such that the buyer will always decide to buy. The prediction then is that allowing feedback withdrawal in the dispute resolution phase of the one-sided feedback market may influence the timing of the seller making good, but beyond this, the market will be as efficient as when feedback withdrawal is absent.

The predicted outcome is very different for treatment 2s-FBW. Under our assumptions, there is a unique SPNE path starting at Stage 2 of the game: 1) A trader agrees to withdraw feedback if and only if either (*i*) the trader received negative feedback herself, or (*ii*) the trader gave negative feedback and the other trader has made good or had cooperated in the first place. 2) Traders make good (i.e. make the payment or send satisfactory quality) when they have received negative feedback while they have given the other trader positive feedback. The reason is that this way, in Stage 5, their negative feedback will be withdrawn. Otherwise, i.e. when both have received negative or both have received positive feedback, there is no make-good, because in case of mutually negative feedback, this feedback will eventually be withdrawn in any case. 3) *Both* traders give negative feedback, because by leaving a negative, a trader can guarantee eventually receiving positive feedback.[10] 4) Since in equilibrium a defecting trader can protect her reputation from being damaged, the feedback system is uninformative about past behavior, and thus cannot support reputation building and cooperation; the buyer will not make the payment and the seller will not deliver quality.[11]

[10] In case both traders cooperated in Stage 2 (which would be out-of-equilibrium), there is an additional SPNE of the subsequent subgame in which both traders leave 'honest' positive feedback in Stage 3. However, even in this case, leaving negative feedback is also in equilibrium, because it is a best response when there is a positive probability that the opponent will leave negative feedback. Nevertheless, even when this probability is zero, negative feedback at least does not hurt as all negative feedback can be withdrawn at Stage 5. This strengthens the robustness of our claim that feedback is given strategically under the shadow of mutual feedback withdrawal.

[11] If in equilibrium feedback is not informative and thus, strictly speaking, worthless in pecuniary terms, our assumption that negative feedback comes at a cost can be interpreted as a robustness check. Even if traders dislike negative feedback for different reasons (e.g. in a broader social and economic context, or relating to psychological values; see Ockenfels and Resnick 2012), they could not overcome the dilemma created by the feedback withdrawal system. That our base game cannot promote any efficient trade if feedback has no (economic or psychological) value is trivial.

Based on the above analysis, we formulate the following hypotheses for the experiment:

H1 [Trading phase; stages 1–3]: Having the feedback withdrawal option available invites gaming. The option leads to a decreased willingness to make the payment or to deliver good quality in the trading phase. In the two-sided market, such schemes lead to unjustified negative feedback in the trading phase (but not in the one-sided case).

H2a [Dispute resolution phase; stages 4–5]: The dispute resolution phase is used to attempt to have negative feedback withdrawn. In particular, traders in the feedback withdrawal treatments make good only when they have received negative feedback, in which case they vote for mutual feedback withdrawal independent of the opponent's make-good behavior. Feedback withdrawal thus leads to distorted final feedback. These effects are more pronounced for markets with two-sided moral hazard than for markets with one-sided moral hazard.

We complement our hypothesis H2a, which is motivated by our strategic model, with hypothesis H2b, which is motivated by our field data together with the behavioral literature on altruistic punishment (Ostrom et al. 1992, Fehr and Gächter 2002, 2002), as discussed in Sections I and II:

H2b [Altruistic punishment; stages 4–5]: In markets with two-sided moral hazard, the leverage of giving negative feedback in negotiating feedback is mitigated to the extent that strategic negative feedback behavior is punished by a refusal to withdraw feedback, even when this comes at a cost—that one's own negative feedback is not withdrawn by the other trader.

H3 [Market implications]: The feedback withdrawal option ultimately hampers the informativeness of feedback and trade efficiency in markets with two-sided moral hazard, but not in markets with one-sided moral hazard. Markets with one-sided moral hazard are not affected by the existence of a feedback withdrawal option.

V. Laboratory data

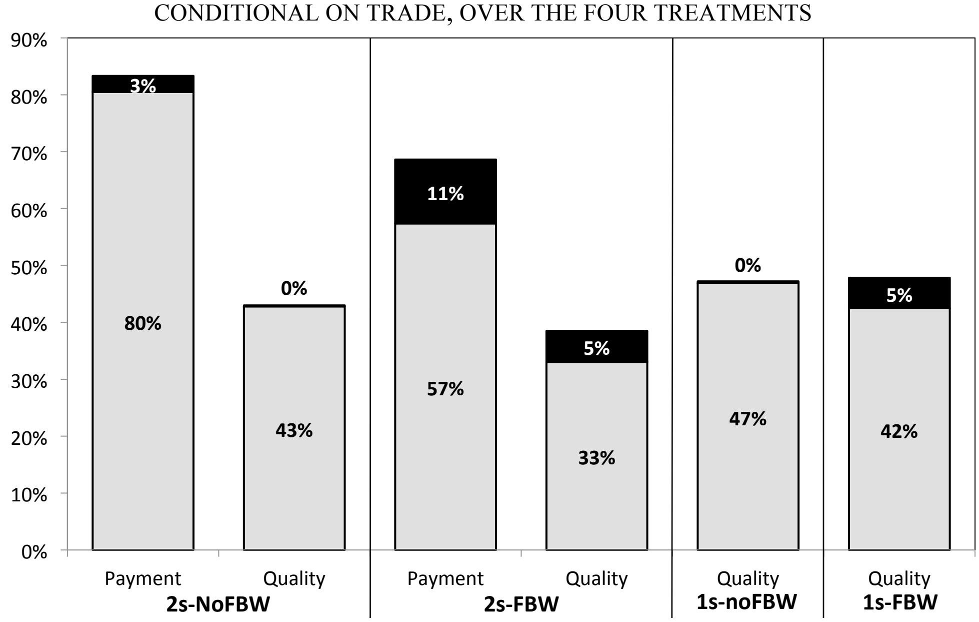
We structure our report of the experimental results according to the four hypotheses described above. As planned when we designed the experiment, all our analyses are based on behavior in rounds 11 to 50 of the 60-round experiment. The first ten rounds in our experiment likely involve some learning (we did not run practice rounds before starting the actual experiment), while the last ten rounds typically involve end-game effects. Discarding both the first ten and the last ten rounds allows us to concentrate on a 'running' feedback system, which is what we are mainly interested in. That said, it turns out that conducting the analyses on all 60 rounds yields the same qualitative results as reported here. We include the all-rounds version of all the tables in the Supplementary Appendix. The experiment uses a small matching group protocol (8 subjects; see section III). Both small and large matching group protocols are vulnerable to session-effects that can confound statistical tests, so one must choose on which side to err (Frechette, 2012). In our case, small matching groups minimize session-effects having to do with sample dependencies, the potential downside being that they might produce more reciprocity than is present in larger matching groups.

V.1. Trading phase

In the next section we will see that behavior in the dispute resolution phase is strongly affected by the feedback withdrawal option. Here, in line with H1, we find that the shadow cast by the withdrawal option impedes trading behavior and feedback giving in the trading phase. As expected, these effects are much stronger for markets with two-sided moral hazard. The following paragraphs describe the trade outcomes in Stage 1, as well as the payment and quality in Stage 2, and feedback giving in Stage 3.

Trade. In our experiment, we do not observe any differences across two-sided treatments in the likelihood to enter trade. In fact, the probability of entering trade is almost identical across all treatments: 80% in 2s-noFBW and 1s-noFBW, 81% in 2s-FBW, and 83% in 1s-FBW.

FIGURE 3: FREQUENCY OF PAYMENT AND AVERAGE QUALITY CONDITIONAL ON TRADE, OVER THE FOUR TREATMENTS



Notes: The grey bars represent the frequency of payment and average quality, respectively, before any make-good. The black bars show the additional frequency of payment and average gains in quality at the make-good stage.

Figure 3 shows payment frequencies and average quality choices across our four treatments. Columns (1) and (2) of Table 4 show the results from the Probit and Tobit analyses, respectively, of the propensity of the buyer to make the payment and the quality delivered by the seller in the trading phase, regressed on treatment dummies. The dummy *FBW* equals 1 in treatments 2s-FBW and 1s-FBW, and is 0 otherwise. *One-sided* takes the value of 1 in the one-sided treatments, and the value of 0 in the two-sided treatments. Finally, the interaction dummy *One-sided*×*FBW* is 1 in treatment 1s-FBW and 0 in all other treatments. The analyses confirm that the withdrawal option significantly and negatively affects payments and qualities in the trading phase of the two-sided market—the FBW variable is significant and negative in Models 1 and 2.

TABLE 4: REGRESSIONS OF THE PROBABILITY OF PAYMENT, QUALITY, AND EFFICIENCY ON TREATMENT DUMMIES

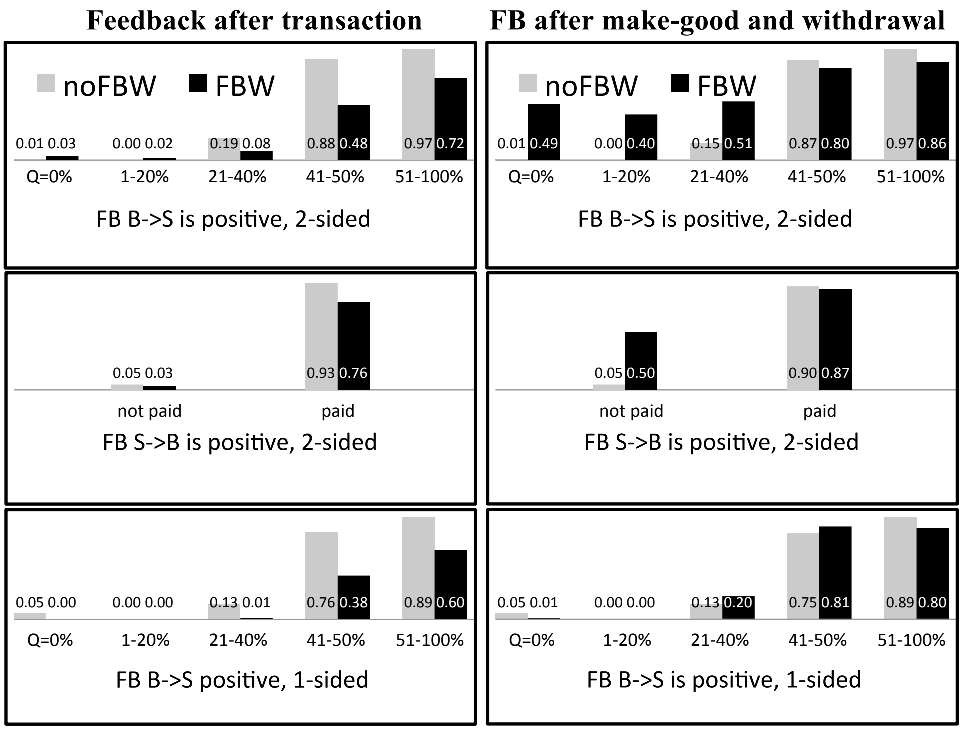
Model	(1)	(2)	(3)	(4)
Model type	Probit	Tobit	Probit	Tobit
Dependent	Payment Trade phase	Quality Trade phase	Payment After MG	Quality After MG
Constant		0.499***		0.492***
		[0.029]		[0.027]
Round	-0.005***	-0.003***	-0.006***	-0.002***
	[0.001]	[0.001]	[0.002]	[0.001]
FBW	-0.232**	-0.118**	-0.146*	-0.054
	[0.106]	[0.052]	[0.086]	[0.048]
One-sided		0.044*		0.045*
		[0.026]		[0.027]
One-sided ×FBW		0.075		0.062
		[0.059]		[0.053]
N	2067	4028	2067	4028
LL	-1195.7	-771.7	-1077.2	-614.7
Censoring Left		636 (3349)		552
(Non) Right		43		(3431) 45

Notes: The table reports average marginal effects dy/dx for the Probit models. Quality and efficiency are censored at 0 and 1. *, **, and ** denote significance at the 10%, 5%, and 1% level, respectively. Regressions are based on data from rounds 11-50 (omitting start and end effects). (Robust) Standard errors are clustered at the level of independent matching groups.

Feedback. The left panel in Figure 4 shows frequencies of positive feedback in Stage 3, conditional on the trading partner's behavior in Stage 2. As predicted, in the two-sided market *without* feedback withdrawal, traders tend to give more 'honest' and positive feedback to trustworthy traders than when there is a withdrawal option. For a buyer, the probability of receiving positive feedback when she paid (so that feedback should be positive) in the treatment without (with) the withdrawal option is 93% (76%), while it is 5% (3%) when the buyer did not pay. For sellers, the probability of receiving a positive in 2s-noFBW (2s-

FBW) when the quality was below 40% is 4% (5%); it is 88% (48%) when quality was between 40% and 50%, and 97% (72%) when quality exceeded 50%. This shows that, for the two-sided markets, the feedback withdrawal option strongly distorts feedback, with much less positive feedback for trustworthy traders.

FIGURE 4: FREQUENCY OF POSITIVE FEEDBACK
CONDITIONAL ON TREATMENT AND OTHER'S STAGE-2-BEHAVIOR



The left panel in Table 5 provides statistical evidence for the described effects, and shows that they are highly significant. We regress the buyers' and sellers' propensity to give positive feedback on the partner's cooperation (quality and payment, respectively), and break out the absolute effect of the feedback withdrawal option (dummy *FBW*) and its impact on the correlation between other's behavior and feedback (cross effect *FBW* × *Other's Quality/Payment*). Feedback is significantly correlated with the other's cooperation level in all treatments.[12] The regression detects a significant negative interaction effect of the feedback withdrawal option and other's behavior. These results are consistent with our interpretation of Figure 4. In particular, the negative interaction effect can be interpreted in two ways. First, it shows that, compared to no feedback withdrawal, feedback with the withdrawal option is more negative for high quality/payment. Second, it implies a lower correlation between trade behavior and feedback in the FBW markets, which means that there is less information value in the traders' reputation score (see also Section

[12] Interestingly, feedback giving is also correlated with one's own behavior: traders who cooperate are also more likely to give positive (i.e. less likely to give strategic negative) feedback (see also Figure B.1 in Appendix B).

V.3). Overall, we conclude that the effect of a feedback withdrawal option is significant for both buyer and seller feedback giving in the two-sided system.

As predicted, *one-sided markets* do not respond as strongly to the feedback withdrawal option.[13] Figure 3 shows that there is only a small, and statistically non-significant, negative effect of the withdrawal option on seller's quality choices (42% vs. 47%; the joint effect of *Has Withdrawal* and *One-sided* × *Has Withdrawal* in Model 2 shown in Table 4 is not significantly different from zero, p=0.101). The lower left panel in Figure 4 also shows that in one-sided markets, feedback is more negative when a withdrawal option is in place.[14] However, as Model 3 in Table 5 shows, unlike in two-sided markets, there is no indication that the *correlation* of feedback with behavior (and thus its informativeness) decreases.

TABLE 5: PROBIT REGRESSIONS OF INITIAL AND EVENTUAL FEEDBACK ON OTHER'S TRADE BEHAVIOR AND EXISTENCE OF THE FEEDBACK WITHDRAWAL OPTION

	Feedback after transaction			Feedback after withdrawal		
Model	(1)	(2)	(3)	(4)	(5)	(6)
Dependent	B->S FB is pos	S->B FB is pos	B->S FB is pos	B->S FB is pos	S->B FB is pos	B->S FB is pos
Condition	2-sided	2-sided	1-sided	2-sided	2-sided	1-sided
Other's Quality / Payment	2.071***	0.532***	1.247***	2.272***	0.792***	1.104***
	[0.507]	[0.043]	[0.165]	[0.571]	[0.019]	[0.191]
FBW	0.335	-0.022	-0.958**	0.444***	0.352***	-0.199*
	[0.256]	[0.04]	[0.393]	[0.027]	[0.037]	[0.117]
FBW × Other's Quality / Payment	-1.224**	-0.106**	1.365*	-1.824***	-0.372***	0.422
	[0.499]	[0.05]	[0.801]	[0.574]	[0.04]	[0.306]
N	2067	2067	1961	2067	2067	1961
LL	-762.8	-636.2	-921.0	-889.2	-806.3	-890.4

Notes: The table reports average marginal effects dy/dx. *, **, and *** denote significance at the 10%, 5%, and 1% level, respectively. "B->S" refers to feedback given by the buyer to the seller, and "S->B" vice versa. Regressions are based on data from rounds 11-50 (omitting start and end effects). (Robust) Standard errors are clustered at the level of independent matching groups.

Overall, we find strong evidence for the mutual withdrawal option to invite gaming in two-sided markets and feedback systems, leading to a decreased willingness to make the payment and to deliver good quality,

[13] The only difference in trading activities in one-sided as opposed to two-sided markets is that we observe somewhat higher quality in the one-sided market (see Figure 3 and Table 4, Model 2). While this is not predicted by our model, it is not implausible. The reason is that two-sidedness creates a different type of reciprocal relationship, one which allows sellers to be exploited and thus gives them more reason to be cautious (as suggested, e.g., by the work of Bohnet and Zeckhauser 2004, and Bohnet et al. 2008).
[14] Without any other strategic reasoning in the one-sided market, buyers may attempt to elicit even higher qualities from sellers by submitting negative feedback, which can then later, at no cost, be turned into 'honest' feedback.

as well as dishonest feedback in the trading phase. These effects are mitigated in markets with one-sided moral hazard.

V.2. Dispute resolution phase

Why does the design of the dispute resolution phase affect behavior in the trading phase? Does altruistic punishment mitigate the leverage of retaliatory feedback? We analyze the dispute resolution behavior according to our model's strategic predictions, as summarized by Hypothesis H2a, and based on our altruistic punishment Hypothesis H2b.

V.2.1 Make-good behavior in two-sided markets

For two-sided markets, our model suggests three effects. First, we should observe less make-good behavior when there is no feedback withdrawal stage, compared to when there is a feedback withdrawal option later on. Second, we would expect players in the feedback withdrawal treatments to make good only when they have received negative feedback. Third, due to the incentive in two-sided systems to vote for withdrawal irrespective of the other's make-good behavior when having received negative feedback oneself, we expect that a player is less likely to make good when she has strategic negotiation power at the feedback withdrawal stage (i.e. has submitted negative feedback herself).

The black-filled parts of the bars in Figure 3 show the extent of observed make-good behavior in the respective treatments. Table 6 lists, for each of our four treatments, the possible outcomes of the feedback stage (column 1) and their frequencies (column 2). It also shows the Stage 2 frequency of payments (P) and average quality delivered (Q) for each of these cases in column 3, and the same numbers *after* traders made good in Stage 4 in column 4. Column 5 shows the frequency of an improvement in payment (conditional on no payment in Stage 2) or quality at the make-good stage. Inspection of Table 6 yields strong support for all three parts of our hypothesis:

(1) Across all feedback outcomes, make-good results in an increase in average quality of 0.2% and in likelihood of payment of 3% in 2s-noFBW, compared to increases of 5% and 11% in 2s-FBW (see also Figure 3).

(2) The make-good in 2s-FBW mainly comes from traders who received negative feedback (3% of sellers improved their quality – by an average of 8% – when they received positive feedback, while 45% of sellers who received negative feedback made good (by an average of 17%). Similarly, previously non-paying buyers made the payment in 7% of the cases in which they received positive feedback, but in 27 % of the cases when they received negative feedback.

(3) Finally, in treatment 2s-FBW, the likelihood of a seller (buyer) to make good after receiving negative feedback is 34% (22%) when he had given a negative himself, and 68% (54%) when he did not.

In order to test whether make-good behavior is strategically influenced by the negotiation power in the subsequent feedback withdrawal stage, we run Probit models on the buyer's and seller's propensity to make good after receiving negative feedback in the two-sided FBW markets, with the only independent being a

TABLE 6: AGGREGATE FEEDBACK, MAKE-GOOD, AND WITHDRAWAL BEHAVIOR IN THE FOUR TREATMENTS

Treatment & Given Feedback	FB Freq.	P & Q before make-good	P & Q after make-good	Freq. of make-good	Withdrawal	Eventual FB Freq.
2-sided no FBW						
B->S pos, S->B pos	64%	P: 0.99 Q: 0.51	P: 0.99 Q: 0.51	P: 13% Q: 7%	-	64%
B->S neg, S->B pos	12%	P: 0.99 Q: 0.28	P: 0.99 Q: 0.28	P: 0% Q: 4%	-	12%
B->S pos, S->B neg	10%	P: 0.30 Q: 0.49	P: 0.50 Q: 0.49	P: 29% Q: 6%	-	10%
B->S neg, S->B neg	14%	P: 0.18 Q: 0.16	P: 0.23 Q: 0.17	P: 6% Q: 6%	-	14%
2-sided FBW						
B->S pos, S->B pos	21%	P: 0.98 Q: 0.59	P: 0.99 Q: 0.59	P: 25% Q: 1%	-	67%
B->S neg, S->B pos	24%	P: 0.96 Q: 0.35	P: 0.96 Q: 0.44	P: 0% Q: 68%	B: 62% S: 99% Both: 61%	9%
B->S pos, S->B neg	8%	P: 0.14 Q: 0.42	P: 0.61 Q: 0.43	P: 54% Q: 9%	B: 97% S: 51% Both: 49%	4%
B->S neg, S->B neg	47%	P: 0.26 Q: 0.19	P: 0.43 Q: 0.26	P: 22% Q: 34%	B: 73% S: 82% Both: 57%	20%
1-sided no FBW						
B->S pos	71%	Q: 0.53	Q: 0.54	Q: 7%	-	71%
B->S neg	29%	Q: 0.31	Q: 0.32	Q: 4%	-	29%
1-sided FBW						
B->S pos	31%	Q: 0.52	Q: 0.52	Q: 7%	-	72%
B->S neg	69%	Q: 0.38	Q: 0.46	Q: 72%	B: 59%	28%

Notes: P and Q stand for 'frequency of payment' and 'average quality', respectively; FB denotes 'feedback', Freq. means 'frequency', and B and S refer to buyer and seller, respectively. Column 2 shows the frequency of buyer/seller feedback outcomes in a treatment (Stage 3), column 3 shows the frequency of payment and average quality underlying this feedback outcome. Column 4 gives the same numbers after make-good (Stage 4) has taken place, and column 5 lists the frequencies of non-zero make-good behavior for payment (when no payment was sent in Stage 2) and quality. Column 6 includes information on how often buyer, seller, and both agreed to withdraw feedback, and column 7 shows the frequency of feedback outcomes after withdrawal has taken place (Stage 5). All aggregates are based on data from rounds 11-50 (omitting start and end effects).

TABLE 7: PROBIT REGRESSIONS OF LIKELIHOOD TO MAKE-GOOD IN FBW TREATMENTS AFTER RECEIVING NEGATIVE FFEDBACK, DEPENDING ON OWN SUBMITTED FEEDBACK

Dependent	B makes good	S makes good
Baseline	B->S pos, S->B neg	B->S neg, S->B pos
B->S neg, S->B neg	-0.276***	-0.321***
	[0.100]	[0.056]
N	429	739
LL	-236.5	-470.0

Notes: The table shows the average marginal effects dy/dx. Buyer make-good is conditional on payment sent at Stage 2. *, **, and ** denote significance at the 10%, 5%, and 1% level, respectively. Regressions are based on data from rounds 11-50 (omitting start and end effects). (Robust) Standard errors are clustered at the level of independent matching groups.

dummy variable on whether the trader has given negative feedback herself. Results are presented in Table 7, and show that having submitted negative feedback herself is negatively correlated with making good when a feedback withdrawal option is present.[15]

V.2.1 Withdrawal behavior and the role of altruistic punishment

In the feedback withdrawal treatments, after observing the negative feedback and make-good choices, feedback givers had the opportunity to vote for feedback withdrawal. If both traders agreed to a withdrawal, then the feedback for both traders would be made positive. We find strong support for our key strategic prediction, that mutually negative feedback is more likely to be withdrawn and is less conditional on make-good behavior in the dispute resolution phase. In Column 6 of Table 6, we show the frequency that a buyer or seller withdraws for all treatments and feedback outcomes where withdrawal is possible. The data show that traders who have received negative feedback but have submitted positive feedback themselves almost always (97–99%) vote for a withdrawal of their negative feedback. This is consistent with our strategic model.

We also argue that if they have not received negative feedback themselves, then feedback givers will condition their withdrawal decision on the observation that the trading partner has made good, whereas when both had submitted negative feedback, then withdrawal will be unconditional, since both have a strong incentive to delete negative feedback from their own history. To test this, we run Probit regressions which estimate the likelihood to vote for feedback withdrawal depending on the other's make-good behavior, and

[15] Once again, results are similar when using the difference between Q1 and Q2 as the dependent in a Tobit model, rather than seller's likelihood to make good. (Coefficients of B->S neg, S->B neg dummy is -0.120, *p*<0.001.) Interestingly, we also observe some make-good behavior of buyers in markets without feedback withdrawal. One explanation is social preferences. These buyers seem in principle willing to cooperate, but attempt to protect themselves against exploitation by delaying their contribution (at the cost of negative feedback), and then using the make-good stage to reciprocate the seller's quality choice.

include a dummy variable for whether the other has also given negative feedback, as well as an interaction effect between this dummy variable and the make-good behavior of the trading partner. The results, shown in Table 8, are mixed. For both buyer and seller (Models 1 and 2, respectively), we find that having received negative feedback significantly increases the likelihood of voting for withdrawal. But withdrawal is also significantly correlated with make-good behavior, even when having received a negative feedback.[16]

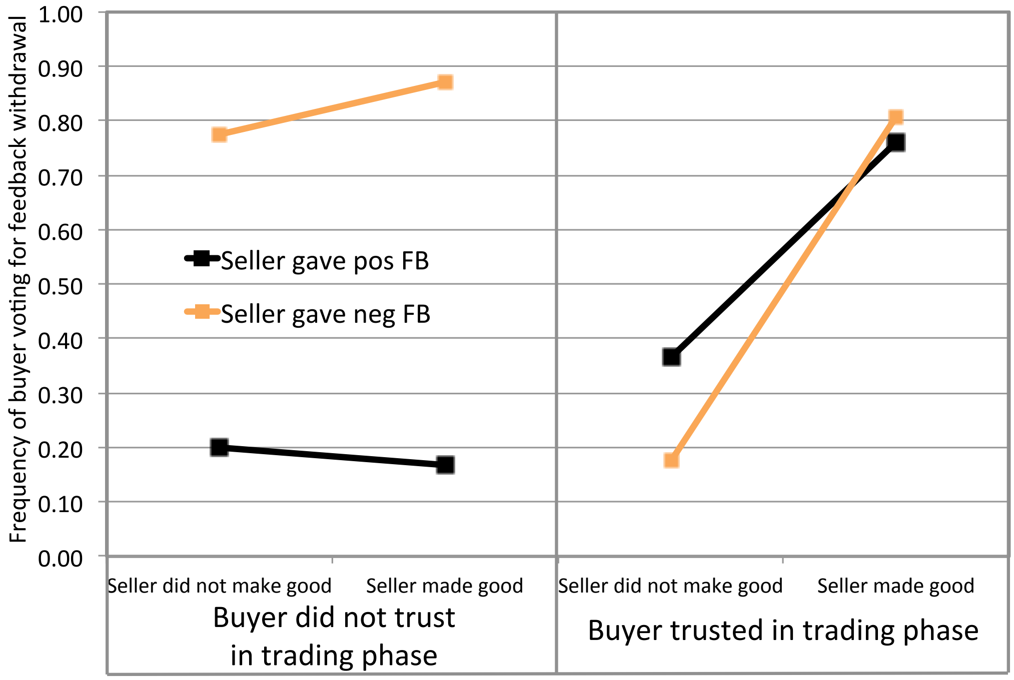
TABLE 8: PROBIT REGRESSION OF THE LIKELIHOOD TO WITHDRAW ON OTHER'S MAKE-GOOD BEHAVIOR AND FEEDBACK CONDITION

Market	2-sided				1-sided
Dependent	B withdraws y/n			S withdraws y/n	B withdraws y/n
Sample	All buyers	Non-coop B.	Coop. B.		
Model	(1)	(1a)	(1b)	(2)	(3)
Baseline	B->S neg, S->B pos	B->S neg, S->B pos	B->S neg, S->B pos	B->S pos, S->B neg	
Quality improved y/n	0.332***	-0.035	0.325***		0.528***
	[0.108]	[0.207]	[0.095]		[0.039]
Payment improved y/n				0.436***	
				[0.126]	
B->S neg, S: neg	0.266**	0.441**	-0.181	0.445***	
	[0.126]	[0.191]	[0.183]	[0.074]	
B->S neg, S->B neg × Quality improved y/n	-0.145	0.140	0.231		
	[0.124]	[0.191]	[0.152]		
B: neg, S->B neg × Payment improved y/n				-0.088	
				[0.176]	
N	739	372	367	429	737
LL	-426.0	-183.4	-200.4	218.1	-366.0

Notes: The table reports average marginal effects dy/dx. *, **, and ** denote significance at the 10%, 5%, and 1% level, respectively. Regressions are based on data from rounds 11-50 (omitting start and end effects). (Robust) Standard errors are clustered at the level of independent matching groups.

[16] The interaction effect is negative but statistically not significant, the joint effects of 'Quality improved y/n + B->S neg, S->B neg × Quality improved y/n' and 'Payment improved y/n + B->S neg, S->B neg × Payment improved y/n' are significantly different from zero (post-estimation F-tests, $p<0.01$ in both cases).

FIGURE 5: FREQUENCY OF BUYERS WITHDRAWING NEGATIVE FEEDBACK, CONDITIONAL ON HAVING TRUSTED THE SELLER IN THE TRADING PHASE, SELLER'S FEEDBACK, AND SELLER'S MAKE-GOOD



That making good has a positive effect beyond giving negative feedback is inconsistent with our strategic model (H2a), but is predicted when taking into account of altruistic punishment (H2b). Figure 5 shows the probability of buyers' willingness to withdraw their negative feedback given to the seller. Each panel shows one type of buyer behavior, and both panels together explain our regression results. The left panel shows the strategic behavior of buyers who do not pay in the trading phase of the market. Their decision to remove negative feedback is almost completely conditional on whether they received negative feedback from the seller. This is in line with our strategic analysis, which emphasizes the sellers' negotiation power in the mutual feedback withdrawal process. However, the right panel shows the opposite behavioral pattern. Buyers who trusted sellers and paid in the trading phase condition their decision to withdraw negative feedback mostly on the strategically irrelevant make-good behavior of the seller, and seem mostly immune to the pressure of having received negative feedback. This latter behavior is inconsistent with strategic behavior, yet consistent with altruistic punishment as has been observed in various studies on norm enforcement (Fehr and Gächter 2000, Balafoutas et al. 2013). Altruistic punishment in these studies is characterized as an emotional response to norm violations, that is executed even when it comes at a cost to oneself. Moreover, altruistic punishment is more likely among those who cooperated most themselves (e.g., Fehr and Gächter 2002). Similarly, in our study buyers who cooperated in the trading phase and subsequently received purely strategically motivated negative feedback are likely to be much more emotionally distressed than buyers who behaved strategically and uncooperatively in the trading phase. They are thus more likely to punish the sellers' attempt to enforce feedback withdrawal, and to insist on making-good.

Models 1a and 1b in Table 8 show results of the same regression model as in Model 1, separately applied to buyers who did pay and buyers who did not pay in the trading phase, respectively. The results statistically support the effects described above, that the withdrawal vote of non-paying buyers is only conditioned on the feedback they received themselves, while the withdrawal behavior of paying buyers is conditioned on make-good and not on the feedback received.[17]

V.2.3 Dispute resolution in one-sided markets

Our theoretical considerations suggest that in one-sided markets, sellers only make good when they received negative feedback and there is feedback withdrawal later on. Our data (also reported in Table 6) is consistent with the comparative statics of this prediction. Across the one-sided treatments, a seller is not likely to make good when there is no feedback withdrawal, or when there is feedback withdrawal and he received positive feedback (about 7% make-good likelihood in all three conditions). However, the seller is much more likely to make good after having received negative feedback in treatment 1s-FBW (72% likelihood, highly significantly different from all other conditions). In the withdrawal process of the one-sided system, the buyer could unilaterally decide to turn negative feedback from Stage 2 into positive feedback, and we predicted that this would only be the case if the seller did indeed make good. In fact, as Model 3 in Table 8 shows, the decision to withdraw negative feedback is highly correlated with the observation of make-good behavior by the seller.

Overall, in line with our model, we find that the dispute resolution phase is used to make good, but mostly so if feedback withdrawal is an option. Feedback withdrawal also leads to distorted feedback information. These effects are much more pronounced for markets with two-sided moral hazard than for markets with one-sided moral hazard.

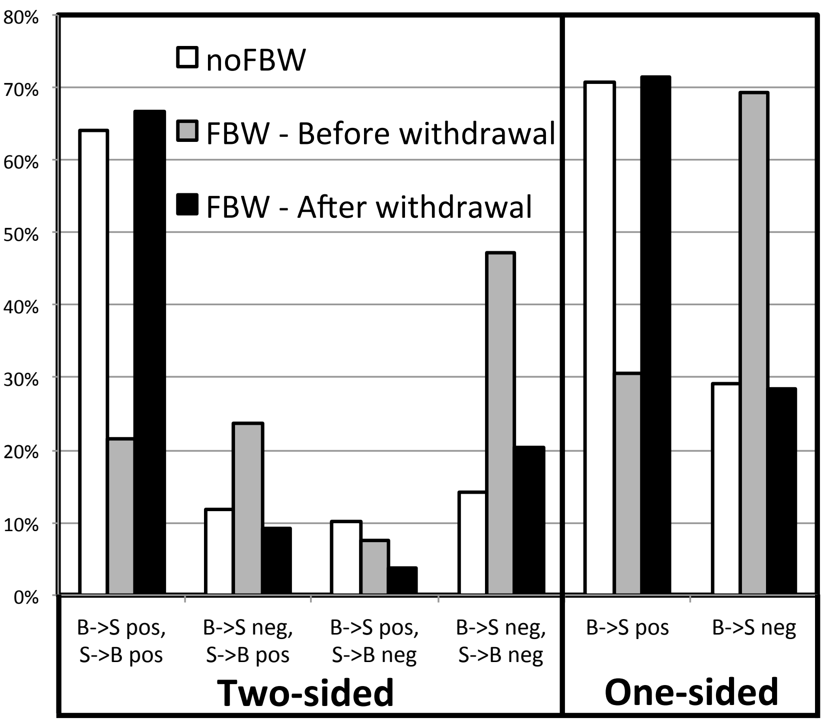
V.3. Market implications

The trading and dispute resolution phases affect the overall market performance. In particular, the effectiveness of the feedback system and its ability to provide incentives to behave in a trustworthy manner may be hampered when a feedback withdrawal option is offered, due to the resulting gaming in the feedback and withdrawal behavior. If feedback as a statistic of the past behavior of traders is less informative, then

[17] We do not find analogous evidence for the existence of different types of behavior in relation to the seller's withdrawal, no matter how we define "cooperation in the trading stage" with respect to the seller's quality level chosen in Stage 2. Part of the reason might be that the 'satisfactory' level of quality is ambiguous. Therefore, when receiving negative feedback, the seller may be unsure whether this feedback is an honest expression of dissatisfaction and expectation of make-good, or a strategic response in order to get negative feedback removed. In contrast, there is no room for such ambiguity when it comes to the question of whether the buyer has made the payment.

reputation building is hampered, which may lead to lower trustworthiness of traders, and eventually lower market efficiency. This section investigates how the feedback collected in the market place and general market performance are affected by the design of the dispute resolution phase.

FIGURE 6: FREQUENCY OF FEEDBACK OUTCOMES
(IN FBW: BEFORE AND AFTER WITHDRAWAL)



Eventual feedback distributions and informativeness. Our analysis of Stage-2 feedback as well as a comparison of white and grey bars in Figure 6 show that we observe very different distributions of *initial feedback* in markets with and without feedback withdrawal options. A comparison of white and black bars in Figure 6 shows that these differences in overall feedback distributions disappear when examining the *final feedback* collected in the market (both at the margin and across buyer/seller feedback interactions, see also the second and the last column of Table 6 in Section V.2). That is, after the withdrawal process, markets with and without feedback withdrawal options look indistinguishable when measured with respect to the distribution of collected feedback. In the following paragraphs, we show that a conclusion that similar feedback distributions imply similar underlying trading behaviors would be misleading.

The right panel of Figure 4 shows the *eventual* frequencies of positive feedback conditional on the trading partner's *eventual* behavior. In the distributions of Stage-2 feedback (left panel, discussed in Section V.1 above), we observe that in the FBW markets, feedback is *more negative for cooperators* compared to markets without FBW. Then after the make-good and withdrawal process in the FBW markets, we see much *more positive feedback for defectors*. In particular, buyers who did not pay end up with positive feedback in

50% of the cases, and sellers who provided quality of less than 40% received positive feedback in 48% of the cases, while the corresponding frequencies in the markets without feedback withdrawal are 5% and 2%, respectively. Thus, feedback in the markets with withdrawal is very strongly biased, and less informative about past behavior.

Statistical evidence is provided in the right panel of Table 5 (Models 4 and 5), which mirror the estimations of the initial feedback giving in the left panel of Table 5, only that now we use the eventual feedback at the end of a round as the dependent in each model, and the eventual payments and qualities as independents. As the right panel of Figure 4 suggests, we observe two significant effects for both buyer and seller feedback. We find a significant positive effect of the *FBW* dummy and a significant negative interaction effect *FBW* × *Other's Quality/Payment*. These two effects correspond to a positive shift in the intercept and a reduced slope of the best-fitting line through feedback in Figure 4. In other words, feedback for no payment / low quality is more positive under FBW, and the correlation between behavior and feedback is lower. We conclude that, while the overall feedback distributions look similar in markets with and without feedback withdrawal options, the feedback in FBW markets is distorted and less informative about past behavior.

Market efficiency. The total size of the payment and quality bars in Figure 3 represents *eventual* payment and quality choices in our treatments after make-good has taken place. Columns (3) and (4) in Table 4 show results from the respective regressions of eventual payment and quality choices on treatment dummies. The make-good behavior observed in the feedback withdrawal treatments reduces the differences in quality and payment between noFBW and FBW markets, which we observed in the initial trade behavior. Correspondingly, the negative coefficients for the *FBW* dummy become smaller (in payment choices) and even non-significant (in quality choices). Final market efficiencies (realized gains from trade divided by maximum gains from trade) in our treatments equal the product of the frequency of trade and the average quality. Since the frequency of trade does not differ between treatments, statistical analysis results on market efficiency almost perfectly mimic our results on the sellers' quality.

In our treatments with one-sided moral hazard, we also observe the same eventual distributions of feedback independent of whether the feedback system allows for feedback withdrawal. After the feedback withdrawal stage, the noFBW markets end up with 71% positive and 29% negative feedback, and in the FBW markets, we observe 72% positive and 29% negative feedback. Different to the two-sided markets, however, these similar distributions actually rely on a similar relation of feedback to behavior. As the lower right panel of Figure 4, and regression model (6) in the right panel of Table 5 show, there is no large difference in the correlation of the buyer's feedback and the seller's quality between the one-sided FBW and noFBW markets. Correspondingly, we do not observe any effect of the feedback withdrawal option on the final market outcomes in the one-sided markets (Model 4 in Table 4).

VI. Conclusions

The option to withdraw negative feedback is a common element in online market dispute resolution systems. Motivated by field data and guided by theoretical considerations, we use laboratory markets to show that this mechanism, while intended to incentivize cooperation, can actually interact with trading behavior and reputation building in ways that escalate dispute rather than resolve it. In particular, in markets with two-sided moral hazard, mutual feedback withdrawal creates a strong incentive to leave negative reputation information on a transaction partner for purely strategic reasons. This leads to less or delayed trust and trustworthiness, and to distorted reputation information.

Our study mainly focuses on the strategic aspects associated with feedback withdrawal dispute options. However, our field data seem to suggest that the leverage of strategic negative feedback is smaller than theoretically expected. Recall that in the field data (section II), we observed that using negative feedback as retaliation for negative feedback from the other trader is not matched by a higher probability of success in terms of getting the feedback withdrawn. We find a similar pattern in the laboratory: the negotiation power in the dispute resolution phase is not improved as expected by strategic negative feedback. In particular, the frequency of cases where both traders vote for withdrawal is similar when one trader gives negative feedback compared to when both give negative feedback (a circumstance where feedback withdrawal constitutes the rational strategy independent of make-good). In the laboratory markets, we can also observe the underlying trading and make-good behavior, which helps explaining these observations. On the one hand, we observe that make-good behavior after a receiving a negative feedback is much more prevalent when the trader did not submit a negative feedback herself, which results in relatively high withdrawal rates in this case. On the other hand, withdrawal after strategically negatively feedback is limited by the existence of altruistic punishment among buyers, who were willing to punish untrustworthy trade and aggressive feedback behaviors, even when this would damage their own reputation and thus comes at a cost. This pattern of behavior, while inconsistent with our strategic model, is in line with the fundamental research on punishing norm violations, which finds that some people are willing to altruistically punish norm violators, and that this pattern is more common among cooperators. By the same mechanism, altruistic punishment can also explain the fact that the value of aggressive feedback giving seems lower than expected in the field. One interesting question for further research is the underlying motivation for altruistic punishment; one natural candidate in our context is a strict preference for truth-telling—giving accurate feedback—as it has been observed by Gneezy (2005) and Gneezy et al. (2013) in other contexts.

In the field, there are likely other strategic and non-strategic aspects of the broader game being played that may have a role when it comes to engineering dispute resolution systems. We speculate, however that the external validity of our laboratory and theoretical studies are robust to these considerations. For instance, there is more uncertainty about the underlying behaviors that lead to specific outcomes (Ockenfels and Resnick 2012), which tend to make traders more forgiving and less punishing (see also Ambrus and Greiner

2012 for the effects of uncertainty on punishment). Yet, uncertainty of this kind would not alter the key insight from our analysis regarding the strategic value of giving negative feedback. In addition, altruistic punishers may be less likely to self-select into market platforms like eBay, and may be more likely to exit these markets after having received a purely strategically motivated negative feedback. This, however, would only strengthen our main conclusion—there is little reason to suppose that a feedback withdrawal system can significantly contribute to dispute resolution and improve the economic performance of two-sided markets. In contrast, our theory and data strongly suggest that there is good reason to believe that such elements of dispute resolution systems actually escalate dispute, hamper trade efficiency, and damage the effectiveness of reputation building systems. At the same time, our study shows that the detrimental effects are mitigated in markets with one-sided moral hazard and thus one-sided feedback withdrawal, because the one-sidedness diminishes the scope for strategic feedback giving.

Our findings also contribute to a growing literature to show that feedback information is prone to distortions that can cause inefficiency and other kinds of problems for markets (e.g., Dellarocas and Wood 2008, Bolton et al. 2013). Standard models of indirect (reputation-based) reciprocity fail to anticipate these problems because they take feedback information as generally accurate. Improving these models and our general understanding of how indirect reciprocity works requires more research into the behavior that drives feedback provision: why people give feedback (it is a public good) and how they choose a rating (Rand and Nowak 2013).

Whether and under which circumstances the feedback withdrawal option can actually improve market efficiency and reputation informativeness, and more generally, what combination of factors make an electronic dispute resolution system effective, is an open question. One might speculate that unilateral feedback withdrawal (only one market side can withdraw feedback) would be a good idea, because it would remove the incentive to strategically retaliate that we study in this paper. However, unilateral feedback withdrawal might create other gaming issues, such as traders using the option to put inappropriate pressure on the transaction partner who can give no feedback, or even blackmail this partner to extort concessions. A potentially more promising approach is to require transaction partners, who are trying to resolve their conflict, to mutually agree on a feedback revision option. Only if both agree, both partners can simultaneously decide whether to revise the initially left feedback on the opponent, or to not change the initially given feedback. This approach would combine the advantages of mutual consensus and unilateral actions: While both traders must work together in order to come to a consensus that feedbacks should be reconsidered, each trader can choose his or her feedback absent the threat of retaliation that is built in the standard mutual feedback withdrawal system. Applying our model, this creates the desired incentive to mutually resolve disputes and to honestly report feedback. We are planning to investigate such ideas in future research.

References

Ambrus, A. and B. Greiner (2012). Imperfect public monitoring with costly punishment: An experimental study. *American Economic Review*, 102(7), 3317–3332.

Ashenfelter, O., J. Currie, H.S. Farber, and M. Spiegel (1992). An experimental comparison of dispute rates in alternative arbitration systems. *Econometrica*, 60(6), 1407–1433.

Balafoutas L., N. Nikiforakis, and B. Rockenbach (2014). Direct and indirect punishment among strangers in the field, *Proceedings of the National Academy of Sciences*. 111(45), 15924–15927.

Bohnet, I., F. Greig, B. Herrmann, and R. Zeckhauser (2008). Betrayal aversion: Evidence from Brazil, China, Oman, Switzerland, Turkey, and the United States. *American Economic Review*, 98(1), 294–310.

Bohnet, I. and S. Huck (2004). Repetition and reputation: Implications for trust and trustworthiness when institutions change. *American Economic Review*, 94(2), 362–366.

Bohnet, I., and R. Zeckhauser (2004). Trust, risk and betrayal. *Journal of Economic Behavior & Organization*, 55(4), 467–484.

Bolton, G., B. Greiner, and A. Ockenfels (2013). Engineering trust: Reciprocity in the production of reputation information. *Management Science*, 59(2), 265–285.

Bolton, G. and E. Katok (1998). Reinterpreting arbitration's narcotic effect: An experimental study of learning in repeated bargaining. *Games and Economic Behavior*, 25(2), 1–33.

Bolton, G. E., A. Ockenfels, and F. Ebeling (2011). Information value and externalities in reputation building. *International Journal of Industrial Organization*, 29(1), 23–33.

Cabral L. (2012). Reputation on the Internet, in M. Peitz and J. Waldfogel (eds.), The Oxford Handbook of the Digital Economy, Oxford, UK: Oxford University Press, 343-354.

Deck, C. A. and A. Farmer (2007). Bargaining over an uncertain value: Arbitration mechanisms compared. *The Journal of Law, Economics, & Organization*, 23(3), 547–579.

Dellarocas C. (2004) Building trust on-line: The design of robust reputation mechanisms for online trading communities. Doukidis G, Mylonopoulos N, Pouloudi N, eds. Social and Economic Transformation in the Digital Era, Hershey, PA: Idea Group Publishing, 95–113.

Dellarocas, C. and C. A. Wood (2008). The sound of silence in online feedback: Estimating trading risks in the presence of reporting bias. *Management Science*, 54(3), 460–476.

Fehr, E. and S. Gächter (2000). Cooperation and punishment in public goods experiments. *American Economic Review*, 90(4), 980–994.

Fehr, E. and S. Gächter (2002). Altruistic punishment in humans. *Nature*, 415, 137–140.

Fortunato, F. (2007). Calling eBay: Negating negative feedback. *Ecommerce Guide*, http://www.ecommerce-guide.com/essentials/ebiz/article.php/3693241/Calling-eBay-Negating--Negative-Feedback.htm, posted 8 August 2007, last retrieved 15 July 2015, not online as of April 2016, available on archive.org or as file on request.

Frechette, G. R. (2012). 'Session-effects in the laboratory', *Experimental Economics*, vol. 15, pp. 485-498.

Gneezy, U. (2005). Deception: The role of consequences. *American Economic Review*, 95(1), 384–394.

Gneezy, U., B. Rockenbach, and M. Serra-Garcia (2013). Measuring lying aversion. *Journal of Economic Behavior & Organization*, 93, 293–300.

Greiner, B. (2015). Subject Pool Recruitment Procedures: Organizing Experiments with ORSEE. *Journal of the Economic Science Association*, 1(1), 114-125.

Houser, D., and J. Wooders (2005). Reputation in auctions: Theory and evidence from eBay. *Journal of Economics & Management Strategy*, 15(2), 353–369.

Hui, X., M. Saeedi, Z. Shen, and N. Sundaresan (forthcoming). Reputation and regulations: Evidence from eBay. *Management Science*.

Li, Lingfang (Ivy), and E. Xiao (2014). Money talks: Rebate mechanisms in reputation system design. *Management Science*, 60(8), 2054–2072.

Livingston, J. A. (2005). How valuable is a good reputation? A sample selection model of Internet auctions. *Review of Economics and Statistics*, 87(3), 453–465.

Lucking-Reiley, D., D. Bryan, N. Prasad, and D. Reeves (2007). Pennies from eBay: The determinants of price in online auctions. *Journal of Industrial Economics*, 55(2), 223–233.

Luca, M., and G. Zervas (forthcoming). Fake it till you make it: Reputation, competition, and Yelp review fraud. *Management Science.*

Nikiforakis, N. (2008). Punishment and counter-punishment in public good games: Can we really govern ourselves? *Journal of Public Economics*, 92, 91-112

Nikiforakis N., C. Noussair and T. Wilkening (2012). Normative conflict and feuds: The limits of self-enforcement. *Journal of Public Economics*, 96(9-10), 797-807.

Nowak, M. A. and K. Sigmund (1998). Evolution of indirect reciprocity by image scoring. *Nature*, 393(6685), 573–577.

Ockenfels, A. and P. Resnick (2012). Negotiating reputations. In: Gary Bolton, and Rachel Croson (eds), The Oxford Handbook of Economic Conflict Resolution. Oxford University Press, 223–237.

Ostrom, E., J. Walker, and R. Gardner (1992). Covenants with and without a sword: Self governance is possible." *American Political Science Review*, 86, 404-417.

Rand, D. G., and Nowak, M. A. (2013), Human cooperation, *Trends in Cognitive Sciences*, 17(8), 413 – 425.

Appendix

A. Detailed theoretical model

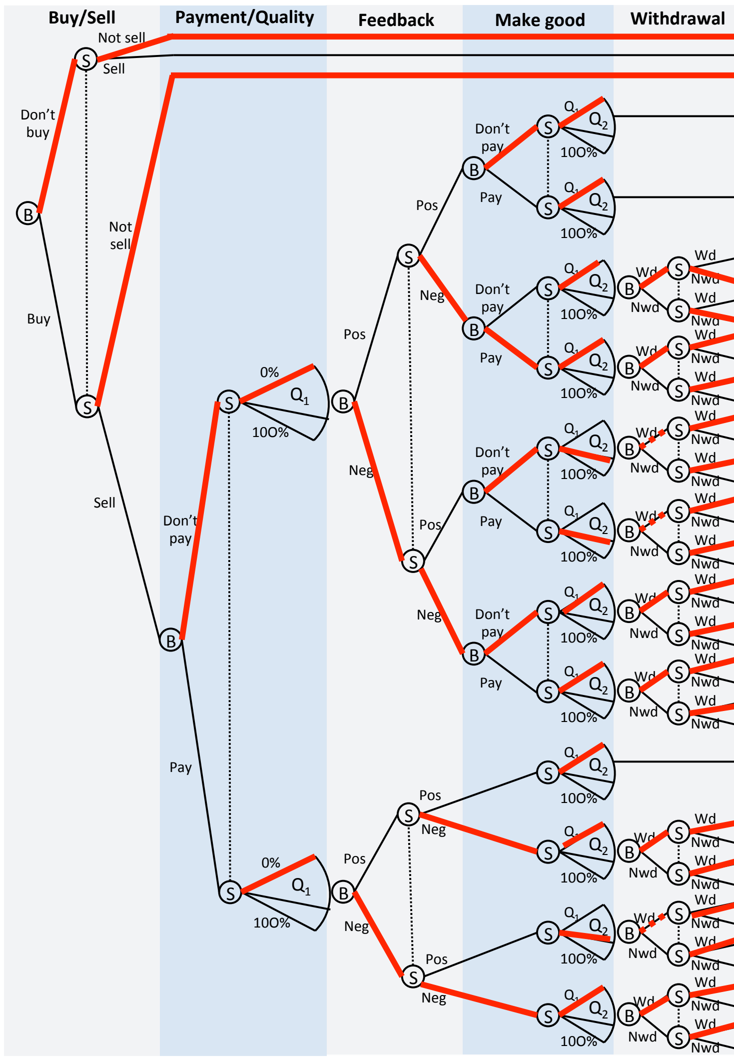
Feedback giving in markets involving interaction with strangers often only affects others' payoffs, and not one's own payoffs. This naturally leads to indifference regarding feedback giving if we assume that traders are purely selfish. For instance, a purely selfish buyer does not care whether the seller's reputation is damaged in the feedback stage, even when the seller turned out to be trustworthy and generous. However, this is implausible and inconsistent with previous laboratory and field data on feedback giving (Ockenfels and Resnick 2012), as well as with recent studies on lying aversion (Gneezy 2005, Gneezy et al. 2013). This is why we assume that traders who are monetarily indifferent prefer to submit 'honest' feedback. (An alternative approach that would yield an equivalent solution would be to assume that both the buyer and the seller receive a small benefit for honest feedback.) For the feedback from the seller to the buyer, feedback is defined as 'honest' if the buyer receives negative feedback when he does not make the payment, and if the buyer receives positive feedback if he does make the payment. For feedback from the buyer to the seller, there is a quality aspiration Q^* of the buyer, above which honest feedback is positive, and below which honest feedback is negative (see below for more on Q^*). Our assumption will make traders submit honest rather than dishonest feedback if there is no reason to decide otherwise.

Figure A.1 shows the extensive form of the stage game for treatment 2s-FBW. The game consists of 5 simultaneous stages: trade choice, payment/quality choice, feedback choice, payment/quality make-good choice, and withdrawal choice. All choices are binary, except for the quality choices, which can range from 0% to 100% in integers. Monetary payoffs at the final nodes are represented by the first terms in the buyer and seller columns in the payoff table next to the extensive form. The payoffs also include the continuation payoffs from negative feedback $-B$ and $-S$, with $B > 100$ and $S > Q^*$.

To further narrow our predictions down, we assume that in any subgame opened by a simultaneous stage in our game, players do not play weakly dominated strategies. For example, at the feedback withdrawal stage, both traders' agreement is needed to withdraw feedback. Thus, there is always an equilibrium of the subgame where both do not withdraw. Our assumption rules out this solution, because if a trader prefers the feedback to be withdrawn, not withdrawing is a weakly dominated strategy in the subgame that follows.

Second, while the exact value of the quality aspiration Q^* is not important for our prediction, we assume this to be larger than 1/3 and smaller than 1 so that trade is beneficial for both buyers and sellers. Moreover, when we compute equilibria, we will assume that Q^* is commonly known. This ensures that false beliefs cannot be a (confounding) source of cooperation failure in our model. Moreover, this will not be critical for our predictions. While, if unknown by the seller, the buyer might try to communicate his Q^* to the seller in the dispute resolution phase, this cannot affect seller behavior in equilibrium. The reason is that if such

FIGURE A.1: EXTENSIVE FORM OF THE ONE-SHOT GAME IN TREATMENT 2S-FBW, MONETARY ONE-SHOT PAYOFFS AND ASSUMED NON-MONETARY AND CONTINUATION PAYOFFS



Buyer payoff (Money, Rep)		Seller payoff (Money, Rep)	
100		100	
100		100	
100		100	
$100+3Q_2$		$100-Q_2$	
$3Q_2$		$200-Q_2$	
$100+3Q_2$		$100-Q_2$	
$100+3Q_2$	-B	$100-Q_2$	
$100+3Q_2$	-B	$100-Q_2$	
$100+3Q_2$	-B	$100-Q_2$	
$3Q_2$		$200-Q_2$	
$3Q_2$	-B	$200-Q_2$	
$3Q_2$	-B	$200-Q_2$	
$3Q_2$	-B	$200-Q_2$	
$100+3Q_2$		$100-Q_2$	
$100+3Q_2$		$100-Q_2$	-S
$100+3Q_2$		$100-Q_2$	-S
$100+3Q_2$		$100-Q_2$	-S
$3Q_2$		$200-Q_2$	
$3Q_2$		$200-Q_2$	-S
$3Q_2$		$200-Q_2$	-S
$3Q_2$		$200-Q_2$	-S
$100+3Q_2$		$100-Q_2$	
$100+3Q_2$	-B	$100-Q_2$	-S
$100+3Q_2$	-B	$100-Q_2$	-S
$100+3Q_2$	-B	$100-Q_2$	-S
$3Q_2$		$200-Q_2$	
$3Q_2$	-B	$200-Q_2$	-S
$3Q_2$	-B	$200-Q_2$	-S
$3Q_2$	-B	$200-Q_2$	-S
$3Q_2$		$200-Q_2$	
$3Q_2$		$200-Q_2$	
$3Q_2$	-B	$200-Q_2$	
$3Q_2$	-B	$200-Q_2$	
$3Q_2$	-B	$200-Q_2$	
$3Q_2$		$200-Q_2$	
$3Q_2$		$200-Q_2$	-S
$3Q_2$		$200-Q_2$	-S
$3Q_2$		$200-Q_2$	-S
$3Q_2$		$200-Q_2$	
$3Q_2$	-B	$200-Q_2$	-S
$3Q_2$	-B	$200-Q_2$	-S
$3Q_2$	-B	$200-Q_2$	-S

Notes: B and S represent the continuation costs of having received negative feedback for the buyer and the seller, respectively. The highlighted branches denote the subgame-perfect equilibrium play under the assumptions described in the main text, where a dotted highlighted branch denotes that behavior is conditional on the previous quality choice, but this branch is chosen in equilibrium.

communication would lead to more quality, all buyers would engage in this, regardless of their true aspirations. As a result, sellers will not condition their behavior on such communication attempts. That is, even with uncertainty about $Q*$ on the seller side, the seller can only rely on his own guess when making his quality choice.

Under these assumptions, for treatment 2s-FBW there is a unique subgame-perfect Nash equilibrium path starting at Stage 2 of the game described by the extensive form in Figure A.1. The red lines in the extensive form graph denote the subgame-perfect equilibrium choices at each node. A dotted line denotes that the choice is conditional on the previous quality choice but that this node is chosen in equilibrium.

Specifically, in equilibrium of the *dispute resolution phase* (CRP): In Stage 5, a trader agrees to withdraw if and only if either (i) having received negative feedback, or (ii) having given negative feedback but the other has made good or had cooperated in the first place. In Stage 4, traders make good (i.e. make the payment or choose a Q_2 equal to Q^* if $Q_1<Q^*$) when they have received negative feedback while the other has received positive feedback. The reason is that this way, in Stage 5, they will get their negative feedback withdrawn. Otherwise, i.e. when both have received negative feedback or both have received positive feedback, there is no make-good, because in case of mutually negative feedback, this feedback will be eventually withdrawn in any case.

In equilibrium of the *trading phase* (TP): In Stage 3, both traders giving negative feedback is the only equilibrium of the subgame whenever at least one trader did not cooperate. A trader gives negative feedback whenever the opponent or oneself defected (or both), since this prevents the need to make good in case the other submits negative feedback, and may incentivize the other to cooperate in case the other submits positive feedback. This is the insight in the quote from Fortunato (2007) cited in the Introduction and our key prediction regarding the detrimental shadow of the dispute resolution phase on the trading phase. Only when both cooperated, the traders will give positive feedback right away.[18] Because feedback can be gamed when a trader did not cooperate, the feedback system has no bite, and the interaction will end up as mutually positive feedback due to feedback withdrawal, regardless of whether traders cooperated at the trading phase. Thus, in Stage 2, both buyer and seller face a classical dilemma game; the buyer will not pay and the seller will not deliver satisfactory quality. Finally, in Stage 1, because all traders will cheat in equilibrium, traders are indifferent toward whether to start the trading procedure. (This prediction relies on the equilibrium expectation of zero probability of payment and zero probability of positive quality. Any epsilon deviation from equilibrium in payments or quality leads to strict incentives to enter – the same incentives that we find for one-sided feedback withdrawal.)

Now consider the game for treatment 2s-noFBW. This corresponds to the extensive form game in Figure A.1 without the feedback withdrawal stage, with the final node payoffs represented by the no-withdrawal outcomes in the payoff table. Under the same assumptions as above, the subgame perfect Nash equilibrium now looks very different. In the dispute resolution phase traders do not make good, because there is no feedback withdrawal where this could have any effect. In the trading phase, this allows both traders to give honest feedback, which in turn incentivizes them to initiate trade, make the payment, and deliver satisfactory quality.

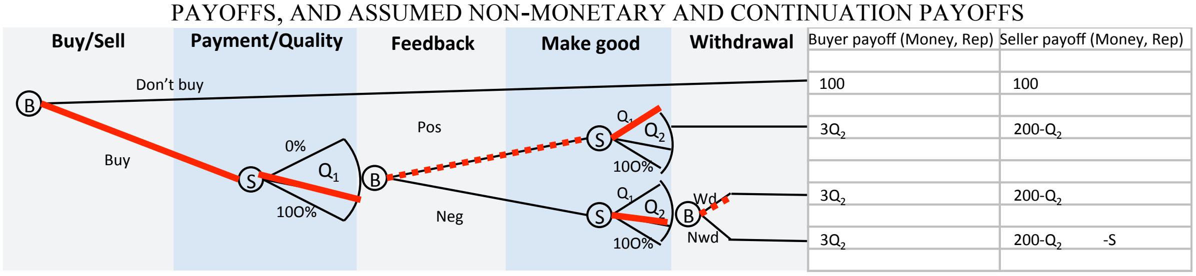
In summary, a two-sided feedback system without a withdrawal option can function effectively. However, the introduction of a withdrawal option makes it a (weakly) dominant strategy to submit negative

[18] In fact, if both traders cooperated, then traders are indifferent between giving positive feedback now and giving negative feedback with the plan to withdraw it later. Moreover, there is always an equilibrium in which both traders give negative feedback under the shadow of mutual feedback withdrawal.

feedback when not having cooperated, in order to get feedback withdrawn and prevent the need to make good. This, in turn, renders the feedback system ineffective, thus hampering trade efficiency.

The stage game played for the treatment with a one-sided feedback system and a feedback withdrawal option is shown in Figure A.2. The buyer decides whether to trade, then the seller chooses a quality, the buy-

FIGURE A.2: EXTENSIVE FORM OF THE ONE-SHOT GAME FOR TREATMENT 1S-FBW, MONETARY ONE-SHOT PAYOFFS, AND ASSUMED NON-MONETARY AND CONTINUATION PAYOFFS



Notes: S represents the continuation costs of having received negative feedback for the seller. The highlighted branches denote the subgame-perfect equilibrium play under the assumptions described in the main text, where a dotted highlighted branch denotes that behavior is conditional on the previous quality choice, but this branch is chosen in equilibrium.

er submits feedback, the seller can choose whether to improve his quality, and the buyer can decide whether to withdraw negative feedback if given. The final node payoffs including the costs are shown in the payoff table next to the extensive form. Under the same assumptions we once again obtain a subgame perfect Nash equilibrium, the strategies of which are displayed as thick red lines (dotted if conditional) in Figure A.2.

In the dispute resolution phase, in Stage 5, the buyer withdraws negative feedback when the seller has made good (or had delivered sufficient quality in the first place). In Stage 4, the seller makes good when she has received negative feedback. In the trading phase, the buyer submits negative feedback whenever the seller does not deliver sufficient quality. Thus, the seller has the choice whether to deliver sufficient quality right away ($Q_1=Q^*$) or to deliver lower quality first and then improve the quality later ($Q_1<Q^*$ and $Q_2=Q^*$). In any case, the seller will deliver sufficient quality by the end of the round, such that the buyer will always decide to buy.

Finally, consider the game for treatment 1s-noFBW, represented by the extensive form in Figure A.2 without the feedback withdrawal stage and with the final node payoff of that branch equal to the no-withdrawal outcome. It is easy to observe that in the subgame perfect Nash equilibrium, the following strategies will be employed: In Stage 4, the seller will not make good when she has received negative feedback, because this cannot change anything in the received feedback. In Stage 3, the buyer gives honest feedback. Because feedback has no impact on the seller's subsequent make-good choice, it does not affect own payoffs other than through the honesty benefit. In Stage 2, the seller will choose $Q_1=Q^*$ in order to prevent negative feedback. In Stage 1, the buyer will buy, since $1/3<Q^*<1$ implies that this will result in positive profits.

Thus, in a one-sided feedback system, the existence of a feedback withdrawal option allows for multiple equilibria (one of which features delayed seller cooperation, an issue specifically listed in eBay's reasons for abandoning feedback withdrawal even after its move to a one-sided system). However, unlike in the two-sided system, it does not hurt market efficiency, because the one-sidedness of feedback renders withdrawal strategically innocent.

B. Additional Tables and Figures

TABLE B.1: CONDITIONAL FEEDBACK GIVING, MUTUAL FEEDBACK WITHDRAWWAL (MFW) INITIATIONS, AND SUCCESS RATES OF MUTUAL FEEDBACK WITHDRAWAL PROCESS

Buyer moves first and gives	Seller responds with			Seller initiates MFW		Feedback is withdrawn		Buyer initiates MFW		Feedback is withdrawn	
Positive	pos	88.0%	(n=256599)	0.0%	(n= 2)	0.0%	(n= 0)	0.0%	(n= 15)	6.7%	(n= 1)
N=291695	neu	0.0%	(n=126)	0.8%	(n= 1)	0.0%	(n= 0)	0.8%	(n= 1)	0.0%	(n= 0)
	neg	0.1%	(n=158)	2.5%	(n= 4)	100.0%	(n= 4)	5.7%	(n= 9)	22.2%	(n= 2)
	no FB	11.9%	(n=34812)	0.0%	(n= 0)	-	(n= 0)	0.0%	(n= 9)	0.0%	(n= 0)
Neutral	pos	9.2%	(n=68)	7.9%	(n= 29)	37.9%	(n= 11)	0.3%	(n= 1)	0.0%	(n= 0)
N=3988	neu	15.4%	(n=614)	8.3%	(n= 51)	45.1%	(n= 23)	1.1%	(n= 7)	85.7%	(n= 6)
	neg	11.2%	(n=445)	**32.4%**	(n=144)	**59.7%**	(n= 86)	4.0%	(n= 18)	83.3%	(n= 15)
	no FB	64.2%	(n=2561)	**10.2%**	(n=262)	**58.8%**	(n=154)	0.7%	(n= 19)	0.0%	(n= 0)
Negative	pos	2.7%	(n=164)	20.1%	(n= 33)	54.5%	(n= 18)	7.9%	(n= 13)	76.9%	(n= 10)
N=6022	neu	0.5%	(n=31)	16.1%	(n= 5)	20.0%	(n= 1)	0.0%	(n= 0)	-	(n= 0)
	neg	46.8%	(n=2821)	**39.1%**	(n=1102)	**54.2%**	(n=597)	4.0%	(n=113)	89.4%	(n=101)
	no FB	49.9%	(n=3006)	**16.2%**	(n= 486)	**75.7%**	(n=368)	5.7%	(n=170)	0.0%	(n= 0)

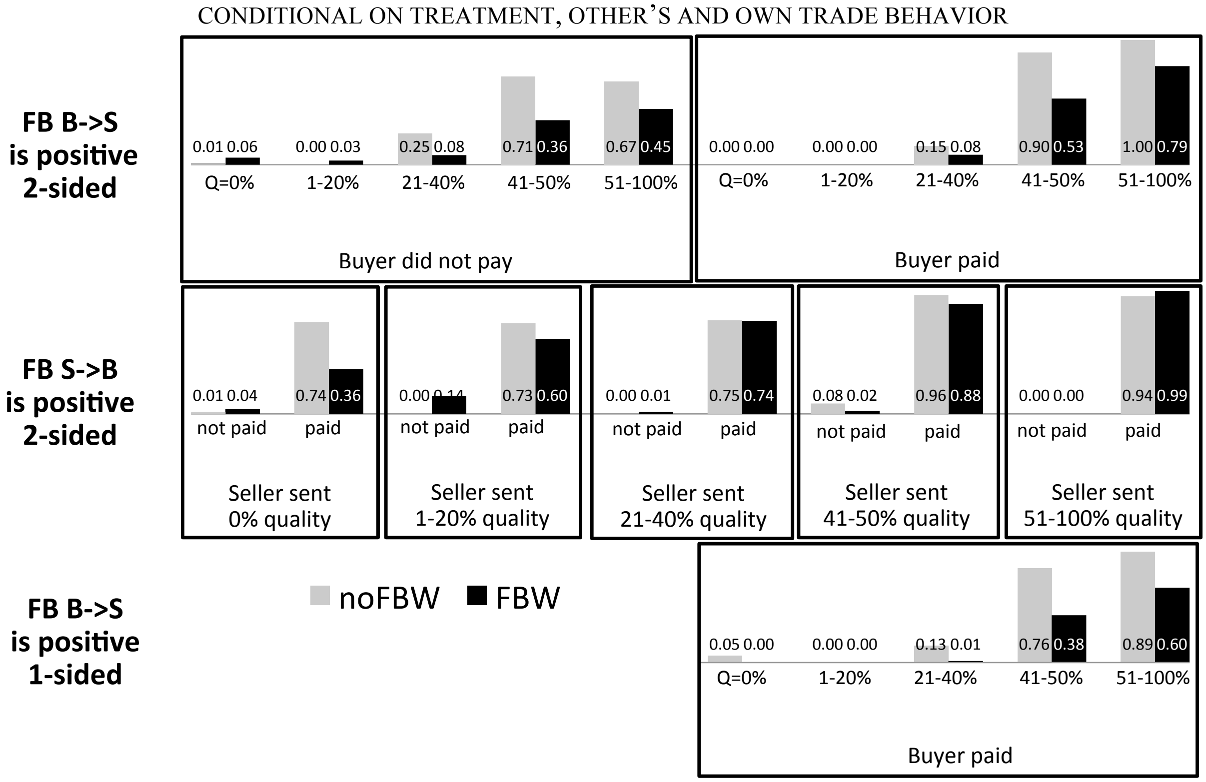
Seller moves first and gives	Buyer responds with			Buyer initiates MFW		Feedback is withdrawn		Seller initiates MFW		Feedback is withdrawn	
Positive	pos	71.9%	(n=98225)	0.0%	(n= 2)	0.0%	(n= 0)	0.0%	(n= 7)	14.3%	(n= 1)
N=136554	neu	0.7%	(n=922)	0.4%	(n= 4)	75.0%	(n= 3)	3.5%	(n= 32)	62.5%	(n= 20)
	neg	0.9%	(n=1204)	2.4%	(n= 29)	93.1%	(n= 27)	13.2%	(n=159)	64.8%	(n=103)
	no FB	26.5%	(n=36203)	0.0%	(n= 0)	-	(n= 0)	0.0%	(n= 9)	0.0%	(n= 0)
Neutral	pos	13.6%	(n=50)	0.0%	(n= 0)	-	(n= 0)	0.0%	(n= 0)	-	(n= 0)
N=368	neu	7.9%	(n=29)	6.9%	(n= 2)	0%	(n= 0)	0.0%	(n= 0)	-	(n= 0)
	neg	6.8%	(n=25)	4.0%	(n= 1)	100%	(n= 1)	20.0%	(n= 5)	80.0%	(n= 4)
	no FB	71.7%	(n=264)	0.8%	(n= 2)	100%	(n= 2)	0.4%	(n= 1)	0.0%	(n= 0)
Negative	pos	2.0%	(n=94)	4.3%	(n= 4)	50.0%	(n= 2)	6.4%	(n= 6)	50.0%	(n= 3)
N=4622	neu	0.5%	(n=22)	9.1%	(n= 2)	50.0%	(n= 1)	4.5%	(n= 1)	100.0%	(n= 1)
	neg	11.4%	(n=525)	**8.6%**	(n=45)	**66.7%**	(n=30)	13.5%	(n=71)	84.5%	(n=60)
	no FB	86.1%	(n=3981)	**0.3%**	(n=13)	**92.3%**	(n=12)	1.5%	(n=61)	0.0%	(n= 0)

TABLE B.2: RESULTS FROM PROBIT REGRESSIONS ESTIMATING DIFFERENCE BETWEEN FBW AND NOFBW TREATMENTS IN PROPENSITY TO MAKE-GOOD AFTER A CERTAIN FEEDBACK OUTCOME

Model	Feedback outcome		N	N	Probit estimates	
	B->S	S->B	noFBW	FBW	dx/dy FBW Dummy	StdErr
2-sided market, buyer make-good						
(1)	pos	pos	8	4	0.116	[0.186]
(2)	neg	pos	1	11	no variation (all 0)	
(3)	pos	neg	72	68	0.243*	[0.146]
(4)	neg	neg	119	361	0.198*	[0.104]
2-sided market, seller make-good						
(5)	pos	pos	656	224	-0.073	[0.066]
(6)	neg	pos	121	248	0.594***	[0.022]
(7)	pos	neg	103	79	0.030	[0.036]
(8)	neg	neg	145	491	0.365***	[0.087]
1-sided market, seller make-good						
(9)	pos		636	325	-0.003	[0.065]
(10)	neg		263	737	0.626***	[0.055]

Notes: The table reports marginal effects dx/dy for the Probit models with only constant and FBW dummy as independents. Estimations for buyer make-good are conditional on no payment in Stage 2, while estimations of seller make-good include all observations in the cell. Regressions are based on data from rounds 11-50 (omitting start and end effects). (Robust) Standard errors are clustered at the level of independent matching groups.

FIGURE B.1: FREQUENCY OF POSITIVE FEEDBACK, CONDITIONAL ON TREATMENT, OTHER'S AND OWN TRADE BEHAVIOR



C. Experimental Instructions[19]

Treatment dimensions are: 2-sided/1-sided and WD/nWD

Instructions

Welcome and thank you for participating in this experiment. In this experiment you can earn money. The specific amount depends on your decisions and the decisions of other participants. From now on until the end of the experiment, please do not communicate with other participants. If you have any questions, please raise your hand. An experimenter will come to your place and answer your question privately. In the experiment we use ECU (Experimental Currency Unit) as the monetary unit. At the end of the experiment your income will be converted from ECUs into Euros according to the conversion rate of 400 ECUs = 1 Euro, and paid out in cash jointly with your show-up fee of 2.50 Euros.

At the beginning of the experiment, you will be randomly assigned to the role of a buyer or a seller. You will keep your role throughout the experiment. The experiment consists of 60 rounds. In each round the computer will randomly match pairs of one buyer and one seller. Additionally the computer will make sure that you are never matched with the same other participant twice in a row. At the beginning of the round, both the buyer and the seller are endowed with an amount of 100 ECU. Each round consists of **[WD:** 5**]** **[nWD:** 4**]** stages:

1. **Trade [2-sided: decision:] [1-sided: decisions:] [2-sided:** Simultaneously, the buyer and the seller decide whether they want to trade with each other. If one of them or both don't want to trade, then the round ends at this stage, and the round income of buyer and seller equals their endowment.**] [1-sided:** The buyer decides whether s/he wants to trade with the seller. If the buyer doesn't want to trade, then the round ends at this stage, and the round income of buyer and seller equals their endowment.**]**
2. **[2-sided: Money transfer and quality decision**: The buyer decides to send his/her 100 ECU to the seller or not.**] [1-sided: Quality decision**: The 100 ECU of the buyer are automatically transferred to the seller.**] [2-sided:** At the same time, the**] [1-sided:** The**]** seller chooses the quality of the product which s/he is sending to the buyer. The quality must be between 0% and 100%. Each quality percent costs the seller 1 ECU, and benefits the buyer by 3 ECU. So, for example,
    - if the quality is 0%, the seller has costs of 0 ECU and the buyer receives a product value of 0 ECU;
    - if the quality is 50%, the seller has costs of 50 ECU and the buyer receives a product value of 150 ECU;
    - and if the quality is 100%, the seller has costs of 100 ECU and the buyer receives a product value of 300 ECU.

    [2-sided: Once the buyer and seller made their decisions, both transaction partners are informed about each other's choice. **] [1-sided:** Once the seller has made his/her choice the buyer will be informed about the choice.**]**
3. **Feedback: [2-sided:** Simultaneously, the buyer and the seller decide which feedback they want to submit on the transaction.**] [1-sided:** The buyer decides which feedback s/he wants to submit on the transaction.**]** The feedback can be either 'negative', or 'positive'. **[2-sided:** After both have given feedback,**] [1-sided:** After the buyer has given feedback, **]** it will be shown on the screen to both transaction partners. The received feedback will also be displayed to **[2-sided:** transaction partners**] [1-sided:** buyers**]** in subsequent rounds (see below).
4. **[2-sided: Money transfer/quality revision**: If the buyer did not send the 100 ECU in Stage 2, then s/he now receives the opportunity to revise this decision, and can once again decide to send the 100 ECU to the seller. Simultaneously, the seller has the opportunity to revise his/her quality decision in Stage 2. The revised quality has to be between the quality chosen in Stage 2 and 100%. Once both

[19] There was one typo in one of the examples in the original instructions, which was corrected verbally in the sessions.

have made their revision decisions, they are informed about each other's choices.] **[1-sided:** **Quality revision**: The seller has the opportunity to revise his/her quality decision in Stage 2. The revised quality has to be between the quality chosen in Stage 2 and 100%. Once the seller has made his/her revision decision, the buyer is informed about that choice.**]**

5. **[WD:** **Feedback revision**: **[2-sided:** This stage is only entered if at least one of the feedback ratings given in Stage 3 was negative. Simultaneously, both the buyer and the seller can decide whether they support to revise the feedback and turn both feedback ratings into 'positive' feedback. If both support the revision, then both feedback ratings will be made 'positive'. If only the buyer or only the seller or none supports the feedback revision, then the feedback given in Stage 3 remain unchanged.**]** **[1-sided:** This stage is only entered if the feedback given by the buyer in Stage 3 was negative. The buyer can decide whether s/he wants to revise the feedback given and turn the feedback rating into a 'positive' feedback. If the buyer supports the revision, then the feedback rating will be made 'positive'. If the buyer does not support the feedback revision, then the feedback given in Stage 3 remains unchanged.**]** **]**

After these **[WD:** 5**]** **[nWD:** 4**]** stages the round ends. In the next round, you will be randomly matched to a new other buyer or seller, respectively.

At the end of the round, both buyer and seller are informed about all the choices they made and their respective round payoffs and feedback.

The round payoff of a buyer is

100 ECU

{ if **[2-sided:** both**]** **[1-sided:** the buyer **]** decided to trade:

– 100 ECU **[2-sided:** if s/he decided to send the 100 ECU to the seller**]**

+ 3 * Q with Q equaling the quality percent the seller has chosen for the product, being between 0 and 100

}

The round payoff of a seller is

100 ECU

{ if **[2-sided:** both**]** **[1-sided:** the buyer **]** decided to trade:

+ 100 ECU **[2-sided:** if the buyer decided to send the 100 ECU to the seller**]**

- Q with Q equaling the quality percent the s/he has chosen for the product, being between 0 and 100

}

Your final payoff from the experiment will be the sum of all round payoffs.

The number of feedback ratings a **[2-sided:** participant**]** **[1-sided:** the seller **]** collected in previous rounds will be shown to **[2-sided:** his transaction partner**]** **[1-sided:** the buyer **]** at the beginning of the next round, before Stage 1. The display will show the number of positive and negative feedback ratings received in previous rounds, like this: "X positive feedback ratings and Y negative feedback ratings received in previous rounds".

Supplementary Appendix

S.1. Reproduction of analyses in Tables 4, 5, 6, 7, 8, and B.2 for all rounds 1-60

TABLE S.4: REGRESSIONS OF PROBABILITY OF PAYMENT, QUALITY, AND EFFICIENCY ON TREATMENT DUMMIES

Model	(1)	(2)	(3)	(4)
Model type	Probit	Tobit	Probit	Tobit
Dependent	Payment Stage 2	Quality Stage 2	Payment after MG	Quality after MG
Constant		0.515***		0.516***
		[0.029]		[0.022]
Round	-0.008***	-0.004***	-0.009***	-0.004***
	[0.001]	[0.001]	[0.001]	[0.001]
FBW	-0.166*	-0.096*	-0.010	-0.040
	[0.096]	[0.049]	[0.079]	[0.046]
One-sided		0.054*		0.056*
		[0.030]		[0.030]
One-sided × FBW		0.053		0.047
		[0.057]		[0.053]
N	3064	5853	3064	5853
LL	-1763.8	-1417.3	-1569.1	-1275.8
Censoring Left		1042		933 (4847)
(Non) Right		(4746) 65		73

Notes: The table reports average marginal effects dy/dx for the Probit models. Quality and efficiency are censored at 0 and 1. *, **, and ** denote significance at the 10%, 5%, and 1%-level, respectively. Regressions are based on data from rounds 1-60. (Robust) Standard errors are clustered at the level of independent matching groups.

TABLE S.5: PROBIT REGRESSIONS OF INITIAL AND EVENTUAL FEEDBACK ON OTHER'S TRADE BEHAVIOR AND EXISTENCE OF FEEDBACK WITHDRAWAL OPTION

	Feedback after transaction			Feedback after withdrawal		
Model	(1)	(2)	(3)	(4)	(5)	(6)
Dependent	B->S FB is pos	S->B FB is pos	B->S FB is pos	B->S FB is pos	S->B FB is pos	B->S FB is pos
Condition	2-sided	2-sided	1-sided	2-sided	2-sided	1-sided
Other's Quality /	1.704***	0.499***	1.276***	1.748***	0.775***	1.101***
Payment	[0.169]	[0.029]	[0.164]	[0.160]	[0.018]	[0.184]
FBW	0.176*	-0.041	-0.645**	0.433***	0.317***	-0.200**
	[0.096]	[0.031]	[0.267]	[0.030]	[0.038]	[0.092]
FBW × Other's	-0.836***	-0.077**	0.763	-1.294***	-0.313***	0.420
Quality / Payment	[0.195]	[0.037]	[0.559]	[0.190]	[0.042]	[0.262]
N	3064	3064	2789	3064	3064	2789
LL	-1167.7	-944.8	-1296.0	-1354.6	-1190.2	-1247.8

Notes: The table reports average marginal effects dy/dx. *, **, and ** denote significance at the 10%, 5%, and 1%-level, respectively. Regressions are based on data from rounds 1-60. (Robust) Standard errors are clustered at the level of independent matching groups.

TABLE S.6: AGGREGATE FEEDBACK, MAKE-GOOD AND WITHDRAWAL BEHAVIOR IN THE FOUR TREATMENTS

Treatment & Given Feedback	FB Freq.	P & Q before make-good	P & Q after make-good	Freq. of make-good	Withdrawal	Eventual FB Freq.
2-sided no FBW						
B->S pos, S->B pos	58%	P: 0.98 Q: 0.51	P: 0.98 Q: 0.51	P: 18% Q: 8%	-	58%
B->S neg, S->B pos	14%	P: 0.95 Q: 0.27	P: 0.95 Q: 0.28	P: 0% Q: 9%	-	14%
B->S pos, S->B neg	10%	P: 0.22 Q: 0.49	P: 0.43 Q: 0.49	P: 27% Q: 7%	-	10%
B->S neg, S->B neg	18%	P: 0.14 Q: 0.12	P: 0.18 Q: 0.12	P: 5% Q: 4%	-	18%
2-sided FBW						
B->S pos, S->B pos	25%	P: 0.97 Q: 0.57	P: 0.98 Q: 0.57	P: 40% Q: 2%	-	66%
B->S neg, S->B pos	24%	P: 0.95 Q: 0.35	P: 0.96 Q: 0.44	P: 27% Q: 68%	B: 59% S: 99% Both: 59%	10%
B->S pos, S->B neg	8%	P: 0.12 Q: 0.42	P: 0.53 Q: 0.43	P: 46% Q:10%	B: 98% S: 47% Both: 46%	4%
B->S neg, S->B neg	44%	P: 0.25 Q: 0.17	P: 0.39 Q: 0.22	P: 18% Q: 30%	B: 71% S: 81% Both: 55%	20%
1-sided no FBW						
B->S pos	69%	Q: 0.53	Q: 0.53	Q: 7%	-	69%
B->S neg	31%	Q: 0.28	Q: 0.29	Q: 10%	-	31%
1-sided FBW						
B->S pos	31%	Q: 0.52	Q: 0.53	Q: 8%	-	69%
B->S neg	69%	Q: 0.36	Q: 0.44	Q: 70%	B: 55%	31%

Notes: P and Q stand for 'frequency of payment' and 'average quality', respectively; FB denotes 'feedback', Freq. means 'frequency', and B and S refer to buyer and seller, respectively. Column 2 shows frequency of buyer/seller feedback outcomes in a treatment (Stage 3), column 3 shows the frequency of payment and average quality underlying this feedback outcome. Column 4 gives the same numbers after make-good (Stage 4) has taken place, and column 5 lists the frequencies of non-zero make-good behavior for payment (conditional on no payment sent in Stage 2) and quality. Column 6 includes information on how often buyer, seller, and both of them agreed to withdraw feedback, while column 7 shows the frequency of feedback outcomes after withdrawal has taken place (Stage 5). All aggregates are based on data from rounds 1-60.

TABLE S.7: PROBIT REGRESSIONS OF LIKELIHOOD TO MAKE-GOOD IN FBW TREATMENTS AFTER RECEIVING NEGATIVE FFEDBACK, DEPENDING ON OWN SUBMITTED FEEDBACK

Dependent	B makes good	S makes good
Baseline	B->S pos, S->B neg	B->S neg, S->B pos
B->S neg, S->B neg	-0.237***	-0.349***
	[0.840]	[0.053]
N	611	1042
LL	-309.3	-641.5

Notes: The table reports average marginal effects dy/dx. Buyer make-good is conditional on payment sent in Stage 2. *, **, and ** denote significance at the 10%, 5%, and 1%-level, respectively. Regressions are based on data from rounds 1-60. (Robust) Standard errors are clustered at the level of independent matching groups.

TABLE S.8: PROBIT REGRESSION OF LIKELIHOOD TO WITHDRAW ON OTHER'S MAKE-GOOD BEHAVIOR AND FEEDBACK CONDITION

Market	2-sided				1-sided
Dependent	B withdraws y/n			S withdraws y/n	B withdraws y/n
Sample	All buyers	Non-coop B.	Coop. B.		
Model	(1)	(1a)	(1b)	(2)	(3)
Baseline	B->S neg, S->B pos	B->S neg, S->B pos	B->S neg, S->B pos	B->S pos, S->B neg	
Quality improved y/n	0.419***	0.336*	0.393***		0.517***
	[0.104]	[0.203]	[0.081]		[0.035]
Payment improved y/n				0.462***	
				[0.108]	
B->S neg, S->B neg	0.338***	0.484**	-0.112	0.442***	
	[0.098]	[0.196]	[0.162]	[0.030]	
B->S neg, S->B neg × Quality improved y/n	-0.242**	-0.216	0.132		
	[0.112]	[0.211]	[0.142]		
B->S neg, S->B neg × Payment improved y/n				-0.097	
				[0.103]	
N	1042	521	521	611	1037
LL	-604.2	-265.2	-283.03	-314.9	-537.8

Notes: The table reports average marginal effects dy/dx. *, **, and ** denote significance at the 10%, 5%, and 1%-level, respectively. Regressions are based on data from rounds 1-60. (Robust) Standard errors are clustered at the level of independent matching groups.

TABLE S.B.2: RESULTS FROM PROBIT REGRESSIONS ESTIMATING DIFFERENCE BETWEEN FBW AND NOFBW TREATMENTS IN PROPENSITY TO MAKE-GOOD AFTER A CERTAIN FEEDBACK OUTCOME

Model	Feedback outcome		N	N	Probit estimates	
	B->S	S->B	noFBW	FBW	dx/dy FBW Dummy	StdErr
2-sided market, buyer make-good						
(1)	pos	pos	17	10	0.207	[0.141]
(2)	neg	pos	10	18	1.110**	[0.528]
(3)	pos	neg	118	108	0.187	[0.123]
(4)	neg	neg	232	503	0.155***	[0.058]
2-sided market, seller make-good						
(5)	pos	pos	888	380	-0.072	[0.054]
(6)	neg	pos	208	367	0.522***	[0.020]
(7)	pos	neg	152	123	0.025	[0.053]
(8)	neg	neg	271	675	0.340***	[0.066]
1-sided market, seller make-good						
(9)	pos		884	473	0.013	[0.056]
(10)	neg		395	1037	0.546***	[0.039]

Notes: The table reports marginal effects dx/dy for the Probit models with only constant and FBW dummy as independents. Estimations for buyer make-good are conditional on no payment in Stage 2, while estimations of seller make-good include all observations in the cell. Regressions are based on data from rounds 1-60. Robust) Standard errors are clustered at the level of independent matching groups.