UNSW Business School

Working Paper

Never Stand Still Business School Economics

UNSW Business School Research Paper No. 2015 ECON 19

Conflict resolution vs. conflict escalation in online markets

Gary Bolton
Ben Greiner
Axel Ockenfels

CONFLICT RESOLUTION VS. CONFLICT ESCALATION IN ONLINE MARKETS

GARY BOLTON, BEN GREINER, AND AXEL OCKENFELS

15 Juli 2015

Abstract. Many online markets encourage traders to make good after an unsatisfactory transaction by offering the opportunity of withdrawing negative reputational feedback in a conflict 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 conflict in the form of strategically distorted reputation information and less trust and trustworthiness in the trading phase. The detrimental impact is mitigated by buyers who refuse to give feedback strategically, even when it comes at a cost to themselves. It is also mitigated in markets with one-sided moral hazard.

Keywords: conflict 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 New South Wales, University of Sydney, University of Technology Sydney, 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 in the conduction of 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-)designing certain aspects of their auction rules and feedback forum, partly in order to mitigate various forms of social conflict; this study is 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. University of New South Wales, School of Economics, Sydney, NSW 2052, Australia, Tel: +61 2 9385 9701, e-mail: bgreiner at unsw.edu.au.
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

Asynchronous and geographically dispersed interaction in online markets can make it hard to resolve conflicts that come up when negotiating the (non-price) terms of trade. Thus, almost all large online markets complement their trading platform with conflict resolution systems that attempt to incentivize and facilitate cooperation. One particularly common element of these systems is the possibility to withdraw negative 'feedback' (the name for reputation ratings posted on online markets) after a make-good attempt and once a conflict has been resolved. The goal is to increase market efficiency through more mutually satisfactory trades. Indeed, a major incentive for making good on a problematic trade is maintaining a good reputation, not only with an otherwise disappointed trading partner, but also with prospective future trading partners who might hear about it.

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, only 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 surveys the use of feedback withdrawal mechanisms for some of the better-known online markets. All of these markets have a feedback system in place, and some of them allow feedback withdrawals (FBW) as a conflict resolution facilitator. Probably depending on the scope for moral hazard and adverse selection on each market side, each market implements a one-sided feedback system (here, usually the buyer can rate the seller) or two-sided feedback system (buyer and seller rate one another).[1] Also

[1] While seller moral hazard is a known and accepted issue on two-sided online platforms, the existence and extent of buyer moral hazard are debatable, given that in most cases buyers pay first. In 2008, eBay changed from a two-sided to what can be interpreted as being essentially a one-sided feedback system, possibly because they judged buyer moral hazard to be a relatively minor problem. In this paper, we study both, two-sided and one-sided systems, and our two- (one-) sided systems include two- (one-) sided moral hazard. Quantitative evidence for buyer moral hazard is rare. Anecdotal evidence from surveys, eBay's discussion forums, and eBay seller conferences can be classified into four categories: (1) The buyer purchases the item, but doesn't pay. (2) The buyer raises unsubstantiated complaints about the item. (3) The buyer blackmails the seller for over-fulfillment by threatening negative feedback. (4) After some time, the buyer requests the credit card provider to retract the payment (PayPal does not provide support in these cases).

observe from Table 1 that some sites allow feedback to be withdrawn unilaterally while others require mutual consent.

In this study we show that such feedback withdrawal systems can create opportunities for strategic gaming that, contrary to the intended purpose, escalate conflict and diminish trust and trustworthiness in the trading phase. While there is presently little in the literature examining online conflict resolution mechanisms, it is well established that binding arbitration, a dispute resolution mechanism in which a third party imposes a settlement, can invite unintended strategic behavior. Binding arbitration saves the negotiating parties the costs of impasse (the intended purpose), yet expectations of what the arbitor will decide can change negotiation strategies and, in turn, both the nature and probability of a 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, owing to the fact that the party accused of doing harm typically has little incentive to agree to it after an impasse is reached (Shavelle 1995). The online dispute mechanisms we study here also aim to reduce impasse costs. But they take a softer, incentive based approach to encouraging participation. There is nevertheless reason to suspect that these systems, too, invite strategic behavior, especially in the case of two-way systems. The underlying argument is evident from advice Frank Fortunato (2007) gave to ecommerce-guide.com readers:

> "'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 strategic feedback retaliation is successful at getting the other trader to remove their negative feedback even when the trader did not make good, then this distorts feedback in a way that hurts future traders ability 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 possibility when moving to a one-sided system. EBay writes:[2]

> "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] http://pages.ebay.com.au/help/feedback-changes.html, last accessed July 22, 2014. The "mutual feedback withdrawal" procedure on eBay was introduced in 2004 and dropped in 2008. Little later, eBay introduced a withdrawal policy allowing sellers to submit "feedback revision requests" to buyers upon which buyers could change their feedback, limited to at most five for every 1,000 feedbacks received by the seller.

In the next section, we review field evidence from eBay before the mutual feedback withdrawal option was removed. The data suggest that feedback giving is in fact manipulated as suggested above. 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), we hypothesize that the reduced leverage is due to some buyers' unwillingness to accept a seller's aggressive trade and feedback behavior, and who 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 investigating our hypotheses in more detail, because it does not allow us to relate conflict resolution behavior to either initial trading or the make-good phase. There are also a number of endogeneity issues associated with judging the field data, which we describe later. Thus, to further examine the influence of online feedback withdrawal on trader actions and market performance, we constructed a model that we then examine both theoretically and behaviorally. The theoretical analysis reflects the strategic hypotheses. Our laboratory observations complement theory and 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, contrary to the intended purpose, produces an escalation of conflict: It invites gaming, leading to distorted reputation information and hampering 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 our model's prediction that the gaming incentives due to a feedback withdrawal option largely vanish in markets with one-sided moral hazard.

II. Field data

To get a closer look at how feedback withdrawal mechanisms influence feedback giving, we turn to field data from eBay. The dataset includes 573,567 transactions in June 2007 over 7 countries and 6 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 has led to a withdrawal of the respective feedback.

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

TABLE 2: FEEDBACK AND WITHDRAWAL FREQUENCIES IN EBAY FIELD DATASET

Feedback given by		Frequency of feedback given	Of this: Withdrawn	and (yet) unresolved	Frequency of eventual feedback
buyer to 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 to 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 as given (before any withdrawal) and four months after the transaction (after withdrawals until this time), respectively. Column 2 displays the percentage of feedback in this category that had been withdrawn (for some reason) until the time of data retrieval, while Column 3 displays the percentage of feedback in this category for which mutual feedback withdrawal was initiated but has not found a resolution until time of data retrieval. About 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]

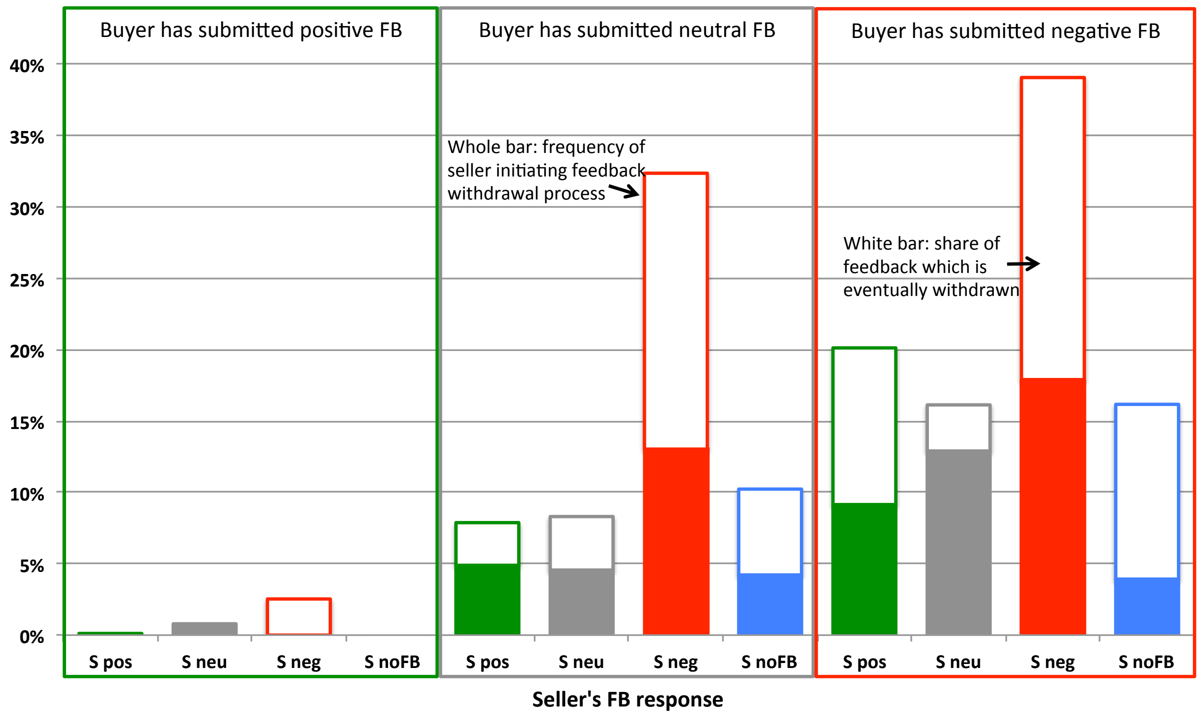
Table 2 displays the distribution of initially submitted feedback on eBay, how much of it is withdrawn, and what the eventual distribution of feedback looks like. About 97% of all submitted feedback is positive (see Bolton et al. 2013 and Dellarocas and Wood 2008 for a discussion of the high frequency of positive feedback). About 16% of the negative feedback submitted by buyers is eventually withdrawn; for another 10% withdrawal was initiated but the process 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% positives), the eventual feedback distribution after withdrawals is even more so (97.7%). As a consequence, it is the most valuable information (the rare information about misbehavior) that is potentially lost by the gaming of the mutual feedback withdrawal system.

Figure 1 displays 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 MFW process initiations by the second mover when the latter has responded with a negative feedback to an observed non-positive feedback, rather than responding with no feedback at all: After having retaliated a negative (neutral) feedback with a negative feedback, the seller challenges the feedback in 39% (32%) of the cases, whereas

[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": a seller might contact eBay and complain that the buyer has not paid an item. If the buyer does not respond to this accusation within a certain timeframe, then eBay will delete any feedback the buyer may have left. A second rule is that eBay deletes feedback of 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 MFW procedure, such that the data are not able to reveal how much feedback would have been withdrawn if MFW 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 got resolved, the buyer is excluded based on her negative feedback score.

only 16.2% (10.2%) of feedbacks are challenged when the seller did not respond with feedback; Fischer 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 represent the success rates of MFW 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 strategically negative feedback leads to more negotiation power in the feedback withdrawal process, we find that the success rate after retaliating with a negative is either the same or even less than not responding with a feedback (Fischer Exact tests, p=0.916 and p<0.0001 for buyer neutral first and buyer negative first, respectively). Thus our data fail to verify the second half of Fortunato's hypothesis that retaliating a negative feedback in-kind increases the chances of getting rid of it.

There are several hypotheses for why, in the field, strategically giving negative feedback does not pay in terms of one's own reputation. One hypothesis is self-selection. For instance, sellers who retaliate a negative in the transaction phase are traders with below average communication abilities or, more generally, below average social skills, which would make it more likely that negotiations in both the transaction and the conflict resolution phase fail. An alternative hypothesis, the one we will focus on in our laboratory study, is

[5] Since we observe very little negative feedback in general, the total number of observations of initiated feedback withdrawal processes in this category becomes quite small, in particular for buyers, even though we started with a large dataset of more than half a million transaction observations. Table B.1 in Appendix B includes all observation counts.

that leaving a negative feedback on the seller reflects the buyer's impression that she has been treated unfairly in the transaction phase. (Less emotional and more rational traders may not submit negative feedback in the first place due to anticipation of a retaliatory feedback; see Bolton et al. 2013 and Ockenfels and Resnick 2012 for evidence along these lines.) But for these already annoyed buyers, receiving an unjustified (strategically) negative feedback plus a request to withdraw one's own (truthfully) negative feedback may activate additional negative emotional responses. This can diminish their willingness to concede and increases their willingness to punish the seller, even when it comes at a cost.

Due to limitations in the available field data, we cannot separate the different mechanisms that may be at work in the field. In particular, we are unable to investigate the role of behavior during conflict 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 conflict 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, too, cannot be investigated with our field data, which is confined to a two-sided system. Thus, in order to investigate the interplay of the conflict resolution phase and the feedback system, we run 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. This way, 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 conflict resolution systems under the two archetype 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 benefits and costs of feedback withdrawal systems as being 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 buyer and seller are asked whether they want to trade or not. If both agree to trade, 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 they do not want to interact with.

Next, buyer and seller engage in a transaction. There is potential moral hazard on both sides of the market. The buyer decides whether or not to pay his endowment of 100 ECU to the seller. At the same time, the seller decides on the quality of the good 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.

After the transaction, both the buyer and seller are informed about each others' 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

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. Trade	Simultaneously buyer and seller decide to trade or not. If one of them decides not to trade, then round ends with round payoffs π_B=100, π_S=100.	Buyer decides to trade or not. If buyer decides not to trade, then round ends with round payoffs π_B=100, π_S=100.
2. Payment/Quality	Buyer decides to pay 100 ECU or not. 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	Both buyer and seller decide simultaneously whether they give positive or negative feedback.	Buyer decides whether he gives positive or negative feedback.
Conflict resolution phase		
4. Make-good	If buyer has not paid the price in Stage 2, then he can now decide again to pay the price or not. 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. Withdrawal *(only in 2s-FBW and 1s-FBW)*	If there was at least one negative feedback in Stage 3, both buyer and seller decide whether to vote for feedback withdrawal. If both vote for withdrawal, both feedbacks 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 after trade	π_B= 100 – PricePaid + Q_2 * 3 π_S= 100 + PricePaid – Q_2	π_B= Q_2 * 3 π_S= 200 – Q_2

opportunity to "make good", i.e. to improve upon their transaction choice. In particular, if the buyer has not paid in the transaction stage, he may do so now, while the seller may increase the quality (but not decrease it). Subsequently, both parties are informed about the make-good behavior of their counterpart.

Finally, only in treatment 2s-FBW and only if at least one of the feedbacks was negative, both seller and buyer simultaneously vote for or against a withdrawal of feedback. If both traders agree to withdraw, then both feedbacks would be made positive. Otherwise feedbacks stayed as they were given.

In the treatments with a one-sided feedback system, only the buyer has to agree to trade or not. If he agrees, the 100 ECU are automatically sent to the seller (so there is no scope for buyer moral hazard), and the seller makes a quality choice. 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, while in treatment 1s-FBW the buyer would be informed about the make-good choice of the seller and may withdraw her negative feedback, so that it is turned into a positive feedback, or not.

Table 3 summarizes all four (five) stages of the buyer-seller interaction and the differences between treatments. The experiment was conducted at the Laboratory for Experimental Research at the University of Cologne. We conducted 8 sessions (2 sessions per treatment) with altogether 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 started. The market was repeated over 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) statistically independent matching groups per treatment. Sessions lasted about two hours. Participants were paid out in cash at the end of the experiment. The average payoff was EUR 20.03 (StdDev 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 last 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 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. This said, there are two reasons to be dissatisfied with such 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.[6] Yet this prediction, too, is clearly at odds with both experiment 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, Milgrom, Roberts and Wilson 1982) can lead to more plausible behavior. 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 is in Appendix A; here we provide an overview.

With regard to the assumptions, we assume that receiving a negative feedback imposes a reduction in continuation profits that outweighs profits from shirking in the single-stage game. We also assume that traders, when indifferent monetary-wise, submit "honest" feedback depending on whether they received satisfactory quality; otherwise, they submit feedback strategically.[7] Finally, we suppose that buyers share a minimum "satisfactory quality" level, one that makes trade profitable (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 phase (1s-noFBW and 2s-noFBW), traders do not make good in the conflict resolution phase, because there is no withdrawal stage where this could have any effect. In the trading phase, both traders will give honest feedback. This in turn incentivizes them to initiate trade, pay the price and deliver satisfactory quality. Hence both two-sided and one-sided markets with no feedback withdrawal option should operate efficiently.

Turning to the SPNE analysis for the market with a one-sided feedback system and a feedback withdrawal option: In the conflict resolution phase, the buyer withdraws a negative feedback when the seller delivers satisfactory quality. The seller therefore makes good if she has received a 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 improve quality to satisfactory later. In any case, the seller will deliver sufficient quality at the end of the round, such that the buyer will always decide to buy. The prediction then is that allowing feedback withdrawal in the conflict

[6] 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 with stranger interaction is often characterized as a public good that, by the usual assumptions, contributes little value to the giver (e.g., Bolton et al. 2011).

[7] With regard to strategic play, we suppose that players do not choose weakly dominated strategies in any subgame.

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 a negative feedback herself, or (*ii*) the trader gave a negative feedback and the other trader has made good or had cooperated in the first place. 2) Traders make good (i.e. pay the price or send satisfactory quality) when they have received a negative feedback while the other has received a 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 a negative or both have received a 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 giving negative feedback is the only equilibrium of the subgame whenever at least one trader did not cooperate. Submitting negative feedback is a (weakly) dominant strategy when both (one) did not cooperate in the trading phase, since it prevents the need to make good in case the other submits a negative feedback, and may incentivize the other to cooperate in case the other submits a positive feedback. 4) This implies, however, that the feedback system cannot support reputation building and thus cannot maintain cooperation; the buyer will not pay the price and the seller will not deliver quality, regardless of whether traders start the trading procedure or not. The prediction then is that allowing feedback withdrawal in the conflict resolution phase of the two-sided feedback market will lower market efficiency.[8]

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

H1 [Trading phase; Stage 1-3]: Having the feedback withdrawal option available invites gaming. The option leads to a decreased willingness to pay the price 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 [Conflict resolution phase; Stage 4-5]: The conflict 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 a negative feedback, and when having received a negative feedback 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.

[8] 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 feedbacks for different reasons (say, in a broader social and economic context, or regarding 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 (be it economic or psychological) value is trivial.

We complement our hypothesis H2a, motivated by our strategic model, with hypothesis H2b, motivated by our field data, as discussed in Sections I and II:

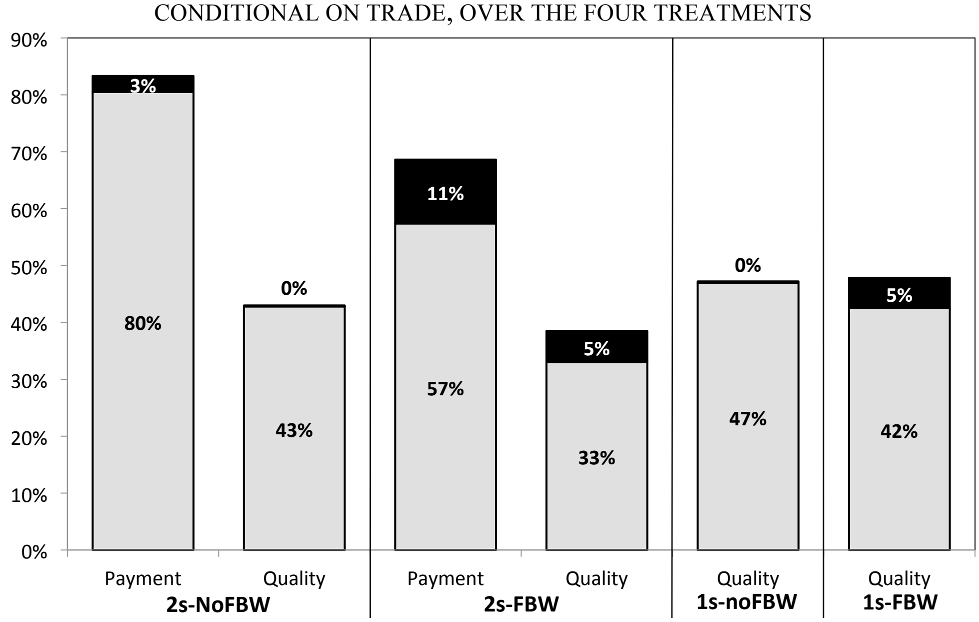
H2b [Altruistic punishment; Stage 4-5]: In markets with two-sided moral hazard, the leverage of giving negative feedback in negotiating feedback is mitigated, because some strategically negative feedback behavior is punished by refusal to withdraw feedback, even though this comes at the cost of keeping ones own negative feedback.

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 conflict resolution phase.

V. Laboratory data

We structure our report of experimental results along 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-rounds experiment. The first ten periods of our experiment likely involve some learning (we did not run practice rounds before starting the actual experiment), while the last ten periods typically involve end-game effects. Discarding both the first ten and the last ten rounds allows us to concentrate on a "running" feedback system, the subject that we are mainly interested in. That said, it turns out that running the analyses on all 60 rounds yields the same qualitative results as reported here. We include all-rounds versions of all tables of this paper in the supplementary appendix.

FIGURE 3: FREQUENCY 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 on show the additional frequency of payment and average gains in quality in the make-good stage.

V.1. Trading phase

In the next section we will see that behavior in the conflict 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 2, payment and quality, 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.

TABLE 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		0.499***		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

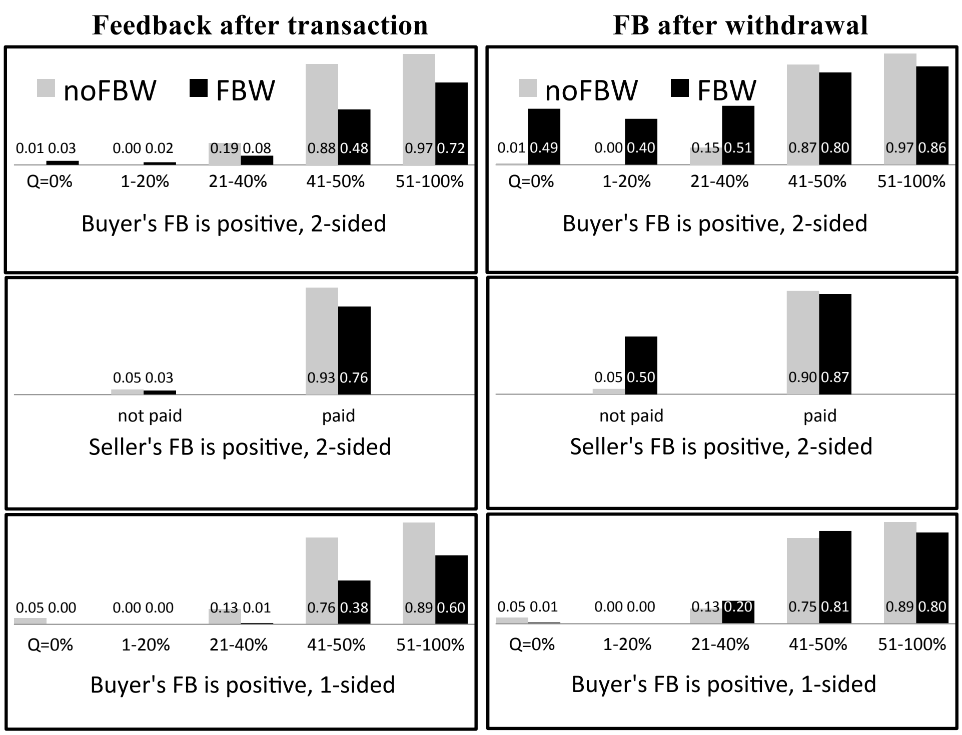
Notes: The table reports average marginal effects dy/dx for the Probit models. Quality and efficiency are censored at 0 and 1. *, **, and ** denote significance at the 10%, 5%, and 1%-level, respectively. Regressions are based on data from rounds 11-50 (omitting start and end effects). (Robust) Standard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

Payment and quality. As predicted, we observe significantly less payments and lower quality in our two-sided feedback markets with feedback withdrawal option compared to the two-sided markets without the option. Figure 3 displays frequencies of payments and average quality levels observed in our four treatments. In the two-sided system, with no feedback withdrawal option present, buyers pay in 80% of the cases, while they do so in only 57% of the cases when there is a feedback withdrawal option. Average seller quality choices decrease from 43% to 33% when the feedback withdrawal option is introduced. For one-sided markets, there is a 5% decrease in average quality sent when moving from the market with no feedback withdrawal to that with withdrawal.

Columns (1) and (2) of Table 4 report results from Probit and Tobit analyses, respectively, of the propensity of the buyer to pay the price and the quality delivered by the seller, 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 Stage 2 of the two-sided market: the FBW variable is significantly negative in Models 1 and 2.

Feedback. The left panel of 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

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



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 a 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.

The left panel of Table 5 below 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.[9] The regression detects a significantly 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 (see also Section 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.[10] Figure 3 above shows that there is only a small, and statistically insignificant, 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 of Table 4 above is not significantly different from zero, p=0.101). The lower left panel in Figure 4 shows that also in one-sided markets, feedback is more negative when a withdrawal option is existent.[11] 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.

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

[10] 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 above). 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).

[11] Absent any other strategic reasoning in the one-sided market, buyers may attempt to elicit even higher qualities from sellers by submitting a negative feedback, which can then later, at no cost, be turned into 'honest' feedback.

TABLE 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	Buyer FB is pos	Seller FB is pos	Buyer FB is pos	Buyer FB is pos	Seller FB is pos	Buyer 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. 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 pay the price and to deliver good quality, and as well as to dishonest feedback in the trading phase. These effects are mitigated in markets with one-sided moral hazard.

V.2. Conflict resolution phase

Why does the design of the conflict resolution phase affect behavior in the trading phase? Does altruistic punishment mitigate the leverage of retaliatory feedback? We analyze the conflict resolution behavior along 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 a negative feedback. And third, due to the incentive in two-sided systems to vote for withdrawal irrespective of the other's make-good behavior when having received a negative feedback oneself, we expect that a player is less likely to make good when she has strategic negotiation power in the withdrawal stage (i.e. has submitted a negative feedback herself).

The black-filled parts of the bars in Figure 3 above 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 displays 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 in the make-good stage. Inspection of Table 6 yields strong support for all three parts of our hypothesis.

(1) Across all feedback outcomes, we observe a make-good increase in average quality of 0.2% and in likelihood of payment of 3% in 2s-noFBW, compared to make-good increases of 5% and 11% in 2s-FBW (see also Figure 3).
(2) The make-good increases in 2s-FBW mainly come from traders who have received a negative feedback (3% of sellers improved their quality (by an average of 0.08) when they had received a positive feedback, while 45% of sellers with a negative feedback made good (by an average of 0.17). Similarly, previously non-paying buyers sent the payment in 7% of the cases in which they received a positive feedback, but in 27 % of the cases when they received a negative.
(3) Finally, in treatment 2s-FBW, the likelihood of a seller (buyer) to make good after having received a negative feedback is 34% (22%) when he had given a negative himself, and 68% (54%) when he did not.

To statistically corroborate these observations, we ran Probit regressions of the likelihood to make-good on a constant and a dummy indicating whether the feedback withdrawal option is present or not, separately for all four (two) possible outcomes of the feedback stage in the two-sided (one-sided) system (column 1 in Table 6). For example, to test whether make-good after mutually negative feedbacks is more likely in 2s-FBW compared to 2s-noFBW, we use all observations from treatments 2s-noFBW and 2s-FBW where we observe mutually negative feedback, and estimate the propensity to make good depending on a constant and a FBW dummy.[12]

We find that when the buyer gave a negative feedback, the seller is more likely to make good under FBW compared to no FBW ($p<0.001$ and $p<0.001$ for seller having submitted positive and negative himself,

[12] This approach allows us to use individual level data on make-good behavior while at the same time robustly clustering all standard errors at the matching group level. Table B.2 in Appendix B contains results from all 10 estimations (2 x 4 for the two-sided treatments and 2 for the one-sided treatment).

respectively),[13] and when the seller gave a negative feedback, the buyer is weakly significantly more likely to make good under FBW compared to no FBW (p=0.096 and p=0.057 for buyer having submitted positive and negative, respectively). In all other comparisons (i.e. after somebody received a positive feedback) we find no statistically significant average marginal effects of FBW on make-good behavior.

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 pos, S pos	64%	P: 0.99 Q: 0.51	P: 0.99 Q: 0.51	P: 13% Q: 7%	-	64%
B neg, S pos	12%	P: 0.99 Q: 0.28	P: 0.99 Q: 0.28	P: 0% Q: 4%	-	12%
B pos, S neg	10%	P: 0.30 Q: 0.49	P: 0.50 Q: 0.49	P: 29% Q: 6%	-	10%
B neg, S neg	14%	P: 0.18 Q: 0.16	P: 0.23 Q: 0.17	P: 6% Q: 6%	-	14%
2-sided FBW						
B pos, S pos	21%	P: 0.98 Q: 0.59	P: 0.99 Q: 0.59	P: 25% Q: 1%	-	67%
B neg, S 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 pos, S 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 neg, S 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						
Buyer pos	71%	Q: 0.53	Q: 0.54	Q: 7%	-	71%
Buyer neg	29%	Q: 0.31	Q: 0.32	Q: 4%	-	29%
1-sided FBW						
Buyer pos	31%	Q: 0.52	Q: 0.52	Q: 7%	-	72%
Buyer 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 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 11-50 (omitting start and end effects).

[13] These results hold true when using the absolute difference between Q_1 and Q_2 (rather than the likelihood of a change) as a dependent in a Tobit regression (p<0.001 and p=0.001 for the seller having submitted positive and negative himself, respectively; p<0.001 in one-sided system; all other comparisons between FBW and noFBW insignificant except for a small *negative* effect of FBW on the difference between Q_1 and Q_2 after mutually positive feedback, p=0.031).

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: pos, S: neg	B: neg, S: pos
B: neg, S: neg	-0.276***	-0.321***
	[0.100]	[0.056]
N	429	739
LL	-236.5	-470.0

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 11-50 (omitting start and end effects). (Robust) Standard errors are clustered at the level of independent matching groups.

In order to test whether make-good behavior is strategically influenced by the negotiation power in the subsequent withdrawal stage, we run Probit models on the buyer's and seller's propensity to make good after receiving a negative feedback in the two-sided FBW markets, with the only independent being a dummy on whether the trader has given a negative feedback herself or not. Results are presented in Table 7, and show that having submitted a negative feedback herself is negatively correlated with making good when a feedback withdrawal option is present.[14]

V.2.1 Withdrawal behavior and the role of altruistic punishment

In the feedback withdrawal treatments, after observing negative feedback and make-good choices, feedback givers had the opportunity to vote for feedback withdrawal. If both traders agreed to a withdrawal, then both feedbacks 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 conflict resolution phase: In Column 6 of Table 6 above we display 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 a negative but have submitted a positive feedback themselves almost always (97-99%) vote for a withdrawal of their negative feedback. This is consistent with our strategic model.

We also argued that if they have not received a negative feedback themselves, then feedback givers will condition their withdrawal decision on the observation that the trading partner has made good, while when both had submitted a negative feedback, then withdrawal will be unconditional, since both have a strong

[14] 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: neg, S: neg dummy is -0.120, p<0.001.) Interestingly, we also observe some make-good behavior of buyers in the 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 a negative feedback) and then using the make-good stage to reciprocate the seller's quality choice.

incentive to delete their own received negative feedback from their 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 include a dummy for whether the other has given a negative feedback, too, as well as an interaction effect of this dummy with the make-good behavior of the trading partner. The results, displayed in Table 8, are mixed. For both buyer and seller (Models 1 and 2, respectively) we find that having received a 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.[15]

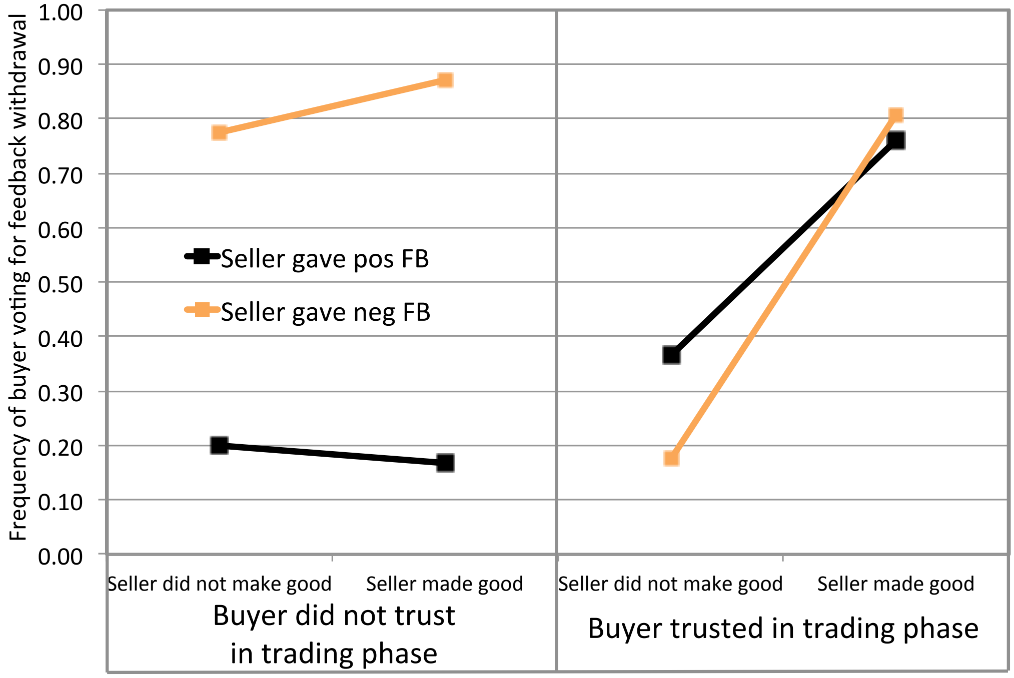
TABLE 8: PROBIT REGRESSION OF LIKELIHOOD TO WITHDRAW ON OTHER'S MAKE-GOOD BEHAVIOR AND FEEDBACK CONDITION

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: neg, S: pos	B: neg, S: pos	B: neg, S: pos	B: pos, S: neg	
Quality improved y/n	0.332*** [0.108]	-0.035 [0.207]	0.325*** [0.095]		0.528*** [0.039]
Payment improved y/n				0.436*** [0.126]	
B: neg, S: neg	0.266** [0.126]	0.441** [0.191]	-0.181 [0.183]	0.445*** [0.074]	
B: neg, S: neg × Quality improved y/n	-0.145 [0.124]	0.140 [0.191]	0.231 [0.152]		
B: neg, S: 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.

[15] The interaction effect is negative but statistically not significant, the joint effects of "Quality improved y/n + B: neg, S: neg × Quality improved y/n" and "Payment improved y/n + B: neg, S: 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 predicted when taking into account 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 near fully conditional on whether they received negative feedback from the seller or not. This is in line with our strategic analysis, which emphasizes the sellers' negotiation power in the mutual feedback withdrawal process. The right panel shows the opposite behavioral pattern, though: Here, we see that 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 a 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, starting with Fehr and Gächter (2000). 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) – as in our study: Buyers who cooperated in the trading phase and subsequently received a 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 who did not pay in the trading phase, respectively. The results corroborate the effects described above statistically: 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.[16]

V.2.3 Conflict resolution in one-sided markets

Our theoretical considerations suggest that in the one-sided markets, sellers only make good when they received a negative feedback and there is feedback withdrawal later on. Our data (also reported in Table 6 above) is consistent with the comparative statics of this prediction. Across the one-sided treatments, a seller is not very likely to make good when there is no feedback withdrawal, or when there is feedback withdrawal and he received a positive feedback (about 7% make-good likelihood in all three conditions). However, the seller is much more likely to make good after having received a 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 a negative feedback from Stage 2 into a positive feedback, and we predicted that this would only be the case if the seller indeed made good. In fact, as Model 3 in Table 8 above shows, the decision to withdraw a 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 conflict 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.

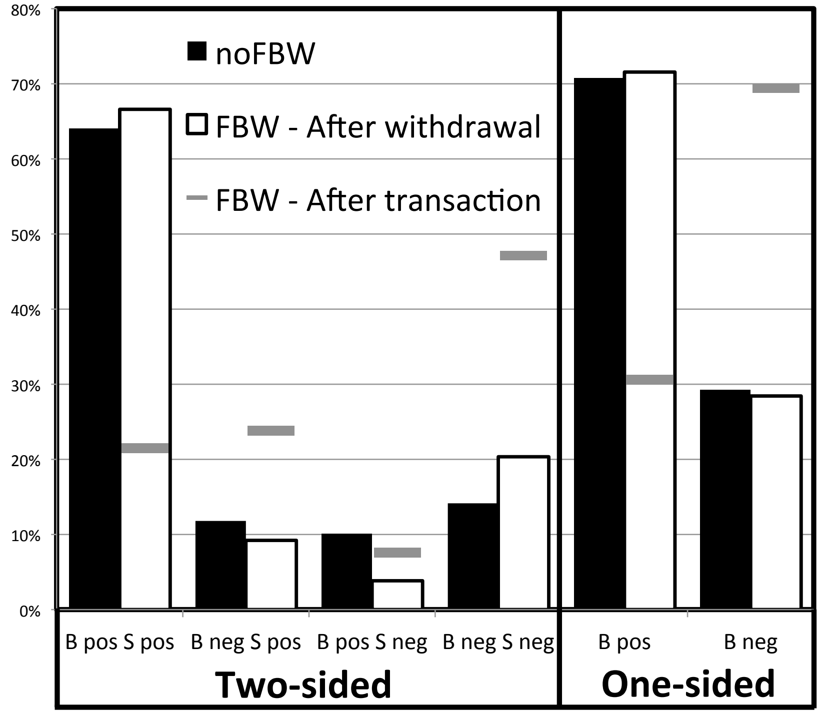
V.3. Market implications

The trading and conflict resolution phases affect the overall market performance. In particular, the effectiveness of the feedback system and its ability to provide incentives to behave trustworthy may be hampered when a feedback withdrawal option is offered, due to the resulting gaming in feedback and withdrawal behavior. If feedback as a statistic of past behavior of traders is less informative, then reputation building is hampered, which may lead to lower trustworthiness of traders, and eventually lower market efficiency. This section investigates how the feedback collected on the market place and general market performance are affected by the design of the conflict resolution phase.

---

[16] We do not find analogous evidence for the existence of different types in seller's withdrawal behavior, no matter how we define "cooperation in the trading phase" with respect to the seller's quality level chosen in Stage 2. Part of the reason might be that the 'satisfactory' quality level is ambiguous. So, when receiving a 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 a negative removed. There is no room for such ambiguity when it comes to the question whether or not the buyer has paid the price.

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



Eventual feedback distributions and informativeness. Our analysis of Stage-2 feedback shows that we observe very different distributions of *initial feedback* in markets with and without feedback withdrawal options. Figure 6 above shows that these differences in overall feedback distributions disappear when looking at the *final feedback* collected in the market (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 we show that a conclusion that similar feedback distributions imply similar underlying trading behaviors would be misleading,.

The right panel of Figure 4 above shows the *eventual* frequencies of positive feedback conditional on the trading partner's *eventual* behavior. In distributions of Stage-2 feedback (left panel, discussed in Section V.1 above) we had observed that in the FBW markets feedback is *more negative for cooperators*, compared to markets without FBW. Now, after the make-good and withdrawal process in the FBW markets took place, we see much *more positive feedback for defectors*. In particular, buyers who did not pay end up with a positive feedback in 50% of the cases, and sellers who provided quality of less than 40% received a positive feedback nevertheless 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 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 above represents *eventual* payment and quality choices in our treatments, after make good has taken place. Columns (3) and (4) of Table 4 above report results from 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 initial trade behavior. Correspondingly, the negative coefficients for the *FBW* dummy become smaller (in payment choices) and even insignificant (in quality choices).

Final market efficiencies (realized gains from trade divided by maximum gains from trade) in our treatments equal the product of frequency of trade and average quality. Since frequency of trade does not differ between treatments, statistical analysis results on market efficiency almost perfectly mimic our results on sellers' quality.

In our treatments with one-sided moral hazard, too, we observe the same eventual distributions of feedback independent of whether the feedback system allows for feedback withdrawal or not. The noFBW markets end up with 71% positive and 29% negative feedback, in the FBW markets we observe 72% positive and 29% negative feedback after the withdrawal stage. Differently 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 buyer's feedback and seller's quality between the one-sided FBW and noFBW markets. Correspondingly, we do not observe any effect of the feedback withdrawal option on final market outcomes in the one-sided markets (Model 4 in Table 4 above).

VI. Conclusions

The option to withdraw negative feedback is a common element of online market dispute resolution systems. Motivated by field data and guided by theoretical considerations, we use laboratory markets to show that this mechanism to incentivize cooperation interacts with trading behavior and reputation building in ways that escalate conflict rather than resolving 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 resolution mechanisms. However, our field data seem to suggest that the leverage of strategically negative feedback is smaller than theoretically expected. Recall that, in the field, we observe that conditional on initiation of feedback withdrawal, having retaliated a negative feedback with a negative feedback was not matched by a higher probability of success in terms of getting the feedback withdrawn. Interestingly, in our laboratory market data we also find that the negotiation power in the conflict resolution phase is not improved as expected by strategically negative feedback. Looking at eventual frequencies of feedback withdrawal (i.e. the frequency of cases where both traders voted for withdrawal), we find that they are similar across feedback outcomes: 61% and 49% if only one of the traders give a negative (where withdrawal should only be observed after make-good), and 59% if both give a negative feedback (where feedback withdrawal constitutes a dominant strategy, independent of make-good). In the laboratory markets, this is explained by the mix of two behavioral buyer types. On the one hand, we observe buyers who were strategically not cooperating during the transaction phase, and are then strategically conceding to the seller's negotiation power earned by strategically negative feedback giving. On the other hand, however, sellers do not always get away with strategic behavior, even when it comes at a cost to the buyers. In particular, those buyers who cooperated in the trading phase were often willing to punish untrustworthy trade and aggressive feedback behaviors, even when this would spoil their own reputation. This pattern of behavior, while inconsistent with our strategic model, is fully in line with 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.

That said, we caution that, in the field, other strategic and non-strategic aspects of the broader game being played may play an additional role when it comes to engineering conflict resolution systems. Most of these would not seem to question the external validity of our laboratory and theoretical studies, though. 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 effects of uncertainty on punishment). Yet, uncertainty of this kind would not much alter the key insight of our analysis regarding the strategic value of giving negative feedback. Also, altruistic punishers may be less likely to self-select themselves to 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 conflict resolution and improve the economic performance of two-sided markets. To the contrary, our theory and data sets strongly suggest that there is good reason to worry that such elements of conflict resolution systems rather escalate conflict, hamper trade efficiency and damages the effectiveness of reputation building systems. At the same time, our study shows that the detrimental effects are much mitigated in a market with one-sided moral hazard and thus one-sided feedback withdrawal, because the one-sidedness diminishes the scope for strategic feedback giving. Yet, whether and under which circumstances feedback withdrawal options can actually improve market efficiency and reputation informativeness, and more generally, what combination of factors make an electronic conflict resolution system effective, must be left to 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., Currie, J., Farber, H. S., and Spiegel, M. (1992). ‘‘An Experimental Comparison of Dispute Rates in Alternative Arbitration Systems,’’ *Econometrica* 60, 1407-1433.

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

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: 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.

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

Bolton, G., B. Greiner, and A. Ockenfels (2013), Engineering Trust - Reciprocity in the Production of Reputation Information. *Management Science*, 59(2), 2013, 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, 1-33.

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. 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.

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

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.

Gneezy, U. (2005), Deception: The Role of Consequences, *American Economic Review* 95, 384-394.

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

Greiner, B. (2004), An Online Recruitment System for Economic Experiments, in: Kurt Kremer, Volker Macho (eds.): Forschung und wissenschaftliches Rechnen 2003. GWDG Bericht 63, Göttingen: Ges. für Wiss. Datenverarbeitung, 79-93.

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, D. Reeves (2007), Pennies from eBay: The determinants of price in online auctions. *Journal of Industrial Economics* 55(2):223–233.

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.

Appendix

A. Detailed theoretical model

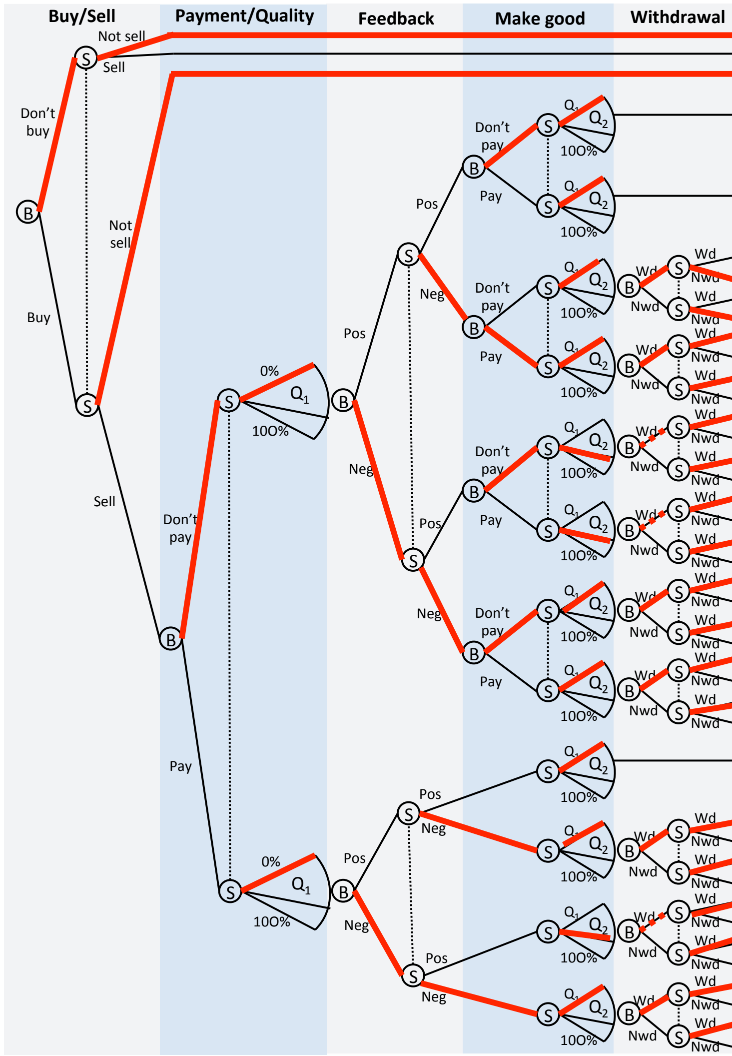
Feedback giving in markets with strangers interaction often only affects others' payoffs, and not one's own payoffs. This naturally yields indifferences regarding feedback giving if we assume that traders are purely selfish. For instance, a purely selfish buyer does not care whether or not to spoil the seller's reputation in the feedback phase, 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 a recent literature on lying aversion (Gneezy 2005, Gneezy et al. 2013). This is why we assume that traders, when indifferent monetary-wise, prefer to submit "honest" feedback. (An alternative way, yielding an equivalent solution, would be to assume that both, buyer and seller, receive a small benefit for honest feedback.) For feedback from seller to buyer, feedback is defined to be "honest" if the buyer received a negative when he did not pay and the buyer received a positive feedback if he paid the price. For feedback from buyer to seller, there is a quality aspiration Q^* of the buyer, from which on 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 displays the extensive form of the stage game of 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 final nodes are represented by the first terms in the buyer and seller columns of 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, in the withdrawal stage, both traders' agreement is needed to withdraw a 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 it to be larger than 1/3 and smaller than 1 so that trade is beneficial to both buyers and sellers. Moreover, when we compute equilibria, we will assume for simplicity that Q^* is correctly guessed by the seller. However, 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 conflict resolution phase, this cannot affect seller behavior in

FIGURE A.1: EXTENSIVE FORM OF 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 a negative feedback for buyer and 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.

equilibrium. The reason is that if such communication would lead to more quality, all buyers would engage in it, 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; there is no role for coordination in equilibrium.

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

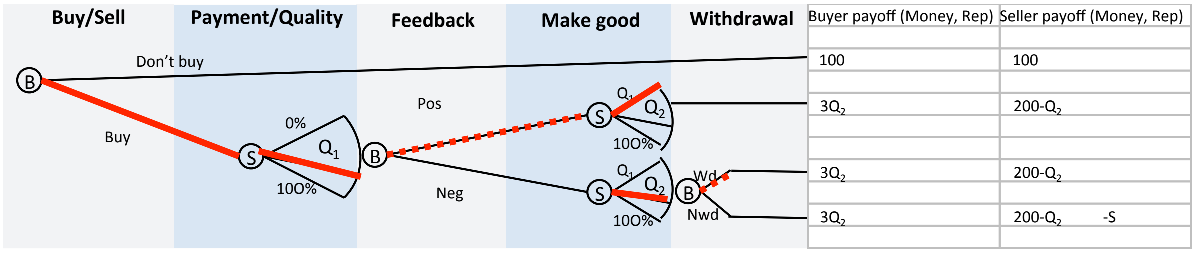
In equilibrium of the *trading phase* (TP): In Stage 3, both traders giving negative feedback is the only equilibrium of the subgame whenever at least one trader did not cooperate. Submitting negative feedback is a (weakly) dominant strategy in the subgame starting when both (one) did not cooperate, since it prevents the need to make good in case the other submits a negative feedback, and may incentivize the other to cooperate in case the other submits a 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 conflict resolution phase on the trading phase. Only when both cooperated, the traders will give positive feedback right away.[17] Because feedback can be gamed when a trader did not cooperate, the feedback system has no bite, and the interaction will end up in mutually positive feedback as a result of feedback withdrawal, no matter whether traders cooperated in the trading stage. Thus, in Stage 2, buyer and seller face a classical dilemma game; the buyer will not pay the price and the seller will deliver no quality. Finally, in Stage 1, because all traders will cheat in equilibrium, traders are indifferent as to start the trading procedure or not.

Now consider the game of our treatment 2s-noFBW. This corresponds to the extensive form game in Figure A.1 without the 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 conflict resolution phase traders do not make good, because there is no withdrawal stage 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, pay the price and deliver satisfactory quality.

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

[17] Actually, if both traders cooperated, then traders are indifferent between giving positive feedback now and giving negative feedback with the plan to withdraw it later.

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



Notes: S represents the continuation costs of having received a 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.

The stage game played in the treatment with a one-sided feedback system and a feedback withdrawal option is displayed in Figure A.2. The buyer decides to trade or not to trade, then the seller chooses a quality, the buyer submits feedback, the seller can improve his quality or not, and the buyer can decide to withdraw a 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 conflict resolution phase, in Stage 5, the buyer withdraws a 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 a negative feedback before. In the trading phase, the buyer submits negative feedback whenever the seller does not deliver sufficient quality. Thus, the seller has the choice to deliver sufficient quality right away ($Q_1=Q^*$) or to deliver lower quality first and improve quality later ($Q_1<Q^*$ and $Q_2=Q^*$). In any case, the seller will deliver sufficient quality at the end of the round, such that the buyer will always decide to buy.

Finally, consider the game of treatment 1s-noFBW, represented by the extensive form in Figure A.2 without the withdrawal stage and final node payoff of that branch equal to the no-withdrawal outcome. It is easy to see 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 a negative feedback before, 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 a 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 why it abandoned feedback withdrawal even after its move to a one-sided system), but 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 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)

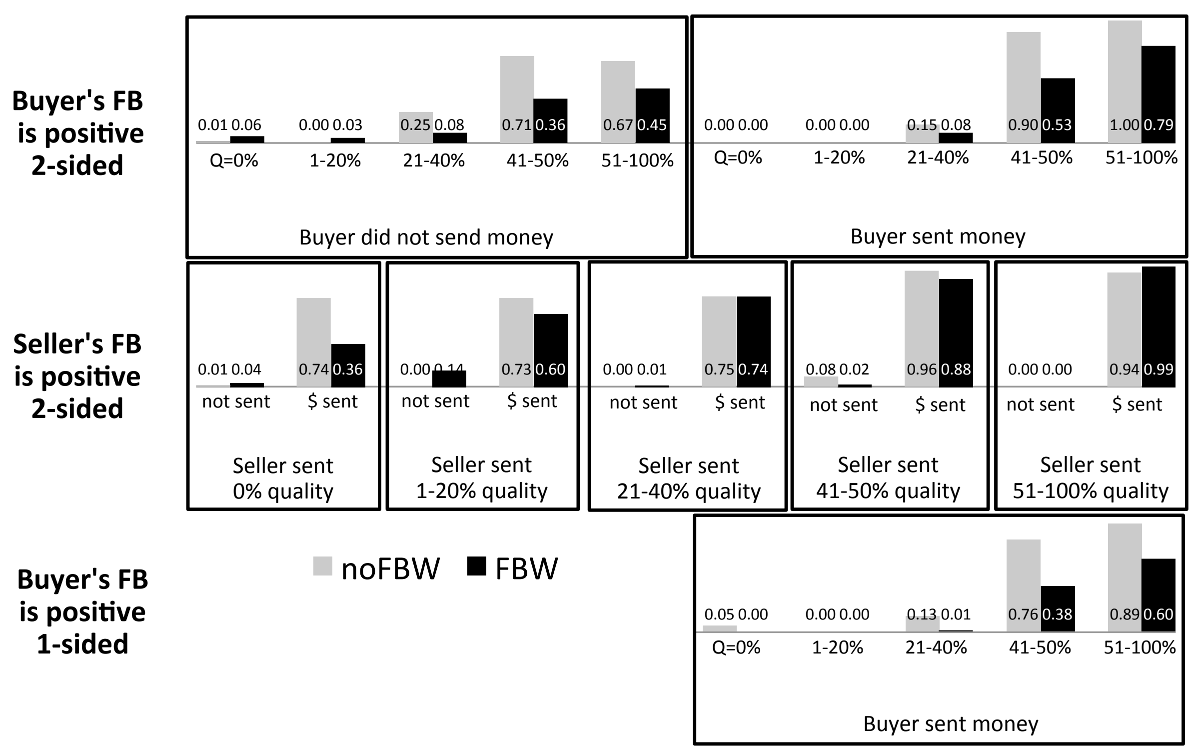
Seller moves first and gives	Buyer responds with			Buyer initiates MFW		Feedback is withdrawn		Seller initiates MFW		Feedback is withdrawn	
Positive	pos	71.9%	(n=98225)	0.0%	(n= 2)	0.0%	(n= 0)	0.0%	(n= 7)	14.3%	(n= 1)
N=136554	neu	0.7%	(n=922)	0.4%	(n= 4)	75.0%	(n= 3)	3.5%	(n= 32)	62.5%	(n= 20)
	neg	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	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

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 100 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 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 feedbacks given in Stage 3 was negative. Simultaneously, both the buyer and the seller can decide whether they support to revise the feedbacks and turn both feedbacks into "positive" feedbacks. If both support the revision, then both feedbacks will be made "positive". If only the buyer or only the seller or none supports the feedback revision, then the feedbacks 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 into a "positive" feedback. If the buyer supports the revision, then the feedback 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 feedbacks 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 feedbacks received in previous rounds, like this: "X positive feedbacks and Y negative feedbacks 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	Buyer FB is pos	Seller FB is pos	Buyer FB is pos	Buyer FB is pos	Seller FB is pos	Buyer 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 pos, S pos	58%	P: 0.98 Q: 0.51	P: 0.98 Q: 0.51	P: 18% Q: 8%	-	58%
B neg, S pos	14%	P: 0.95 Q: 0.27	P: 0.95 Q: 0.28	P: 0% Q: 9%	-	14%
B pos, S neg	10%	P: 0.22 Q: 0.49	P: 0.43 Q: 0.49	P: 27% Q: 7%	-	10%
B neg, S neg	18%	P: 0.14 Q: 0.12	P: 0.18 Q: 0.12	P: 5% Q: 4%	-	18%
2-sided FBW						
B pos, S pos	25%	P: 0.97 Q: 0.57	P: 0.98 Q: 0.57	P: 40% Q: 2%	-	66%
B neg, S 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 pos, S 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 neg, S 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						
Buyer pos	69%	Q: 0.53	Q: 0.53	Q: 7%	-	69%
Buyer neg	31%	Q: 0.28	Q: 0.29	Q: 10%	-	31%
1-sided FBW						
Buyer pos	31%	Q: 0.52	Q: 0.53	Q: 8%	-	69%
Buyer 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: pos, S: neg	B: neg, S: pos
B: neg, S: 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: neg, S: pos	B: neg, S: pos	B: neg, S: pos	B: pos, S: 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: neg, S: neg	0.338***	0.484**	-0.112	0.442***	
	[0.098]	[0.196]	[0.162]	[0.030]	
B: neg, S: neg × Quality improved y/n	-0.242**	-0.216	0.132		
	[0.112]	[0.211]	[0.142]		
B: neg, S: 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	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.