Department of Strategy and Innovation
Working Paper No. 01/2020

Fixing feedback revision rules in online markets

Gary Bolton
Kevin Breuer
Ben Greiner
Axel Ockenfels

January 2020

WIRTSCHAFTS UNIVERSITÄT WIEN VIENNA UNIVERSITY OF ECONOMICS AND BUSINESS

Fixing feedback revision rules in online markets

By GARY BOLTON, KEVIN BREUER, BEN GREINER AND AXEL OCKENFELS*

Feedback withdrawal mechanisms in online markets aim to facilitate the resolution of conflicts during transactions. Yet, frequently used online feedback withdrawal rules are flawed and may backfire by inviting strategic transaction and feedback behavior. Our laboratory experiment shows how a small change in the design of feedback withdrawal rules, allowing unilateral rather than mutual withdrawal, can both reduce incentives for strategic gaming and improve coordination of expectations. This leads to less trading risk, more cooperation, and higher market efficiency.

Keywords: dispute resolution system, market design, reputation, trust

JEL classification: C73, C9, D02, L14

* Bolton: Managerial Economics, Naveen Jindal School of Management, University of Texas at Dallas, Richardson, Texas 75080, e-mail: gbolton AT utdallas.edu; Breuer: Department of Economics, University of Cologne, D-50923 Cologne, Germany, e-mail: kevin.breuer AT wiso.uni-koeln.de; Greiner: Institute for Markets and Strategy, Vienna University of Economics and Business (WU Vienna), 1020 Vienna, Austria and University of New South Wales, Australia, e-mail: bgreiner AT wu.ac.at; Ockenfels: Department of Economics, University of Cologne, D-50923 Cologne, Germany, e-mail: ockenfels@uni-koeln.de. All authors acknowledge financial support from the German Science Foundation (DFG) through the research unit Design and Behavior (FOR 1371). Ockenfels gratefully acknowledges funding from the DFG under Germany's Excellence Strategy – EXC 2126/1– 390838866, as well as by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement No 741409; the results reflect only the authors' views, the ERC is not responsible for any use that may be made of the information it contains). Greiner gratefully acknowledges funding from the Australian Research Council through Discovery Grant DP130104557.

I. Introduction

Most online market and sharing places rely on reputation building systems to foster trust and trustworthiness on their platforms. However, such systems are less than perfect and conflicts still arise (Ockenfels and Resnick, 2012). Many online marketplaces employ conflict resolution systems to manage such conflicts. A widely used example are feedback withdrawal mechanisms, which exploit the infrastructure of existing feedback systems, and offer feedback revision if one or both trading partners are dissatisfied with the trading outcome. The idea is that the possibility of having one's received negative feedback removed incentivizes make-good behavior, and thus may eventually lead to everybody's satisfaction. Feedback withdrawal rules, however, may also invite strategic gaming. Using data from the laboratory and the field, Bolton et al. (2018) show how the standard feedback withdrawal mechanism can backfire and hamper feedback informativeness and market efficiency.

The question that arises is how to design a feedback withdrawal mechanism that provides incentives to resolve conflict without inviting strategic gaming and distorting feedback information. Starting from the commonly used – yet flawed – feedback withdrawal mechanism, we propose a minimal design change, making the final decision to withdraw feedback unilateral instead of mutual. Our laboratory experiment demonstrates how the slightly adapted mechanism undoes the original finding that withdrawal mechanisms significantly reduce feedback informativeness and market efficiency. The reason is that the new mechanism substantially curbs incentives to give feedback strategically, and in this way allows traders to use the feedback revision option as a device to more successfully coordinate expectations between buyers and sellers.

We contribute to a growing theoretical, experimental, and empirical literature on reputation building and the market design of feedback systems. A quickly increasing number of studies investigate the role of feedback systems for trader cooperation and market efficiency (see Chen at al., 2019 and Tadelis, 2016, for surveys). Yet, much less attention has been given to the design of conflict resolution mechanisms. Exceptions include Deck and Farmer (2007) who look into arbitration in bargaining over an uncertain value, Bolton and Katok (1996) who link the negative effect of arbitration on negotiation outcomes to slower learning, Ashenfelter et al. (1992) who investigate how different arbitration procedures affect bargaining outcomes, and Shavell (1995) who looks into binding arbitration as an alternative to trial before court. All these studies are concerned with offline arbitration, e.g. labor market disputes. With respect to online dispute resolution, Vasalou et al. (2008) investigate the effect of apologies to repair trust in one-off online interactions. Bolton et al. (2018), the departure point of our study, explore strategic behavior in eBay's mutual feedback withdrawal mechanism. We complement this literature by showing how a small tweak in the market design of a feedback-based conflict resolution mechanisms can achieve the objective of coordination and trade facilitation without distorting incentives in feedback giving.

In Section I we describe our experimental design and procedures. Section II develops our two main hypotheses. Section III presents our experimental results, and Section IV concludes. The Appendix includes robustness checks, experiment instructions, and additional analyses of the experimental data.

II. Experimental Design and Procedures

We compare three feedback withdrawal mechanisms. Each mechanism is placed in the same market place with two-sided moral hazard (both buyer and seller) and a two-sided feedback system. Participants interact as buyers and sellers for 60 rounds. Table 1 below displays the sequence of stage decisions taken in each round. Each round starts with a choice to engage in a trade (or not) by both traders. If one or both trading partners decide not to trade, seller and buyer receive their outside option of 100 ECU and the round ends. Otherwise, buyer and seller enter the transaction phase. The buyer decides whether or not to make the payment $P_1 \in \{0,100\}$ while simultaneously the seller decides on the level of quality Q_1 of the product (between 0% and 100%). Then both parties are informed about the decisions of their respective trading partner and submit feedback on the transaction. Feedback can be either positive or negative. After both trading partners are informed about their feedback, they receive the opportunity to make good. Specifically, if the buyer had not paid yet ($P_1 = 0$), then he receives another chance to pay, $P_2 \in \{0,100\} \geq P_1$. The seller may improve upon her initial quality with $Q_2 \geq Q_1$.

TABLE 1: PROCEDURE IN EACH ROUND OF THE EXPERIMENT

Stage	**Feedback System**
Feedback displayed	Sum of transaction partner's positive and negative feedback in previous rounds
Trade	Buyer and seller simultaneously decide whether to trade or not. If one doesn't agree, the round ends with $\pi B = 100$ and $\pi S = 100$.
Transaction	Buyer decides whether or not to pay 100 ECU. Seller simultaneously decides on Quality Q_1 with $0 \leq Q_1 \leq 100\%$.
Feedback	Buyer and seller simultaneously give either positive or negative feedback.
Make-good	If buyer has not made the payment yet, then he can pay now; seller simultaneously decides on quality Q_2 with $Q_1 \leq Q_2 \leq 100\%$.
Feedback withdrawal	*noFW*: No feedback withdrawal/revision.
	muFW: Both trading partners are asked to vote for feedback withdrawal. If both traders agree, negative feedback is changed to positive feedback.
	uniFW: Both trading partners are asked to vote for feedback revision option. If both traders agree, they simultaneously and independently decide whether they want to change their negative feedback to a positive feedback.
Payoffs	$\pi_B = 100 - \text{PricePaid} + Q_2 * 3$, $\pi_S = 100 + \text{PricePaid} - Q_2$

The three treatments of the experiment differ only in the last stage, concerning feedback withdrawal. In treatment *noFW*, there is no such stage. In treatment *muFW*, if there was at least one negative feedback, both trading partners are asked whether they agree to withdraw any negative feedback and make it positive. If, and only if, both agree, then both feedbacks are made positive. In treatment *uniFW*, both trading partners are asked whether they agree to allow a revision of feedback. If both agree, then both trading partners can *unilaterally* withdraw their negative feedback, or not. (In no treatment can positive initial feedback be made negative.)

All data was collected in the Cologne Laboratory for Economic Research, and participants were students from the University of Cologne recruited via ORSEE (Greiner 2015). The experiment was programmed in zTree (Fischbacher 2007). Average payoffs were about EUR 20 plus a show-up fee of EUR 2.50. The original Bolton et al. (2018) sessions involved 128 participants, with 2 sessions each for conditions *noFW* and *muFW*. The new sessions used 192 participants, with 3 sessions each for treatments *noFW* and *uniFW*. Sessions comprised 32 participants each, who were assigned to matching groups of 8 participants. Thus, our analysis relies on 20 independent markets/matching groups in the baseline *noFW*, 8 matching groups in *muFW*, and 12 matching groups in *uniFW*. In our analysis we pool data from Bolton et al. (2018) with data from the new experiment sessions conducted between June and November 2017.

III. Two hypotheses

The main flaw in *muFW* stems from the feedback withdrawal being required to be mutual, such that either all or none of the negative feedbacks are withdrawn. As long as negative feedback is costly, all traders who receive a negative feedback in the feedback stage will rationally and selfishly agree to mutually withdraw feedback, *irrespective of whether this distorts feedback information*, in order to make sure that one's own reputation does not get spoiled. Yet, at the same time, the incentive to cooperate vanishes, because even defecting traders can evade negative feedback by leaving a negative feedback themselves and thus making the opponent agree on feedback withdrawal. As a result, reputation information becomes less informative thereby reducing incentives for cooperation.[1]

Unilateral feedback withdrawal (*uniFW*) eliminates this flaw, because one's decision to withdraw feedback cannot affect one's own reputation. As a result, the incentives for creating 'honest' reputation information are the same in *uniFW* and *noFW*, as summarized by hypothesis H1.

[1] A model in Bolton et al. (2018), section 2, formalizes this line of reasoning. In synopsis: Even under most favorable conditions for cooperation, there can be no cooperation in equilibrium under mutual feedback withdrawal (*muFW*). The main assumptions of the model are three: (1) the future is sufficiently important, so that traders want to avoid receiving negative feedback; (2) traders' feedback is 'honest' as long as there are no monetary incentives to strategically submit biased feedback; and (3) conflict cannot arise due to coordination problems (which can happen when, for example, the buyer and seller differ in their expectations about what constitutes a 'satisfactory quality level').

H1: uniFW repairs muFW: The negative effects of mutual feedback withdrawal on trading behavior and feedback informativeness vanish if feedback withdrawal becomes unilateral.

If we establish that *uniFW* can repair *muFW*, we can then ask whether it serves to improve the performance of an otherwise identical reputation system with no withdrawal (*noFW*). This is an important question because simple models of reputation giving, including the one presented in Bolton et al. (2018), predict equivalent trading and feedback behavior in the *noFW* and *uniFW* conditions: Because feedback in both conditions is equally predicted to be honest, the reputation systems should, in theory, yield the same incentives to be cooperative (see also Footnote 1).

To put the quandary in a more empirical context, there is ample evidence to show that making information about past trade behavior public effects an increase in trust (e.g., Duffy and Feltovich 2002, 2006, Bolton et al. 2004, Bohnet and Huck 2004, Bohnet et al. 2005, Brown and Zehnder 2007, Bracht and Feltovich 2009, Charness et al. 2011, Huck et al. 2010, 2012, Duffy et al. 2013). So, if the *noFW* and *uniFW* systems offer the same incentive to give honest feedback, what reason is there to expect better trading outcomes in the latter system?

The answer is that the theoretical arguments rely on an implicit assumption, that there is no coordination failure: Traders' beliefs about what trading patterns to expect from each other to obtain a positive feedback are assumed to be mutually consistent. This, however, appears unlikely (see Bolton et al., forthcoming, for a discussion), and indeed one could argue that coordination of expectations is one of the major benefits of any successful conflict resolution system. In our context, for instance, coordination failure may arise with respect to a seller's expectation about what quality level makes the buyer sufficiently happy to leave a positive feedback. Some might think that any positive quality level signals some level of trust and kindness and thus should be reciprocated by a positive feedback; others may believe that any level below the quality that guarantees an equal split of payoffs is unfair and must thus be punished; others might argue that any level that does not maximize total payoffs deserves a negative feedback; and still others might take a hybrid perspective. A chance to revise one's behavior and feedback in an organized conflict resolution process, even as minimalistic as implemented by *uniFW*, might help traders to better coordinate these expectations. Doing so might reduce future trading risk and improve cooperation.

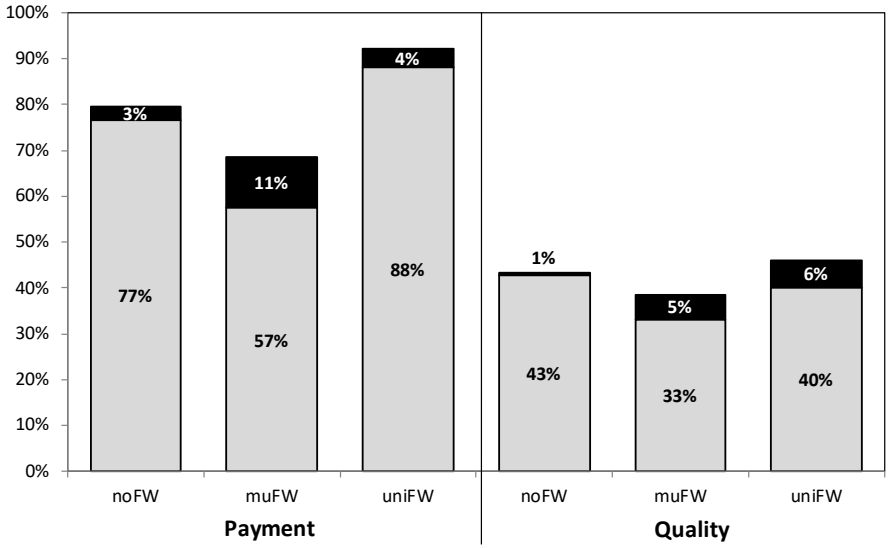
H2: uniFW improves coordination over noFW: uniFW reduces uncertainty and facilitates coordination of expectations, implying positive effects on trader cooperation.

IV. Results[2]

IV.1 uniFW does not reduce payments and quality like muFW does

Figure 1 below shows payment frequencies and average quality choices across our three treatments (conditional on there being trade).[3] Payments represent market merchandise revenues and are often a major concern of real-world market platforms which typically earn a share of these. The level of product quality traded scales the gains from trade, determining market efficiency. We observe strong treatment effects on the frequency of payments/market revenue. Compared to no feedback withdrawal, the feedback withdrawal mechanism *muFW* used in practice reduces the likelihood of initial (eventual) payment by 20 (12) percentage points. In contrast, the proposed *uniFW* mechanism, which implements but a small change compared to *muFW*, *increases* the likelihood of initial (eventual) payment by 11 (12) percentage points.

FIGURE 1: AVERAGE PAYMENT/REVENUES AND QUALITY/EFFICIENCY CONDITIONAL ON TRADE, ACROSS THE THREE TREATMENTS



[2] We focus our analysis on rounds 11 to 50, studying a running system rather than start-up or end-game effects. We provide more in-depth analyses in Appendix A and refer to them in this text where appropriate. In particular, in Appendix A.1 we present a direct comparison of the *noFW* baseline condition between the original Bolton et al. (2018) data and our new replication. We observe very similar behavioral pattern across original and replication. We do not find statistically significant differences at the 5% level for any of the major variables of interest (Wilcoxon matched pairs tests based on independent matching groups). We detect a weakly significant effect (at the 10% level) for seller profits as well as the likelihood to agree to trade, both being lower in the replication sessions than in the original sessions. All statistics provided below rely on the pooled data. Conclusions are largely the same when using only original baseline sessions, and somewhat more favorable for the *uniFW* system when using only the new replication sessions.

[3] The probability of entering trade in the three treatments is 74% in *noFW*, 81% in *muFW*, and 81% in *uniFW*. The lower number for the *noFW* control condition is mainly driven by the (weakly significantly) lower likelihood of trade in the new replication sessions compared to the older sessions (see previous footnote). When considering payment and quality unconditional on trade, these differences in trade likelihood somewhat mitigate the negative effects of *muFW* and increase the positive effects of *uniFW*. The comparison of *uniFW* and *muFW* however is unaffected, in particular since they show almost identical trade frequencies.

TABLE 2: REGRESSIONS OF PROBABILITY OF PAYMENT AND QUALITY

Model	(1)	(2)	(3)	(4)
Model type	Probit	Tobit	Probit	Tobit
Dependent	Initial Payment	Initial Quality	Final Payment	Final Quality
Constant		0.517***		0.512***
		[0.024]		[0.021]
Round	-0.005***	-0.004***	-0.005***	-0.003***
	[0.001]	[0.001]	[0.001]	[0.001]
muFW	-0.156**	-0.120**	-0.086	-0.058
	[0.074]	[0.053]	[0.059]	[0.047]
uniFW	0.130***	-0.028	0.148***	0.032
	[0.049]	[0.045]	[0.044]	[0.025]
N	4945	4945	4945	4945
LL	-2520.1	-1546.8	-2209.4	-937.7
Censoring Left		874		660
(Non) Right		(3965) 106		(4179) 106
Post-estimation test muFW = uniFW, p-value	0.0002	0.1544	0.0002	0.0598

Notes: Probit coefficent estimates are stated as average marginal effects dy/dx. Quality is censored at 0 and 1. 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. *, **, and ** denote significance levels 10%, 5%, and 1% level, respectively.

The Probit regressions reported in Table 2 Models (1) and (3) support these observations statistically. The differences in initial payment frequencies are highly significant. For eventual payment frequency, the differences between *uniFW* and the other two treatments reach significance at the 1% level, while the comparison between *noFW* and *muFW* is not statistically significant. (Non-parametric Wilcoxon Ranksum tests based on independent matching group averages support these conclusions.[4])

For initial product quality (market efficiency), we observe a reduction by 11 percentage points with the *muFW* mechanism compared to no feedback withdrawal, which Model (2) in Table 2 shows to be statistically significant. The small reduction by 3 percentage points in treatment *uniFW* is statistically not significant. For eventual quality, the negative effect of treatment *muFW* is 6 percentage points, while

[4] P-values for *noFW* vs. *muFW*, *noFW* vs. *uniFW*, and *muFW* vs. *uniFW* are 0.075, 0.011, and 0.003, respectively, in terms of initial payment frequencies, and 0.309, 0.006, and 0.004, respectively, in terms of eventual payment frequencies.

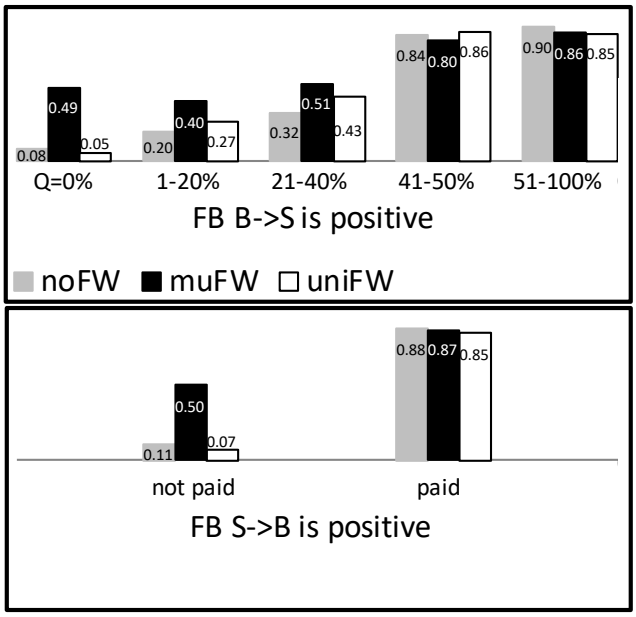
treatment *uniFW* yields an increase in quality of 4 percentage points. Both differences are not statistically significant. However, the eventual 10%-difference between *muFW* and *uniFW* is statistically weakly significant at the 10%-level (see post-estimation test in Table 2 Model (4)).[5]

In summary, while *muFW* creates negative effects on market revenues and market efficiency (though the latter effect is not significant when considering eventual quality), *uniFW* does not come with these costs, and even has a considerable positive effect in terms of payments/market revenues. In direct comparisons, *uniFW* outperforms *muFW* both in terms of payment and quality. We interpret this evidence as strong support for the trading terms portion of Hypothesis 1. We now turn to evaluating the second part of that hypothesis, regarding strategic feedback behavior and information distortion.

IV.2 uniFW does not distort feedback like muFW does

Figure 2 displays the frequency of positive feedback conditional the trading partner's behavior (eventual payment or quality choice), for all three treatments. The gray bars display data from the *noFW* treatment. We observe that the higher the quality, the larger is the likelihood of positive feedback, with a zero-quality yielding a positive feedback in only 8% of the transactions and a 51-100% quality resulting in a positive feedback in 90% of the cases. A similar trend is observed for sellers' feedback to buyers, where no payment receives a positive feedback only in 11% of the cases while a payment results in positive feedback in 88% of the cases.

FIGURE 2: EVENTUAL FEEDBACK CONDITIONAL ON TRADING PARTNER'S BEHAVIOR



[5] Results from non-parametric Wilcoxon Ranksum tests based on independent matching group averages are mostly consistent with the regression results. P-values for *noFW* vs. *muFW*, *noFW* vs. *uniFW*, and *muFW* vs. *uniFW* are 0.025, 0.559, and 0.076, respectively, in terms of initial quality, and 0.075, 0.350, and 0.076, respectively, in terms of eventual quality.

The black bars show the distortion in feedback informativeness resulting from *muFW*. In the face of incentives for strategic feedback behavior, 49% of the sellers who delivered a zero quality and 50% of buyers who do not pay nevertheless end up with a positive feedback. Thus, feedback in *muFW* is less informative in the sense of being less correlated with actual behavior than feedback in *noFW*. In the *uniFW* system (white bars), which mitigates the strategic feedback gaming incentives, this information distortion disappears, and eventual feedback conditional on eventual payment and quality resembles the data from a system without any feedback withdrawal possibilities.

Regressions reported in Table 3 below statistically support these conclusions. In *noFW*, feedback by the transaction partner is strongly correlated with the trader's behavior (quality/payment). In *muFW*, however, the probability of an unconditional positive feedback increases significantly, while the relation to the underlying quality and payment decisions is significantly reduced. No such effects are observed in treatment *uniFW*. In other words, the correlations between feedback and trader behavior are significantly reduced in treatment *muFW* but not in treatment *uniFW*.[6] This confirms the informativeness part of Hypothesis 1, in that *muFW* distorts feedback information but *uniFW* does not.

TABLE 3: PROBIT REGRESSIONS OF POSITIVE FEEDBACK ON QUALITY/PAYMENT AND TREATMENT INDICATORS

Dependent: Positive feedback	B->S FB is pos (1)		S->B FB is pos (2)	
Round	0.001	[0.001]	-0.001	[0.001]
Quality/Payment	0.011***	[0.001]	0.749***	[0.036]
muFW	0.308***	[0.064]	0.270***	[0.048]
muFW × Quality/Payment	-0.007***	[0.002]	-0.280***	[0.055]
uniFW	-0.055	[0.112]	-0.051	[0.061]
uniFW × Quality/Payment	0.002	[0.003]	0.022	[0.077]
N	4945		4945	
LL	-2205.2		-1965.3	
Post-estimation test muFW = uniFW, p-value	0.0003		0.0000	

Notes: The table reports average marginal effects dy/dx. 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. *, **, and ** denote significance at the 10%, 5%, and 1% level, respectively.

[6] In Appendix A.2 we provide a similar analysis, using non-parametric Wilcoxon rank sum tests based on correlations between feedback and quality/payment calculated at the independent matching group level.

The mitigated distortion of feedback in *uniFW* as compared to *muFW* is due to reduced incentives for strategic gaming of the feedback and withdrawal rules. To further illustrate this, Appendix A.3 shows that both *muFW*- and *uniFW*-traders condition withdrawal of negative feedback on make-good behavior when not threatened by a negative feedback themselves. When having received a negative feedback themselves, behavior becomes different in the two markets. In *muFW* making up does not matter anymore, and traders agree to withdrawal unconditionally, making feedback and withdrawal losing its bite. In *uniFW*, however, the conditionality is coming back in the unilateral withdrawal stage, preserving incentives to make-good in all cases. As one result, *muFW*-traders are more likely to give preemptive negative feedback in order to extort a withdrawal decision, something that is not possible under *uniFW*.

IV.3 uniFW reduces variance in payoffs compared to noFW and muFW

In order to assess the strategic uncertainty faced by buyers and sellers – and thus the scope for coordination failure – in our different markets, we calculate the standard deviation of buyer and seller round profits (conditional on entering trade) within each matching group. We also calculate these numbers for trusting buyers, who sent payment in the initial transaction phase, and trusting sellers, who delivered a quality of more or equal to 50% in the initial phase. We then conducted Wilcoxon Ranksum tests to assess whether the distributions of these standard deviations differ between treatments. Table 4 below lists the averages of the calculated standard deviations across all matching groups of the respective treatments along with the corresponding p-values.

TABLE 4: BUYER AND SELLER PROFITS AND THEIR AVERAGE STANDARD DEVIATIONS ACROSS INDEPENDENT MATCHING GROUPS, AND WILCOXON RANKSUM RESULTS

Average round payoffs	*noFW*	*muFW*	*uniFW*
All buyers	143.0	142.6	142.3
All sellers	131.1	127.0	140.3
Buyers who paid initially	141.9	142.1	144.8
Sellers who sent initial quality >=50%	128.6	123.5	136.2
Average Standard Deviation	*noFW*	*muFW*	*uniFW*
All buyers	55.89	60.16	40.34
All sellers	35.67	37.35	25.77
Buyers who paid initially	48.42	56.36	34.20
Sellers who sent initial quality >=50%	30.81	31.46	18.45
P-values from Wilcoxon rank sum tests	*noFW* vs. *muFW*	*noFW* vs. *uniFW*	*muFW* vs. *uniFW*
All buyers	0.222	0.013	0.006
All sellers	0.576	0.007	0.001
Buyers who paid initially	0.204	0.032	0.017
Sellers who sent initial quality >=50%	0.799	0.002	0.017

As the middle part of Table 4 shows, we find that *uniFW* leads to a lower variation in (expected) round payoffs not just in comparison to the strategically problematic *muFW* mechanism, but also to the system without any feedback withdrawal mechanism (*noFW*). And this particularly holds for initially trusting buyers and sellers. The test results presented in the lower part of Table 4 confirm that standard deviations for all inspected groups are lower in *uniFW* than in *noFW* and *muFW*, with no statistically significant difference between the latter two. At the same time, we observe that buyer and seller profits in *uniFW* are equal to or even larger than in *noFW* and *muFW*. Specifically, seller profits (over all sellers) in *uniFW* outperform both *noFW* and *muFW* ($p = 0.0176$ and 0.0136, respectively), while the other differences are not significant at the $p = 0.1$ level.

We conclude that the strategic uncertainty of a trader with respect to expected profits from entering a transaction with a trading partner is significantly reduced in *uniFW* compared to when no feedback withdrawal system is present (or when *muFW* is at work). Thus, we confirm Hypothesis 2 that *uniFW* can reduce uncertainty and facilitate expectation coordination.

V. Conclusion

We experimentally compare two-sided markets with three different conflict resolution systems: one where no such system exists (*noFW*), one that employs a standard mutual feedback withdrawal (*muFW*, where only all negative feedback can be withdrawn at once upon mutual agreement), and one that uses a slightly modified system (*uniFW*, where both trading partners mutually agree to let each other withdraw feedback unilaterally). We find that in contrast to *muFW*, the *uniFW* option neither reduces market efficiency nor distorts feedback informativeness. Rather, it facilitates the coordination of expectations by reducing traders' strategic uncertainty. It also positively affects market merchandise revenues, which are often important to real-world market platform profitability.

While the work here speaks directly to a problem with online dispute resolution mechanisms, the results speak indirectly to problems common to many offline dispute resolution mechanisms, a problem long known to researchers studying offline arbitration (Ashenfelter 2009); namely, having dispute resolution available tends to reduce the incentives for actors to solve a problem in the first place (prior to using dispute resolution). In other words, the availability of dispute resolution tends to reduce the number of voluntary settlements we would otherwise see, and the additional arbitrated outcomes may be distorted relative to the voluntary settlements they displace. The results here show that a careful assessment of the dispute resolution rules can turn up design modifications in those rules that mitigate the incentive distortion that causes these problems in the first place. Whether such design modifications can be successfully employed in offline mechanisms is therefore an interesting avenue for further research.

REFERENCES

Ashenfelter, O., Currie, J., Farber, H.S., Spiegel, M., 1992. An Experimental Comparison of Dispute Rates in Alternative Arbitration Systems. Econometrica 60, 1407–1433. https://doi.org/10.2307/2951527

Ashenfelter, O., Iyengar, R., 2009. Economics of commercial arbitration and dispute resolution. Edward Elgar, Cheltenham, UK. ISBN: 978 1 84720 332 8.

Ba, S., Whinston, A.B., Zhang, H., 2003. Building trust in online auction markets through an economic incentive mechanism. Decision Support Systems 35, 273–286. https://doi.org/10.1016/S0167-9236(02)00074-X

Behnk, S., Barreda-Tarrazona, I., García-Gallego, A., 2019. Deception and reputation – An experimental test of reporting systems. Journal of Economic Psychology, Uncovering Dishonesty 71, 37–58. https://doi.org/10.1016/j.joep.2018.10.001

Belleflamme, P., Peitz, M., 2018. Inside the engine room of digital platforms: Reviews, ratings, and recommendations. Working Paper.

Bohnet, I., and Huck, S. (2004). Repetition and reputation: Implications for trust and trustworthiness when institutions change. American Economic Review, Papers and Proceedings, 94(2), 362-366. https://doi.org/10.1257/0002828041301506

Bohnet, I., Harmgart, H., Huck S., Tyran, J.-R., (2005). Learning Trust. Journal of the European Economic Association 3(2-3), 322–329. https://doi.org/10.1162/jeea.2005.3.2-3.322

Bolton, Gary E., Ferecatu, A., Kusterer, D.J., 2019. Rate this transaction: Design principles for coordinating mappings in feedback systems. Working Paper.

Bolton, G., Greiner, B., Ockenfels, A., 2018. Dispute Resolution or Escalation? The Strategic Gaming of Feedback Withdrawal Options in Online Markets. Management Science 64, 4009–4031. https://doi.org/10.1287/mnsc.2017.2802

Bolton, G., Greiner, B., Ockenfels, A., 2012. Engineering Trust: Reciprocity in the Production of Reputation Information. Management Science 59, 265–285. https://doi.org/10.1287/mnsc.1120.1609

Bolton, G.E., Katok, E., 1998. Reinterpreting Arbitration's Narcotic Effect: An Experimental Study of Learning in Repeated Bargaining. Games and Economic Behavior 25, 1–33. https://doi.org/10.1006/game.1997.0633

Bolton, G., Katok, E., and Ockenfels, A., 2004. How effective are electronic reputation mechanisms? An experimental investigation. Management Science 50(11), 1587-1602. https://doi.org/10.1287/mnsc.1030.0199

Bolton, Gary E., Kusterer, D.J., Mans, J., 2019. Inflated Reputations: Uncertainty, Leniency, and Moral Wiggle Room in Trader Feedback Systems. Management Science. https://doi.org/10.1287/mnsc.2018.3191

Bolton, G.E., Mans, J., Ockenfels, A., forthcoming. Norm Enforcement in Markets: Group Identity and the Volunteering of Feedback. The Economic Journal.

Bracht, J., and Feltovich, N., 2009. Whatever you say, your reputation precedes you: observation and cheap talk in the trust game. Journal of Public Economics 93(9-10), 1036–1044. https://doi.org/10.1016/j.jpubeco.2009.06.004

Brown, M., and C. Zehnder, 2007. Credit reporting, relationship banking, and loan repayment. Journal of Money, Credit, and Banking 39(8), 1884–1918. https://doi.org/10.1111/j.1538-4616.2007.00092.x

Burtch, G., Hong, Y., Bapna, R., Griskevicius, V., 2017. Stimulating Online Reviews by Combining Financial Incentives and Social Norms. Management Science 64, 2065–2082. https://doi.org/10.1287/mnsc.2016.2715

Cabral, L., Li, L. (Ivy), 2015. A Dollar for Your Thoughts: Feedback-Conditional Rebates on eBay. Management Science 61, 2052–2063. https://doi.org/10.1287/mnsc.2014.2074

Charness, G., Du, N., and Yang, C-L., 2011. Trust and trustworthiness reputation in an investment game. Games and Economic Behavior 72(2), 361–375. https://doi.org/10.1016/j.geb.2010.09.002

Chen, Y., Cramton, P., List, J., and Ockenfels, A., 2019. Market Design, Human Behavior and Management. Working paper.

Cui, R., Li, J., Zhang, D.J., 2019. Reducing Discrimination with Reviews in the Sharing Economy: Evidence from Field Experiments on Airbnb. Management Science. https://doi.org/10.1287/mnsc.2018.3273
Deck, C.A., Farmer, A., 2007. Bargaining over an Uncertain Value: Arbitration Mechanisms Compared. J Law Econ Organ 23, 547–579. https://doi.org/10.1093/jleo/ewm012
Dellarocas, C., Wood, C.A., 2007. The Sound of Silence in Online Feedback: Estimating Trading Risks in the Presence of Reporting Bias. Management Science 54, 460–476. https://doi.org/10.1287/mnsc.1070.0747
Duffy, J., and Feltovich, N., 2002. Do actions speak louder than words? An experimental comparison of observation and cheap talk. Games and Economic Behavior 39(1), 1–27. https://doi.org/10.1006/game.2001.0892
Duffy, J., and Feltovich, N., 2006. Words, deeds and lies: Strategic behavior in games with multiple signals. Review of Economic Studies 73(3), 669–688. https://doi.org/10.1111/j.1467-937X.2006.00391.x
Duffy, J., Xie, H., and Lee, Y-J., 2013. Social norms, information and trust among strangers: theory and evidence. Economic Theory 52(2): 669–708. https://doi.org/10.1007/s00199-011-0659-x
Fischbacher, U., 2007. z-Tree: Zurich toolbox for ready-made economic experiments. Exp Econ 10, 171–178. https://doi.org/10.1007/s10683-006-9159-4
Greiner, B., 2015. Subject pool recruitment procedures: organizing experiments with ORSEE. Journal of the Economic Science Association 1, 114–125. https://doi.org/10.1007/s40881-015-0004-4
Huck, S., Lünser, G. K. and Tyran, J.-R., 2010. Consumer networks and firm reputation: A first experimental investigation. Economics Letters 108(2), 242–244. https://doi.org/10.1016/j.econlet.2010.04.017
Huck, S., Lünser, G., and Tyran, J.-R., 2012. Competition fosters trust. Games and Economic Behavior 76(1), 195-–209. https://doi.org/10.1016/j.geb.2012.06.010
Hui, X., Saeedi, M., Shen, Z., Sundaresan, N., 2016. Reputation and Regulations: Evidence from eBay. Management Science 62, 3604–3616. https://doi.org/10.1287/mnsc.2015.2323
Kim, K., Chung, K., Lim, N., 2019. Third-Party Reviews and Quality Provision. Management Science 65, 2695–2716. https://doi.org/10.1287/mnsc.2018.3082
Lafky, J., 2014. Why do people rate? Theory and evidence on online ratings. Games and Economic Behavior 87, 554–570. https://doi.org/10.1016/j.geb.2014.02.008
Li, L. (Ivy), Xiao, E., 2014. Money Talks: Rebate Mechanisms in Reputation System Design. Management Science 60, 2054–2072. https://doi.org/10.1287/mnsc.2013.1848
Livingston, J.A., 2005. How Valuable Is a Good Reputation? A Sample Selection Model of Internet Auctions. The Review of Economics and Statistics 87, 453–465. https://doi.org/10.1162/0034653054638391
Luca, M., Zervas, G., 2016. Fake It Till You Make It: Reputation, Competition, and Yelp Review Fraud. Management Science 62, 3412–3427. https://doi.org/10.1287/mnsc.2015.2304
Ockenfels, A., Resnick, P., 2012. Negotiating reputations, in: Rachel Croson and Gary E. Bolton (eds.). The Oxford Handbook of Economic Conflict Resolution. Oxford University Press.
Rice, S.C., 2011. Reputation and Uncertainty in Online Markets: An Experimental Study. Information Systems Research 23, 436–452. https://doi.org/10.1287/isre.1110.0362
Shavell, S., 1995. Alternative Dispute Resolution: An Economic Analysis. The Journal of Legal Studies 24, 1–28. https://doi.org/10.1086/467950
Tadelis S., 2016. Reputation and Feedback Systems in Online Platform Markets. Annual Review of Economics 8(1), 321–340. https://doi.org/10.1146/annurev-economics-080315-015325
Vasalou, A., Hopfensitz, A., Pitt, J.V., 2008. In praise of forgiveness: Ways for repairing trust breakdowns in one-off online interactions. International Journal of Human-Computer Studies 66, 466–480. https://doi.org/10.1016/j.ijhcs.2008.02.001

Appendix A. Further analysis, tables, and figures

Appendix A.1 Comparison between noFW original and replication sessions

In Table A.1.1 we compare the baseline data from original Bolton et al (2018) sessions with the data from our baseline replication sessions, using Wilcoxon Ranksum tests based on independent matching group averages. We generally find no statistically significant differences between the two sets of sessions, with two exceptions: The frequency of trade and the average seller profit are weakly significantly lower in the replication sessions than in the original sessions (p=0.076 and p=0.070, respectively). These significant results however have to be evaluated on the background of multiple hypothesis testing. A Bonferroni correction for eight concurrent hypothesis tests would render all p-values insignificant.

TABLE A.1.1: MAIN AGGREGATE OUTCOME VARIABLES OF INTERESTS FOR ORIGINAL AND REPLICATION NOFW SESSIONS, AND RESULTS FROM WILCOXON RANKSUM TESTS

	Original sessions	Replication sessions	Wilcoxon Ranksum p-value
Frequency of trade	0.801	0.693	0.076*
Frequency of initial payment	0.785	0.722	0.355
Frequency of eventual payment	0.816	0.754	0.165
Avg. initial quality	42.2%	42.3%	0.488
Avg. eventual quality	42.4%	43.3%	0.316
Avg. profit	134.4	130.2	0.247
Avg. buyer profit	136.5	137.5	0.758
Avg. seller profit	132.3	122.9	0.070*

Appendix A.2 Assessing the informativeness of eventual feedback across treatments

A further series of non-parametric statistical tests supports the conclusion that compared to *noFW*, *muFW* leads to information distortion while *uniFW* does not. Wilcoxon Ranksum tests based in independent matching group averages yield that the frequencies of positive feedbacks after zero quality / no payment are significantly different in the *muFW* treatment compared to *noFW* and *uniFW*, with no differences between the latter. Obtained p-values for *noFW* vs. *muFW*, *noFW* vs. *uniFW*, and *muFW* vs. *uniFW* are 0.0001, 0.4676, and 0.0001, respectively, for 0% quality, and 0.0001, 0.3514, and 0.0002, respectively, for no payment. Probit models regressing the likelihood of a positive feedback when there is 0% quality / no payment on treatment dummies yield exactly the same conclusions.

For further support, we compute and compare the correlations between behavior (payment/quality) and eventually received feedback. The point-biserial correlation between the continuous variable quality and the dichotomous variable feedback equals 0.822, 0.731, and 0.891 in treatments *noFW*, *muFW*, and *uniFW*, respectively. Cramer's V as a measure of correlation between the two dichotomous variables payment and feedback is 0.698, 0.401, and 0.519 for treatments *noFW*, *muFW*, and *uniFW*, respectively. Wilcoxon Ranksum tests that are based on these correlations at the matching group level confirm that these differences in correlations are statistically significant (except for the comparison of the payment-feedback correlation between treatments *muFW* and *uniFW*). P-values for *noFW* vs. *muFW*, *noFW* vs. *uniFW*, and *muFW* vs. *uniFW* are 0.0933, 0.0391, and 0.0055, respectively, for comparing correlations between quality and feedback, and 0.0008, 0.0390 and 0.1167, respectively, for comparing correlations between payment and feedback.

Appendix A.3 Detailed description and analysis of feedback behavior

In this appendix, we discuss the observed pattern of feedback and withdrawal behavior in the three treatments (in particular in the two feedback withdrawal treatments *muFW* and *uniFW*) in more detail. Figure A.3.1 shows the pattern of *initial* feedback behavior (as opposed to *eventual* feedback patterns) in the three treatments. We observe that buyers in *muFW* are more likely to withhold positive feedback for high quality, compared to the other two systems, presumably in order to not give away their negotiation power in the subsequent withdrawal stage. For sellers, we observe that they are more likely to withhold positive feedback for an initial payment in both withdrawal systems. While in *muFW* sellers may have similar strategic reasons as buyers for that, sellers in *uniFW* may also want to protect themselves against buyers extracting "unfairly high" quality from them in the next stage.

FIGURE A.3.1: INITIAL FEEDBACK CONDITIONAL ON TRADING PARTNER'S BEHAVIOR

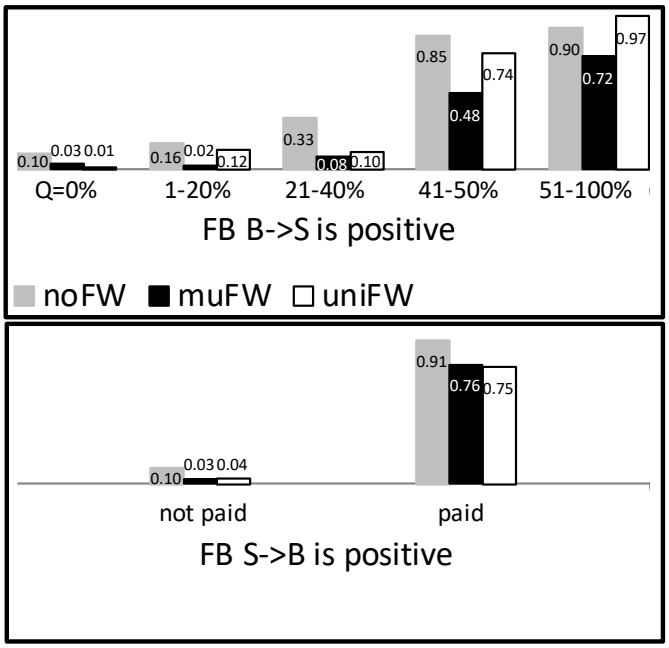


Table A.3.1 show the detailed pattern of feedback giving and underlying buyer and seller behavior. Figure A.3.2 visualizes the feedback conversion through withdrawal in the three different treatments. We observe that first, initial feedback distributions vary markedly between the three treatments, especially in the share of mutually positive feedback, reflecting the different initial payment and quality choices in the three treatments, as well as strategic considerations. Second, while there is little make-good in the system without feedback withdrawal (*noFW*), there is considerable make-good behavior in two feedback withdrawal

treatments (*muFW* and *uniFW*), almost exclusively by those who have received an initial negative feedback. Make-good is more prevalent for the traders with negative feedback when feedback was asymmetric, i.e. the other trader received a positive feedback. And third, in both *muFW* and *uniFW*, when initial feedback was asymmetric, the "weak party" (i.e. the one who received the negative feedback) almost always also agrees to (the option of) have feedback withdrawn. In *uniFW*, when feedback was asymmetric, those who agree to withdraw also typically withdraw in the end, such that there is no big difference between the agreement to allow withdrawal and the act of withdrawal. When feedback was symmetrically negative, however, then agreement to withdrawal is much higher (since also the own feedback is on the line), but only half of affected buyers and sellers then also unilaterally withdraw the other's feedback. Thus, the mutual agreement to withdraw and the unilateral act to withdraw are indeed treated differently. The fourth observation, highlighted both by Table A.3.1 and Figure A.3.2, is that even though the three different treatments yield very different initial feedback distributions and feature quite different make-good and feedback-withdrawal pattern, the final distributions of feedback (see last column of Table A.3.1 and right side of Figure A.3.2) are very similar across the three treatments. That is, the actually observed distribution of feedback in a feedback system may tell us very little about underlying market and feedback behavior, a caveat to keep in mind when examining empirical data collected on real-world platforms.

FIGURE A.3.2: FEEDBACK TRANSFORMATION THROUGH WITHDRAWAL, FOR ALL THREE TREATMENTS

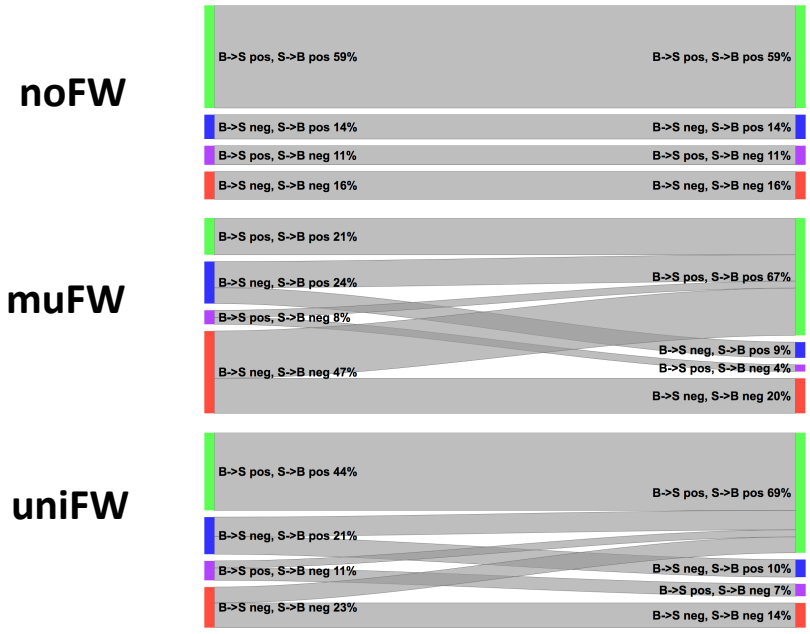


TABLE A.3.1: FEEDBACK, MAKE-GOOD AND WITHDRAWAL FREQUENCIES AS WELL AS INITIAL AND EVENTUAL FEEDBACK DISTRIBUTIONS ACROSS THE THREE TREATMENTS

Treatment & Given FB	FB Freq.	P & Q before make-good	P & Q after make-good	Freq. of make-good	Vote for withdrawal (opportunity)	Revise	Evtl. FB Freq.
noFW							
B->S pos, S->B pos	59%	P: 0.97 Q: 0.52	P: 0.97 Q: 0.52	P: 8% Q: 8%	-	-	59%
B->S neg, S->B pos	14%	P: 0.94 Q: 0.31	P: 0.94 Q: 0.31	P: 10% Q: 7%	-	-	14%
B->S pos, S->B neg	11%	P: 0.27 Q: 0.46	P: 0.41 Q: 0.46	P: 18% Q: 8%	-	-	11%
B->S neg, S->B neg	16%	P: 0.21 Q: 0.17	P: 0.29 Q: 0.18	P: 10% Q: 7%	-	-	16%
muFW							
B->S pos, S->B pos	21%	P: 0.98 Q: 0.59	P: 0.99 Q: 0.59	P: 25% Q: 1%	-	-	67%
B->S neg, S->B pos	24%	P: 0.96 Q: 0.35	P: 0.96 Q: 0.44	P: 0% Q: 68%	B: 62% S: 99% Both: 61%	-	9%
B->S pos, S->B neg	8%	P: 0.14 Q: 0.42	P: 0.61 Q: 0.43	P: 54% Q: 9%	B: 97% S: 51% Both: 49%	-	4%
B->S neg, S->B neg	47%	P: 0.26 Q: 0.19	P: 0.43 Q: 0.26	P: 22% Q: 34%	B: 73% S: 82% Both: 57%	-	20%
uniFW							
B->S pos, S->B pos	44%	P: 1.00 Q: 0.52	P: 1.00 Q: 0.52	P: - Q: 0%	-	-	69%
B->S neg, S->B pos	22%	P: 0.98 Q: 0.28	P: 0.98 Q: 0.46	P: 14% Q: 68%	B: 66% S: 98% Both: 66%	B: 95%	10%
B->S pos, S->B neg	11%	P: 0.75 Q: 0.46	P: 0.94 Q: 0.46	P: 75% Q: 6%	B: 91% S: 64% Both: 60%	S: 96%	7%
B->S neg, S->B neg	23%	P: 0.62 Q: 0.26	P: 0.71 Q: 0.34	P: 23% Q: 49%	B: 78% S: 80% Both: 63%	B: 53%, S: 48% Both: 34%	14%

Table A.3.2 examines strategic behavior in the feedback withdrawal process after having given a negative feedback in the treatments *muFW* and *uniFW*. The regressions reported in Table A.3.3 provide supporting statistical evidence.

Under *muFW*, when a buyer did not cooperate at all (i.e. did not pay initially and also did not make good), then a seller who has received a positive feedback herself withdraws only in 16% of the cases, while she withdraws in 71% of the cases when she also has a negative feedback on her back. Under *uniFW*, this difference disappears, with only 3% / 10% of sellers with a positive/negative feedback eventually withdrawing, respectively. We observe very similar pattern for the opposite side of the market, for withdrawal behavior towards a seller who delivered quality of less than 50% and did not improve upon this

in the make-good stage. The negative feedback of these sellers is withdrawn in *muFW* in only 14% of the cases when the buyer had received a positive feedback, but in 60% of the cases when the buyer had received a negative feedback himself. Once again, this difference disappears in treatment *uniFW* where towards such an uncooperative seller it does not make a difference whether the buyer has received a positive or negative feedback herself (with withdrawal frequencies of 0% and 4%, respectively). These data are strong evidence that withdrawal behavior is strategic and highly dependent on own received feedback in *muFW*, while such considerations do not play a role in *uniFW*.

However, we also observe a difference in withdrawal behavior from (at least initially) uncooperative traders towards *cooperative* trading partners. A seller with an initially negative feedback agrees to withdraw an (unfair) negative feedback towards an initially paying buyer in 96% of the cases in *muFW*, but eventually withdraws only in 49% of the cases in *uniFW*. Once again we see similar patterns on the other market side, with the corresponding eventual withdrawal frequencies of a buyer with negative feedback towards a cooperative seller with an unfair negative feedback being 100% and 41% in treatments *muFW* and *uniFW*, respectively. This indicates that *uniFW* may also have some caveats, something that was not anticipated by our theoretical reasoning where we assumed honest feedback behavior absent any other monetary motives. However, as our analysis of aggregate behavior shows, this caveat may not have much weight on overall market behavior.

TABLE A.3.2: FEEDBACK WITHDRAWAL FREQUENCIES
CONDITIONAL ON INITIAL COOPERATION AND MAKE-GOOD

Frequency of seller's withdrawal	muFW		uniFW	
	Seller received		Seller received	
	neg	pos	neg	pos
Buyer paid initially	96%	82%	49%	71%
Buyer did not pay and made good	96%	70%	45%	40%
Buyer did not pay and did not make good	71%	16%	3%	10%
Frequency of buyer's withdrawal	**muFW**		**uniFW**	
	Buyer received		Buyer received	
	neg	pos	neg	pos
Seller delivered $Q_1 \geq 50\%$ and improved	100%	77%	41%	85%
Seller delivered $Q_1 \geq 50\%$ and did not improve	98%	52%	28%	34%
Seller delivered $Q_1 < 50\%$ and improved	83%	73%	64%	80%
Seller delivered $Q_1 < 50\%$ and did not improve	60%	14%	4%	0%

In Table A.3.3 we report results from Probit regressions predicting the decision of (agreeing to) withdraw a negative feedback based on whether the trading partner had improved their initial payment/quality choice and whether the trader had received a negative feedback herself. For *muFW*, we find (as reported in Bolton et al., 2018) that when the trader has not received a negative feedback herself, then withdrawal is strongly conditioned on whether the partner has made good or not. On the other hand, when the trader has received a negative feedback herself, then there is a higher likelihood that the feedback is withdrawn unconditionally, with the correlation between withdrawal and make-good being significantly reduced.

Under *uniFW*, we have a two-step decision: the choice to agree to allow feedback withdrawal, and the choice to actually unilaterally withdraw the given feedback. As we observed above when discussing Table A.3.1, when the trader has received a positive feedback herself under *uniFW* (such that only the other has received a negative feedback), the trader mainly conditions the agreement to withdraw on make-good behavior, and then follows through with the actual unilateral withdrawal (such that the latter is not correlated with make-good behavior anymore; see coefficients on "Quality improved" and "Payment improved" in the four right-hand side regressions in Table A.3.3). When the trader has received a negative feedback, we see slightly different patterns for buyers and sellers. For buyers, the agreement to allow withdrawal is significantly higher and less conditioned on make-good when the buyer had received a negative herself. Instead, the conditionality is moved to the unilateral second stage of the withdrawal decision, where the buyer is now more critical and more likely to condition on withdrawal. For sellers, we also observe a higher likelihood to (unconditionally) allow feedback withdrawal when the seller had received a negative herself, but we do not detect significant effects on conditionality and second-stage behavior. However, our analysis for sellers also relies on a much lower number of data points (see Table A.3.3).

To sum up the detailed analysis of feedback and withdrawal behavior in this Appendix section, we find more detriment strategic behavior in the *muFW* as compared to the *uniFW* feedback systems. In both *muFW* and *uniFW*, when traders have given a negative feedback but have received a positive feedback themselves, they largely condition their agreement to feedback withdrawal on the make-good behavior of the other trader. When the traders have received a negative feedback themselves, then in both systems they are more likely to unconditionally agree to feedback withdrawal. However, while in *muFW* the mutual agreement automatically leads to the actual withdrawal, in *uniFW* traders have a second stage in the withdrawal process, where they unilaterally decide to actually withdraw the feedback or not. There we observe (at least for buyers, with too few observations for sellers) that traders move the conditionality to this second stage, preserving its incentive impact on make-good and cooperation. As a result of strategic anticipation of these

withdrawal behaviors, under *muFW* traders are more likely to give preemptive negative feedback in order to extort a withdrawal decision, something that is not possible under *uniFW*.

TABLE A.3.3: PROBIT REGRESSION OF THE LIKELIHOOD TO WITHDRAW ON OTHER'S MAKE-GOOD BEHAVIOR AND FEEDBACK CONDITION

Treatment	muFW		uniFW			
Dependent	B withdraws y/n	S withdraws y/n	B votes for WD option	S votes for WD option	B revises y/n	S revises y/n
Model	(1)	(2)	(3)	(4)	(5)	(6)
Baseline	B->S neg, S->B pos	B->S pos, S->B neg	B->S neg, S->B pos	B->S pos, S->B neg	B->S neg, S->B pos	B->S pos, S->B neg
Quality improved y/n	0.332***		0.441***		0.110	
	[0.108]		[0.047]		[0.098]	
Payment improved y/n		0.436***		0.257**		0.267
		[0.126]		[0.106]		[0.290]
B->S neg, S: neg	0.266**	0.445***	0.308***	0.342**	-0.443***	-0.387
	[0.126]	[0.074]	[0.058]	[0.137]	[0.086]	[0.252]
B->S neg, S->B neg × Quality improved y/n	-0.145		-0.281***			
	[0.124]		[0.096]			
B: neg, S->B neg × Payment improved y/n		-0.088		0.207	0.187*	0.110
		[0.176]		[0.127]	[0.104]	[0.324]
N	739	429	696	178	449	78
LL	-426.0	218.1	-345.4	-105.7	-167.6	-34.8

Notes: The table reports average marginal effects dy/dx with robust standard errors clustered at the matching group level, based on data from rounds 11-50 (omitting start and end effects). *, **, and ** denote significance at the 10%, 5%, and 1% level, respectively.

Appendix B. Experimental instructions
(translated from German original)

Treatments: noFW/muFW/uniFW

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 **[uniFW:** 6**][muFW:** 5**] [noFW:** 4**]** stages:

1. **Trade decision:** 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.
2. **Money transfer and quality decision**: The buyer decides to send his/her 100 ECU to the seller or not. At the same time, the seller chooses the quality of the product which s/he is sending to the buyer. The quality must be between 0% and 100%. Each quality percent costs the seller 1 ECU, and benefits the buyer by 3 ECU. So, for example,
   - if the quality is 0%, the seller has costs of 0 ECU and the buyer receives a product value of 0 ECU;
   - if the quality is 50%, the seller has costs of 50 ECU and the buyer receives a product value of 150 ECU;
   - and if the quality is 100%, the seller has costs of 100 ECU and the buyer receives a product value of 300 ECU.

   Once the buyer and seller made their decisions, both transaction partners are informed about each other's choice.
3. **Feedback:** Simultaneously, the buyer and the seller decide which feedback they want to submit on the transaction. The feedback can be either 'negative', or 'positive'. After both have given feedback, it will be shown on the screen to both transaction partners. The received feedback will also be displayed to transaction partners in subsequent rounds (see below).
4. **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.
5. **[muFW:** **Feedback revision**: This stage is only entered if at least one of the feedback ratings given in Stage 3 was negative. Simultaneously, both the buyer and the seller can decide whether they support to revise the feedback and turn both feedback ratings into 'positive' feedback. If both support the revision, then both feedback ratings will be made 'positive'. If only the buyer or only the seller or none supports the feedback revision, then the feedback given in Stage 3 remain unchanged.**] [uniFW:** **Feedback revision option**: This

stage is only entered if at least one of the feedback ratings given in Stage 3 was negative. Simultaneously, both the buyer and the seller can decide whether they support the option to revise feedback and turn 'negative' feedback ratings into 'positive' feedback. If both transaction partners support the revision option, both transaction partners can revise their feedback rating in stage 6 to "positive". If only the buyer or only the seller or none supports the feedback revision option, then the feedback given in Stage 3 remain unchanged.**]**

6. **[uniFW: Feedback revision**: This stage is only entered if both transaction partners support the option to change feedback to "positive". Simultaneously, both the buyer and the seller can decide whether they they want to revise their feedback to "positive". Transaction partners who have already given a "positive" feedback rating in stage 3, cannot revise their feedback rating. Following the feedback revision, both transaction partners are informed of the other's decision.**]**

After these **[uniFW:** 6**]| [muFW:** 5**] [noFW:** 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 both decided to trade:
– 100 ECU 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 both decided to trade:
+ 100 ECU if the buyer decided to send the 100 ECU to the seller
- Q with Q equaling the quality percent the s/he has chosen for the product, being between 0 and 100
}
Your final payoff from the experiment will be the sum of all round payoffs.
The number of feedback ratings a participant collected in previous rounds will be shown to his transaction partner at the beginning of the next round, before Stage 1. The display will show the number of positive and negative feedback ratings received in previous rounds, like this: "X positive feedback ratings and Y negative feedback ratings received in previous rounds".