Received: 16 February 2021 | Revised: 5 October 2022 | Accepted: 19 December 2022

DOI: 10.1111/jems.12512

ORIGINAL ARTICLE

Journal of Economics & Management Strategy WILEY

Fixing feedback revision rules in online markets

Gary Bolton[1] | Kevin Breuer[2] | Ben Greiner[3,4] | Axel Ockenfels[2]

[1]Managerial Economics, Naveen Jindal School of Management, University of Texas at Dallas, Richardson, Texas, USA

[2]Department of Economics, University of Cologne, Cologne, Germany

[3]Institute for Markets and Strategy, Vienna University of Economics and Business (WU Vienna), Vienna, Austria

[4]School of Economics, University of New South Wales, Sydney, Australia

Correspondence

Axel Ockenfels, Department of Economics, University of Cologne, Albertus Magnus Platz, D-50923 Cologne, Germany.

Email: ockenfels@uni-koeln.de

Funding information

Australian Research Council, Grant/Award Number: DP130104557; Deutsche Forschungsgemeinschaft, Grant/Award Numbers: EXC 2126/1—390838866, FOR 1371

Abstract

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.

1 | 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 & Resnick, 2012). Many online marketplaces employ conflict resolution systems to manage such conflicts. A widely used example is 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, G. Bolton et al. (2018) show how a feedback withdrawal mechanism that was widely used backfired and hampered 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 with the feedback withdrawal mechanism studied in G. Bolton et al. (2018), 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.

After completing our laboratory investigation, we surveyed platforms that motivated our initial studies in G. Bolton et al. (2018). We observed that all the marketplaces which used problematic mutual withdrawal feedback rules have abandoned that mechanism and adapted in the direction we propose here. Some, such as eBay, Etsy, and Discogs, have moved to a one-sided feedback system which makes strategic withdrawal issues obsolete (they often do not allow feedback withdrawal at all). Some, such as AirBnB, do not facilitate the revision of feedback. Others, such as ricardo.ch or Uber, use unilateral feedback withdrawal in their two-sided feedback system, similar to the new proposal evaluated in this paper. The feedback rules that come closest to our proposed mechanism seem to be in place on Upwork. This quick evolution of feedback withdrawal design suggests that many of these platforms may have indeed experienced the kind of problems that we previously identified, and consideration of design alternatives led them to similar conclusions that we reached. Our new study validates those choices in a highly controlled laboratory context.

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 et al., 2021; Tadelis, 2016; for surveys). Yet, much less attention has been given to the design of online market conflict resolution mechanisms. A number of studies have examined how a related conflict resolution mechanism, third-party arbitration, changes bargaining outcomes, for example, labor market disputes. These include Deck and Farmer (2007) who look into arbitration in bargaining over an uncertain value, G. E. Bolton and Katok (1998) 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. With respect to online dispute resolution, Vasalou et al. (2008) investigate the effect of apologies to repair trust in one-off online interactions. G. 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 mechanism can achieve the objective of coordination and trade facilitation without distorting incentives in feedback giving.

In Section 2 we describe our experimental design and procedures. Section 3 develops our two main hypotheses. Section 4 presents our experimental results, and Section 5 concludes. The Supporting Information Appendix includes robustness checks, experiment instructions, and additional analyses of the experimental data.

2 | 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, with fixed role assignment but random trader matching in each round. Table 1 displays the

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 does not agree, the round ends with $\pi_B = 100$ and $\pi_S = 100$
Transaction	Buyer decides whether or not to pay, $P_1 \in \{0 \text{ ECU}, 100 \text{ ECU}\}$. 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, $P_2 \in \{P_1, 100 \text{ ECU}\}$; 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	$\pi_B = 100 - P_2 + Q_2 * 3$, $\pi_S = 100 + P_2 - Q_2$

sequence of stage decisions taken in each round. Each round starts with a choice to engage in a trade (or not) by both traders, after observing each other's past feedback numbers. 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 (an integer number between 0% and 100%). Then both parties are informed about the decisions of their respective trading partner and submit feedback on the transaction. Feedback is a binary variable that 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 $Q_2 \geq Q_1$. The eventual round payoff of the buyer equals his endowment minus the price paid plus the value of the product scaled by the product's quality, that is, $\pi_B = 100 - P_2 + Q_2 * 3$. The seller's round income results from endowment plus received price minus costs for chosen quality provision, that is, $\pi_S = 100 + P_2 - Q_2$. Since a buyer's valuation for each percent product quality is three times higher than the seller's cost to produce that percent product quality, trade is efficiency enhancing, but subject to moral hazard on both sides.

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* change their negative feedback to positive, or not. In no treatment can positive initial feedback be made negative.

All data were collected in the Cologne Laboratory for Economic Research, between June and November of 2017. Participants were students from the University of Cologne recruited via the Online Recruitment System for Economic Experiments (Greiner, 2015). (For a survey of results on the generality of student behavior see Fréchette, 2015; Snowberg & Yariv, 2021.) The experiment was programmed in z-Tree (Fischbacher, 2007). Average payoffs were about EUR 20 plus a show-up fee of EUR 2.50. The original G. Bolton et al. (2018) sessions involved 128 participants, with two sessions each for conditions *noFW* and *muFW*. The new sessions used 192 participants, with three sessions each for treatments *noFW* and *uniFW*. Sessions comprised 32 participants each, who were assigned to matching groups of eight participants. Thus, our analysis relies on 20 independent markets/matching groups in the baseline *noFW*, eight matching groups in *muFW*, and 12 matching groups in *uniFW*. In our analysis we pool data from G. Bolton et al. (2018) with data from the new experiment sessions conducted between June and November 2017.

3 | TWO HYPOTHESES

As shown by G. Bolton et al. (2018), 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 is 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*, 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. The theory assumes that—absent a monetary or strategic motive to distort their feedback—people will report feedback honestly, and if they do have such a motive, they may in fact distort what they report.[1]

The hypothesis we study here is whether 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.

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 G. 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., Bohnet & Huck, 2004; Bohnet et al., 2005; G. E. Bolton et al., 2004; Bracht & Feltovich, 2009; Brown & Zehnder, 2007; Charness et al., 2011; Duffy & Feltovich, 2002, 2006; Duffy et al., 2013; Huck et al., 2010, 2012). 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.*

As a final note, the theoretical arguments for all the hypotheses above would be unchanged if the vote to allow feedback revision were dropped from the *uniFW* mechanism while maintaining the opportunity for unilateral revision (Table 1). We include this step in the *uniFW* treatment for purposes of control: If we observe a difference between *muFW* and *uniFW*, the sole attribution will be to the difference discussed here in the consequence of choosing to revise a given negative feedback.

4 | RESULTS[2]

4.1 | *uniFW* does not reduce payments and quality like *muFW* does

Figure 1 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 with 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 with *muFW*, *increases* the likelihood of initial (eventual) payment by 11 (12) percentage points.

The regressions reported in Table 2 Models (1) and (3) support these observations statistically. Here we report ordinary least-squares (OLS) models (Gillen et al., 2019). (Probit and Tobit specifications yield very similar results; see Supporting Information Appendix A.5.) The differences in initial payment frequencies are 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. (Nonparametric Wilcoxon rank-sum tests based on independent matching group averages support these conclusions.[4])

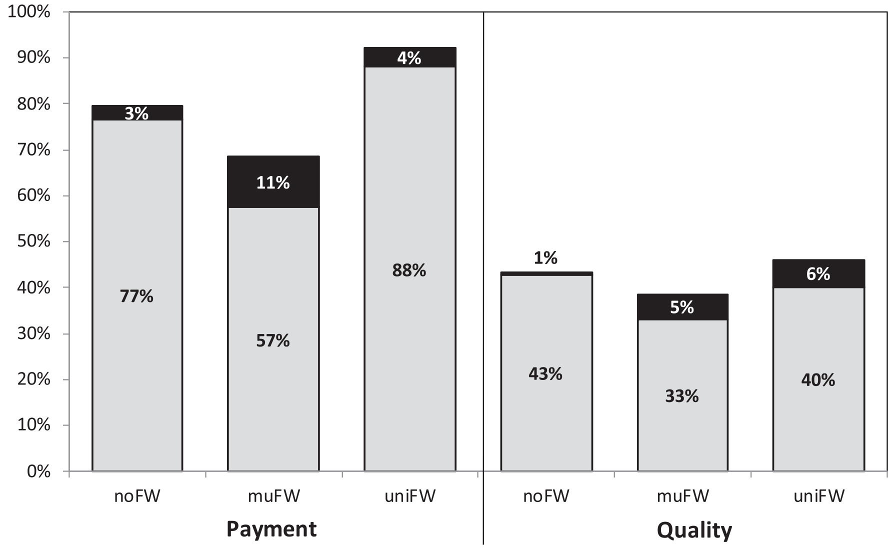

FIGURE 1 Average payment/revenues and quality/efficiency conditional on trade, across the three treatments. The figure reports initial payment and quality in the transaction stage (gray share of the bars), as well as additionally provided payments and quality in the make-good stage (black share of the bars). Numbers are based on rounds 11–50 in the experiment.

TABLE 2 OLS regressions of probability of payment and quality

Model dependent	(1) Initial payment	(2) Initial quality	(3) Final payment	(4) Final quality
Constant	0.905	0.509***	0.953	0.509***
	[0.035]	[0.019]	[0.036]	[0.018]
Round	−0.005***	−0.003***	−0.005***	−0.003***
	[0.001]	[0.001]	[0.001]	[0.000]
muFW	−0.191*	−0.096**	−0.108	−0.047
	[0.096]	[0.042]	[0.078]	[0.039]
uniFW	0.115***	−0.026	0.125***	0.026
	[0.042]	[0.037]	[0.036]	[0.022]
N	4945	4945	4945	4945
R^2	0.081	0.048	0.072	0.037
Postestimation test *muFW* = *uniFW*, *p* value	0.0024	0.1853	0.0031	0.0733

Notes: Regressions are based on data from rounds 11 to 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.

Abbreviation: OLS, ordinary least-squares.

For initial product quality (market efficiency), we observe a reduction of 11 percentage points with the *muFW* mechanism compared with 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 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 postestimation test in the last row of Table 2 Model 4). Results from nonparametric Wilcoxon rank-sum tests based on independent matching group averages are mostly consistent with the regression results.[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*

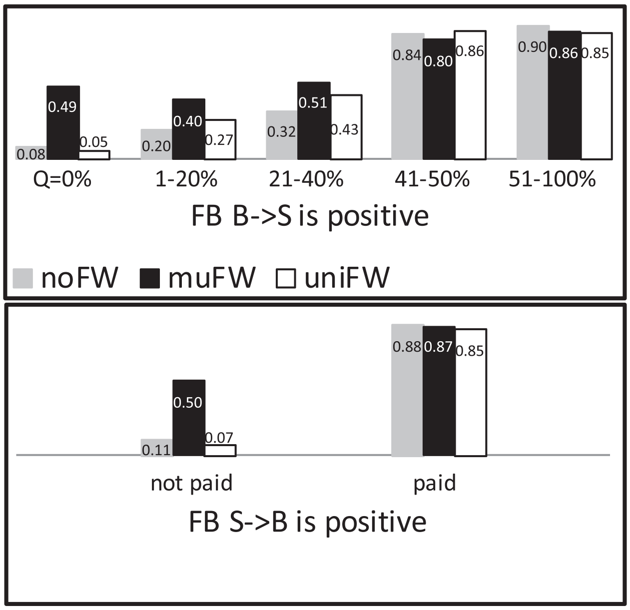

FIGURE 2 Eventual feedback conditional on trading partner's behavior. The figure reports the eventual share of positive feedback given by the buyers (sellers) conditional on the quality (payment) decision of the transaction partner, after make-good and withdrawal decisions. FB, feedback.

both in terms of payment and quality (Table 2, postestimation test). 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.

4.2 | *uniFW* does not distort feedback like *muFW* does

Figure 2 displays the frequency of positive feedback conditional on 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 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.

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 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.

The mitigated distortion of feedback in *uniFW* as compared with *muFW* (Table 2, postestimation test) is due to reduced incentives for strategic gaming of the feedback and withdrawal rules. To further illustrate this, Supporting Information 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,

TABLE 3 OLS regressions of positive feedback on quality/payment and treatment indicators

Dependent: Positive feedback	*B* -> *S* FB is positive (1)		*S* -> *B* FB is positive (2)	
Constant	0.108*	[0.055]	0.132***	[0.037]
Round	0.001	[0.001]	−0.001	[0.001]
Quality	0.013***	[0.001]		
Payment			0.769***	[0.040]
muFW	0.342***	[0.064]	0.395***	[0.051]
muFW × *Quality*	−0.007***	[0.001]		
muFW × *Payment*			−0.404***	[0.059]
uniFW	0.017	[0.093]	−0.037	[0.047]
uniFW × *Quality*	0.000	[0.002]		
uniFW × *Payment*			0.009	[0.065]
N	4945		4945	
R^2	0.280		0.362	
Postestimation test *muFW* = *uniFW*, *p* value	0.0011		0.0000	

Notes: Regressions are based on data from rounds 11 to 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% levels, respectively.

Abbreviation: FB, feedback; OLS, ordinary least-squares.

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 to extort a withdrawal decision, something that is not possible under *uniFW*.

4.3 | *uniFW* reduces variance in payoffs compared with *noFW* and *muFW*

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 rank-sum tests to assess whether the distributions of these standard deviations differ between treatments. Table 4 lists the averages of the calculated standard deviations across all matching groups of the respective treatments along with the corresponding *p* values.

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* ($p = .0176$ and .0136, respectively), while the other differences are not significant at the $p = .1$ level. Additional analysis shows that these patterns are stable across early rounds (11–30) and late rounds (31–50). In particular, we find no evidence that higher coordination failure in treatments *noFW* and *muFW* compared with *uniFW* disappears with experience (see Supporting Information Appendix A.6).

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 with when no feedback withdrawal system is present (or when *muFW* is at work). Thus, we find evidence in favor of Hypothesis 2 that *uniFW* can reduce uncertainty and facilitate expectation coordination.

TABLE 4 Buyer and seller profits and their average standard deviations across independent matching groups, and Wilcoxon rank-sum 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

Notes: The table reports average payoffs and standard deviations of collapsed data on the independent matching group level.

5 | CONCLUSION

How can a previously identified flaw in feedback revision rules in online markets be fixed? 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 the previously commonly used *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. Our new mechanism is thus the preferred choice. In fact, since we published our studies that identified the design flaws of the previous standard mechanism, many market platforms chose to adapt new mechanisms that are similar to the one proposed here (see Introduction).

While the work here focuses directly on 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 & Iyengar, 2009); namely, having dispute resolution available tends to reduce the incentives for actors to solve a problem in the first place (before 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.

ACKNOWLEDGMENTS

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. Greiner gratefully acknowledges funding from the Australian Research Council through Discovery Grant DP130104557. Open Access funding enabled and organized by Projekt DEAL.

DATA AVAILABILITY STATEMENT

Upon acceptance, all data and analysis code that yield the findings reported in this study will be made available in the supplementary material of the article.

ORCID

Axel Ockenfels http://orcid.org/0000-0003-1456-0191

ENDNOTES

[1] A model in G. 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, e.g., the buyer and seller differ in their expectations about what constitutes a "satisfactory quality level").

[2] We focus our analysis on rounds 11–50, as in G. Bolton et al. (2018), studying a running system rather than start-up or end-game effects. We provide more in-depth analyses in Supporting Information Appendix A and refer to them in this text where appropriate. In particular, in Supporting Information Appendix A.1 we present a direct comparison of the *noFW* baseline condition between the original G. 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. Further, in Supporting Information Appendix A.4 we replicate all tables and figures in the main text using all rounds 1–60, with qualitatively the same results. Supporting Information Appendix A.5 provides, as a robustness check, the respective results from Probit/Tobit models, which yield the same results as our linear (probability) models reported in the main text.

[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 with the older sessions (see the 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.

[4] p Values for *noFW* versus *muFW*, *noFW* versus *uniFW*, and *muFW* versus *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.

[5] p Values for *noFW* versus *muFW*, *noFW* versus *uniFW*, and *muFW* versus *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.

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

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. ISBN: 978 1 84720 3 32 8.

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

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

Bolton, G., Ferecatu, A., & Kusterer, D. (Forthcoming). Rate this transaction: Coordinating mappings in market feedback systems. *Management Science*.

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. E., Katok, E., & 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, 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

Bracht, J., & 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., & Zehnder, C. (2007). Credit reporting, relationship banking, and loan repayment. *Journal of Money, Credit, and Banking*, *39*(8), 1883–1918. https://doi.org/10.1111/j.1538-4616.2007.00092.x

Charness, G., Du, N., & Yang, C.-L. (2011). Trust and trustworthiness reputations 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. A., & Ockenfels, A. (2021). Market design, human behavior and management. *Management Science*, *67*(9), 5317–5348. https://doi.org/10.1287/mnsc.2020.3659

Deck, C. A., & Farmer, A. (2006). Bargaining over an uncertain value: Arbitration mechanisms compared. *Journal of Law, Economics, and Organization*, *23*(3), 547–579. https://doi.org/10.1093/jleo/ewm012

Duffy, J., & 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., & Feltovich, N. (2006). Words, deeds, and lies: Strategic behaviour 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., & 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. *Experimental Economics*, *10*, 171–178. https://doi.org/10.1007/s10683-006-9159-4

Fréchette, G. R. (2015). Experimental economics across subject populations. In J. H. Kagel, & A. E. Roth (Eds.), *The handbook of experimental economics* (Vol. 2, pp. 435–480). Princeton University Press.

Gillen, B., Snowberg, E., & Yariv, L. (2019). Experimenting with measurement error: Techniques with applications to the Caltech Cohort Study. *Journal of Political Economy*, *127*(4), 1826–1863. https://doi.org/10.1086/701681

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., & 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

Huck, S., Lünser, G. K., & 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

Ockenfels, A., & Resnick, P. (2012). Negotiating reputations. In R. Croson, & G. E. Bolton (Eds.), *The Oxford handbook of economic conflict resolution*. Oxford University Press.

Shavell, S. (1995). Alternative dispute resolution: An economic analysis. *The Journal of Legal Studies*, *24*, 1–28. https://doi.org/10.1086/467950

Snowberg, E., & Yariv, L. (2021). Testing the waters: Behavior across participant pools. *American Economic Review*, *111*(2), 687–719. https://doi.org/10.1257/aer.20181065

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

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Bolton, G., Breuer, K., Greiner, B., & Ockenfels, A. (2023). Fixing feedback revision rules in online markets. *Journal of Economics & Management Strategy*, 32, 247–256. https://doi.org/10.1111/jems.12512