Essays on Market Design

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

2022

vorgelegt von

Kevin Breuer, M.Sc.

aus Köln

Referent: Axel Ockenfels

Korreferenten: Peter Cramton, Alexander Westkamp

Datum der Promotion:

Acknowledgments

I am very grateful to my supervisor, Axel Ockenfels, for his continuous support, guidance and advice throughout my doctorate. This also applies to my second supervisor, Alexander Westkamp, who was always available for thoughtful feedback.

Furthermore, I am indebted to my co-authors Dirk Bergemann, Gary Bolton, Peter Cramton, Christoph Feldhaus, Ben Greiner, and Axel Ockenfels. I am very grateful for a trusting, helpful, and insightful collaboration on my joint research projects.

I would also like to thank my (former) colleagues who contributed to making my doctorate a joyful experience. Viola Ackfeld, Rebekka Cordes, Max Großmann, Kiryl Khalmestki, Felix Kölle, Melisa Kurtis, Susanne Ludewig-Greiner, Mark Marner-Hausen, Yero Ndiaye, Andreas Pollak, Uta Schier and Peter Werner helped me not only with advice, suggestions and feedback, but also with moral support and good company.

Moreover, I want to thank the research assistants at the chair. Fabian Balmert, Malte Bonart, Tobias Dahint, Celina Decker, Lena Holtmeyer, Nikolai Jäger, Aenne Läufer, Kirsten Marx, Markus Möller, Jan Prusvic, Carl Riemann, Robin Schmidt, Benedikt Tomaschko, Wolfram Uerlich and Rosa Wolf helped me with countless tasks and duties with eagerness, diligence and competence.

Last but not least, I would like to thank my wife, family and friends for their emotional support, for believing in me and, most importantly, for distracting me when necessary.

Contents

Introduction

"*Market Design* […] strives to understand how the design of marketplaces influences the functioning of markets" (Roth 2018, p. 1609). The simple but powerful rationale of market design is to improve markets by actively designing them, guided by economic theory, empirical data, and carefully designed economic experiments. In recent years, economists have been successful in designing a variety of institutions, including spectrum auctions, electricity markets, feedback systems, kidney exchanges, and school choice (Chen et al., 2020). This thesis consists of four chapters, all devoted to different aspects and areas of market design. Another unifying element of this thesis is the methodology. In all chapters, laboratory experiments are conducted, data are analyzed, and the results are linked to real-world applications. Laboratory experimental studies are a particularly useful tool in the context of market design. They are often compared to a wind tunnel, where the performance of existing designs is studied in a simplified environment or even new design ideas are tested in a controlled environment (Chen et al., 2020).

The first chapter looks at auction design. We investigate the puzzle behind the popularity of a non-binding soft reserve price in practice. Here, we use the laboratory as a "wind tunnel" to compare the performance of different existing auction designs in a controlled environment. Chapter two focuses on the design of feedback systems. In this chapter, we propose a small but very effective modification to existing feedback withdrawal mechanisms. Therefore, we use the possibility of laboratory experiments to test a new design idea that has not yet been implemented in practice and for which, of course, no field data are available. The third chapter is concerned with the area of school choice. Here, I investigate the value of fairness to participants in school choice markets, which can guide a market designer in choosing an appropriate algorithm. A laboratory experiment allows for the observation and control of student preferences that are typically unobservable in field data. Finally, chapter four focuses on norm information acquisition. When designing real-world institutions, incentives must be aligned with behavior in terms of underlying goals (Bolton and Ockenfels 2012). Therefore, social norms, which are known to be a powerful force influencing behavior, are of great importance for market design. We study how economic agents choose between different types of norm information in a social choice context with uncertainty. In the following, I provide a short summary of the four chapters of the thesis.

Chapter 1. The first chapter, entitled "Why is the soft reserve so popular? An experimental study", is joined work with Dirk Bergemann, Peter Cramton and Axel Ockenfels and investigates experimentally the soft reserve price auction that is particularly popular in auctions practice. All authors contributed equally to the project.[1] In this chapter, we show that a soft reserve auction may outperform the standard optimal auction in both efficiency and revenues. A soft reserve auction asks bidders to accept an opening price to participate in an ascending auction, but also allows alternative bids below the opening price if the opening is unacceptable. If no bidder accepts the opening price, the alternative bids are considered applying first-price auction rules. Soft reserves are commonly used in practice, yet under standard auction theory they do not affect the auction outcome. We show in a laboratory experiment that softening the reserve of the standard optimal auction with a binding hard reserve price indeed improves efficiency and revenue. Moreover, many bidders accept any profitable opening price. As a result, the soft reserve also outperforms the first-price auction. Robustness checks confirm the attractiveness of soft reserves to auctioneers.

Chapter 2. The chapter "Fixing feedback revision rules in online markets" is joined work with Gary Bolton, Ben Greiner and Axel Ockenfels and studies feedback withdrawal rules in online markets. All authors contributed equally to the project.[2] 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 minor change in the design of feedback withdrawal rules, allowing unilateral rather than mutual withdrawal, can reduce incentives for strategic gaming and improve coordination of expectations. This leads to less trading risk, more cooperation, and higher market efficiency.

Chapter 3. The third chapter is single authored and entitled "You don't get what I deserve – Priorities and Fairness in School Choice Problems". This chapter explores the value of

[1] All authors were equally involved in generating the ideas regarding research question, experimental design, paper design and statistical analyses. The experiment was planned and conducted by Kevin Breuer and statistical analyses were carried out by Kevin Breuer. All authors were equally involved in writing this draft.

[2] All authors were equally involved in generating the ideas regarding research question, experimental design, paper design and statistical analyses. The experiment was planned and conducted by Kevin Breuer and statistical analyses were carried out by Kevin Breuer and Ben Greiner. All authors were equally involved in writing this draft.

fairness in the context of school choice problems. In school choice problems the properties of fairness and efficiency are incompatible. Efficiency is undisputedly a desirable property, but in practice it is often sacrificed in favor of fairness. I investigate if participants in experimental matching markets value fairness and how the valuation of fairness depends on specific characteristics of the matching market. I find that a significant share of subjects refuses to consent to fairness violations and subjects endowed with a low priority consent more often in a matching market with random priorities compared to a market with performance-based priorities. This indicates the importance of fairness in school choice markets.

Chapter 4. The final chapter, named "Choosing norm information: what other people do or what they think is ought to be done", is joined work with Christoph Feldhaus and investigates norm information acquisition. All authors contributed equally to the project.[3] We experimentally investigate people's choice between information regarding what others do and what others think is ought to be done in a social choice context in which a decision-maker can take away money from a charity. This choice is made just before they themselves act in the same social choice context, but without knowing the charities identity. We vary (i) the role of the person that chooses between the two types of information, she is either a decision-maker or a choice-architect, and (ii) the probability that her decision is actually implemented. We observe that most participants choose to be informed about what others do rather than what they think is ought to be done, irrespective of their role. However, we further observe that the share of decision-makers and choice-architects that choose information on what is ought to be done increases when their own decision is more likely to be implemented. We discuss a potential hypothesis in line with this observation.

[3] All authors were equally involved in generating the ideas regarding research question, experimental design, paper design and statistical analyses. The experiment was planned and conducted by Kevin Breuer and statistical analyses were carried out by Kevin Breuer with help from Christoph Feldhaus. All authors were equally involved in writing this draft.

Chapter 1

Why is the soft reserve so popular? An experimental study

Co-Authors: Dirk Bergemann, Peter Cramton and Axel Ockenfels*

Abstract

We show that a soft reserve auction may outperform the standard optimal auction in both efficiency and revenues. A soft reserve auction asks bidders to accept an opening price to participate in an ascending auction, but also allows alternative bids below the opening price if the opening is unacceptable. If no bidder accepts the opening price, the alternative bids are considered with first-price auction rules. Soft reserves are commonly used in practice, yet under standard auction theory they are irrelevant to the auction outcome. A laboratory experiment shows that softening the reserve of the standard optimal auction indeed improves efficiency and revenue. Moreover, many bidders accept any profitable opening price. As a result, the soft reserve also outperforms the first-price auction. Robustness checks confirm the attractiveness of soft reserves.

* Bergemann: Department of Economics, Yale University, New Haven, CT 06511, dirk.bergemann@yale.edu; Breuer, Cramton and Ockenfels: Department of Economics, University of Cologne, D-50923 Cologne, Germany, kevin.breuer@wiso.uni-koeln.de, cramton@umd.edu, ockenfels@uni-koeln.de. Support from Center for Social and Economic Behavior (C-SEB), the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany´s Excellence Strategy – EXC 2126/1– 390838866, and from the European Research Council (ERC, European Union's Horizon 2020 research and innovation programme, GA No. 741409) is gratefully acknowledged.

I. Introduction

This chapter is a preliminary version of Bergemann et al. (2021). It contains the general research idea, an initial set of data collection and analysis, but the model and theorical foundation were not completed at the time of submission of this dissertation and are therefore not included in this chapter.

Standard auction theory recommends the seller to set a reserve price that exceeds her valuation for the item being auctioned (Myerson 1981, Riley and Samuelson 1981). If the seller can credibly commit to the reserve price, she can, in expectation, extract a profit from the highest bidder that goes beyond the profit which would result from mere competition among bidders. While theory requires the reserve price to be binding, in practice it is typically not (Burguet and Sákovics 1996). Ascending auctions often start at an opening price, which is however lowered if no bidder accepts the opening. In many other settings, re-auctioning is common. For instance, governments set reserve prices in spectrum auctions, but if the spectrum goes unsold the spectrum is re-auctioned later. Sometimes bids below the reserve are explicitly allowed and part of the auction rules. One example is eBay's best offer program (Huang et al. 2013), which allows sellers to utilize a reserve price while simultaneously allowing offers below the reserve. Informal auctions often also have this feature. For example, in selling a home, the seller sets a listing price. In the event multiple buyers accept the listing price, an ascending auction results, whereas, if the listing price is unacceptable to all, the seller entertains bids below the listing price.

The predominant use of non-binding reserve prices is puzzling, because a non-binding reserve price renders the whole idea of a reserve useless (Zeithammer 2019). In this paper we look into this puzzle by showing experimentally that a non-binding, "soft" reserve can be behaviorally superior to a conventional hard reserve. More specifically, in our framework, a soft reserve auction asks bidders to accept an opening price to participate in an ascending auction, but also allows alternative bids below the opening price if the opening is unacceptable. If no bidder accepts the opening price, the alternative bids are considered with first-price auction rules.

The soft reserve obviously dominates the standard optimal auction with respect to efficiency: With a hard reserve, the item may be retained by the seller, even when the seller's valuation is lower than a bidder's valuation. This cannot happen with a soft reserve auction, because a bidder who is unwilling to accept the soft reserve may bid a lower amount. But does a hard reserve increase the revenue? Previous empirical and theoretical research has shown that many people feel regret, which can in turn substantially affect bidding and auction revenues. In particular, based on Engelbrecht-Wiggans (1989) who introduced bidder regret into an auction framework, Ockenfels and Selten (2005), Filiz-Ozbay and Ozbay (2007) and Engelbrecht-Wiggans and Katok (2006, 2008, 2009) provide evidence that regret explains overbidding and high revenues in first-price auctions (see Kagel and Levin 2016 for a survey).[1] Following this literature, softening the reserve price in an ascending auction therefore may increase efficiency and revenue if bidders feel sufficient regret when they lose at an affordable price. We confirm the prediction in a laboratory experiment.

The previous literature does not offer a clear prediction regarding the comparison of the first price auction and soft reserve price auction. We find our laboratory first-price auction to be clearly outperformed by soft reserve auctions with an intermediate opening price. The reason appears to be the specific kind of regret from rejecting a profitable opening price—learning a competitor accepted the opening price. This regret appears to be so salient that many bidders accept *any* profitable opening price.

Finally, as additional behavioral robustness checks, we investigate the performance of a soft reserve price in high and low competitive scenarios, and we show that sellers select a soft reserve to a hard reserve if given the choice.

Our experiment helps explaining the puzzle of widespread use of the soft reserve in practice. It does so in a framework that gives the hard reserve the best shot, because there are at least three important advantages of softening the auction reserve price that our analysis intentionally abstracts away. First, in practice, a seller who wishes to use a hard reserve may find it difficult to commit to not selling the item if the reserve is not met (Skreta 2015, Caillaud and Mezzetti 2004). With a soft reserve, on the other hand, no such commitment is needed. In our theoretical and laboratory comparisons, this advantage of the soft reserve is controlled away because all hard reserves are credible and costless commitments. Second, a hard reserve tends to discourage participation (Bajari and Hortacsu 2003). Some auction houses, such as Ritchie Brothers, even require no reserve at all in order to attract more bidders and provide certainty of sale. With a soft reserve, on the other hand, even buyers with a valuation below the opening price can

[1] Greenleaf (2004) and Davis et al. (2011) find evidence for regret on the sellers' side, and that this may lower the reserve price. Other studies relate the optimal choice of reserve prices to behavioral phenomena. Rosenkranz and Schmitz (2007) analyze the effect of reference-point utility on optimal reserve prices and find, contrary to standard auction theory, that the optimal hard reserve increases in the number of bidders. Crawford et al. (2009) explore how the optimal reserve may be affected by "level-k" thinking.

participate. Because in our theoretical and laboratory comparison participation is exogenous and kept constant across treatments, this advantage of a soft reserve cannot contribute to our findings. Third, in practice, bidders often act on behalf of others and are driven by a desire to avoid blame, which may come with regretful outcomes. Accepting a profitable opening price safely avoids any regret and blame from others. There are no such agency issues in our model and experiments. Because our model and our experiment are designed to identify the potential role of regret, abstracting away from these other advantages, they tend to underestimate the attractiveness of soft reserves.

The reminder of the paper is organized as follows. Section II describes the experimental design and procedures. Next, Section III analyzes the laboratory results and Section IV discusses the role of regret in explaining our results. Finally, Section V concludes.

II. Experiment design and procedures

Table 1 illustrates our laboratory auction design choices, which resemble the way soft reserves have been modeled in the context of selling display advertisement (Zeithammer 2019). A soft reserve auction asks bidders to accept an opening price to participate in an ascending auction, but also allows alternative bids below the opening price if the opening is unacceptable. If no bidder accepts the opening price, the alternative bids are considered with first-price auction rules.

TABLE 1: LABORATORY SOFT RESERVE AUCTION

Stage I	Auctioneer announces an opening price of s.
Stage II	Bidders decide if they want to participate in a second-price auction with a minimum bid of s.
Stage III	Participating bidders compete in the second-price auction by submitting a bid equal to or larger than s. Non-participating bidders submit a bid between 0 and smaller than s.
Payment Rule	The highest bid in the second-price auction wins and pays the second highest bid (or, in case of a single bidder, the opening price). If no bidder participates, the highest bid smaller than s wins, and the winner pays her winning bid.

In practice, an ascending auction instead of a sealed-bid second-price auction would follow if multiple bidders accept the opening price. Our use of a second-price auction in Stage III is for convenience in the laboratory. Second-price auctions can be conducted much more quickly in the laboratory than the same number of ascending auctions. In our private value setting, a second-price auction yields nearly the same outcome as an ascending auction, both in theory and in the laboratory, when properly explained (see Ariely et al. 2005, Shachat and Wei 2012). Indeed, bidding in our second-price auction is consistent with what we would expect, both theoretically and from previous laboratory studies of ascending auctions. Thus, we anticipate that our results should apply when Stage III is an ascending auction, rather than a second-price auction. If at all, the common preference among bidders for ascending auctions (Cramton 1998) suggests the use of an ascending auction may strengthen our results.

The laboratory bidding environment otherwise follows standard procedures. Two bidders compete for a single item. We focus on the two-bidder case, because reserve prices are especially important when there are few bidders. Private valuations of the bidders are independently and uniformly distributed between 0 and 100 ECU (Experimental Currency Units). Valuations are randomly drawn before the experiment and are identical across all sessions and treatments. Each subject participates only in a single treatment (between subject design) and competes for 50 rounds within a strangers' matching protocol. Sellers are computerized and the set of auction rules is exogenously given. Subjects accumulate profits during all 50 rounds and the final payoff is converted to Euros and paid out immediately after the experiment. Sessions lasted around 90 minutes and the average payoff was approximately 9.90 € with a standard deviation of 3.38 €. After each round, bidders learned whether they won the auction, the final price of the item and their own profit.

Within this framework, we have four different treatments, each corresponding to a specific set of auction rules. We compare two different soft reserve auctions (SOFT50 and SOFT66) to two benchmark cases (HARD50 and FPA). Table 2 summarizes the parameters of the auctions along with null hypotheses based on standard behavioral assumptions.

Our first benchmark, HARD50, is a second-price auction with a hard reserve of 50 ECU. In a second-price auction, the highest bid wins the auction but pays only the second-highest bid. Under standard assumptions, a hard reserve h = 50 ECU maximizes revenue; the expected revenue is 41.67 ECU. In comparison, under the same assumptions, all other auction formats yield a smaller expected revenue of 33.33 ECU.

From a standard economic theory perspective, a soft reserve is inferior to hard reserve in terms of revenue (Zeithammer 2019). Nevertheless, the popularity of the soft reserve in the field indicates advantages which may not be captured by the theory. To allow direct comparison with the 'optimal' hard reserve in HARD50, we chose a soft reserve s = 50 ECU in SOFT50.

Our second benchmark treatment, FPA, is a standard first-price auction. Bidders submit bids between 0 and 100 ECU. The bidder submitting the highest bid wins the auction and pays her bid. The FPA is an interesting benchmark for the performance of our soft reserve auctions for two reasons. First, under standard assumptions and by the revenue equivalence theorem, it should lead to the same revenue and efficiency as our soft reserve auctions. (Indeed, because in standard equilibrium analysis bidders are expected to bid one-half of their valuation in our auction environment, standard theory predicts that *no* bidder in our soft reserve auctions should ever be willing to accept the opening price, so that we should observe the same bidding and outcome as in the FPA.) Second, because the FPA is known to produce high revenues by exploiting regret-averse bidders.

Finally, we compare SOFT50 with SOFT66, with s = 66.67 ECU, to see whether revenues increase in s. While like in SOFT50, no bidder is ever expected in standard equilibrium to accept the opening price in SOFT66, we note that s = 66.67 ECU maximizes revenue if we make the extreme assumption that all bidders with a valuation exceeding the soft reserve accept it (see Appendix B for details). As we will see later, many bidders behave as if they follow this simple strategy.

TABLE 2: TREATMENTS

	HARD50	**SOFT50**	**SOFT66**	**FPA**
Payment rule	2nd Price	1st and 2nd Price	1st and 2nd Price	1st Price
Type of reserve	Hard	Soft	Soft	Soft
Level of reserve	50 ECU	50 ECU	66.67 ECU	100 ECU
Revenue*	41.67 ECU	33.33 ECU	33.33 ECU	33.33 ECU
Efficiency *	75%	100%	100%	100%

Notes: *Refers to standard auction theory predictions assuming rational and risk neutral profit-maximizing bidders.

Besides our four main treatments, we conducted an additional treatment (ENDO) in which we endogenize sellers' choices of their preferred type of reserve. This treatment serves as an additional robustness check: Even when the soft reserve format empirically outperforms the other formats, sellers might not correctly anticipate this, or for some other reason might still want to choose one of the prominent textbook auction formats.

In each period of ENDO, sellers choose the preferred type of reserve, either a hard or soft reserve, and the level of the reserve in $\{0; 33.33; 50; 66.67; 100\}$. This menu allows for a broad range of different auction setups. It includes pure first-price (s = 100), pure second-price ($s = h$ = 0), and auctions with soft and hard reserve prices at intermediate levels.

All sessions were conducted between December 2016 and October 2017 in the Cologne Laboratory for Economic Research (CLER). Participants were students from the University of Cologne invited via ORSEE (Greiner 2015). The experiment was programmed with z-tree (Fischbacher 2007). We conducted two sessions for each of the main treatments, with exogenous sellers, consisting of 32 (with one exception of 28) participants in each session. Participants were randomly matched within matching groups utilizing a stranger's matching protocol. One matching group consisted of four bidders, thus we collected 16 independent observations for each treatment. In HARD50 we only have 15 independent observations because some invited participants failed to attend. For ENDO, we collected data from 96 students in four sessions.[2] One matching group consisted of four bidders and two sellers. Thus, we have 16 independent observations for the endogenous seller treatment. In total, we collected 15,800 bids and 1,600 reserve price decisions from 348 subjects.

III. Laboratory results

We first investigate the impact of softening the reserve on efficiency and revenue. Next, we compare FPA and the SOFT50 and Soft66 auctions. We finally provide robustness checks. If not explicitly stated otherwise, all reported p-values for our main treatments come from non-parametric Wilcoxon rank-sum tests based on independent matching group averages. Test statistics in the ENDO treatment are, if not stated otherwise, based on Wilcoxon signed-rank test based on independent matching groups.[3]

III.1. Softening the reserve increases efficiency and revenue

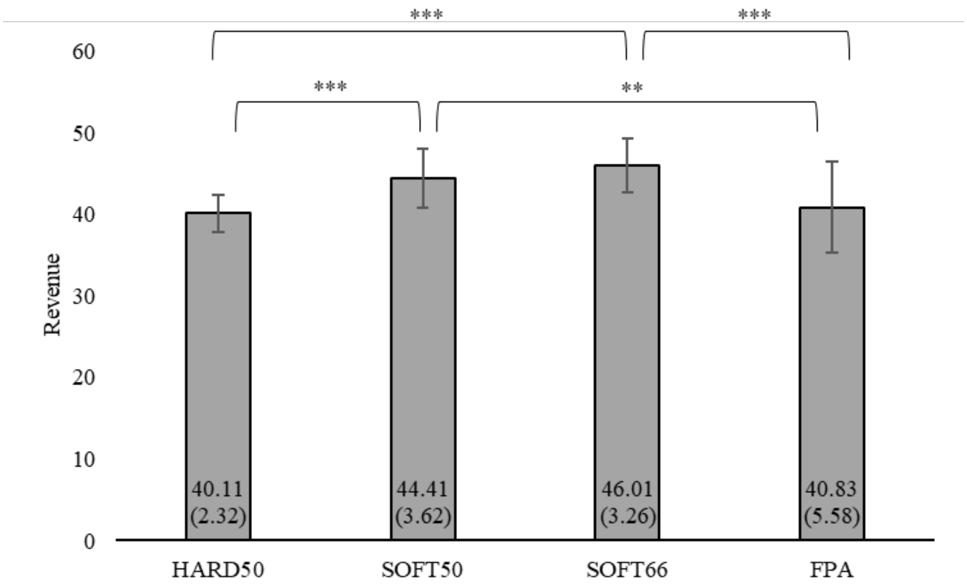
Figure 1 compares revenues across our treatments. While the revenue from HARD50 is with 40.11 ECU slightly lower than the predicted 41.67 ECU ($p = 0.036$ Wilcoxon signed-rank test), the predicted 33.33 ECU for the soft reserve auctions very strongly underestimate the actual revenues of 44.41 ECU in SOFT50 and 46.01 ECU in STOFT66 ($p < 0.001$ for both

[2] We conducted one more session which crashed after 15 minutes and had to be restarted. We do not use the data of this session.

[3] We only note here that our results do not change as subjects gain experience over the 50 rounds of play, and refer the interested reader to our analyses in Appendix A.

treatments). Comparing HARD50 to SOFT50, we find that softening the reserve does not decrease, but in fact increase revenue by 10% ($p = 0.003$). This finding provides strong support against the prediction of standard economic theory.

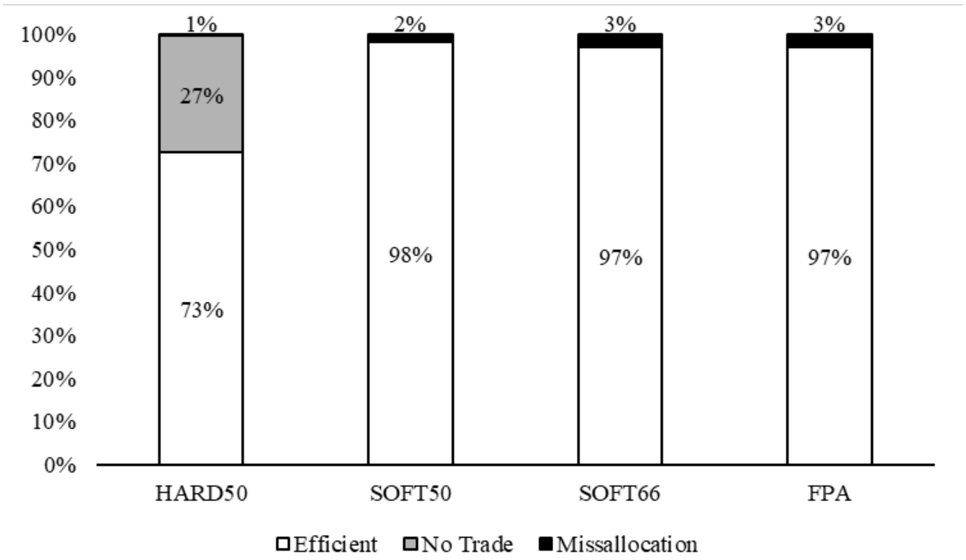
FIGURE 1: REVENUE



Notes: The figure reports average revenue and standard errors for each treatment on the observation group level. Significance levels are based on pair-wise rank-sum tests and *, **, and *** denote significance at the 10%, 5%, and 1% level, respectively.

Regarding efficiency, standard theory predicts that the soft reserve auctions, including FPA, increase efficiency over hard reserve auctions. Figure 2 shows efficiency levels as a percentage of maximal efficiency. The prediction is strongly confirmed ($p < 0.001$ for SOFT50, SOFT66 and FPA). As illustrated in the figure, the main reason for the large differences in efficiency is that a hard reserve – unlike a soft reserve – prevents the object from being sold when the hard reserve exceeds all bidders' valuations. This is predicted to happen in 25% of all HARD50 auctions, and actually happens in 26% of the relevant cases in our laboratory auctions. Inefficiency can also occur when the object goes to the bidder with the lower valuation. Similar to previous evidence, however, this happened rarely and it occurs less often for bidding under second-price rules than for bidding under first-price rules in our auctions, because second-price bidding is predicted to be in dominant strategies (Pezanis-Christou 2002). Indeed, the share of efficient allocations (95% in HARD50, 91% in SOFT50, 86% in SOFT66 and 87% in FPA) decreases monotonically as the share of bids submitted under first-price auction rules decreases (0% in HARD50, 59% in SOFT50, 81% in SOFT66 and 100% in FPA). The corresponding efficiency loss in percentage points of maximum welfare are displayed in black in Figure 3. All treatment differences are significant at least at the 5% level, except for FPA vs. SOFT66 (p = 0.8802) (see Appendix Table A.1 and A.2 for all efficiency rates and p-values).

FIGURE 2: EFFICIENCY



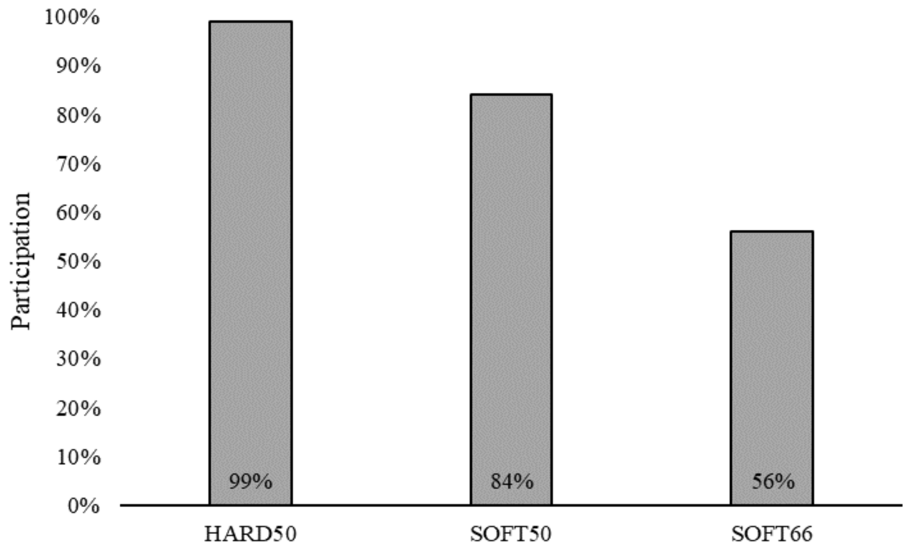
Notes: The figure reports for each treatment achieved efficiency as share of maximum achievable efficiency (white), efficiency loss due to no trade (grey) and efficiency loss due to allocation to the low valuation bidder (black).

Summing up, our evidence shows that softening the reserve auction can improve both efficiency and revenue.

Comparing SOFT50 with SOFT66 we find that SOFT66 indeed increases revenue by 3.6 percentage points, yet the difference is not statistically significant ($p > 0.227$). However, unlike predicted by simple regret, SOFT50 and SOFT66 both outperform FPA (which performs similar to HARD50) in terms of revenue ($p = 0.029$ for SOFT50 and $p = 0.003$ for SOFT66).

The reason for the dominant performance of SOFT50 and SOFT66 is many bidders' very strong willingness to accept the soft reserve whenever the valuation exceeds the reserve. To see this, recall that standard theory predicts no bidder ever to commit to the reserve in SOFT50, let alone SOFT66. As Figure 3 shows, 84 % of subjects in SOFT50 and 56 % of subjects in SOFT66 whose valuation exceed soft reserve accept the opening and the prediction of no participation can be strongly rejected for both treatments ($p < 0.001$).

FIGURE 3: PARTICIPATION



Notes: The figure reports for HARD50, SOFT50 and SOFT66 the share of subjects with a valuation above the hard, respectively soft, reserve price who participate in the auction.

III.2. Robustness checks

We conduct two robustness checks. First, we analyze if the advantages of a soft reserve hold both, in high and low competitive settings. Second, we study whether sellers prefer a soft to a hard reserve if given the choice.

The revenue equivalence theorem tells us the special circumstances in which auction formats are *ex ante* revenue equivalent. Auction revenues still may differ *ex post*. Which auction format yields the highest revenue depends on the realization of valuations. For example, the second-price auction performs especially well if the two valuations are close while the first-price auction is particularly useful if the two valuations are far apart. We will call the former scenario "high competition" and the latter "low competition." In practice, valuations of bidders might still be unknown, but sellers often have some idea of how competitive they expect the auction to be. Sellers could tailor the chosen auction format accordingly which potentially implies moving away from a soft reserve. Hence it is important to explore if the soft reserve price still holds pace in high and low competitive scenarios.

In our framework the expected difference between valuations of two bidders is 33.33 ECU. Thus, we define bidding in auctions with a valuation difference smaller than 33.33 ECU as high competition, and bidding in auctions with a valuation difference greater than 33.33 ECU as low competition. The following analyses also includes simulated revenue for the SPA.[4] The SPA is an important benchmark because of its high revenue in case of high competition. Table 3 displays mean revenues for the different auction formats and scenarios.

TABLE 3: REVENUE IN CASE OF HIGH AND LOW COMPETITION

Mean	HARD50	SOFT50	SOFT66	FPA	SPA
All	40.11	44.41	46.00	40.83	32.87
Low competition	47.90	48.15	53.93	48.23	23.18
High competition	34.08	41.55	39.81	35.14	40.41

Notes: The table reports average revenue on the observation group level across all periods (Full), subset of periods with |valuation bidder 1 – valuation bidder 2| > 33.33 (low competition) and subset of periods with |valuation bidder 1 – valuation bidder 2| ≤ 33.33 (high competition). The data for the second-price auction (SPA) is simulated based on the assumption bidders follow their weakly dominant strategy of bidding valuation.

With low competition, the FPA revenue is slightly yet insignificantly higher than the HARD50 revenue (p = 0.843) and the SOFT50 revenue (p = 0.999), yet significantly smaller than the SOFT66 revenue (p < 0.001). In the case of high competition, the SPA outperforms the FPA (p = 0.002) as well as HARD50 (p < 0.001) but is not significantly different from the soft reserve treatments (p = 0.477 for SOFT50 and p = 0.453 for SOFT66). That is, SOFT50 and SOFT66 perform at least as good as the best competitor in each given scenario, and in that sense the soft reserve shows robust performance in terms of revenue, regardless of competition.

Our second robustness check concerns the seller side of the market. In the additional treatment ENDO, we let sellers make a choice in each period about the auction format as explained in Section 3. Table 4 summarizes the menu of choices, which includes the first-price auction, the second-price auction and auctions with soft and hard reserve prices at intermediate levels and it also shows descriptive statistics.

[4] We simulated revenue based on the valuation distribution of the experiment and the assumption of bidders following their weakly dominant strategy of bidding valuation (consistent with our observed bidding behavior).

TABLE 4: SELLER DECISION

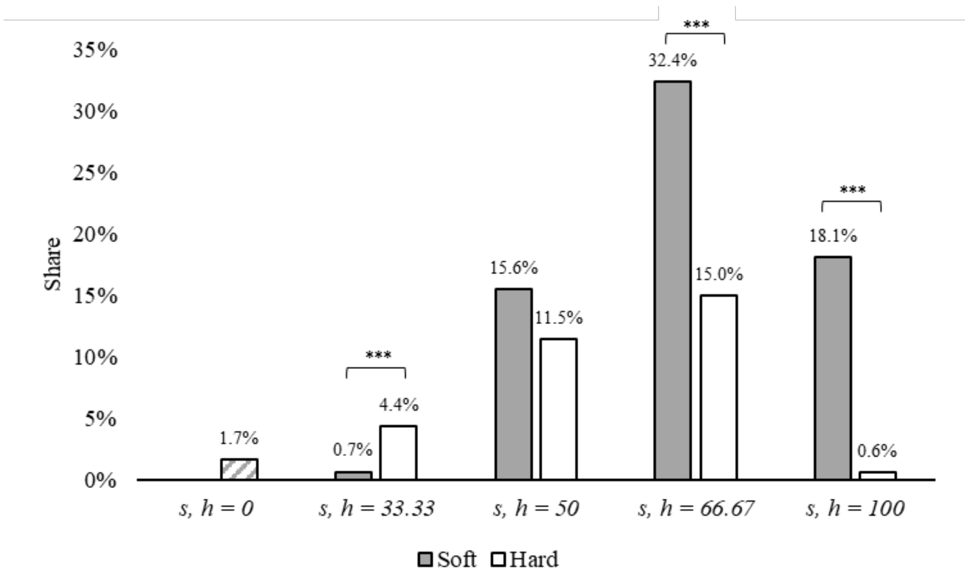
Type of Reserve	Level of Reserve				
Soft	$s = 0$	$s = 33.33$	$s = 50$	$s = 66.67$	$s = 100$
	SPA	SOFT33	SOFT50	SOFT66	FPA
Share	0%	0.69%	15.56%	32.44%	18.13%
Revenue	-	32.55	43.86	51.18	49.28
Efficiency	-	0.987	0.977	0.987	0.987
Hard	$h = 0$	$h = 33.33$	$h = 50$	$h = 66.67$	$h = 100$
	SPA	HARD33	HARD50	HARD66	-
Share	1.69%	4.38%	11.50%	15.00%	0.63%
Revenue	27.61	33.22	39.16	38.59	0
Efficiency	0.970	0.820	0.725	0.563	-

Notes: The table reports summary statistics for the different available auction type choices. In brackets the specific auction menus are labeled either as first-price auction (FPA), second-price auction (SPA) or either auctions with a soft or hard reserve price. Revenue and efficiency are displayed on the individual level and efficiency is measured as share of maximal achievable efficiency.

Across all auctions, a large majority of sellers prefer the soft reserve (66.81%) over the hard reserve (33.19%) ($p < 0.001$). The same holds if we restrict ourselves to intermediate reserve price levels (s,h ={33.33; 50; 66.67}), where 60.83% of our sellers prefer the soft reserve while 39.17% prefer the hard reserve ($p = 0.049$).

Not only do sellers prefer the soft reserve, they also choose higher soft than hard reserves. The average hard reserve is 53.70 ECU, slightly above the optimal reserve for risk neutral bidders of 50 ECU, while the average soft reserve is 71.80 ECU ($p < 0.001$). If we restrict ourselves to intermediate reserve levels, the average hard reserve is 55.36 ECU and the average soft reserve is 60.87 ECU ($p = 0.08$).

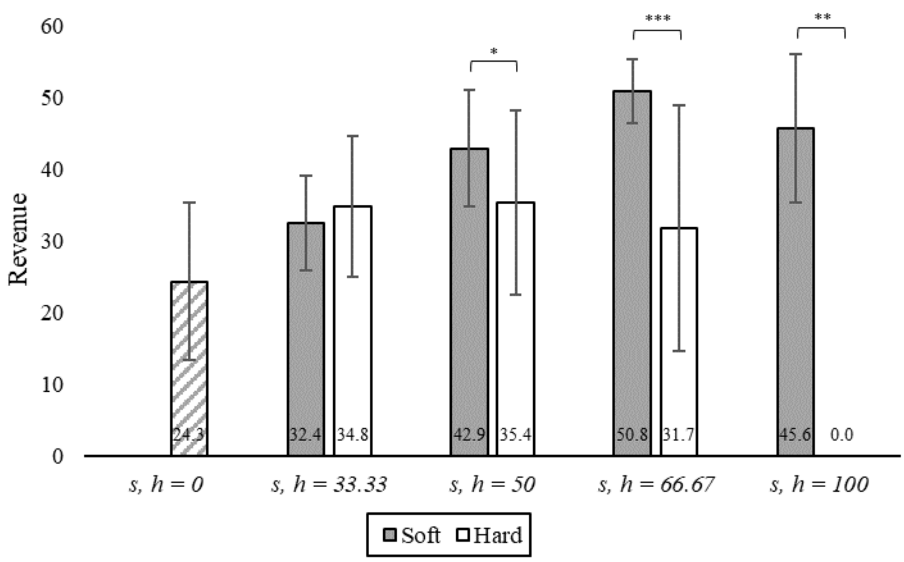
FIGURE 4: POPULARITY OF RESERVE TYPES IN THE ENDO TREATMENT



Notes: The figure reports how often different auctions types have been chosen by sellers. For $s,h = 0$ we only report a single bar since both auctions are identical. Significance levels are based on Wilcoxon signed-rank tests on the observation group level and *, **, and *** denote significance at the 10%, 5%, and 1% level, respectively.

Figure 4 shows the relative frequencies with which sellers choose soft and hard reserves across reserve price levels. Here, we exclude the very rarely chosen SPA because they can be equally implemented by a soft and a hard reserve of $s = h = 0$. The figure shows that a low hard reserve of 33.33 ECU is more popular than a correspondingly low soft reserve, yet with about only 5% of all cases, low reserves are hardly chosen at all. The absolute and relative attractiveness of the soft reserve increases in s. For a reserve price of 50 ECU, the soft reserve is more popular, although the effect is not statistically significant ($p = 0.736$). For $h = s = 66.67$ the difference becomes highly significant in economic (more than twice as many sellers choose the soft reserve) and in statistical terms ($p = 0.035$). Not surprisingly, the difference of the popularity between corresponding soft and hard reserves becomes even larger for $s = h = 100$, because the $h = 100$ auction is obviously dominated by all other auction formats.

FIGURE 5: REVENUE AND RESERVE PRICE LEVEL IN THE ENDO TREATMENT



Notes: The figure reports average revenue and standard errors for different reserve price levels on the observation group level. For $s=h=0$ we only report a single bar since both auctions are identical. Significance levels are based on Wilcoxon signed-rank tests and *, **, and *** denote significance at the 10%, 5%, and 1% level, respectively.

As suggested by our theoretical and experimental results in our main experiment, sellers' choices reflect that soft reserves generally lead to higher revenues. Figure 5 displays the average revenue for the different levels of soft and hard reserves in our ENDO treatment. Soft reserve auctions never perform significantly worse and perform better than hard reserve auctions for high enough reserves. Comparing reserve types among the different reserve price levels we find weakly significantly higher revenue for a soft reserve of 50 ECU (p = 0.063) and highly significant higher revenue for 66.67 ECU (p = 0.001), but no significant difference for a reserve level of 33.33 (p = 0.399). Computing the average revenue across all reserve price levels, we find the soft reserve price revenue (48.29 ECU) trumps the hard reserve revenue (36.11 ECU) by a significant margin ($p < 0.001$). Restricting the data to intermediate levels of reserves only, the numbers are 48.26 ECU for the soft reserve and 37.31 ECU for the hard reserve ($p < 0.001$).

IV. The SRA and regret

Famously, the revenue equivalence theorem postulates the first-price, second-price and soft-reserve auction all to yield the same revenue under standard assumptions. Increasing revenue beyond this threshold requires a hard reserve (Myerson 1981, Riley and Samuelson 1981). We have shown experimentally, that contrary to theory predictions, a soft reserve outperforms a hard reserve in terms of efficiency as well as revenue. But what is the mechanism behind the success of the soft reserve? Previous research has shown, unlike predicted by standard theory, the FPA yields higher revenues than the SPA in the laboratory (Cox et al., 1982 and Kagel and Levin, 2016). The literature has argued in favor of regret as a driver of this observation. Participants in the FPA dislike losing at an affordable price (loser's regret) as well as leaving money on the table (winner's regret). Importantly, winner's regret seems to outweigh loser's regret, therefore increasing bids in the FPA compared to the regret free SPA (Ockenfels and Selten 2005, Filiz-Ozbay and Ozbay, 2007). The relative strength of both types of regret depends on exposure to corresponding feedback. Providing bidders with feedback about the highest loosing bid, makes winners regret salient, while informing bidders about the winning bid shifts attention to loser's regret (Isaac and Walker 1985, Dufwenberg and Gneezy 2002, Ockenfels and Selten 2005). We extend the existing literature by applying the framework of salient regret to reserve prices.

In Bergemann et al. (2021) we develop a model of salient regret and theoretically investigate the performance of the SRA. In our model, bidders maximize a convex combination of profit and regret. We find that in a standard symmetric independent private value auction model with N bidders the SRA with an optimal soft reserve price is efficient and yields strictly higher revenue than the efficient SPA and FPA. A bidder in the SRA with a valuation below the soft reserve has a fairly straight forward bidding strategy and should never enter the SPA stage. The first-price bid depends on the number of bidders and the level of regret and is smaller or equal than the valuation. In contrary, bidders with a valuation above the soft reserve face a difficult trade-off. Entering the SPA on the one hand may drive up the price, but on the other insures against regret. Contrarily, refusing to commit to the reserve may decrease the price in case of winning, but opens up the possibility of regret. A higher valuation increases the risk of regret and the willingness to pay a higher price in the SPA. The necessary amount depends positively on the soft reserve level and negatively in the weight of regret in the utility function. Intuitively, holding the soft reserve price level constant, an increase of the weight on regret decreases the valuation necessary to persuade a bidder to commit to the reserve. Vice versa, holding the weight on regret constant, an increase in the soft reserve price level increases the valuation necessary to convince a bidder to enter the SPA stage. From the auctioneer's perspective, a soft reserve price compared to a FPA increases revenue in case of high competition in the SPA stage, while simultaneously having a price floor in case of low competition. If no bidder is willing to accept the opening, revenue and bids remain unchanged in comparison to the FPA.

In summary, by committing to the soft reserve, bidders are able to insure against the risk of regret and are willing to pay a premium and sellers are able to exploit bidder's regret aversion by employing a soft reserve price. The model in Bergemann et al. (2021) based on salient regret is in line with our experimental results.

V Conclusion

Reserve prices play an important role in improving auction revenues, especially in instances with few bidders and weak competition. Yet the use of a reserve price has some downsides. The reserve price may reduce bidder participation, since the reserve eliminates some profitable opportunities for bidders. Also, the reserve price prevents efficient trade when the highest bidder's valuation is above the seller's valuation but less than the reserve price.

We explore whether by adopting a soft reserve the seller can achieve the revenue gain of the traditional hard reserve without incurring the undesirable efficiency loss. The answer from standard economic theory is no. To enhance revenues, the seller must impose a hard reserve: a binding commitment to not sell below the reserve price. This leads to a loss in efficiency and reduced incentives to participate in the auction. However, there may be behavioral reasons, outside the standard theory, why softening the reserve price may improve auction performance. The chief candidate is regret aversion which has proven to be important in auctions.

The soft reserve we propose is intended to exploit bidders' regret avoidance. The seller announces an opening price. Bidders can accept the opening price or place an alternative bid below the opening price. These alternative bids are only considered if no bidder accepts the opening, in which case the highest alternative bids wins at the price bid. Thus, not accepting the opening price excludes the bidder from competing in the subsequent auction. To maximize the regret from rejecting the opening price, we want to make the auction as attractive as possible. In the field, this would be an ascending auction which bidders tend to prefer. However, in the lab, we use a second-price auction which is easier and faster in our setting, but otherwise strategically similar.

We test the soft reserve auction in the laboratory. First, we observe the soft reserve price to significantly increases efficiency and revenue compared to a conventional hard reserve price. Additionally, a soft reserve significantly increases revenue compared to a first-price auction and matches the best performing auction type in instances of both low and high competition.

Finally, in line with our main results sellers also prefer a soft reserve to a hard reserve and have a good understanding of choosing the reserve price.

These results have been demonstrated in an experimental lab in a specific auction setting with student subjects and low stakes. Can we expect similar results to hold in practice? We do not know.[5] Nonetheless, we believe that the soft reserve has to offer much to sellers who are interested in both revenues and efficiency. The strong lab performance primarily hinges on regret aversion. Bidders tend to accept the opening price whenever it is profitable since rejecting the opening price means exclusion from the subsequent auction. Regret from exclusion appears more salient than the profits that may be attained from making a lower offer. In practice, we would expect bidders to exhibit at least some degree of regret aversion.

There are other advantages of the soft reserve in practice. It delivers certainty of sale. This is valued by both sellers and buyers. It motivates bidder participation and eliminates sellers' concerns of no sale. Another advantage of the soft reserve is that it is more robust in at least two dimensions. First, it is more robust to mistakes in setting the reserve. The soft reserve determines the range of prices over which the competition is first price vs. second price; the hard reserve determines the range of prices over which there is no trade vs. second-price competition, a more extreme dichotomy. Second, the soft reserve is more robust to the level of competition. The soft reserve performs well when the highest valuation is both near (high competition) and far (low competition) from the second-highest valuation.

[5] Among other important differences, in practice bidders often are teams representing a company or an institution rather than single individuals. As some experimental studies have shown that teams tend to behave more strategic than individuals (for example Cooper and Kagel (2005) in signalling games), this could potentially affect the external validity of our results. For discussion on teams vs. individuals see Charness and Sutter (2012).

Appendix A. Further analysis, tables and figures

TABLE A1.1: EFFICIENCY

	FPA	Hard50	Soft50	Soft66
Efficient allocation	86.94 %	95.15 %	91.06 %	85.88 %
Significance level	-/***/**/-	***/-/**/***	**/**/-/***	-/***/***/-
Good sold	100 %	73.87 %	100 %	100 %
Significance level	-/***/-/-	***/-/***/***	-/***/-/-	-/***/-/-
Total welfare	96.69 %	73.37 %	98.17 %	97.03 %
Significance level	-/***/**/-	***/-/***/***	**/***/-/**	-/***/**/-

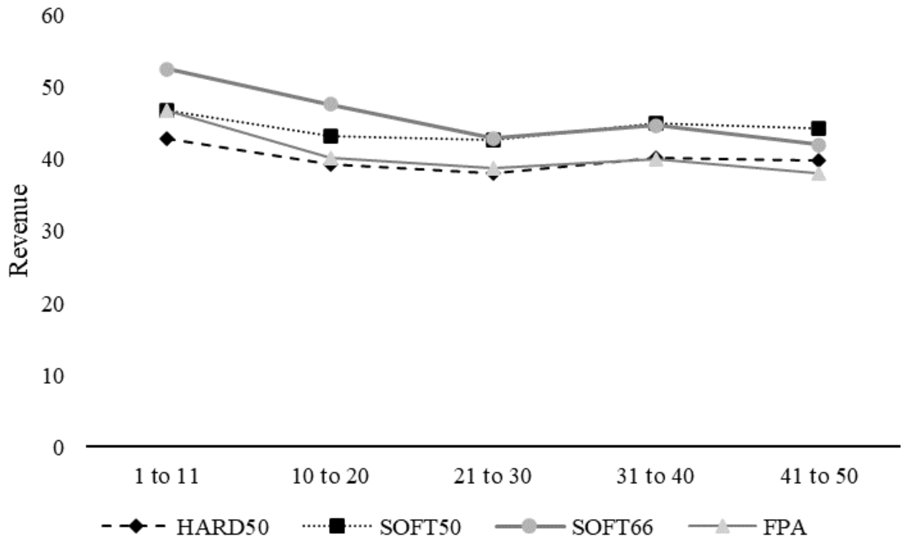
Notes: The table reports results from non-parametric Wilcoxon Ranksum tests based on independent matching group averages. Significance levels are pairwise for each treatment: FPA/Hard50/Soft50/Soft66. *, **, and *** denote significance levels 10%, 5%, and 1% level.

TABLE A1.2 *P*-VALUES EFFICIENCY

Efficient allocation				
	FPA	Hard50	Soft50	Soft66
FPA	-	0.0000***	0.0217**	0.3529
Hard50	0.0000***	-	0.0158**	0.0000***
Soft50	0.0217**	0.0158**	-	0.0027***
Soft66	0.3529	0.0000***	0.0027***	-
Good sold				
	FPA	Hard50	Soft50	Soft66
FPA	-	0.0000***	.	.
Hard50	0.0000***	-	0.0000***	0.0000***
Soft50	.	0.0000***	-	
Soft66	.	0.0000***	.	-
Overall efficiency				
	FPA	Hard50	Soft50	Soft66
FPA	-	0.0000***	0.0116**	0.8802
Hard50	0.0000***	-	0.0000***	0.0000***
Soft50	0.0116**	0.0000***	-	0.0348**
Soft66	0.8802	0.0000***	0.0348**	-

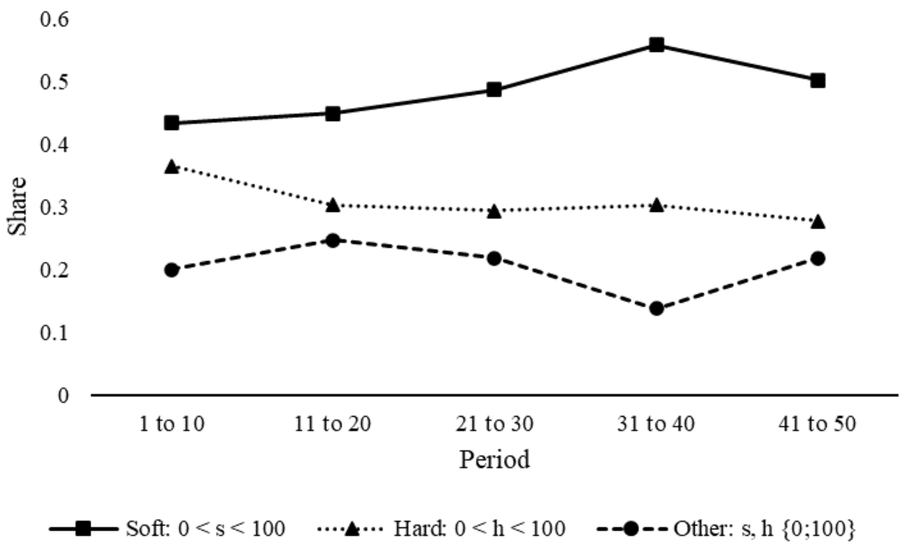
Notes: Results from non-parametric Wilcoxon Ranksum tests based on independent matching group averages. Tests are pairwise for each treatment: FPA/Hard50/Soft50/Soft66. *, **, and *** denote significance levels 10%, 5%, and 1%.

FIGURE A1.1 DYNAMICS REVENUE



Notes: This figure reports average revenue across periods and treatment.

FIGURE A1.2 DYNAMICS REVENUE ENDO



Notes: This figure reports revenue across periods and different types of reserve prices, *s* refers to soft reserve prices, *h* to hard reserve prices and *others* to auctions without a reserve price.

TABLE A1.3 TREATMENT EFFECTS ACROSS PERIODS

Average revenues				
	HARD50	SOFT50	SOFT66	FPA
Period 1 to 10	42.90 (4.73)	46.76 (3.93)	52.58 (4.01)	46.76 (3.93)
Period 11 to 20	39.42 (3.69)	43.27 (5.31)	47.70 (6.24)	40.27 (5.90)
Period 21 to 30	38.0 (4.93)	42.72 (4.78)	38.09 (4.93)	42.72 (7.34)
Period 31 to 40	38.09 (4.93)	42.72 (4.78)	42.96 (6.18)	38.88 (7.33)
Period 41 to 50	39.79 (3.23)	44.30 (3.76)	42.04 (3.35)	38.03 (4.76)
p-values Wilcoxon pair-wise rank -sum tests Period 1- 10				
	HARD50	SOFT50	SOFT66	FPA
HARD50	-	0.024**	< 0.001***	0.027**
SOFT50	0.024**	-	0.001**	0.821
SOFT66	< 0.001***	0.001**	-	0.006***
FPA	0.027**	0.821	0.006***	-
p-values Wilcoxon pair-wise rank -sum tests Period 11- 20				
	HARD50	SOFT50	SOFT66	FPA
HARD50	-	0.040**	< 0.001***	0.782
SOFT50	0.040**	-	0.038**	0.821
SOFT66	< 0.001***	0.038**	-	0.142
p-values Wilcoxon pair-wise rank -sum tests Period 21- 30				
	HARD50	SOFT50	SOFT66	FPA
HARD50	-	0.014**	0.027**	0.969
SOFT50	0.014**	-	0.763	0.065**
SOFT66	0.027**	0.763	-	0.077**
FPA	0.969	0.065**	0.077**	-
p-values Wilcoxon pair-wise rank -sum tests Period 31- 40				
	HARD50	SOFT50	SOFT66	FPA
HARD50	-	0.004***	0.020**	0.527
SOFT50	0.004***	-	0.188	0.012**
SOFT66	0.020**	0.188	-	0.018**
FPA	0.527	0.012**	0.018**	-
p-values Wilcoxon pair-wise rank -sum tests Period 41- 50				
	HARD50	SOFT50	SOFT66	FPA
HARD50	-	0.003***	0.097*	0.553
SOFT50	0.003***	-	0.083*	< 0.001***
SOFT66	0.097*	0.083*	-	0.013**
FPA	0.553	< 0.001***	0.013**	-

Notes: Results from non-parametric Wilcoxon rank-sum tests based on independent matching group averages. Tests are pairwise for each treatment: FPA/Hard50/Soft50/Soft66. *, **, and *** denote significance levels 10%, 5%, and 1%.

Appendix B. A stylized model

The following simplified stylized model serves as a heuristic to identify useful treatment parameters for the experiment design and is not related to our theoretical analysis in Bergemann et al. (2021).

Assume an auction framework with n bidders with independently and identically distributed valuations v_i according to the increasing distribution function $F(v_i)$. For simplification and in line with our experiment design we assume further valuations are uniformly distributed $v_i \sim U[0{,}100]$. Bidders compete in an action with a soft reserve price (as described in section II) for a single object. There is exogenously given a soft reserve price of $s \epsilon (0{,}100)$. Now we assume the following bidding strategy for all bidder i:

Bidders pursue the following strategy:

- If $v_i < s$ do not participate in the auction and make an alternative bid of $b_i(v_i) = v_i \frac{n-1}{n}$.
- If $v_i \geq s$ do participate in the auction and bid your valuation $b_i(v_i) = v_i$.

There are three possible cases depending on the realized valuations and the soft reserve:

Case 1: All valuations are below the soft reserve. This happens with probability $F(s)^N$ and expected revenue is the second highest valuation among those bidders.

Case 2: Only one valuation is greater or equal to the soft reserve. This happens with probability $nF(s)^{n-1}(1-F(s))$ and the expected revenue is s.

Case 3: At least two bidders have a valuation greater or equal to the soft reserve. This happens with probability $\sum_{a=2}^{n}\binom{n}{a}F(s)^{n-a}(1-F(s))^a$ expected revenue is the second highest valuation among those bidders.

This leads to the following expected revenue formula depending on s and n:

$$E[R(s,n)]$$
$$= F(s)^N \int_0^s yn(n-1)\big(1-F_L(y)\big)\,F_L(y)^{n-2} f_L(y)\,dy + nF(s)^{n-1}\big(1-F(s)\big)s$$
$$+\sum_{a=2}^{n}\binom{n}{a} F(s)^{n-a}(1-F(s))^a \int_T^{100} ya(a-1)\big(1-F_H(y)\big)\,F_H(y)^{a-2} f_H(y)\,dy$$

For a given number of bidders n this term can be maximized for the optimal soft reserve price. Table B1 displays the optimal soft reserve price of s and the corresponding expected revenue for a given number of bidders n. As benchmarks we also provide expected revenue in this auction framework for a hard reserve ($E[R_{Hard}]$) and no reserve ($E[R_{FPA}]$).

TABLE B.1 OPTIMAL SOFT RESERVE

N	Opt. Soft Reserve	$E[R_{SOFT}]$	$E[R_{SOFT}]$	$E[R_{FPA}]$
1	50	25	25	0
2	66.67	48.15	41.67	33.33
3	75	60.55	53.13	50
4	80	68.19	61.25	60
5	83.33	73.36	67.19	66.67
6	85.71	77.09	71.65	71.43
7	87.5	79.91	75.10	75
8	88.89	82.11	77.82	77.78
9	90	83.87	80.02	80
10	90.91	85.32	81.83	81.82
20	95.24	92.27	90.47	90.47

Notes: This table reports levels and revenues of optimal soft reserve prices and benchmark cases for a stylized model.

Interestingly, but in line with for example Rosenkranz and Schmitz (2007) and Davis et al. (2011), in contrast to a hard reserve price the optimal soft reserve price in this framework increases with the number of bidders. But we also observe the difference in expected revenue between all three auction types vanishes as the number of bidders grows large. The choice of soft or hard reserve price becomes less important with more bidders.

Appendix C

Appendix C.1. Experimental instructions: Main treatments

Treatment dimensions are: FPA / Soft50 / Soft66 / Hard50

General part

Welcome to our experiment!

Please read the following instructions carefully. If you have a question, please raise your hand. We will then come to you and answer your question. Communication with other participants is not allowed during the whole experiment. If you violate this rule, we might exclude you from the experiment and all payouts.

All participants receive 4,00 Euro show-up fee. In addition, you can gain further payoffs depending on your and the other participants' decisions.

The currency used in this experiment is experimental currency units (ECU). At the end of the experiment all ECUs will be converted into Euro and paid out in cash. The conversion rate is 75 ECU = 1 Euro. All decisions and payoffs in this experiment will be treated anonymously. All participants receive identical instructions.

Experiment

The experiment consists of a total of 50 rounds. In each round, the participants are randomly matched into groups of two. We ensure that nobody is matched with the same participant in two consecutive rounds. All rounds are identical and independent of each other. All round profits are added up and paid out at the end of the experiment. Possible losses will be offset against the 4.00 Euro show-up fee.

In each round the participants can bid in an auction. A round consists of two stages: At the first stage, the bidders make a decision about participation in the auction **[Soft50/Soft66/Hard50:** with a minimum bid**]**. On the second level, bidders can bid on a fictitious good if they have decided to participate in the auction **[Soft50/Soft66/Hard50:** with a minimum bid**]**.

The personal value of the fictitious good (the valuation) of each bidder varies between the bidders and the rounds. At the beginning of each round, each bidder is told how much the good is worth to him in that round. The valuations are determined randomly and independently for each bidder and each ECU-Amount between 0.00 ECU and 100.00 ECU (with two decimal places) is equally probable.

Stage 1: Participation decision

[FPA, Hard50: In the first stage, both bidders decide whether they wish to participate in the auction **[Hard50:** with a minimum bid of 50 ECU**]**. If both bidders decide not to participate in the auction, none of the bidders will receive the good.**]**

[Soft50, Soft66: At the first stage, both bidders decide whether they wish to participate in the auction with a minimum bid of 50 ECU **[Soft66:** 66.67 ECU**]**. If a bidder does not wish to participate in the auction, he makes an alternative bid between 0.00 and 50 ECU **[Soft66:** 66.67 ECU**]**. If a bidder does not wish to make an alternative bid, he has the option of making an alternative bid of 0.00 ECU.

If both bidders decide not to participate in the auction with a minimum bid, the bidder with the higher alternative bid will receive the good. In this case, the bidder pays his alternative bid. If both bidders propose the same alternative bid, the buyer will be determined at random.**]**

Stage 2: Auction

[FPA: Bidders who participate in the auction place their bids for the good at the second stage. Bidders who have decided not to participate do not submit a bid. The bid must be a minimum of 0 ECU and a maximum of 100,00 ECU. The bidder who places the higher bid wins the auction. The price he has to pay corresponds to his own bid. If both bidders place the same bid, a random decision is made as to who wins the auction.**]**

[Hard50, Soft50, Soft66: Bidders who participate in the auction with minimum bid place their maximum bid for the good at the second stage. Bidders who have decided not to participate do not submit a bid. The maximum bid is the maximum price the bidder is willing to pay for the good. This maximum bid must be at least 50 ECU **[Soft66:** 66.67 ECU**]** and may not exceed 100.00 ECU. The bidder who places the higher maximum bid wins the auction. The price he has to pay equals the second highest bid plus 0.01 ECU.

Exceptions:

- ⇨ *If only one bidder participates in the minimum bid auction, the price will be equal to the minimum bid of 50 ECU* **[Soft66:** *66.67 ECU***]** *regardless of his own bid.*
- ⇨ *If both bidders place the same bid, a random decision will be made as to who wins the auction. In this case the price is exactly the maximum bid of the auction winner.*

The bidding system of this auction can be imagined as a representative who bids for you at the auction. The representative always offers exactly enough to make you the highest bidder. He does this until your maximum bid is reached. Therefore, the price never exceeds the second highest bid plus 0.01 ECU.**]**

Round payoff

The round payoff depends on whether the bidders receive the good. If a bidder receives the good, the round payoff corresponds to his personal valuation subtracted the price. For a bidder who does not receive the good, the round payoff is 0.00 ECU.

$$\text{Round payoff } = \begin{cases} \text{Valuation} - \text{ Price, if the bidder receives the Good} \\ 0\text{, if the bidder does not receive the Good} \end{cases}$$

Feedback

At the end of each round, all bidders receive information about the price, the bidder who received the good and their own round payoff.

Appendix C2. Experimental instructions: Endogenous seller

General part

Welcome to our experiment!

Please read the following instructions carefully. If you have a question, please raise your hand. We will then come to you and answer your question. Communication with other participants is not allowed during the whole experiment. If you violate this rule, we might exclude you from the experiment and all pay-outs.

All participants receive 4,00 Euro show-up fee. In addition, you can gain further payoffs depending on your and the other participants' decisions.

The currency used in this experiment is experimental currency units (ECU). At the end of the experiment all ECUs will be converted into Euro and paid out in cash. The conversion rate is 75 ECU = 1 Euro. All decisions and payoffs in this experiment will be treated anonymously. All participants receive identical instructions.

Experiment

At the beginning of the experiment all participants are randomly assigned the role of a seller or a bidder. All participants keep this role for the whole experiment. The experiment consists of a total of 50 rounds. In each round, the participants are randomly divided into groups, each consisting of a seller and two bidders. This ensures that no bidder forms a group with the same bidder in two consecutive rounds. All rounds are identical and independent of each other. All round profits are added up and paid out at the end of the experiment. Possible losses will be offset against the 4.00 Euro show-up fee.

A round consists of three stages: in the first stage, the seller chooses a form of auction. In the second stage, the bidders make a decision about participation in the auction with a minimum bid. In the third stage, bidders can bid on a fictitious good if they have decided to participate in the auction.

The personal value of the fictitious good (the valuation) of each bidder varies between the bidders and the rounds. At the beginning of each round, each bidder is told how much the good is worth to him in that round. The valuations are determined randomly and independently for each bidder and each ECU amount between ECU 0.00 and ECU 100.00 (with two decimal places) is equally probable.

Stage 1: Choice of type of auction

In the first stage, the seller decides on the type of auction. The seller makes two decisions. First, the seller determines the level of the minimum bid and second, whether alternative purchase prices below the minimum bid can be proposed.

- Level of the minimum bid: The seller chooses a minimum bid of 0 ECU, 33,33 ECU, 50 ECU, 66,67 ECU or 100 ECU.
- Alternative purchase price: The seller decides whether bidders who do not participate in the auction with a minimum bid may propose an alternative purchase price below the minimum bid.

Stage 2: Participation decision

At the second stage, the two bidders are informed about the form of auction chosen by the seller and decide whether they wish to participate in the auction with a minimum bid. If a bidder does not wish to participate in the auction and the seller has accepted alternative purchase prices, the bidder may propose an alternative purchase price between 0.00 ECU and the minimum bid. If a bidder does not wish to propose an alternative purchase price, he has the option of offering a purchase price of 0.00 ECU. The bidder may then offer an alternative purchase price between 0.00 ECU and the minimum bid.

If both bidders decide not to participate in the minimum bid auction while the seller has allowed alternative purchase prices to be proposed, the bidder with the higher proposed purchase price will receive the good. In this case he pays his proposed purchase price. If both bidders propose the same purchase price, the bidder will be determined at random. If the seller has not allowed alternative purchase prices, the property will not be sold and neither bidder will receive the good.

Stage 2: Auction with minimum bid

Bidders who participate in the auction with a minimum bid submit their maximum bid for the good at the second stage. Bidders who have decided not to participate do not submit a bid. The maximum bid is the maximum price the bidder is willing to pay for the good. This maximum bid must be at least equal to the minimum bid chosen by the seller and may not exceed 100.00 ECU. The bidder who places the higher maximum bid wins the auction. The price he has to pay corresponds to the second highest bid plus 0.01 ECU.

Exceptions:

⇨ *If only one bidder participates in the auction with a minimum bid, the price will be equal to the minimum bid chosen by the seller, regardless of his own bid*

⇨ *If both bidders place the same bid, a random decision will be made as to who wins the auction. In this case the price is exactly the maximum bid of the auction winner.*

The bidding system of this auction can be imagined as a representative who bids for you at the auction. The representative always offers exactly enough to make you the highest bidder. He does this until your maximum bid is reached. Therefore, the price never exceeds the second highest bid plus 0.01 ECU.

Round payoff

The round payoff depends on whether the seller sells the good. If a seller sells the good, the round payoff corresponds to the realized price subtracted by sales fee of 75%. For a seller who does not sell the good, the round payoff is 0.00 ECU.

$$\text{Round payoff } = \begin{cases} 0.25 * \text{ Price, if the seller receives the good} \\ \text{0, if the seller does not receive the good} \end{cases}$$

The round payoff depends on whether the bidders receive the good. If a bidder receives the good, the round payoff corresponds to his personal valuation subtracted the price. For a bidder who does not receive the good, the round payoff is 0.00 ECU.

$$\text{Round payoff } = \begin{cases} \text{Valuation} - \text{ Price, if the bidder receives the good} \\ \text{0, if the bidder does not receive the good} \end{cases}$$

Feedback

At the end of each round, seller and bidder receive information about the chosen auction type, the price and their own round payoff. In addition, the seller receives information about the bids and alternative purchase prices of the bidders. Bidders will also know whether they have received the good.

Chapter 2

Fixing feedback revision rules in online markets

CO-AUTHORS: GARY BOLTON, BEN GREINER AND AXEL OCKENFELS*

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.

* 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. 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 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 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 Bolton et al. (2018), and observed that, in fact, all marketplaces we identified there which used such 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). 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, or do not facilitate the revision of feedback, such as AirBnB. The feedback rules that come closest to our proposed and slightly more sophisticated 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 careful 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 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 (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. 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.

As a matter of fact, market places seem to have adapted their feedback mechanisms in the direction we propose here. Since the publication of our previous study on the problematic aspects of mutual feedback withdrawal (Bolton et al., 2018), the market places we identified there which used such feedback rules have all abandoned that mechanism. 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). 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, or do not facilitate the revision of feedback, such as AirBnB. The feedback rules that come closest to our proposed more sophisticated procedure (mutual agreement to allow withdrawal, then unilateral withdrawal) seem to be in place on only one platform, Upwork.

In Section II we describe our experimental design and procedures. Section III develops our two main hypotheses. Section VI presents our experimental results, and Section V 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, 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 (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 $Q_2 \geq Q_1$. 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, i.e. $\pi_B = 100 - \text{PricePaid} + Q_2 * 3$. The seller's round income results from endowment plus received price minus costs for chosen quality provision, i.e. $\pi_S = 100 + \text{PricePaid} - Q_2$. 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* withdraw their negative feedback, or not. (In no treatment can positive initial feedback be made negative.)

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	$\pi_B = 100 - \text{PricePaid} + Q_2 * 3$, $\pi_S = 100 + \text{PricePaid} - Q_2$

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.

> *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

[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').

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

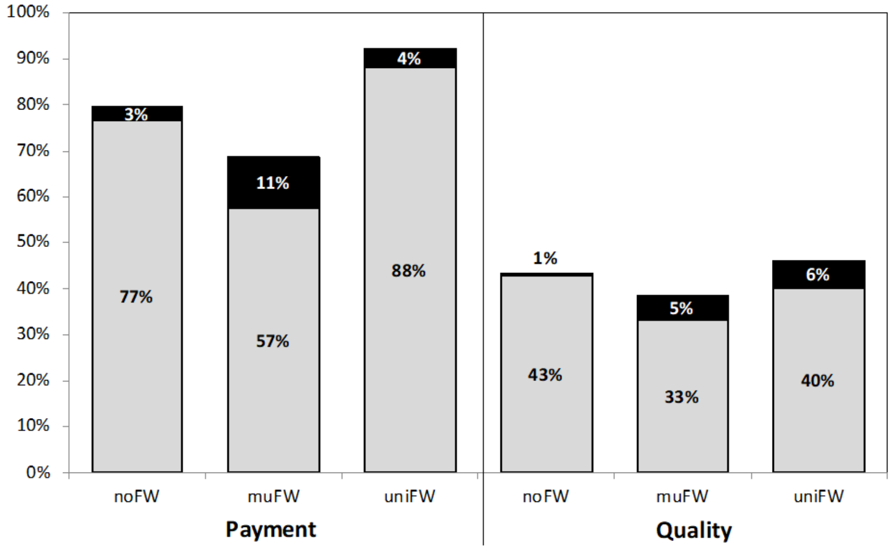
---

[2] We focus our analysis on rounds 11 to 50, as in Bolton et al. (2018), 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. Further, in Appendix A.4 we replicate all tables and figures in the main text using all rounds 1-60, with qualitatively the same results.

[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.

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



Notes: The figure reports initial payment and quality in the transaction stage (grey 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.

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 treatment *uniFW* yields an increase in quality of 4

[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.

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.

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.

TABLE 2: REGRESSIONS OF PROBABILITY OF PAYMENT AND QUALITY

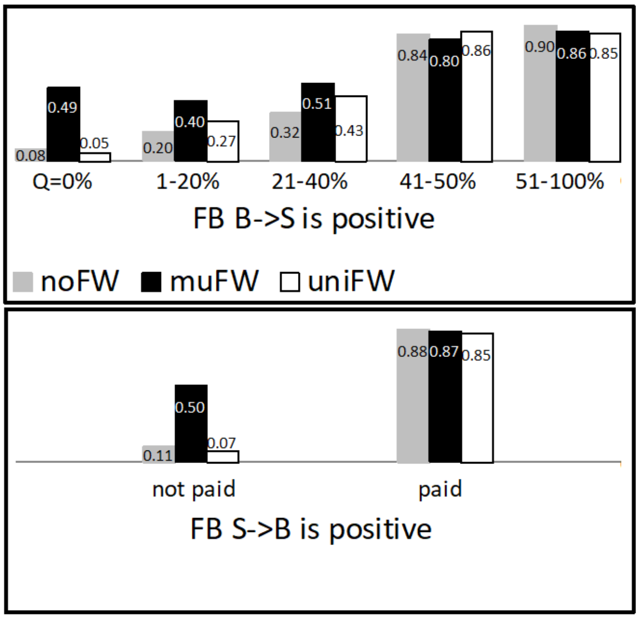
Model	(1)	(2)	(3)	(4)
Model type	Probit	Tobit	Probit	Tobit
Dependent	Initial Payment	Initial Quality	Final Payment	Final Quality
Constant		0.517***		0.512***
		[0.024]		[0.021]
Round	-0.005***	-0.004***	-0.005***	-0.003***
	[0.001]	[0.001]	[0.001]	[0.001]
muFW	-0.156**	-0.120**	-0.086	-0.058
	[0.074]	[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.

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



Notes: 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.

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	0.011***	[0.001]		
Payment			0.749***	[0.036]
muFW	0.308***	[0.064]	0.270***	[0.048]
muFW × Quality	-0.007***	[0.002]		
muFW × Payment			-0.280***	[0.055]
uniFW	-0.055	[0.112]	-0.051	[0.061]
uniFW × Quality	0.002	[0.003]		
uniFW × Payment			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.

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-

[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.

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

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

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

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 and Iyengar 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.

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

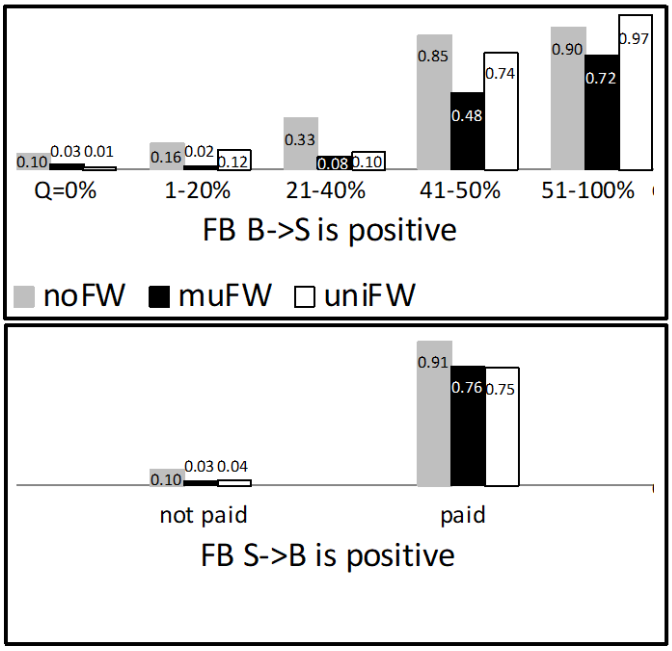


Table A.3.1 show the detailed pattern of feedback giving and underlying buyer and seller behavior. Figure A.3.2 visualizes the feedback conversion through withdrawal in the three different treatments. We observe that first, initial feedback distributions vary markedly between the three treatments, especially 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

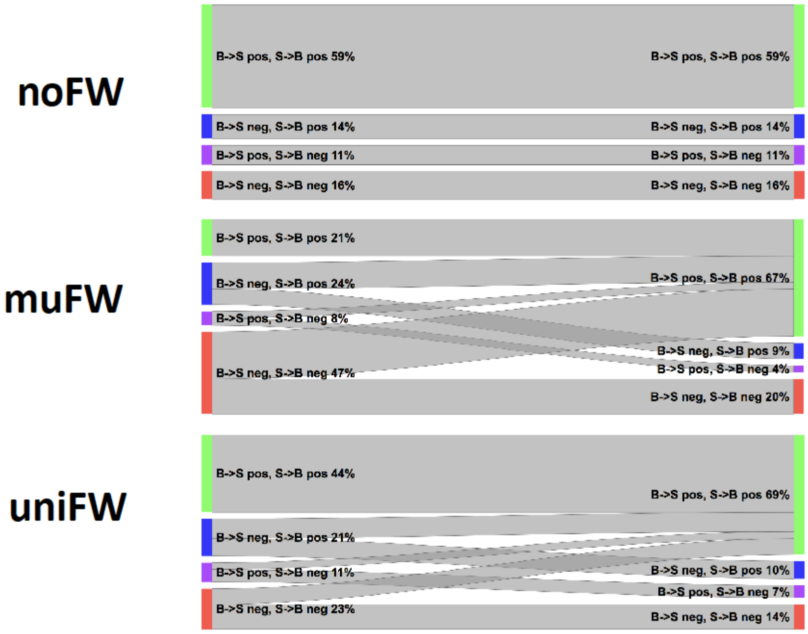


TABLE A.3.1: FEEDBACK, MAKE-GOOD AND WITHDRAWAL FREQUENCIES AS WELL AS INITIAL AND EVENTUAL FEEDBACK DISTRIBUTIONS ACROSS THE THREE TREATMENTS

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 Seller received		uniFW 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 Buyer received**		**uniFW 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*.

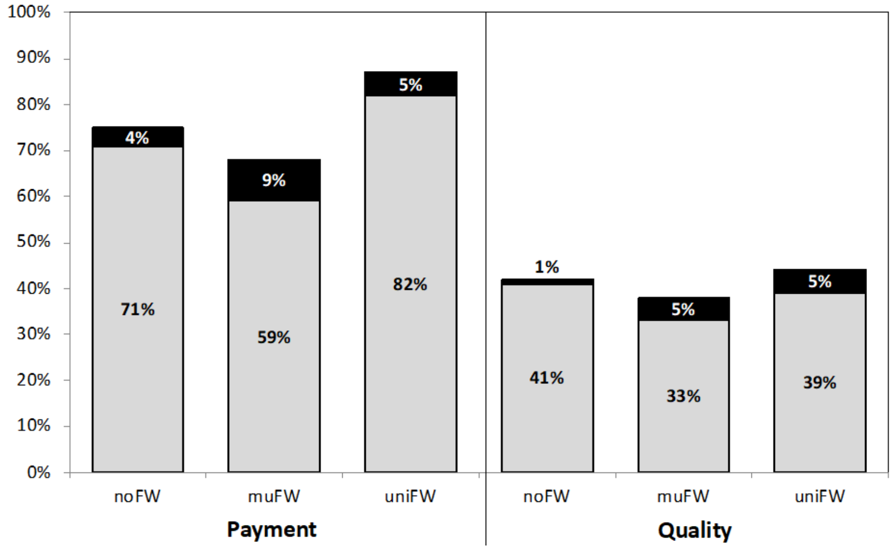
TABLE A.3.3: PROBIT REGRESSION OF THE LIKELIHOOD TO WITHDRAW ON OTHER'S MAKE-GOOD BEHAVIOR AND FEEDBACK CONDITION

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 A.4 Main text tables and figures based on all rounds 1-60

FIGURE A.4.1: AVERAGE PAYMENT/REVENUES AND QUALITY/EFFICIENCY CONDITIONAL ON TRADE, ACROSS THE THREE TREATMENTS



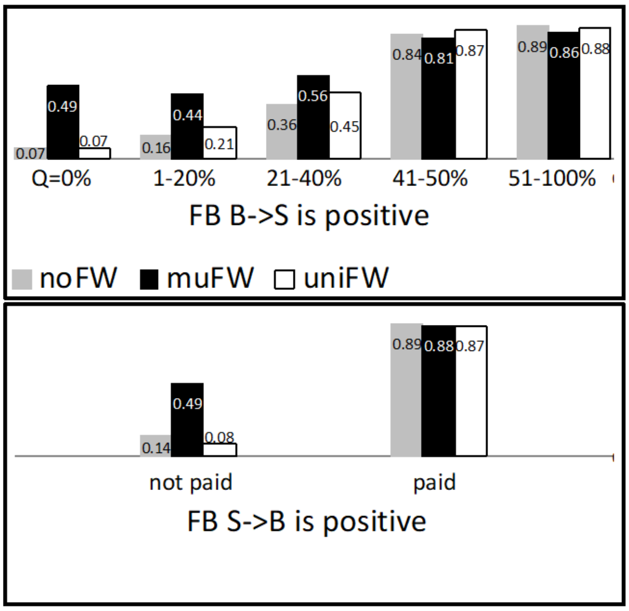
Note: The figure reports initial payment and quality in the transaction stage (grey share of the bars), as well as additionally provided payments and quality in the make-good stage (black share of the bars). Numbers are based on all rounds 1-60 in the experiment.

TABLE A.4.1: 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		0.612***		0.614***
		[0.024]		[0.022]
Round	-0.010***	-0.008***	- 0.010***	- 0.008***
	[0.008]	[0.001]	[0.001]	[0.001]
muFW	-0.113**	-0.100**	-0.056	-0.047
	[0.070]	[0.050]	[0.056]	[0.045]
uniFW	0.113***	-0.022	0.136***	0.032
	[0.044]	[0.041]	[0.040]	[0.027]
First 10 rounds	-0.141***	-0.080***	- 0.105***	- 0.082***
	[0.035]	[0.017]	[0.033]	[0.015]
Last 10 rounds	0.081***	0.097***	0.078***	0.094***
	[0.016]	[0.024]	[0.014]	[0.013]
N	7377	7377	7377	7377
LL	-3884.8	-2509.4	-3332.5	-1905.7
Censoring Left (Non) Right		1495 (5714) 168		1226 (5976) 175
Post-estimation test muFW = uniFW, p-value	0.0011	0.1977	0.0006	0.0899

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 all rounds 1-60. “First 10 rounds” and “Last 10 rounds” are dummy variables indicating round 1-10 (51-60). Robust standard errors are clustered at the level of independent matching groups. *, **, and *** denote significance levels 10%, 5%, and 1% level, respectively.

FIGURE A.4.2: EVENTUAL FEEDBACK CONDITIONAL ON TRADING PARTNER'S BEHAVIOR



Notes: 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. Numbers are based on all rounds 1-60 in the experiment.

TABLE A.4.2: 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	-0.002	[0.001]	-0.001	[0.001]
Quality	0.011***	[0.001]		
Payment			0.711***	[0.038]
muFW	0.321***	[0.053]	0.227***	[0.040]
muFW × Quality	-0.007***	[0.001]		
muFW × Payment			-0.231***	[0.051]
uniFW	-0.038	[0.090]	-0.066	[0.047]
uniFW × Quality	0.002	[0.002]		
uniFW × Payment			0.052	[0.059]
First 10 rounds	-0.002	[0.020]	0.044**	[0.020]
Last 10 rounds	0.017	[0.016]	0.009	[0.012]
N	7377		7377	
LL	-3271.5		-2872.8	
Post-estimation test				
muFW = uniFW, p-value	0.0000		0.0000	

Notes: The table reports average marginal effects dy/dx. Regressions are based on data from all rounds 1-60. "First 10 rounds" and "Last 10 rounds" are dummy variables indicating round 1-10 (51-60). Robust standard errors are clustered at the level of independent matching groups. *, **, and *** denote significance at the 10%, 5%, and 1% level, respectively.

TABLE A.4.3: 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	142.6	139.9	139.2
All sellers	128.1	126.9	137.5
Buyers who paid initially	142.4	135.4	141.6
Sellers who sent initial quality >=50%	124.8	120.0	130.3
Average Standard Deviation	*noFW*	*muFW*	*uniFW*
All buyers	58.02	60.81	45.33
All sellers	36.29	37.83	29.87
Buyers who paid initially	51.08	56.11	38.14
Sellers who sent initial quality >=50%	31.77	32.83	23.47
P-values from Wilcoxon rank sum tests	*noFW* vs. *muFW*	*noFW* vs. *uniFW*	*muFW* vs. *uniFW*
All buyers	0.360	0.009***	0.007***
All sellers	0.510	0.005***	0.002***
Buyers who paid initially	0.509	0.052*	0.045**
Sellers who sent initial quality >=50%	0.722	0.002***	0.021**

Notes: The table reports average payoffs and standard deviations of collapsed data on the independent matching group level for all rounds 1-60. *, **, 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".

Chapter 3

You don't get what I deserve – Priorities and Fairness in School Choice Problems*

Abstract

In school choice problems the properties of fairness and efficiency are incompatible. Efficiency is undisputedly a desirable property, but in practice it is often sacrificed in favor of fairness. I investigate if participants in experimental matching markets value fairness and how the valuation of fairness depends on specific characteristics of the matching market. I find that a significant share of subjects refuses to consent to fairness violations and subjects endowed with a low priority consent more often in a matching market with random priorities compared to a market with performance-based priorities.

* Department of Economics, University of Cologne, D-50923 Cologne, Germany, kevin.breuer@wiso.uni-koeln.de.
Financial support from the Center for Social and Economic Behavior (C-SEB) is gratefully acknowledged. I thank Alexander Westkamp, Axel Ockenfels, Christoph Feldhaus, and Markus Möller for very helpful comments and suggestions

I. Introduction

Famously, in school choice problems the properties of fairness and efficiency are incompatible (Balinsky and Sönmez, 1999). Achieving efficiency may come at the cost of violating priorities thereby inducing justified envy and sacrificing fairness.[1] A market designer choosing a matching algorithm therefore engages in a trade-off between fairness and efficiency. A big branch of literature in matching theory is concerned with compromising between the two properties. In practice, the fair deferred-acceptance algorithm (DA) (Gale and Shapley, 1962) has historically been more popular[2] than the efficient top trading cycle (TTC) (Shapley and Scarf, 1974). Instead of aiming to reconcile fairness and efficiency, this paper takes a different approach by investigating in which specific market environments the property of fairness is important. The value of efficiency is clear, but why does fairness matter? On a practical level, unfair matchings raise legal and political concerns and parents may file lawsuits as a response to priority violations (see Erdil and Ergin 2008 and Ehlers and Morrill 2019). Besides those concerns, fairness might also matter *per se* to participants in matching algorithms, especially to the victims of unfair matching outcomes, the participants who are being exposed to justified envy. This paper discusses if (1) participants in matching algorithms value fairness and (2) the value of fairness depends on the specific characteristics of the matching market. More precisely, this paper investigates experimentally how the origin of the priority order shapes the willingness of being exposed to justified envy.

The incompatibility relates to the notion that no algorithm is able to produce a fair and efficient outcome for every possible set of priorities and preferences. Nevertheless, algorithms exist which produce fair and efficient outcomes for certain sets of priorities and preferences. Ergin (2002) and Ehlers and Erdil (2010) characterize such priority structures for which the efficient and fair outcomes coincide. A different approach found in the literature has been to let go of achieving both properties, but rather stick to one and improve on the other as much as possible. For example, Kesten (2010) develops an algorithm to improve the efficiency of the inherently fair Deferred Acceptance Algorithm (DA). Taking the opposite direction, Hakimov and Kesten (2014) have worked on improving the fairness of the efficient Top Trading Cycle (TTC). In a related manner, Erdil and Ergin (2008) and Abdulkadiroğlu et al. (2015) try to increase

[1] In the context of school choice justified envy is commonly defined as a property a particular student might be exposed to. This student is assigned a certain school, but there exists another school she prefers to her assignment and there exists another student who is assigned to the school she prefers, even though she has a higher priority at this school than him. A matching is fair it eliminates justified envy (Abdulkadiroğlu and Sönmez, 2003).

[2] See for example Abdulkadiroğlu et al. (2005a) for the New York City High School Match.

efficiency by improving tiebreaking rules if priorities are clustered in classes. Another way of approaching the incompatibility is to define weaker notions of fairness which are compatible with efficiency. In this spirit Morrill (2015) characterizes "just" and Alcade and Romero-Medina (2017) "alpha-equity" assignments which relax the requirements on fairness. More recently, Dur et. al (2019) and Troyan et al. (2020) introduce the notions of "partial stability" and "essentially stable", which allow for priority violations of certain types.

Close to this paper, in Schmelzer (2016) the author elicits preferences over an institutional feature of the matching market, randomization of allocation priorities. Interestingly, he finds that even though being ex-ante equivalent in expectation multiple-tie breaking is preferred by participants over single randomization and is perceived as fairer. This indicates that participants also care about the priority structure beyond their own matching outcome.

This paper is structured as follows. Section II derives the hypotheses and discusses the concept of fairness. Section III introduces the experimental design and section IV presents the experimental results. Finally, Section V concludes and discusses the role of behavioral effects in the field of matching.

II. Fairness and priorities

In the school choice problem, a finite set of students competes for assignment to a finite set of schools. Each student can at most be assigned to a single school and has strict preferences over schools and each school has strict priorities over students. Each school has a finite number of seats and can only accept students up to the capacity. A school problem consists of the collection of preference profiles and priority orders and an algorithm solves the school choice problem by producing a matching of students to schools.

A matching algorithm is defined as fair if it exclusively produces fair matching outcomes. A matching outcome is referred to as fair in the absence of justified envy. Finally, a student has justified envy, if she prefers another student's assignment to her own assignment, and also has a higher priority than this other student at the school he is assigned to. In other words, fairness forbids these specific priority violations and thereby eliminates exposure to justified envy for all students.

This paper employs the Efficiency Adjusted Deferred Acceptance Algorithm (EDA)[3] (Kesten, 2010), in a setting known as school choice with consent, as a tool to reveal the valuation of fairness. EDA takes the DA as a starting point and improves efficiency by allowing for justified envy if participants give their consent. Crucially, in the EDA no participant is ever disadvantaged by consenting himself nor by the consent of any other student. The algorithm only considers consent decision if participants turn out to be so called "interrupters". Interrupters get accepted to a school to the detriment of another student, thereby triggering a chain of rejections which ultimately leads to themselves being rejected at this particular school. An interrupter's consent decision allows another student to ignore the interrupters priority at a school the interrupter has no chance of being assigned to, thereby inflicting justified envy on himself. The experiment measures the value of fairness by eliciting the consent decision of participants in the EDA.

[3] See Table 1 for an informal description.

TABLE 1: MATCHING ALGORITHMS

Deferred Acceptance Algorithm (DA)	
Step 1:	Each student proposes to her most preferred school. Each school tentatively accepts out of all proposals the students with the highest priority up to capacity and rejects all others.
Step k, k≥ 2:	Each rejected student proposes to her most preferred school which she has not already proposed to. Each school tentatively accepts out of all proposals (new as well as tentatively accepted) the students with the highest priority up to capacity.
	The algorithm terminates if there are no more proposals.
Efficiency Adjusted Deferred Acceptance Algorithm (EDA)	
Step 0:	Elicit binary consent decision.
Step 1:	Run DA. Identify the last step of the DA in which a consenting interrupter was rejected by a school she is an interrupter for. Remove this school from the preferences of the interrupting student.
Step k, k≥ 2	Rerun DA. Identify the last step of the DA in which a consenting interrupter was rejected by a school she is an interrupter for. Remove this school from the preferences of the interrupting student.
	The algorithm terminates if there no more interrupters are identified.

Notes: For a formal and more detailed description of the Efficiency Adjusted Deferred Acceptance Algorithm (EDA) see Kesten (2010) and see Dur et. al. (2019) for a generalization.

From a purely rational and selfish perspective, students are indifferent regarding their consent decision (Kesten, 2010). The decision might positively influence the matching outcome of fellow students, but never their own. The EDA is efficient if all students consent, which corresponds to the null hypothesis. If the null cannot be rejected, this would be a strong argument against the idea that participants value fairness and in favor of using efficient algorithms in the first place.

Hypothesis 1: No full consent: A significant share of pooled participants refuses to consent.

In contrast, I hypothesize that a significant fraction of participants refuses to consent. From a behavioral economics standpoint, there are several reasons to refuse to consent. The literature on other regarding preferences proposes that people do not only care about their own outcomes, but also about the outcomes of others (see Bolton and Ockenfels, 2000 and Fehr and Schmidt, 1999). People tend to dislike inequality in general, but crucially there is imbalance between disliking being worse off compared to being better off than the reference group. Consent allows another student to be assigned to a school she prefers to her own match, making the other student potentially better off than her.[4] Other behavioral effects such as entitlement (discussed in more detail below) and default effects (discussed in the conclusion) may work in the same direction.

School choice mechanisms have two main inputs: (1) preferences of students and (2) priorities of students at schools. In contrast to the college admission problems, the priorities are assumed to be exogenously imposed to schools (for example by legislation). Therefore, schools are regarded merely as objects assigned to students and the welfare of schools is not considered. In practice, the origin of the priority structure might widely differ. Typical examples include specific student characteristics, student performance, random draws or some combination of those (see table 2).

TABLE 2: ORIGIN OF PRIORITY STRUCTURE

Student characteristics:	Walk zone distance, siblings, minorities etc. *Example: Boston public school admission (see Abdulkadiroğlu et al. 2005b)*
Student performance:	Entrance exams, grades *Example: Chinese College admission system (see Chen and Kesten, 2017)*
Random:	Fully random, partly random *Example: Tie breaking within priority classes in Boston*

From a student's perspective, priorities can be interpreted as a specific form of endowment. Previous experimental research in psychology as well as economics has shown that performance or merit-based endowments induce entitlement and are highly influential on

[4] Because preferences are not common knowledge interrupters do not know if they will help students to improve who are better or worse off than themselves in terms of the other student's preferences ranks. But, given the imbalance in disliking inequality students might prefer not helping in the first place or project their own preferences on other students due to the lack of information.

behavior. Most prominently, it has been shown that performance-based endowments induce selfish behavior in dictator games (see Cherry et al. 2002 and Korenok et al. 2017). In line with this research, I hypothesize that the origin of the priority structure affects the valuation of fairness and therefore the willingness to consent.

Hypothesis 2: Priority Structure: If the priority structure is based on performance, consent decreases compared to priorities based on random draws.

III. Experiment design

The experiment consists of two independent parts with separate instructions. In part 1 participants work individually on a task and in part 2 they compete one-shot in a subsequent matching market.

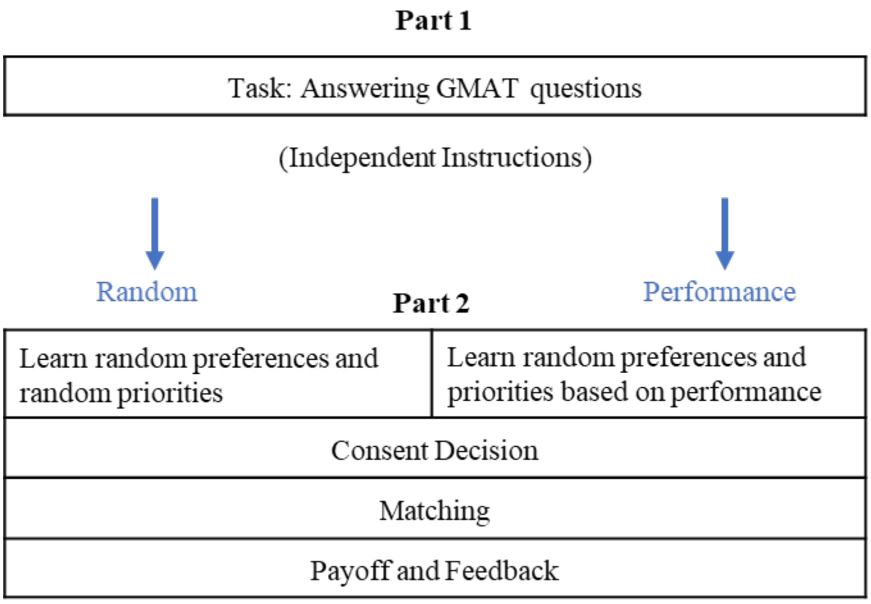
In the first part subjects answer randomly ordered questions taken from the Graduate Management Admission Test (GMAT™).[5] GMAT is a standardized admission test employed by more than 1500 universities and colleges worldwide, primarily in the United Kingdom and the USA. All questions are multiple-choice and have five possible answers one of which exactly is correct. Subjects have 10 minutes to answer as many questions as they can and are paid out a fixed fee, independently of performance.[6] Questions fall into one of two categories: verbal or quantitative. After 10 minutes participants receive feedback on performance, including the total number, number of verbal and number of quantitative questions answered.

The second part of the experiment consists of 4 distinct stages. Subjects are randomly matched into groups of four and assigned the role of students competing for assignment to four schools, each with the capacity of accepting a single student.

[5] GMAT questions are an established paradigm in experimental economics and have been employed for example in Cherry, Frykblom and Shogren (2002) to induce entitlement in dictator games.

[6] I do not pay according to performance to avoid income effects in part 2 of the experiment.

FIGURE 1: EXPERIMENT DESIGN



In the first stage subjects learn about their own preferences and the priorities of schools (see Table 3). Preferences are drawn randomly and are displayed as a rank order list. In both treatments, priorities of school A and school B, as well as school C and school D are identical. This ensures that priorities are impactful, but still allows for variance, which is important because perfectly correlated priorities imply the absence of interrupters. Depending on the treatment, priorities are either based on performance or on random draws. In PERFORMANCE, school A and school B rank students according to the number of correctly answered verbal questions and school C and school D according to quantitative questions[7]. This mimics schools with different specializations. In RANDOM, priorities at school A and B are based on a random draw, and priorities at school C and D are based on another random draw. The next stage elicits the students binary consent decision. Subjects are informed that the consent decision will never affect their own outcome but may positively affect the outcome of other students. In the matching stage the EDA algorithm assigns students to schools taking as an input preferences, priorities and consent decisions.[8] A description of the algorithm for participants was provided in the instructions (see Appendix B).

[7] Ties are broken randomly.

[8] I do not elicit student preferences and the algorithm assumes preferences are reported truthfully even though EDA is not strategy-proof (an algorithm is strategy-proof if it is a weakly dominant strategy for students to report preferences truthfully). However, this paper focuses on the consent decision and therefore abstracts from the problem of strategy-proofness.

Finally, in the last stage students are informed about the assignments, consent decisions and corresponding payoffs. Payoffs are based on assignments and participants receive 7 € for being assigned to their most, 5 € for their second most, 3 € for their third most and 1 € for their least preferred school.

TABLE 3: PREFERENCES AND PRIORITIES

Student 1	**School A**	**School B**	**School C**	**School D**
School A	Student 1	Student 1	Student 4	Student 4
School B	Student 2	Student 2	Student 3	Student 3
School C	Student 3	Student 3	Student 1	Student 1
School D	Student 4	Student 4	Student 2	Student 2

Notes: An example for the preferences and priority structure provided to subjects during the experiment. See also Appendix B for the full instructions.

All sessions were conducted in April 2018 in the Cologne Laboratory for Economic Research (CLER). Participants were students from the University of Cologne invited via ORSEE (Greiner 2015) and the experiment was programmed with z-tree (Fischbacher 2007). I conducted three sessions for each of the two treatments, amounting to 96 participants and independent observations for each treatment. Sessions lasted for around 45 minutes and participants earned on average 13.82 € with a minimum of 9 € and a maximum of 15€.

IV. Results

In the first part of the experiment subjects answered a minimum of 0 and a maximum of 8 questions in RANDOM and 9 questions in PERFORMANCE. On average 3.29 question and 3.56 were answered respectively. There are no statistically significant differences between treatments or the average number of correctly answered questions in the verbal or quantitative category.

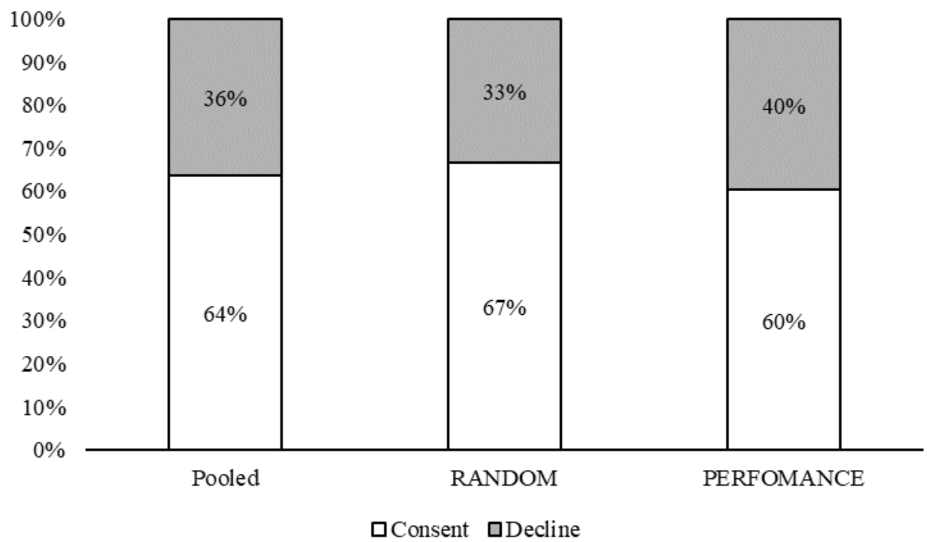
TABLE 4: GMAT QUESTIONS

Random	**Minimum**	**Maximum**	**Average**
Verbal	0	4	1.81
Quantitative	0	5	1.75
Total	1	9	3.29
Random	**Minimum**	**Maximum**	**Average**
Verbal	0	4	1.62
Quantitative	0	5	1.68
Total	0	8	3.56

Notes: The table provides summary statistics for part 1 of the experiment. The data is pooled for both treatments and consists of 96 independent observations.

In line with hypothesis 1 I find that for the pooled data 36.46 % of subjects refuse to consent (see Figure 2). Thereby, I can reject the null hypothesis of full consent (p<0.001***, Wilcoxon sign-rank test). This is an indication that the EDA might not be able to achieve full efficiency and some particpants might indeed care about fairness in this context.[9]

FIGURE 2: TREATMENT EFFECT



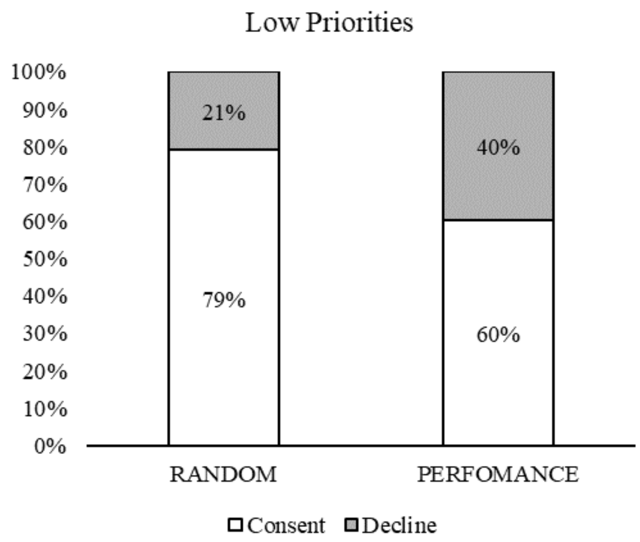
Notes: This figure displays share of subjects who consent/ decline to consent pooled, as well as for both treatments individually. The pooled data consists of 96 and both treatment of 48 independent observations.

[9] Refusal to consent does not need to be driven by fairness concerns. Alternatively, subjects could simply decide randomly or do not trust the instructions that consent will never harm them.

Comparing consent across treatments I find in line with hypothesis 2 a more than six percentage points higher consent share in Random (66.67 %) compared to PERFORMANCE (60.42 %). However, employing a Wilcoxon rank-sum test I find the difference not to be significant on the five percent level (p=0.369).[10]

Next, I will focus on students who are endowed with "low priorities". Those students are more important for the outcome of the EDA because they are more likely to be interrupters.[11] Students are classified by adding up the rank of their two most preferred schools and afterwards a median split classifies half of the subjects as having "high priorities" and the other half as being endowed with "low priorities". Now comparing consent across treatments for subjects with low priorities (see Figure 3), I find a close to 20 percentage points higher consent share in RANDOM (79.17 %) compared to PERFORMANCE (60.42). The treatment effect is significant (*p*=0.0466*, Wilcoxon rank-sum test), but does not necessarily have to be driven by entitlement.

FIGURE 3: TREATMENT EFFECT BY PRIORITY GROUPS



Notes: This figure displays the share of subjects with low priority who consent/ decline to consent for both treatments. The pooled data consists of 24 independent observations for both treatments.

[10] Out of the independent 48 school choice markets, in 5 markets the DA was not efficient (3 markets in treatment RANDOM and 2 markets in treatment PERFORMANCE). In all 5 instances the EDA reached the efficient outcome.

[11] Interrupters need to be rejected by a school in the algorithm which is more likely with low priorities. For example, a student endowed with the highest priority at her most preferred school can never be an interrupter.

A potential confound would be that subjects with "low priorities" also differ in some other relevant dimension from subjects with "high priorities". One obvious candidate would be ability, measured by the number of solved GMAT questions. As a control I do a median split for the treatment RANDOM, categorizing the subjects as "high ability" or "low ability". The results show that "high ability" subjects in RANDOM indeed consent more than "low ability subjects" (75 % and 58.33 %), but not significantly (p=0.0849). Disregarding the underlying mechanism, in practice participants endowed with low priorities will also be of the "low ability" type if priorities are performance based.

V. Conclusion

This paper extends the literature on behavioral matching in general, by introducing entitlement and social preferences into the discussion. Also, the paper contributes to the discussion on the incompatibility of fairness and efficiency and the EDA as a potential solution. From a purely rational and selfish point of view participants are indifferent regarding their consent decision. I find in the experimental study that a substantial share of participants refuses to consent and that the origin of the priority order influences the consent decision of students with "low priority". This has implications for market design. First, it strengthens the argument for avoiding justified envy as currently happens in practice. If participants in matching algorithms do indeed value fairness, one should be cautious to employ efficient algorithms such as the Top Trading Cycle. Additionally, this indicates that mechanisms allowing for voluntarily waving of priorities such as EDA probably improve efficiency but are unlikely to reach full efficiency. Second, the results point to the importance of the origin of the priority order for the value of fairness. If priorities are earned rather than just randomly assigned, participants with low priorities seem to consent less. A market designer might therefore want to take the origin of the priority order into account when choosing a matching algorithm.

In the field economic decision makers do not necessarily behave rationally and selfishly. Designers of real world-institutions and mechanisms are therefore concerned not only with incentives, but also behavior (Bolton and Ockenfels, 2012). The importance of robustness towards behavioural biases has been acknowledged in the mechanism design literature (see for example Bierbrauer and Netzer 2016, Bierbrauer et. al 2017) and guided design ideas in procurement (see for example Engelbrecht-Wiggans, Haruvy and Katok, 2007) as well as auction design in general (see for example Bergemann et al. 2021).

Similarly, also in the field of matching theory behaviour has been receiving more attention in the recent years, specifically around the property of strategy-proofness. Acknowledging the possibility of participants in matching markets failing to understand highly complex mechanisms, which may result in strategic errors, new properties such as obvious[12] strategy-proofness (see Li, 2017) have been introduced to the literature. Seeing cognitive limitations rather as an opportunity than a problem, Troyan and Morril (2020) propose a weaker notion of strategy-proofness, restricting fail cases only to "obviously" manipulatable mechanisms. EDA also falls in the class of not obviously manipulatable mechanisms.

Fernandez (2020) proposes regret-free truth-telling as a weaker incentive property. Interestingly, Chen and Möller (2021) show that EDA, even though not being strategy-proof, is regret-free truth-telling under incomplete information. This strengthens the argument for the EDA as a promising candidate for employment in the field.

Given related studies in the behavioral economics literature, there are potentially other behavioral effects interacting with the value of fairness in matching markets. The literature on status-quo and default effects (see for example Tversky and Kahneman 1974) suggests that the framing of the EDA, endowing participants with the fair assignment and asking for consent to waive afterwards, might be inferior to a framing in which participants are endowed with the efficient outcome and given veto rights afterwards. Another feature of matching markets, the correlation of preferences might also affect behavior. If preferences are highly correlated, the matching market becomes more competitive. A number of studies have shown that (perceived) competitiveness affects behavior. For example, Goette et al. (2012) show how competition between groups increases anti-social motivations or Ellingsen et al. (2012) demonstrate how framing a prisoner's dilemma as Community-Game increases cooperation compared to a Stock Market Game label. Exploring how default effects, framing and competitiveness interact with matching algorithms in general and the value of fairness in particular will be an interesting field for future research.

[12] The term "obvious" refers to recognizable for cognitively limited agents.

Appendix A. Experimental instructions

Treatments: PEFORMANCE/RANDOM

General Part

Welcome to our experiment!

Please read the following instructions carefully. If you have a question, please raise your hand. We will come to you and answer your question. Private communication with other participants is not allowed throughout the experiment. If you violate this rule, we may exclude you from the experiment and all payouts.

For showing up, all participants will receive €4. In addition, you can achieve other payouts that depend on your decisions and the decisions of other participants. The experiment consists of two parts. At the end of the experiment, the payouts for both parts and the €4 for showing up will be added up and paid to you in cash. All decisions and payoffs in this experiment will be treated anonymously. All participants will receive identical instructions.

You will first receive the instructions for the first part. After the first part is completed, you will receive the instructions for the second part of the experiment.

Part 1

In the first part of the experiment, you have 10 minutes to answer multiple-choice questions in English from the areas of language comprehension and numerical reasoning. There are five answer options for each question, of which exactly one answer option is correct. You can skip questions and return to them later. For taking the multiple-choice test you will receive an amount of €4.

At the end of part 1, all participants are informed about the number of correctly answered questions in language comprehension and the number of correctly answered questions in numerical reasoning. Afterwards, Part 2 of the experiment begins with new instructions.

Part 2

In part 2 of the experiment, each participant takes on the role of a student who is assigned to a school. The payoffs in part 2 depend on this assignment. At the beginning, the participants are randomly divided into groups of four students each. Each group member is randomly assigned the role of student 1, student 2, student 3 or student 4. In addition to the four students, there are four schools, School A, School B, School C and School D, each of which assigns only one place. The students each have a random personal ranking across the four different schools. The students' rankings indicate the order in which they rank the schools. The top school is their first choice, then come the second and third choices, and the bottom school is the last choice.

Student 3
School A
School B
School C
School D

Example: the first choice of student 3 is school A, the second choice of student 3 is school B, the third choice of student 3 is school c and the last choice of student 3 is school D.

The payouts in part 2 of the experiment depend on which schools the students are assigned to and can be seen in the table below. If a student is assigned to their first choice, their payout is €7, if a student is assigned to their second choice, their payout is €5, if a student is assigned to their third choice, their payout is €3 and if assigned to their fourth choice, their payout is €1.

RANKING	**PAYOUT**
1.	7 €
2.	5 €
3.	3 €
4.	1 €

The schools each have a ranking list of students. The ranking lists indicate which student has which priority at a particular school. For each school, the students are sorted from highest to lowest priority. The computer takes these priorities into account when assigning them. The higher the priority of a student at a school, the better the chances of being assigned to that school. The exact meaning of the priorities is described at stage 2.

School A	School B	School C	School D
Student 1	Student 1	Student 4	Student 4
Student 2	Student 2	Student 3	Student 3
Student 3	Student 3	Student 1	Student 1
Student 4	Student 4	Student 2	Student 2

Example: The highest priority at school A is student 1, the second highest priority is student 2, the third highest priority is student 3 and the lowest priority is student 4.

[PERFORMANCE: The rankings of the schools are based on the results from the multiple-choice test from the first part of the experiment. School A and School B have a language focus and sort the students according to the number of correctly answered questions in language comprehension. School C and School D have a mathematical focus and sort the students by the number of correctly answered numerical reasoning questions. In this example, student 1 has answered the most language comprehension questions correctly, student 2 the second most, student 3 the third most and student 4 the least. If two or more than two students have answered the same number of questions correctly, the sorting of these pupils is decided randomly.**]**

[RANDOM: The ranking lists of the schools are formed randomly. Hereby applies that the rankings of school A and school B and the rankings of school C and school D are the same. The two different ranking lists are determined randomly and independently of each other by the computer. In the example, student 1 was randomly assigned the highest priority at school A and school B, student 2 the second highest, student 3 the third highest and student 4 the lowest.**]**

The second part of the experiment consists of two stages. At the first stage, you will be informed about your personal ranking and the rankings of all schools and decide whether to waive strict adherence to your priorities. At the second stage, you will be informed about the assignment of students to schools and your associated payout.

In the following, the two stages are explained in more detail.

Stage 1: Decision

The computer uses an algorithm to assign each student to the best school according to their ranking. Before the computer makes the assignment, students decide whether they want to waive strict adherence to their priorities. Not strictly adhering to one's own priorities never changes one's own assignment but can in some cases improve the assignment of other students. Whether the assignment of other students can be improved in the case of a waiver depends on the ranking lists of the students and schools and is not known to the students.

Strict adherence to a student's priorities means that the priority of a student with a higher priority over a student with a lower priority is always respected. Strict adherence to your priorities therefore ensures that the following does **not** happen:

You are assigned to a school. Another student is assigned to a school that you prefer over the school assigned to you, *and* you have a higher priority at that school than the other student.

Student 3	**School A**
School A	Student 1
School B	Student 2
School C	Student 3
School D	Student 4

Example: You are student 3 and are assigned to your second choice school B. You prioritize your first choice school A over your second choice school B. Student 4 is assigned to your first choice school A. You have the third highest priority at school A and student 4 has the fourth highest priority. So you have a higher priority at school A than student 4, and you prefer school A over your assigned school B. So when you assign student 4 to school A, your priorities are not strictly followed

The decision to waive strict adherence to your priorities never changes your own assignment. It allows the computer in some cases to disregard your priority in the assignment, and thus improve the assignment of other students. Therefore, the computer will only ever disregard your priority at a school if it can improve the assignment of other students without changing your own assignment.

Student 1	Student 2	Student 3	Student 4
School C	**School D**	School A	School A
School A	School C	**School B**	**School C**
School D	School A	School C	School B
School B	School B	School D	School D

Example: Student 1 is assigned to his second choice school A and student 4 is assigned to his second choice school C. Student 1 and student 4 would like to switch schools because student 1 prefers school C over school A and student 4 prefers school A over student C. However, student 3 also prefers school A over his assigned school B and has a higher priority at school A than student 4. Thus, in the swap, student 3's priorities would not be strictly adhered to. If student 3 waives strict adherence to his priority, the swap is performed by the computer and the assignment of student 1 and student 4 is improved (see arrows). Student 3's assignment is not changed by waiving strict adherence to his priority.

Stage 2: Assignment

The computer uses an algorithm to determine the assignment of students to schools, considering student and school rankings and students' decisions to waive strict adherence to their priorities. A detailed description of the algorithm can be found in the appendix. Each school is assigned exactly one student. The assignment is shown in the ranked lists of students. The school to which a particular student is assigned is underlined in the student's ranking list.

Student 1	Student 2	Student 3	Student 4
School C	**School D**	School A	School A
School A	School C	**School B**	**School C**
School D	School A	School C	School B
School B	School B	School D	School D

Example: student 1 is assigned to school A here, student 2 is assigned to school D, student 3 is assigned to school B, and student 4 is assigned to school C.

Appendix

Description of the computer algorithm

The algorithm is performed automatically by the computer using the ranking lists of schools and students. It does not require any further input from the students. At the end of the algorithm is the assignment of students to schools. The algorithm is presented below in two parts that are linked together:

Part 1

- *Each student applies at his or her first choice.*
- *Each school provisionally accepts the student with the highest priority and rejects the remaining students. Tentatively accepted students may be rejected again as the algorithm progresses.*
- *Each student who is rejected applies to the school that is highest in his or her personal ranking and has not previously rejected him or her.*
- *Each school provisionally accepts the student with the highest priority and rejects the remaining students. Both new applications and the provisionally accepted student are considered. Thus, the provisionally accepted students may also be rejected again.*
- *Each student who has been rejected applies to the school that is highest in his or her personal ranking and has not previously rejected him or her.*
- *Each school provisionally accepts the student with the highest priority and rejects the remaining students. Both new applications and the provisionally accepted student are considered. Thus, the provisionally accepted students may also be rejected again.*
- *This process is repeated until there are no more rejections and the students are then assigned to the schools that provisionally accepted them.*

The assignment resulting from Part 1 of the algorithm, under strict adherence to priorities, assigns each student to the best possible school.

Part 2

- *The computer searches for a student who has waived strict adherence to his priority and for whom non-adherence to his priority can improve the assignment of other students without changing his assignment.*
- *If the computer finds a student, the school that the student's priority blocks from other students (without getting a spot at that school itself) is removed from the found student's personal ranking.*

- *Part 1 of the algorithm is then rerun with the student's new ranking. When rerun, the assignment of other students is then improved without changing the assignment of the found student.*
- *This process is repeated until the computer finds no more students. Then the algorithm ends and the assignment from part 1 of the algorithm is finalized.*

If multiple students are found by the computer, the student identified at the later time in the algorithm is selected.

A student who can improve the assignment of other students by not strictly adhering to their priorities is identified by the following characteristics in Part 1 of the algorithm:

- *Another student is rejected in favor of this student at a school.*
- *At a later point in the algorithm, this student is also rejected by this school due to another application (the point of identification is the rejection).*

Chapter 4

Choosing norm information: what other people do or what they think is ought to be done

CO-AUTHOR: CHRISTOPH FELDHAUS*

Abstract

We experimentally investigate people's choice between information regarding what others do and what others think is ought to be done in a social choice context in which a decision-maker can take away money from a charity. This choice is made just before they themselves act in the same social choice context, but without knowing the charities identity. We vary (i) the role of the person that chooses between the two types of information, she is either a decision-maker or a choice-architect, and (ii) the probability that her decision is actually implemented. We observe that most participants choose to be informed about what others do rather than what they think what is ought to be done, irrespective of their role. However, we further observe that the share of decision-makers and choice-architects that choose information on what is ought to be done increases when their own decision is more likely to be implemented. We discuss a potential hypothesis in line with this observation.

* Breuer: kevin.breuer@wiso.uni-koeln.de, Department of Economics, University of Cologne, D-50923 Cologne, Germany; Feldhaus: christoph.feldhaus@rub.de, Faculty of Management and Economics, Ruhr-University Bochum, D-40780 Bochum, Germany. Financial support from the Center for Social and Economic Behavior (C-SEB) is gratefully acknowledged. We thank Kiryl Khalmetski, Felix Kölle and Axel Ockenfels for very helpful comments and suggestions. Breuer and Feldhaus gratefully acknowledge support from the European Research Council (ERC) under the European Union's Horizon 2020 Research and Innovation Programme (grant741409); the results reflect only the authors' views; the ERC is not responsible for any use that may be made of the information it contains.

I. Introduction

Social norms are a crucial factor when it comes to decision making in social contexts. Norms have been found to influence behavior across a wide variety of situations, such as donation behavior, energy consumption, tax payments or punishment. However, in order to become relevant for their choices, economic agents have to search for or have to be confronted with the respective norm information. Additionally, first-party decision-makers are typically observed by third-parties, who judge behavior or sometimes even interfere in the choice-set of decision-makers and act as choice-architects (for example other people, legislators or ethic committees). In this paper, we investigate how economic agents, first-party decision-makers and third-party choice-architects, decide between different types of norm information to learn about their underlying motivation for acquiring information.

Norm information may come in two different types; it either provides people with data on 'what other people do' (*often referred to as the descriptive norm*) or with data on 'what other people think is ought to be done' (*often referred to as the injunctive norm*) in a given context (Cialdini et al. 1990, 1991, Bicchieri 2005, Jacobson et al. 2011). In this paper, we let experimental participants decide between these two types of information regarding a social choice context in which a decision-maker can take away money from a charity. Specifically, they choose between the two types right before acting in the same social context, but with three major differences. First, a choice-architect observes the decision-maker and, depending on the treatment, may have the option to interfere in the choice set of the decision-maker. Second, actions are implemented only with a certain probability, and third, the identity of the charity remains unknown. This allows us to investigate which type of norm information participants decide to learn about and why. Thereby we contribute to the literature on social norms in general and to the literature on (strategic) information acquisition in particular. We observe that the relevance of a decision, in terms of the probability that the decision is actually implemented, affects people's information choice in favor of learning about 'what is ought to be done'.

In our online experiment, we have two types of participants in each of our conditions, the "decision-makers (DMs)" and the "choice-architects" (CAs). DMs choose between taking money away from an unknown charity or not. They know that their taking decision is either implemented with low or high probability, or with certainty. The other type of participants, the CAs, can, depending on the treatment, either only observe the decision of the DM or decide to costly restrict the choice set of the DM to remove the option to take money away from the charity. The restriction is also either implemented with low or with high probability. If a CA restricts the choice-set of the DM and the restriction is implemented, a taking decision by the DM is ex-post overruled. Before deciding on whether or not to take money away from the charity (DMs) or restricting the decision to take money away from the charity (CAs), all participants have to choose between two types of norm information collected in a pre-study: they can either receive information regarding the share of participants that decided to take money away from a charity or information regarding the share of participants that said that it was appropriate to take money away from a charity. In contrast to the participants of the main study, the participants of the pre-study knew about the identity of the charity, which was also known by the participants of the main study. This allows us to make the norm information meaningful, as only pre-study participants could assess the extent to which the charity is worthy of support. Moreover, at the beginning of the main study, we implicitly point out to participants that they are acting in a social choice context in which people's actions falls short of what they think is right, i.e. people tend to act more selfishly than they think is right. This will be important for the hypotheses later on. Anecdotally, many interesting cases of norm conflict, such as tax evasion, environmental pollution, or energy waste, fall into this class of norm environment as people tend to fall short of their ideals (see for example Wenzel 2005).

We aim to study three factors that may be related to people's choice regarding the type of norm information. First, they may self-servingly choose the type of information that best enables them to receive the most money for themselves, i.e. they may choose the type of information that tells them that it is legitimate to take the money (*Self-serving hypothesis*). Second, they may choose the type of information that best suits their role in the respective decision context, such that DMs may choose to know what others do and CAs may choose to know how others judge (*Role hypothesis*). Third, their choice may be affected by the relevance of their decision in terms of the probability that their choice will actually be implemented, i.e. they may be more eager to learn what is the right thing to do as their responsibility increases (*Responsibility hypothesis*).

In our experiment, we observe that overall a majority of participants prefers to learn about the behavior of the participants in a previous experiment rather than people's assessment of which behavior is deemed appropriate. However, we observe that participants tend to be more likely to choose information on what is appropriate when their decision gets more relevant. This observation holds for both, the DMs and the CAs. We neither find evidence supporting the idea that people choose self-servingly between the two types of information nor that they do so conditional on their assigned role.

Our study adds to different streams of literature. First, we add to the literature on social norms by showing how and why people choose between the two types of norm information before making a related decision. Previous literature has established that people's behavior is affected by both types of social norms, descriptive as well as injunctive norms. For example, social norms matter for individual decisions in the context of charitable giving (Andreoni and Scholz 1998; List and Lucking-Reiley 2002; Frey and Meier 2004), public-good contributions (Fehr and Fischbacher 2004; Rege and Telle 2004), environmental-friendly behavior (Allcott 2011) or tax compliance (Paetzold and Winner 2016; Hallsworth et al. 2017). The relevance of the equal split norm for sharing has been discussed theoretically (Fehr and Schmidt 1999, Bolton and Ockenfels 2000, Andreoni and Bernheim 2009). In addition, and more closely related to our study, some previous papers compare the effects of descriptive and injunctive norms (Bicchieri and Xiao 2009; Bicchieri et al. 2021). It has been observed that descriptive norms tend to have a stronger influence on behavior in a dictator game, especially when norms are in conflict (Bicchieri and Xiao 2009), and it has been shown that injunctive norms are crucial when punishment is involved (Bicchieri et al. 2021). However, the research question which has so far been neglected is how people choose between these two types of norm information and what may cause the corresponding choices.

Second, previous research has demonstrated that people often make use of moral wiggle room (Dana et al. 2007, Exley 2016, Spiekermann and Weiss 2016, Andreoni et al. 2017) by avoiding information that may make them choose more morally and thereby decrease their individual payoff. Similarly, people tend to choose the information that helps them maximize their gains (Ambuehl 2017), and interpret information about norms in a way that evidence suggests they can keep the most money for themselves (Bicchieri et al. 2019, Kassas and Palma 2019, Foerster and van der Weele 2021). We add to this literature by investigating whether the choice between information on norm types is also subject to self-serving behavior.

Third, the literature on social cognition has proposed that the perception of norm violation depends critically on how information about the corresponding group is processed (Fiske and Taylor 2008; Macrae and Bodenhausen 2000). In this context, Mussweiler and Ockenfels (2013) argue that people care in particular about norms being broken by similar others, leading to more altruistic punishment in cooperation games. We investigate if this line of reasoning also extends to acquiring norm information, e.g. if DMs and CAs look for norm information about what similar others (participants in the same role) did in the pre-studies.

Fourth, the literature on the determinants of moral behavior shows that these are related to the individual responsibility and the costs of acting morally (Andreoni and Miller 2002, Fisman et al. 2007). Here, most closely related to our experiment is the study by Falk et al. (2020) who show that people tend to act more morally in a social choice context as their pivotality for the respective decision increases. This is related to our observation that people are more interested in being informed about the right thing to do as the probability that their decision actually gets relevant increases.

Finally and more generally, spectator designs have become popular recently. The respective literature investigates how third parties judge social situations or choose on behalf of others. This includes paternalistic interventions (Ambuehl et al. 2021), interventions in moral behavior (Ackfeld and Ockenfels 2021) or the distribution of resources (Almås et al. 2020). In the present research, we add to this literature by providing evidence on how people that take intervention decisions chose between different types of information and by investigating if their choices differ systematically from corresponding choices by DMs.

The remainder of the paper is organized as follows. The second section presents the experimental design, the corresponding hypotheses and the experimental procedure. The third section discusses the results and the fourth section concludes.

II. Experiment

Design. The experiment consists of a pre-study and a main study. In the pre-study, we collect the injunctive and descriptive norm information that are available for selection in the main study. We used the strategy method to ask DMs if they want to take money away from six different charities, and CAs judged the appropriateness of taking money away from these charities. Endowments are the same as in the main study. For the main study, we selected a charity that matched the norm environment we were aiming for (see Table 2, Step I). Five out of six DMs (83.3 %) took the money away, and only one out of seven CAs (14.3 %) did think it was appropriate to take the money away from the selected charity, DRK e.V. (German Red Cross). The main study aims to better understand under which circumstances people choose to get informed about either the descriptive or the injunctive norm. To this end, we set up a between-subjects experiment that involves three parties – a decision-maker (DM), a choice-architect (CA) and an unknown charity (UC) (see Table 1). In all treatments, the involved parties receive an initial endowment. The DM receives 1.00 €, the CA receives 1.00 € (plus in some treatments additional 0.50 € that may be used to restrict the choice set of the DM) and the UC is endowed with 2.00 €.

TABLE 1: ROLES AND ENDOWMENTS

	Decision-Maker (DM)	**Choice-Architect (CA)**	**Unknown Charity (UC)**
Endowment	1 €	1 € + 0.5 €	2 €

After the participants are informed about this basic experimental setup, the experiment proceeds in 4 steps, which are depicted in Table 2 and differ depending on the treatment. Our main variable of interest is the binary choice of the norm information in Step II. We investigate how this choice is related to the role of the person taking the decision (either DM or CA) and the probability that the decision is actually implemented, which is either high (99 %) or low (1 %). Thus, the experiment includes four treatments in a between-subjects design, DM_H, DM_L, CA_H, and CA_L.

TABLE 2: TREATMENT DESIGN AND EXPERIMENTAL PROCEDURE

	Treatments			
	$\boldsymbol{DM_H}$	$\boldsymbol{DM_L}$	$\boldsymbol{CA_H}$	$\boldsymbol{CA_L}$
Step I: Norm environment	Establishing norm setting by scale manipulation			
Step II: Norm choice	Binary choice between norm information on the injunctive or descriptive norm			
Step III: Taking decision	Decision implemented with 99 % prob.	Decision implemented with 1 % prob.	Decision implemented with 100 % prob.	
Step IV: Observing/Choice set restriction	Only observing taking decision		Choice set restriction implemented with 99 % prob.	Choice set restriction implemented with 1 % prob.

In Step I, we introduce participants to the basic norm environment of the paradigm. That is, we hint to participants that in our decision setting, the injunctive norm lags behind the descriptive norm. In other words, we establish a prior that fewer people think it is appropriate to take the money away from the charity than actually do so. This prior is implemented using two scales to elicit the beliefs of all participants regarding the injunctive and the descriptive norm. Specifically, they answer the following two questions in a random order and check off exactly one item on the scale that corresponds to their belief.

The first question is “What do you think, how many participants thought it would be appropriate for the decision-maker to take the money?” and the second question is “What do you think, how many participants took the money?”. While the first question could be answered on a scale with relatively low values, the second question was answered on a scale indicating higher values (see Table 3). Previous literature has shown that such a variation strongly affects people’s norm-beliefs and corresponding behavior (e.g. Ockenfels and Werner 2014, Feldhaus et al. 2019). In this way, we want to ensure that participants believe that actual behavior falls short of what people think should be done, and that information about behavior is more in line with their own monetary incentives than information about assessments.

TABLE 3: SCALE MANIPULATION

What do you think, how many participants thought it would be appropriate for the decision-maker to take the money?	*What do you think, how many participants took the money?*
◯ $\leq 10\ \%$	◯ $\leq 50\ \%$
◯ $20\ \%$	◯ $60\ \%$
◯ $30\ \%$	◯ $70\ \%$
◯ $40\ \%$	◯ $80\ \%$
◯ $\geq 50\ \%$	◯ $\geq 90\ \%$

The identity of the charity is unknown to the participants in the main experiment. However, in Step II, both the DM and the CA receive information about the pre-study participants prior to their own actions so that they can act in a more informed manner. Specifically, they learn that they can choose between two types of information: either about the behavior or about the assessments of the pre-study participants. The information about the behavior includes the proportion of participants in the pre-study that decided to take money away from the same unknown charity, but without a CA to observe or constrain their behavior. In terms of assessments, they may choose to obtain information on the proportion of participants who thought it was appropriate to take money away from the same unknown charity. Importantly, unlike participants in the main study, participants in the pre-study were aware of the identity and goals of the charity. This was also communicated to participants in the main study. In this sense, participants in the main study choose between information about a descriptive (behavior) or injunctive (assessment) norm in relation to the social decision context in which they are about to act.

In Step III, in all treatments the DM chooses between taking the 2.00 € away from the charity or not. In the decision-maker treatments (DM_H and DM_L) the decision is implemented with either high probability (99 %) or low probability (1 %). In the choice-architect treatments (CA_H and CA_L) the decision is implemented with certainty, if it is not ex-post overruled. If the decision is not implemented, either by chance or an ex-post restriction by the CA, the 2.00 € remains with the charity.

In Step IV, depending on the treatment the CA either only observes the decision of the DM or may ex-post restrict the choice-set of the DM. In the decision-maker treatments (DM_H and DM_L) the CA only observes the decision by the DM. The norm choice of CAs as mere observers will be an important benchmark for investigating the responsibility hypothesis. In the choice-architect treatments (CA_H and CA_L) the CA chooses whether to invest his additional 0.50 €, that he could also keep for himself, into ex-post overruling the DM in case she decides to take the money from the charity. Depending on the treatment, the restriction is either implemented with high probability (99 %) or low probability (1 %). DMs only know that restrictions are implemented with some probability unknown to them. The costs are incurred only if the constraint is actually implemented, but regardless of the DM's decision. That is, no costs are incurred if the choice set restriction happens to fail by chance, but they are incurred if the restriction is implemented but the DM decided not to take the money away in the first place.

Hypotheses. Our experiment is set up to test four hypotheses regarding factors that may influence the norm choice given that we induce a specific relation between the descriptive and the injunctive norm, i.e. participants likely have the impression that the descriptive behavior falls short off the injunctive norm. The null hypothesis is that participants solely care about their own monetary outcome and do not care about the norm information regardless of their role and the implementation probability. In other words, across all treatments and roles the null hypothesis expects that they choose the norm information at random or with 50 % probability respectively.

Hypothesis 1a: Self-serving Decision-Maker (Higher implementation probability leads to more choices of the descriptive norm.): Previous research has shown that people prefer to receive information that enables them to choose selfishly and maximize their payoff. In our context, we argue that this means that DMs are more inclined to search for information that helps them to choose selfishly. Hence, in order to feel better with their choice to maximize their payoff people may prefer to choose the descriptive norm information.

Hypothesis 1b: Self-serving Choice-Architect (Higher implementation probability leads to more choices of the descriptive norm.): Similarly, also the CAs can earn more by not restricting the DMs' choice set. In order to feel better by not restricting the choice set of the DM, the CAs may prefer to receive information about how many people acting selfishly so that there may be less need to restrict the choice set as taking the money away is common anyway.

Hypothesis 2: Role-based norm choice (DMs choose the descriptive norm more often than CAs if implementation probability is low.): The type of information that people choose may be related to the role they fulfill in a given choice context, i.e. we suggest that the DMs may care more about what others do while CAs, as judges, may be more interested in what is ought to be done in a choice context. In other words, extending the reasoning of Mussweiler and Ockenfels (2013), people are more interested in following the norm established by similar others.

Hypothesis 3: Responsibility-dependent norm choice (DMs and CAs choose the injunctive norm more in case of high implementation probability): When people are involved in a social choice context with uncertainty, they may be the more inclined to learn about what is ought to be done in case their choice is more relevant, in our case more likely to get implemented. This is in line with Falk et al. (2020), who show that people tend to act more moral in a social choice context as their pivotality for the respective decision increases. This could also imply that they are more eager to learn about what people think is ought to be done than what people actually do when their decision gets more relevant.

Procedure. All sessions of the experiment were conducted online in December 2020 in the Cologne Laboratory for Economic Research (CLER). Participants were students from the University of Cologne invited via ORSEE (Greiner 2015). The experiment was programmed with z-Tree (Fischbacher 2007) and conducted online with z-Tu (Duch et al. 2020). In total, we collected 939 observations, between 114 and 122 per treatment and role. Participants were randomly matched into groups of two. Sessions lasted for about 20 minutes and participants earned on average 4.00 € with a minimum of 3.50 € and a maximum of 5.50 €. A total 576 € was donated to the charity.

III. Results

In the following, we present our data descriptively and the results along the lines of our four hypotheses. In total, 939 participants took part in our experiment. In terms of their beliefs, the participants think that 66.3 % of the participants of the pre-study took the money whereas they think that only 19 % of the participants of the pre-study think that it is appropriate to take the money. This indicates that we were successful in implementing the norm environment that we aimed for, i.e. an environment where behavior falls short of moral assessment.[1]

Overall, information on the descriptive norm is in higher demand than information on the injunctive norm: among all participants 73.3 % choose the former whereas only 26.7 % choose the latter ($p < 0.001$; t-test). Among the DMs, 62.6 % decide to take the money and 47.2 % of the CAs decide to restrict the choice set of the DMs.

Hypotheses 1a and 1b. We suggested that a self-serving DM or CA may prefer to receive information that enables them to choose selfishly, especially when the expected payoff for a selfish decision increases. Our reasoning implies that looking for information on what others are doing rather than what others think is ought to be done serves this purpose. In our sample, 35.5 % of the DMs choose the injunctive norm in case of high implementation probability whereas 25.4 % of the DMs choose the injunctive norm when the implementation probability is low ($p = 0.09$; chi-squared). Regarding the CAs, we observe that 30.7 % choose the injunctive norm when the implementation probability is high whereas 24.4 % choose the injunctive norm when the implementation probability is low ($p = 0.282$; chi-squared). Both comparisons provide no evidence that participants in our experiment choose the descriptive norm particularly when they want to enable themselves to act selfishly in case of the high implementation probability. In fact, the results point in the opposite direction suggesting that both the DMs and the CAs tend to be more likely to choose the injunctive norm in case of the higher implementation probability (see below).

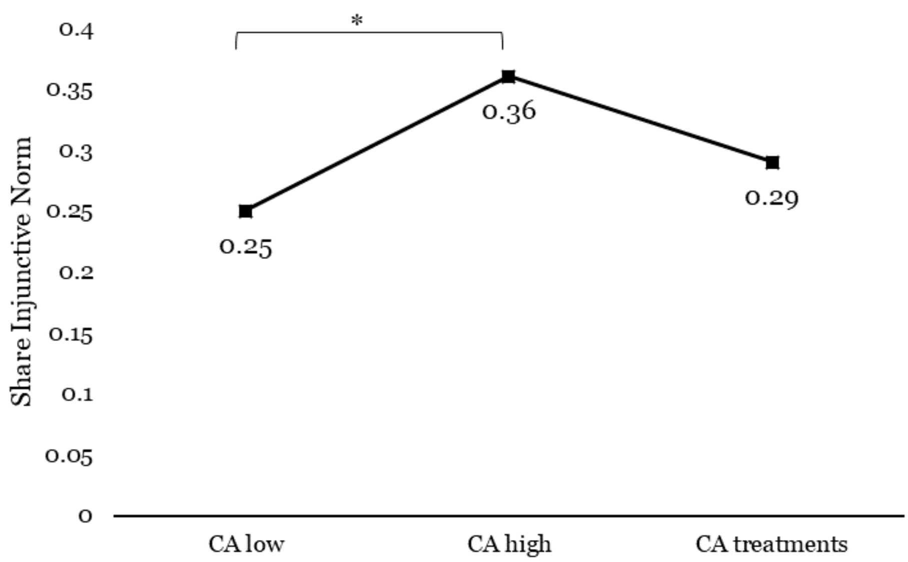
Hypothesis 2. In the second hypothesis, we suggested that the role of the participants in the experiment, either DM or CA, may influence their choice of the norm type. DMs may be more

[1] We find a highly significant difference comparing choices of greater than 50 % for the descriptive scale with choices of greater or equal 50 % for the injunctive scale (p < 0.001, t-test).

interested in what others do, having to take decision themselves, whereas CAs, as judges, may be more interested in what is ought to be done in order to decide whether the choice set should be restricted or not. This is not what we observe in our experiment. In order to have a clear measure of the role-based hypothesis, we only compare the norm choices in the two treatments with low implementation probability so that the financial incentives are unlikely to have an influence on people's choices (the effects of the decision in terms of expected payoffs are very low in both cases). We observe that 25.4 % of the DMs choose the injunctive norm information whereas 24.4 % of the CAs choose the injunctive norm information ($p = 0.849$; chi-squared). This result provides no evidence supporting the hypothesis that people differ in their inclination to choose either the descriptive or the injunctive norm conditional on their role in the experiment.

Hypothesis 3. Finally, we evaluate our third hypothesis that suggests that learning about what is ought to be done may depend on how relevant a decision is. In our case, we argue that the relevance of a decision is based on the probability that it is actually implemented. We first investigate how DMs choose a norm conditional on the implementation probability and then we look at the CAs to find how they choose their norm conditional on the implementation probability.

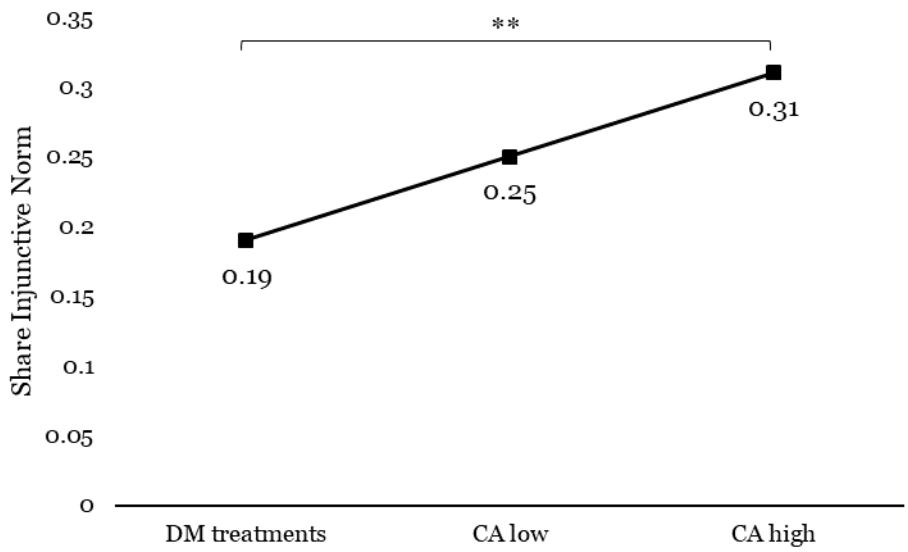
FIGURE 1: DMS NORM CHOICE ACROSS TREATMENTS



In case of the DMs, we observe that they tend to choose the injunctive norm more often when the probability of implementation is high (see Figure 1). In fact, 35.5 % of the DMs choose the injunctive norm when the implementation probability is high whereas 25.4 % of the DMs choose the injunctive norm when the implementation probability is low ($p = 0.09$; chi-squared).

These are the cases where the DMs can decide while only being observed by the CA that cannot restrict their choice set. In the pooled CA treatments (CA_H and CA_L), where the choice architect can restrict the DM's choice set (which likely implies less responsibility on side of the DM as the CAs as judges look over the decision), 28.8% of DMs choose the injunctive norm, and we observe no significant difference between the likelihood of choosing the injunctive norm and DM_H and DM_L. (p = 0.50 for $DM_L$and p = 0.20 for DM_H; chi-squared). This seems to suggest that DMs actually get more likely to choose the injunctive norm as their behavior and their corresponding norm choice gets more relevant for the outcome, i.e. as their responsibility increases.

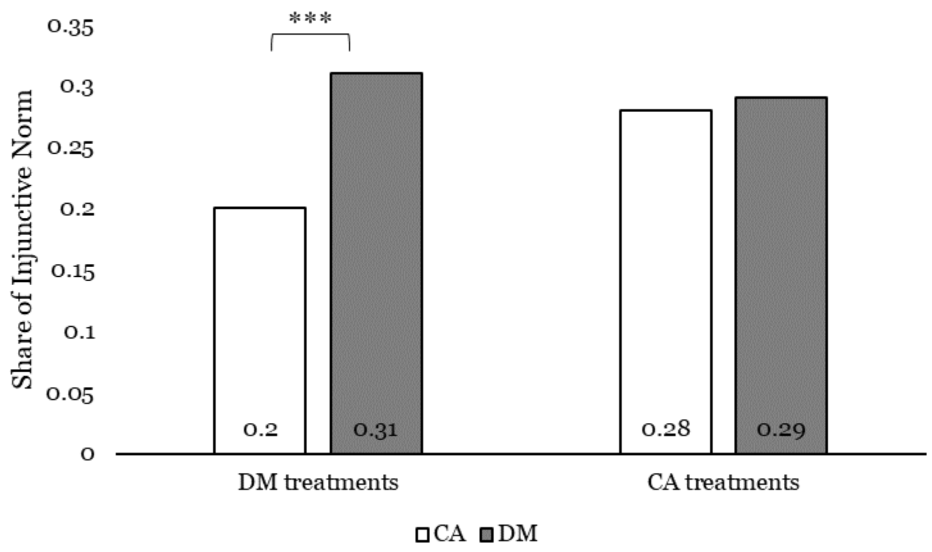
FIGURE 2: CAS NORM CHOICE ACROSS TREATMENTS



We further take the norm choice of the CAs into account (see Figure 2). Here, we compare not only the CAs with low and high implementation probability but also the CAs that only serve as observers of the DMs' behavior without the power to restrict the choice set of the DMs (DM treatments). CAs serving as mere observers also had to choose between the two types of norm information and their behavior may serve as the 'natural' interest in the two types of norms because they do not have any stakes in the game as they are not making a payoff-relevant decision. In fact, we see descriptively that the CAs in the DM treatments choose the injunctive norm less often than CAs who take a restriction decision that is either implemented with low or high probability. In case of the comparison between no restriction decision and the high probability of being implemented this difference is statistically significant (p = 0.029, chi-squared).[2] This provides further evidence that the experimental participants are more interested in the injunctive norm in cases where their decision is more relevant.

Finally, we compare the DMs and their CAs in the treatments where the CAs actually do have the ability to affect outcomes and in the treatments where they are just observers who cannot affect outcomes (see Figure 3). We find that CAs who are mere observers and DMs strongly differ in their willingness to choose information regarding the injunctive norm in case that the spectators cannot affect outcomes which are hence completely determined by the DMs (p = 0.009, chi-squared test). This difference vanishes when both the CAs and the DMs may affect outcomes. In fact, in this case the probability that the injunctive norm is chosen is almost identical across the two types (27.51 % of the CAs and 28.82 % of the DMs choose the injunctive norm respectively, p = 0.755, chi-squared). This finding supports the idea that the relevance of the own decision is an important factor for the choice between the descriptive and the injunctive norm.

FIGURE 3: DIFFERENCE BETWEEN DMS AND CAS IN DM TREATMENTS AND CA TREATMENTS



[2] We find no statistically significant difference between low probability and no control (20.17 % in no control and 24.35 % in low probability of the CAs choose the injunctive norm; p = 0.371; chi-squared) and between low probability and high probability (30.70 % in high probability and 24.35 % in low probability of the CAs choose the injunctive norm; p = 0.282; chi-squared).

IV. Conclusion

Previous research has established the importance of social norms for individual decision making. This holds both for descriptive social norms, what other people actually do and for injunctive social norms, what people think is ought to be done. In the present research, we aim to better understand which of the two types of norm information people care about when making decisions. We suggest three potential factors that may affect people's norm choice. First, participants may strategically choose the type of information that enables them to choose selfishly. Second, their role in a respective context may affect their choice and third, their impact in the situation may render what they want to learn about.

In our experiment, we do not find any evidence suggesting that descriptive norm information is chosen strategically to justify a selfish decision or that the individual's role in a given choice context matters for the norm choice. However, we do observe that the relevance of the upcoming decision in terms of the probability of being implemented affects the norm choice. Specifically, we observe that more participants tend to choose the injunctive norm as their decision gets more relevant for the outcome. This suggests that people may be more inclined to know what is the right thing to do in an unknown choice context as the relevance of their decision increases.

While there are many studies investigating the relevance of both descriptive and injunctive norms for individual decision making, this study is the first study to systematically investigate how people choose to be informed about the descriptive or the injunctive norm before making a decision in a social context. This is important because in contrast to many laboratory studies information on norms is not provided exogenously in the real world, but rather has to be looked for. In norm environments in which descriptive norms fall short of injunctive norms, it may matter a lot which type of information individuals acquire. This is especially true for contexts in which third-parties are involved, such as ethic committees or legislators.

An interesting avenue for future research is to investigate norm choice in different norm environments and different incentive schemes. For example, a comparison between a CA with a low implementation probability and high implementation probability who may restrict the choice set without incurring cost would provide a further test for our responsibility hypothesis.

Appendix A. Additional tables and results

TABLE A.1: SHARE OF INJUNCTIVE NORM CHOICES ACROSS TREATMENTS AND ROLES

	Decision-Maker (DM)	**Choice-Architect (CA)**
Pooled	0.2970 (*N*=468)	0.2379 (*N*=467)
DM treatments	0.3054 (*N*=239)	0.2017 (*N*=238)
DM high	0.3554 (*N*=121)	0.1917 (*N*=120)
DM low	0.2542 (*N*=118)	0.2119 (*N*=118)
CA treatments	0.2882 (*N*=229)	0.2751 (*N*=229)
CA high	0.2895 (*N*=114)	0.3070 (*N*=114)
CA low	0.2870 (*N*=115)	0.2435 (*N*=115)

TABLE A.2: OLS REGRESSIONS OF TAKING AND RESTRICTION

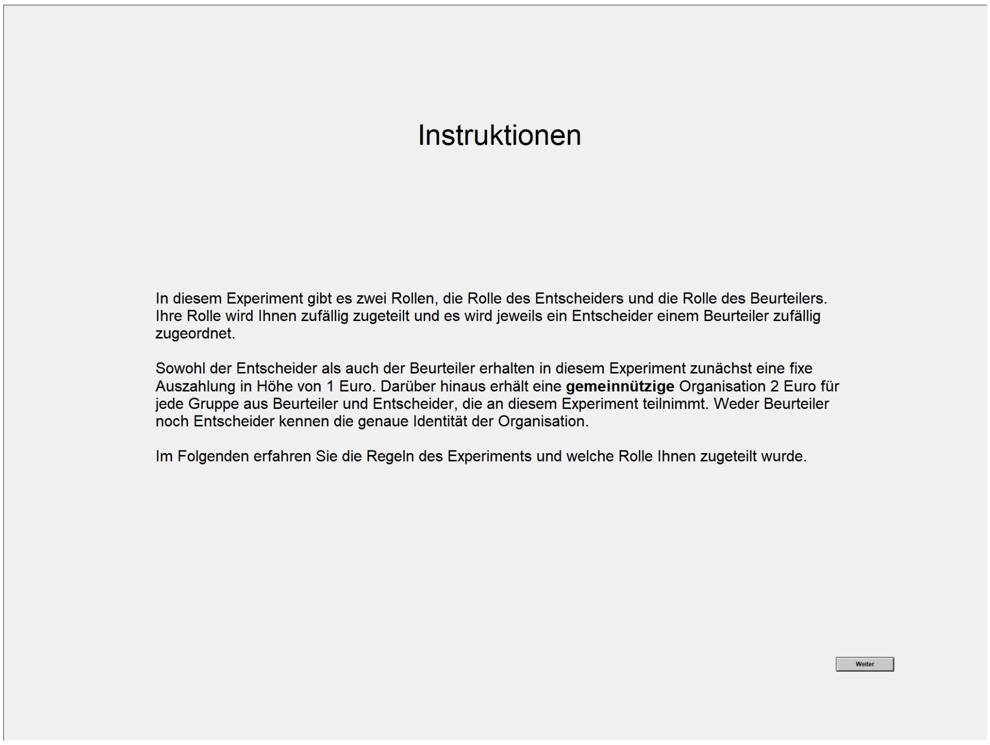
Regression	(1) Taking in DM treatments (Baseline DM_H)	(2) Restricting in CA treatments (Baseline CA_H)
Constant	0.422*** (0.136)	0.326** (0.136)
Exp. inj.	0.193*** (0.066)	-0.178** (0.080)
Exp. des.	0.176*** (0.058)	-0.076 (0.070)
Female	0.035 (0.058)	0.154** (0.076)
Risk taking	0.033*** (0.010)	0.005 (0.015)
Patience	0.007 (0.010)	0.006 (0.013)
Altruism	-0.056*** (0.011)	-0.001 (0.017)
Norm choice des.	0.338*** (0.063)	0.044 (0.072)
Low Impl. Prob.	-0.085*** (0.054)	0.097 (0.67)
N	239	229
R-squared	0.3024	0.0558

Notes: Robust standard errors in parenthesis. Exp. Inj. and Exp- Des. refer to choices in Step 1, scale manipulation. Gender is taken from the subsequent questionnaire, and risk-taking, patience, and altruism are asked about the global preference survey (see Falk et al. 2018). Norm choice descriptive is a binary variable indicating if the descriptive norm was chosen in Step two. Low Impl. Prob. is a treatment dummy referring to the respective treatment with low implementation probability. * p<0.10, ** p<0.05, *** p<0.01.

Appendix B. Screenshots: Instructions and experimental design (relevant translations are below the respective screenshot)

Screenshot 1. General instructions

Standard screen welcoming the participants and explaining general rules.



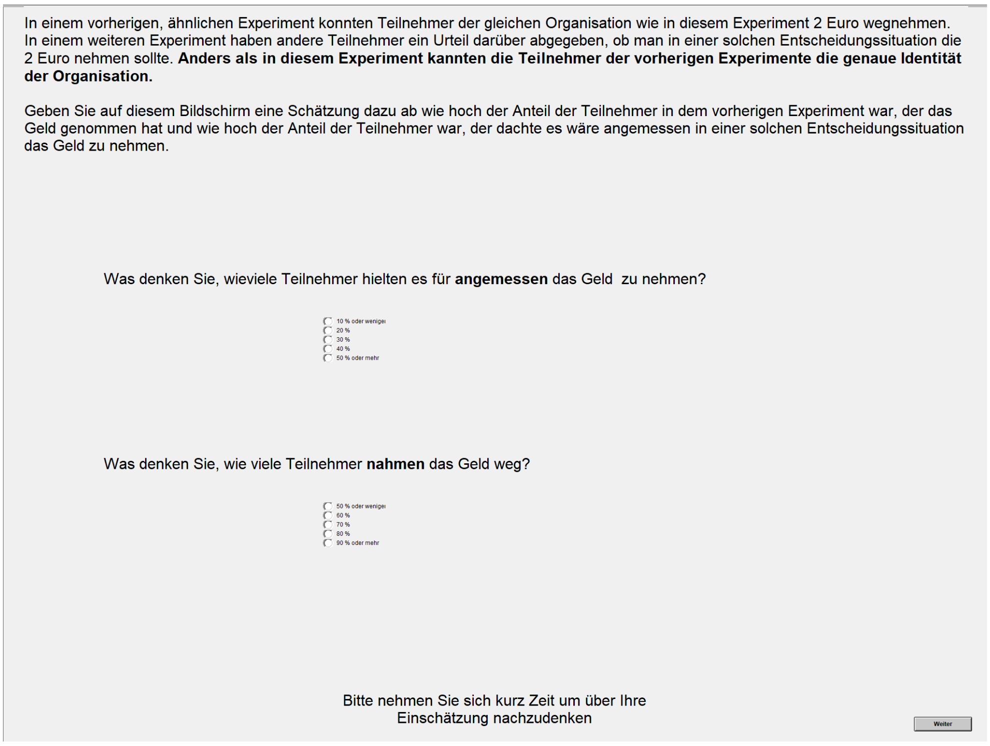
Instructions (translation)

In this experiment there are two roles, the role of the decision maker and the role of the judge. Your role will be randomly assigned to you and one decision maker will be randomly assigned to one judge.

Both the decision maker and the judge in this experiment initially receive a fixed payout of 1 euro. In addition, a nonprofit organization receives 2 euros for each group of judge and decision maker that participates in this experiment. Neither judge nor decision maker knows the exact identity of the organization.

Below you will learn the rules of the experiment and what role you have been assigned to.

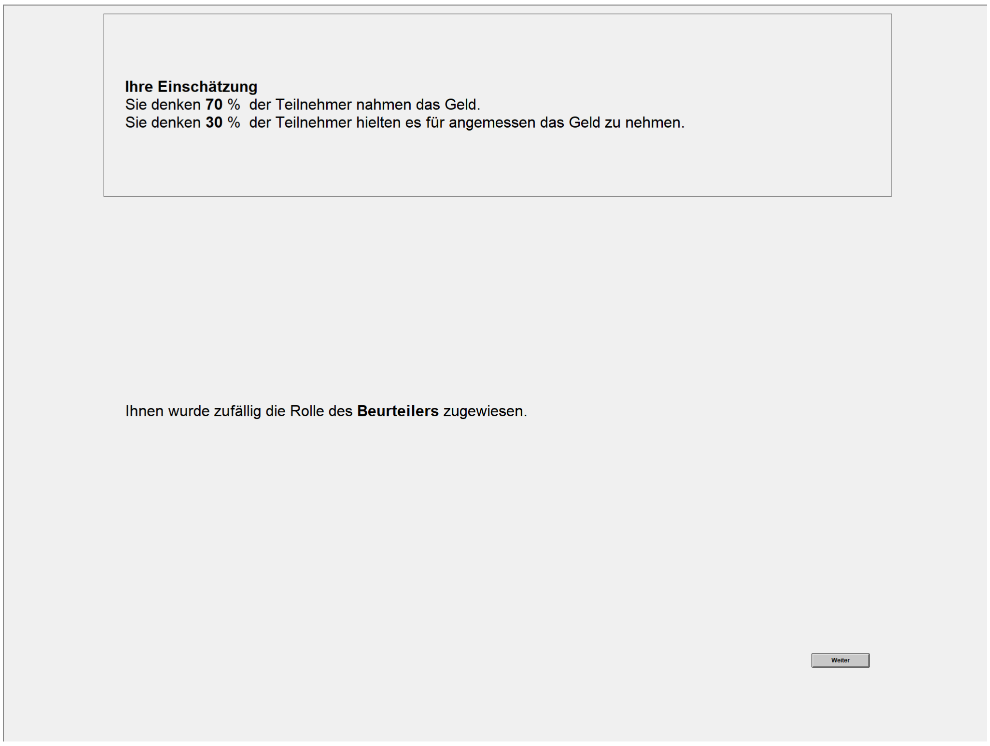
Screenshot 3. Norm Guess



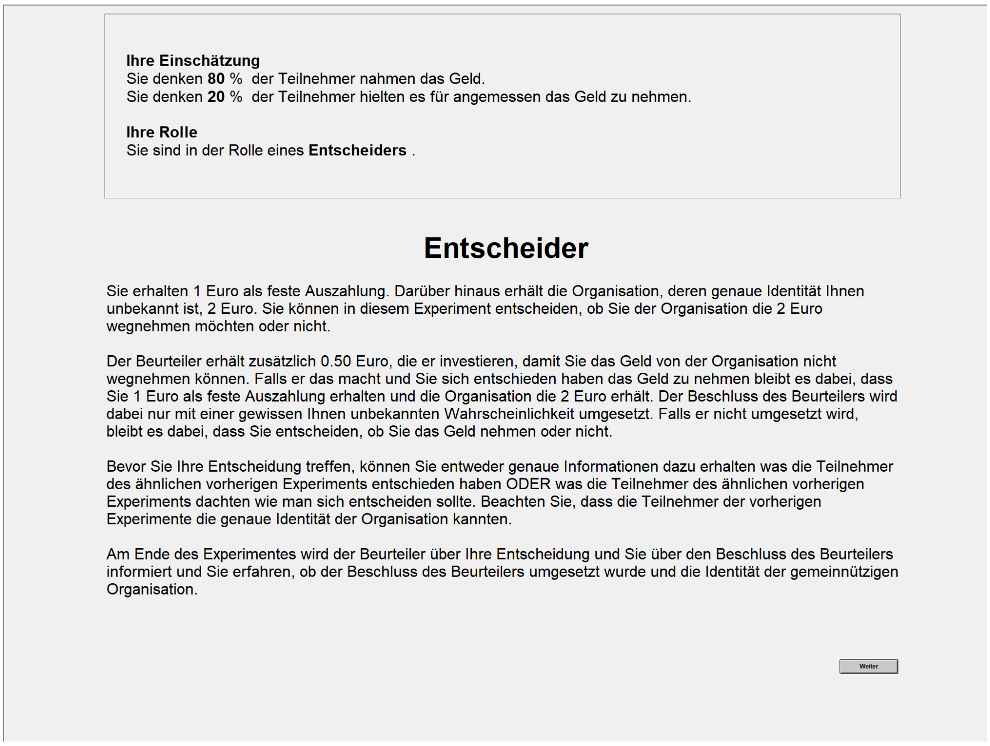
In a previous similar experiment, participants were able to take 2 euros away from the same organization as in this experiment. In another experiment, other participants made a judgment about whether to take the 2 euros in such a decision situation. Unlike in this experiment, participants in the previous experiments knew the exact identity of the organization.

On this screen, participants have to provide an estimate of what proportion of participants in the previous experiment took the money and what proportion of participants thought it would be appropriate in such a decision situation to take the money.

Screenshot 4. Role



Screenshot 5. Role Instructions (decision maker)



Role Instructions (translation)

You will receive 1 euro as a fixed payout. In addition, the organization whose exact identity is unknown to you receives 2 euros. In this experiment, you can decide whether or not to take the 2 euros away from the organization.

[DM_H; DM_L: Ihre Entscheidung wird dabei nur mit einer gewissen Wahrscheinlichkeit umgesetzt. Falls sie nicht umgesetzt wird, bleibt es dabei, dass Sie 1 Euro als feste Auszahlung erhalten und die Organisation die 2 Euro erhält.

Your decision will be implemented with a probability of [DM_H: 99%] [DM_L: 1%].

Before making your decision, you can either get accurate information about what the participants of the similar previous experiment decided OR what the participants of the similar previous experiment thought how to decide. Note that the participants in the previous experiments knew the exact identity of the organization.

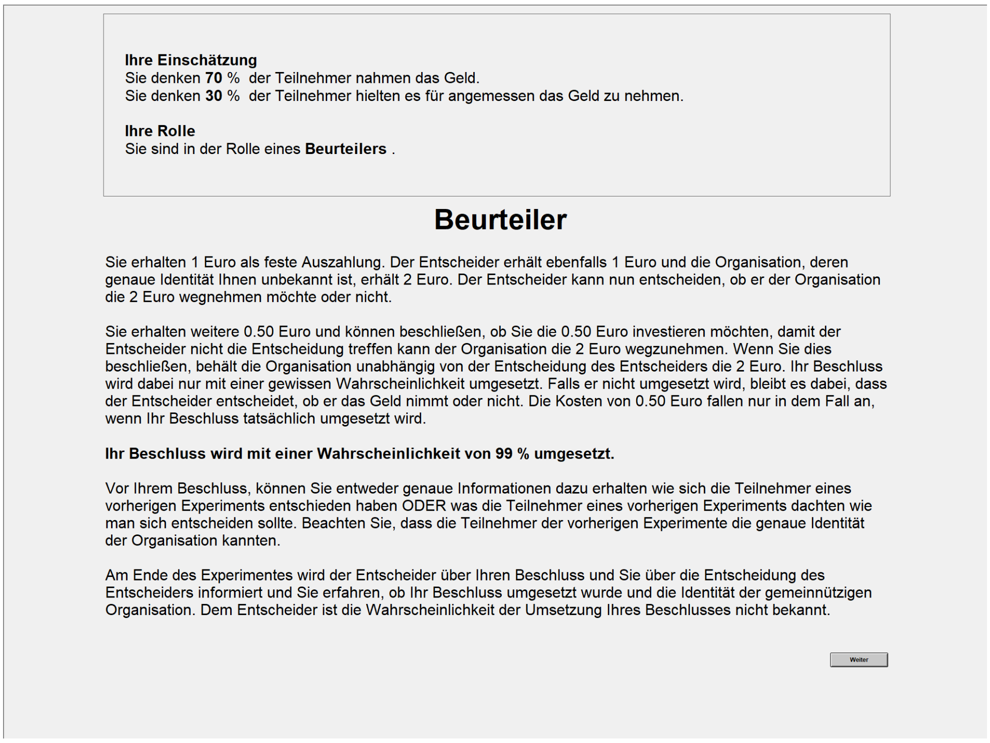
At the end of the experiment, the judge is informed of your decision and you learn whether your decision was implemented and the identity of the nonprofit organization. The judge is not aware of the probability of your decision being implemented.]

[CA_H; CAL_L: The judge will receive an additional 0.50 euro to invest so that you cannot take the money away from the organization. If he does that and you decided to take the money it remains that you get 1 euro as a fixed payout and the organization gets the 2 euro. The decision of the judge is thereby implemented only with a certain probability unknown to you. If it is not implemented, it remains that you decide whether to take the money or not.

Before making your decision, you can either get accurate information about what the participants of the similar previous experiment decided OR what the participants of the similar previous experiment thought how to decide. Note that the participants in the previous experiments knew the exact identity of the organization.

At the end of the experiment, the judge is informed of your decision and you are informed of the judge's decision, and you learn whether the judge's decision was implemented and the identity of the nonprofit organization.]

Screenshot 6. Role Instructions (Judge)



Role Instructions (translation)

[DM_H;DM_L: In the following, you can observe how the decision maker assigned to you decides. The decision of the decision maker will be implemented with a probability unknown to you. If it is not implemented, the decision maker will still receive 1 euro as a fixed payout and the organization will receive 2 euros.

You can either get exact information about how the participants of the previous similar experiment decided OR what the participants of the previous similar experiment thought how to decide. Note that the participants of the previous experiments knew the exact identity of the organization.

At the end of the experiment, you will learn whether the decision maker's decision was implemented and the identity of the nonprofit organization.]

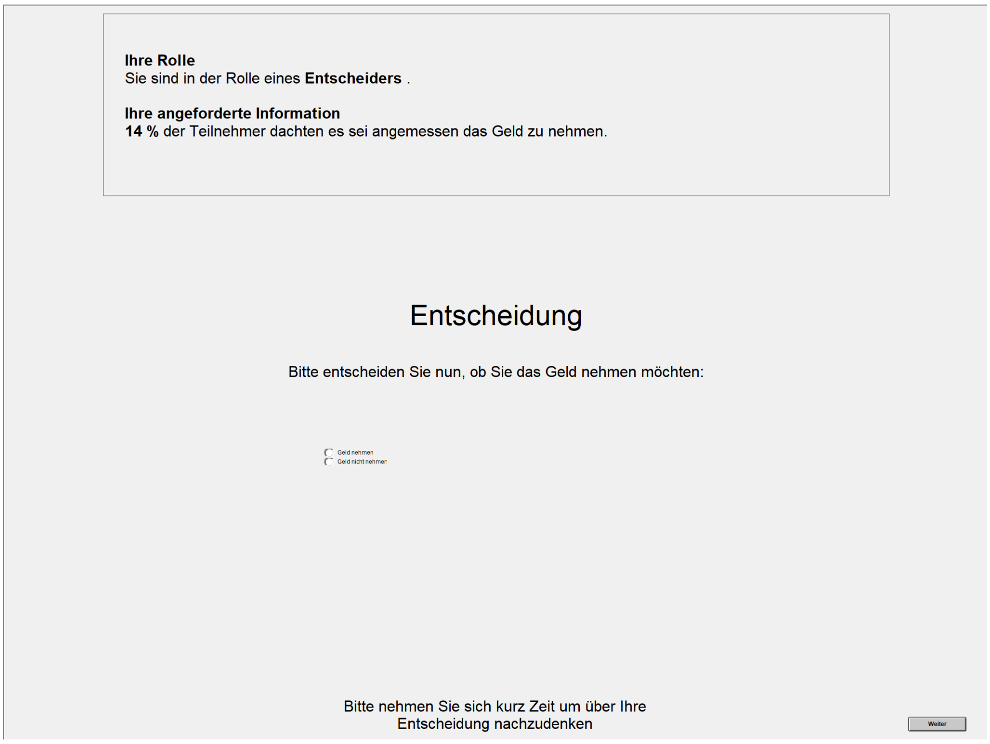
[CA_H;CAL_L: You will receive another 0.50 euro and can decide if you want to invest the 0.50 euro so that the decision maker cannot make the decision to take the 2 euros away from the organization. If you decide to do so, the organization will keep the 2 euros regardless of the decision of the decision maker. Your decision will only be implemented with a certain probability. If it is not implemented, it remains that the decision maker decides whether to take the money or not. The cost of 0.50 euro is incurred only in the case when your resolution is actually implemented.

Your decision will be implemented with a probability of [CA_H: 99%] [CA_L: [1%].

Prior to your decision you can either get accurate information about how the participants in a previous experiment decided OR what the participants in a previous experiment thought how to decide.". Note that participants in previous experiments knew the exact identity of the organization.

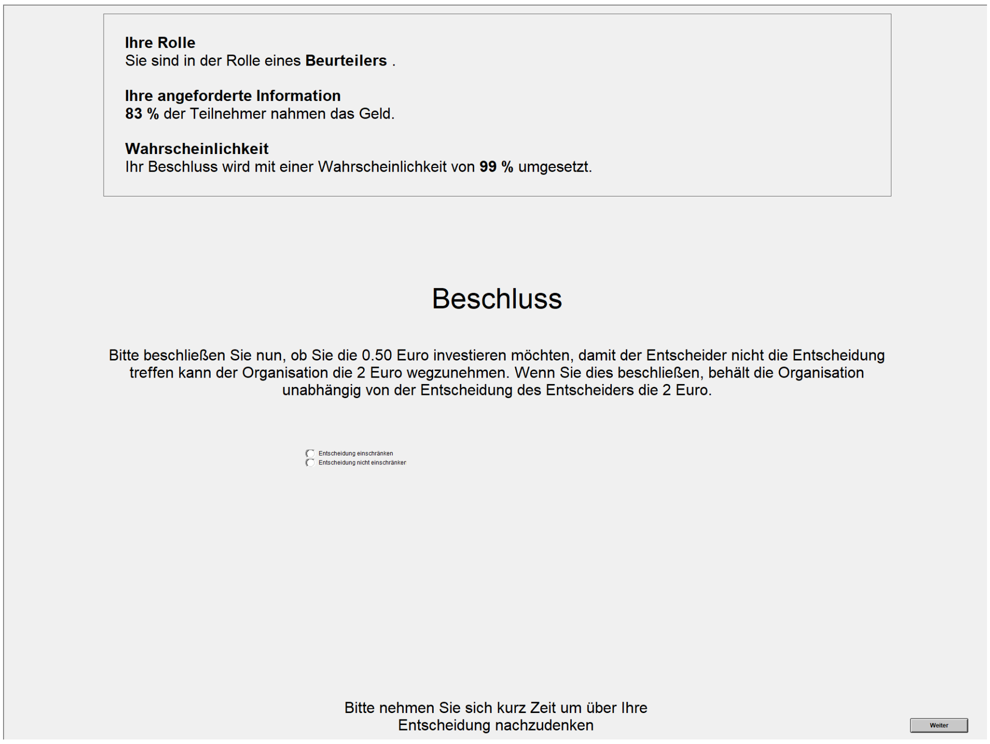
At the end of the experiment, the decision maker is informed about your decision and you about the decision maker's decision, and you learn whether your decision was implemented and the identity of the nonprofit organization. The decision maker is not aware of the probability of your decision being implemented.]

On this screen, participants can now decide whether they want to get accurate information about what the participants of the similar previous experiment decided OR what the participants of the similar previous experiment thought how one should decide.

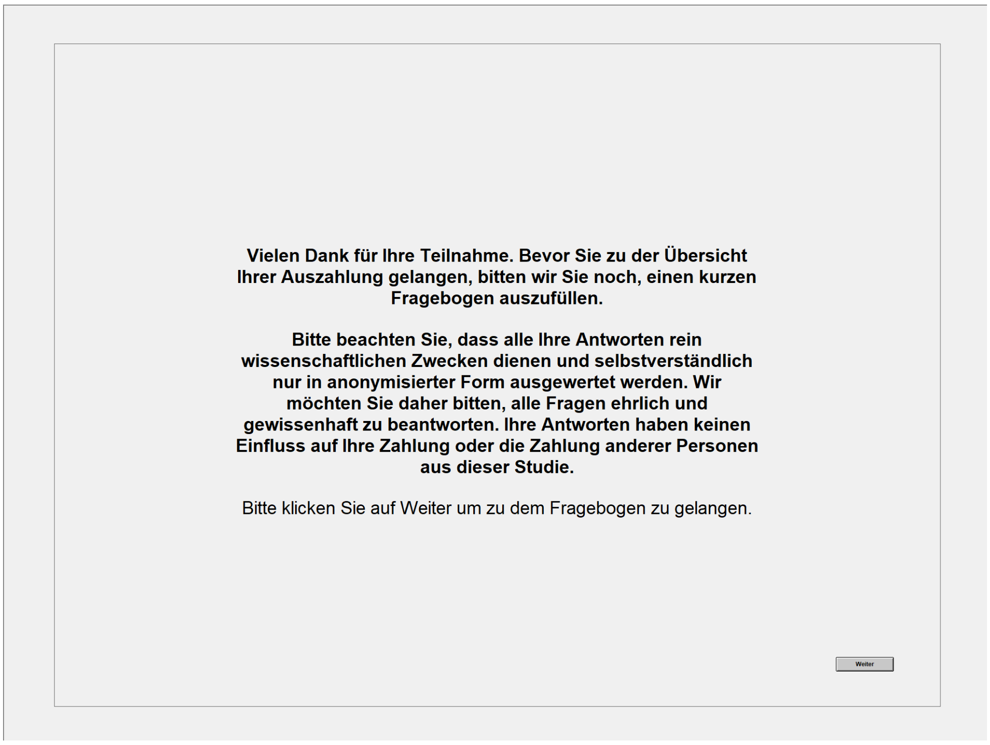


The decision maker sees the requested information and can decide between taking the money or not.

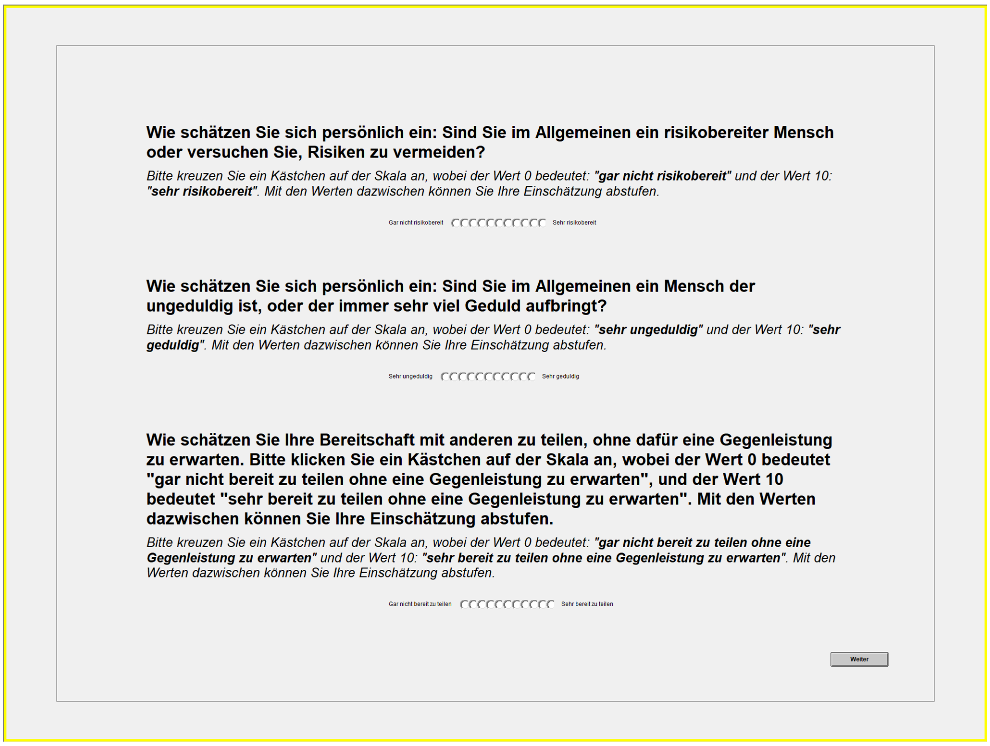
Screenshot 9. Norm Choice Result and Decision (Judge, only in CA_H and CAL_L)



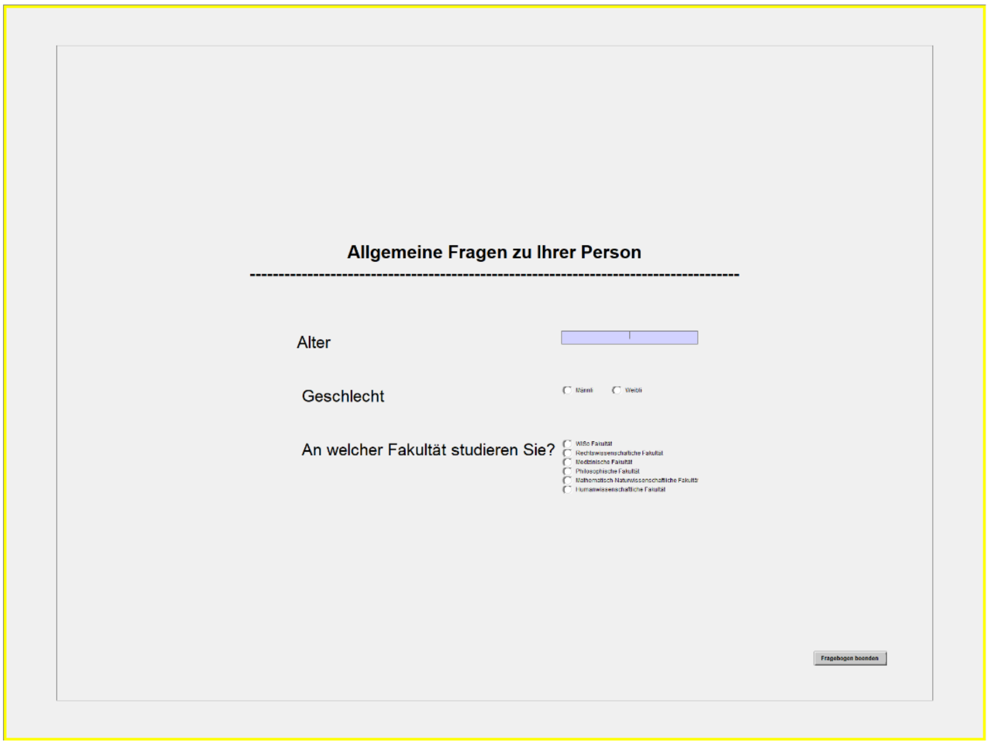
The judge sees the requested information and can decide to invest 0.50 euro so that the decision maker cannot make the decision to take the 2 euros away from the organization.



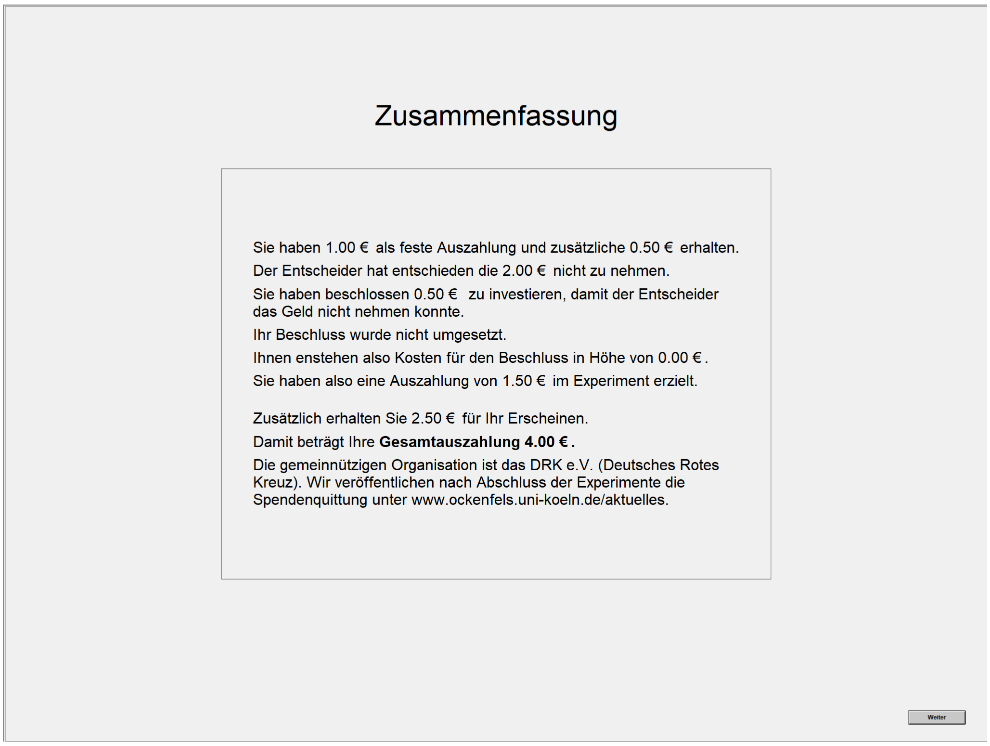
This screen announces a short questionnaire.



The questionnaire asks to assess yourself (from 1 to 10) regarding the willingness to take risks, the level of impatience and the willingness to share something without expecting anything in return.



The payoff of the decision maker consists of endowment (1 euro), decision to take the 2 euros or not (if decision was implemented) and show up fee (2.50 euros).



The payoff of the judge consists of the endowment (1 + 0.50 euro), the decision to invest the 0.50 euro or not (if decision was implemented) and the show up fee (2.50 euros).

References

Abdulkadiroğlu, A., Che, Y.-K., & Yasuda, Y. (2015). Expanding “choice” in school choice. *American Economic Journal: Microeconomics*, *7(1)*, 1–42.

Abdulkadiroğlu, A., Pathak, P.A., Roth, A.E., & Sönmez, T. (2005). The Boston public school match. *American Economic Review, 95(2),* 368–371.

Abdulkadiroğlu, A., Pathak, P.A., & Roth, A.E. (2005). The New York City high school match. *American Economic Review, 95(2)*, 364–367.

Abdulkadiroğlu, A., & Sönmez, T. (2003). School choice: A mechanism design approach. *American Economic Review, 93(3)*, 729–747.

Ackfeld, V., & Ockenfels, A. (2021). Do people intervene to make others behave prosocially? *Games and Economic Behavior*, *128*, 58-72.

Alcalde, J., & Romero-Medina, A. (2017). Fair student placement. *Theory and Decision, 83*, 1–15.

Allcott, H. (2011). Social norms and energy conservation. *Journal of Public Economics, 95 (9)*, 1082-1095.

Almås, I., Cappelen, A. W., & Tungodden, B. (2020). Cutthroat capitalism versus cuddly socialism: Are Americans more meritocratic and efficiency-seeking than Scandinavians? *Journal of Political Economy, 128(5)*, 1753-1788.

Ambuehl, S. (2017). An offer you can’t refuse? Incentives change how we inform ourselves and what we believe. CESifo Working Papers, (6296).

Ambuehl, S., Bernheim, B. D., & Ockenfels, A. (2021). What motivates paternalism? An experimental study. *American Economic Review*, *111*(3), 787-830.

Andreoni, J., & Bernheim, B. D. (2009). Social image and the 50–50 norm: A theoretical and experimental analysis of audience effects. *Econometrica, 77(5)*, 1607-1636.

Andreoni, J., & Miller, J. (2002). Giving according to GARP: An experimental test of the consistency of preferences for altruism. *Econometrica*, *70*(2), 737-753.

Andreoni, J., & J.-K. Scholz (1998). An econometric analysis of charitable giving with interdependent preferences. *Economic Inquiry, 36(3)*, 410-428.

Andreoni, J., Rao, J. M., & Trachtman, H. (2017). Avoiding the ask: A field experiment on altruism, empathy, and charitable giving. *Journal of Political Economy, 125(3)*, 625-653.

Ariely, D., Ockenfels, A., & Roth, A.E. (2005). An experimental analysis of ending rules in internet auctions. *RAND Journal of Economics, 36*, 890–907.

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

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

Bajari, P., & Hortacsu, A. (2003). The winner's curse, reserve prices, and endogenous entry: Empirical insights from eBay auctions. *Rand Journal of Economics, 34(2)*, 329-355.

Balinski, M., & Sönmez, T. (1999). A tale of two mechanisms: Student placement. *Journal of Economic Theory, 84(1)*, 73–94.

Bergemann, D., Breuer, K., Cramton, P., & Ockenfels, A. (2021). Why are soft reserve auctions so popular? *Working Paper*.

Bicchieri, C. (2005). The Grammar of Society: The nature and dynamics of social norms. Cambridge University Press.

Bicchieri, C., Dimant, E., & Sonderegger, S. (2019). It's not a lie if you believe it: On norms, lying, and self-serving belief distortion (No. 2019-07). CeDEx Discussion Paper Series.

Bicchieri, C., Dimant, E., & Xiao, E. (2021). Deviant or wrong? The effects of norm information on the efficacy of punishment. *Journal of Economic Behavior & Organization, 188(5970)*, 209-235.

Bicchieri, C., & Xiao, E. (2009). Do the right thing: but only if others do so. *Journal of Behavioral Decision Making, 22(2)*, 191-208.

Bierbrauer, F., & Netzer, N. (2016). Mechanism design and intentions. *Journal of Economic Theory, 163*, 557–603.

Bierbrauer, F., Ockenfels, A., Pollak, A., & Rückert, D. (2017). Robust mechanism design and social preferences. *Journal of Public Economics, 149*, 59–80.

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

Bohnet, I., Harmgart, H., Huck S., & Tyran, J.-R. (2005). Learning trust. *Journal of the European Economic Association, 3(2-3)*, 322–329.

Bolton, G., Greiner, B., & Ockenfels, A. (2018). Dispute resolution or escalation? The strategic gaming of feedback withdrawal options in online markets. *Management Science, 64(9)*, 4009–4031.

Bolton, G., Greiner, B., & Ockenfels, A. (2012). Engineering trust: Reciprocity in the production of reputation information. *Management Science, 59(2)*, 265–285.

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)*, 1–33.

Bolton, G., Katok, E., & Ockenfels, A. (2004). How effective are electronic reputation mechanisms? An experimental investigation. *Management Science, 50(11)*, 1587-1602.

Bolton, G. E., Mans, J., & Ockenfels, A. (2020). Norm enforcement in markets: Group identity and the volunteering of feedback. *Economic Journal*, 130(629), 1248-1261.

Bolton, G.E., & Ockenfels, A. (2012). Behavioral economic engineering. *Journal of Economic Psychology, 33(3)*, 665–676.

Bolton, G. E., & Ockenfels, A. (2000). ERC: A theory of equity, reciprocity, and competition. *American Economic Review, 90(1)*, 166-193.

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.

Brown, M., & C. Zehnder (2007). Credit reporting, relationship banking, and loan repayment. *Journal of Money, Credit, and Banking, 39(8)*, 1884–1918.

Burguet, R., & Sákovics, J. (1996). Reserve Prices without commitment. *Games and Economic Behavior, 15(2)*, 149–164.

Caillaud, B., & Mezzetti, C. (2004). Equilibrium reserve prices in sequential ascending auctions. *Journal of Economic Theory*, *117(1)*, 78–95.

Charness, G., Du, N., & Yang, C-L. (2011). Trust and trustworthiness reputation in an investment game. *Games and Economic Behavior, 72(2)*, 361–375.

Charness, G., & Sutter, M. (2012). Groups make better self-interested decisions. *Journal of Economic Perspectivesm, 26(3)*, 157–176.

Chen, Y., Cramton, P., List, J. A., & Ockenfels, A. (2021). Market design, human behavior, and management. *Management Science*, *67*(9), 5317-5348.

Chen, Y., & Kesten, O. (2017). Chinese college admissions and school choice reforms: a theoretical analysis. *Journal of Political Economy, 125(1)*, 99–139.

Chen, Y. & Möller, M. (2021). Regret-Free Truth-telling in School Choice with Consent (SSRN Scholarly Paper No. ID 3896306). Social Science Research Network, Rochester, NY.

Cherry, T.L., Frykblom, P., & Shogren, J.F. (2002). Hardnose the dictator. *American Economic Review, 92(4)*, 1218–1221.

Cialdini, R.B., Kallgren, C.A., & Reno, R.R. (1991). A focus theory of normative conduct: A theoretical refinement and reevaluation of the role of norms in human behavior, in: Zanna, M.P. (Ed.), Advances in Experimental Social Psychology. Academic Press, pp. 201-234.

Cialdini, R.B., Reno, R.R., & Kallgren, C.A. (1990). A focus theory of normative conduct: Recycling the concept of norms to reduce littering in public places. *Journal of Personality and Social Psychology, 58(6)*, 1015-1026.

Cooper, D.J., & Kagel, J.H. (2005). Are two heads better than one? Team versus individual play in signaling games. *American Economic Review, 95(3)*, 477–509.

Cox, J.C., Roberson, B., & Smith, V.L. (1982). Theory and behavior of single object auctions. *Research in Experimental Economics, 2*, 1–43.

Cramton, P. (1998). Ascending auctions. *European Economic Review, 42(3-5)*, 745-756.

Crawford, V.P., Kugler, T., Neeman, Z., & Pauzner, A. (2009). Behaviorally optimal auction design: examples and observations. *Journal of the European Economic Association, 7(2-3)*, 377–387.

Dana, J., Weber, R. A., & Kuang, J. X. (2007). Exploiting moral wiggle room: experiments demonstrating an illusory preference for fairness. *Economic Theory, 33(1)*, 67-80.

Davis, A.M., Katok, E., & Kwasnica, A.M. (2011). Do auctioneers pick optimal reserve prices? *Management Science, 57(1)*, 177–192.

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

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.

Duffy, J., & Feltovich, N. (2006). Words, deeds and lies: Strategic behavior in games with multiple signals. *Review of Economic Studies, 73(3)*, 669–688.

Duffy, J., Xie, H., & Lee, Y-J. (2013). Social norms, information and trust among strangers: theory and evidence. *Economic Theory, 52(2)*, 669–708.

Dufwenberg, M., & Gneezy, U. (2002). Information disclosure in auctions: an experiment. *Journal of Economic Behavior & Organization, 48(4)*, 431-444.

Dur, U., Gitmez, A.A., & Yılmaz, Ö. (2019). School choice under partial fairness. *Theoretical Economics, 14(4)*, 1309–1346.

Duch, M. L., Grossmann, M. R., & Lauer, T. (2020). z-Tree unleashed: A novel client-integrating architecture for conducting z-Tree experiments over the Internet. *Journal of Behavioral and Experimental Finance*, *28*, 100400.

Ehlers, L., & Erdil, A. (2010). Efficient assignment respecting priorities. *Journal of Economic Theory, 145(3)*, 1269–1282.

Ehlers, L., & Morrill, T. (2019). (Il)legal Assignments in school choice. *Review of Economic Studies, 87(4)*, 1837-1875.

Ellingsen, T., Johannesson, M., Mollerstrom, J., & Munkhammar, S. (2012). Social framing effects: Preferences or beliefs? *Games and Economic Behavior, 76(1)*, 117–130.

Engelbrecht-Wiggans, R. (1989). The effect of regret on optimal bidding in auctions. *Management Science, 35(6)*, 685–692.

Engelbrecht-Wiggans, R., Haruvy, E., & Katok, E. (2007). A comparison of buyer-determined and price-based multiattribute mechanisms. *Marketing Science, 26(5)*, 629–641.

Engelbrecht-Wiggans, R., & Katok, E. (2006). Regret in auctions: theory and evidence. *Economic Theory, 33*, 81–101.

Engelbrecht-Wiggans, R., & Katok, E. (2008). Regret and feedback information in first-price sealed-bid auctions. *Management Science, 54(4)*, 808–819.

Engelbrecht-Wiggans, R., & Katok, E. (2009). A direct test of risk aversion and regret in first price sealed-bid auctions. *Decision Analysis, 6(2)*, 75–86.

Erdil, A., & Ergin, H. (2008). What's the matter with tie-breaking? Improving efficiency in school choice. *American Economic Review, 98(3)*, 669–689.

Ergin, H.I. (2002). Efficient resource allocation on the basis of priorities. *Econometrica, 70(6)*, 2489–2497.

Exley, C. L. (2016). Excusing selfishness in charitable giving: The role of risk. *Review of Economic Studies, 83(2)*, 587-628.

Falk, A., Becker, A., Dohmen, T., Enke, B., Huffman, D., & Sunde, U. (2018). Global evidence on economic preferences. *Quarterly Journal of Economics*, *133(4)*, 1645-1692.

Falk, A., Neuber, T., & Szech, N. (2020). Diffusion of being pivotal and immoral outcomes. *Review of Economic Studies*, *87(5),* 2205-2229.

Fehr, E., & Schmidt, K. M. (1999). A theory of fairness, competition, and cooperation. *Quarterly Journal of Economics*, *114(3)*, 817-868.

Feldhaus, C., Sobotta, T., & Werner, P. (2019). Norm uncertainty and voluntary payments in the field. *Management Science*, *65(4)*, 1855-1866.

Fernandez, M.A. (2020). Deferred acceptance and regret-free truth-telling. *Working Paper.*

Filiz-Ozbay, E., & Ozbay, E.Y. (2007). Auctions with anticipated regret: Theory and experiment. *American Economic Review, 97(4)*, 1407–1418.

Fischbacher, U. (2007). z-Tree: Zurich toolbox for ready-made economic experiments. *Experimental economics*, *10(2)*, 171-178.

Fiske, S. T., & Taylor, S. E. (2008). Social cognition: From brains to culture. McGraw-Hill, New York.

Fisman, R., Kariv, S., & Markovits, D. (2007). Individual preferences for giving. *American Economic Review*, *97(5)*, 1858-1876.

Foerster, M., & van der Weele, J. J. (2021). Casting doubt: Image concerns and the communication of social impact. *Economic Journal, 131*, 2887-2919.

Frey, B. S., & Meier, S. (2004). Social comparisons and pro-social behavior: Testing "conditional cooperation" in a field experiment. *American Economic Review, 94(5)*, 1717-1722.

Gale, D., & Shapley, L.S. (1962). College admissions and the stability of marriage. *American Mathematical Monthly, 69(1)*, 9–15.

Goette, L., Huffman, D., Meier, S., & Sutter, M. (2012). Competition between organizational groups: Its impact on altruistic and antisocial motivations. *Management Science, 58(5)*, 948–960.

Greenleaf, E.A. (2004). Reserves, regret, and rejoicing in open english auctions. *Journal of Consumer Research*, *31(2)*, 264–273.

Greiner, B. (2015). Subject pool recruitment procedures: organizing experiments with ORSEE. *Journal of the Economic Science Association*, 1(1), 114-125.

Hallsworth, M., J. A. List, R. D. Metcalfe, & Vlaev, I. (2017). The behavioralist as tax collector: Using natural field experiments to enhance tax compliance. *Journal of Public Economics, 148(C)*, 14-31.

Hakimov, R., & Kesten, O. (2018). The equitable top trading cycles mechanism for school choice. *International Economic Review*, 59(4), 2219-2258.

Huang, C., Chen, J., & Lee, C. (2013). Buyer behavior under the best offer mechanism: a theoretical model and empirical evidence from eBay Motors. *Journal of Economic Behavior and Organization, 94*, 11-33.

Huck, S., Lünser, G. K., & Tyran, J.-R., (2010). Consumer networks and firm reputation: A first experimental investigation. *Economics Letters, 108(2)*, 242–244.

Huck, S., Lünser, G., & Tyran, J.-R., (2012). Competition fosters trust. *Games and Economic Behavior, 76(1)*, 195–209.

Isaac, R. M., & Walker, J. M. (1985). Information and conspiracy in sealed bid auctions. *Journal of Economic Behavior & Organization, 6(2)*, 139-159.

Jacobson, R.P., Mortensen, C.R., & Cialdini, R.B. (2011). Bodies obliged and unbound: Differentiated response tendencies for injunctive and descriptive social norms. *Journal of Personality and Social Psychology, 100(3)*, 433-448.

Kagel, J.H., & Levin, D. (2016). Auctions a survey of experimental research, in: Kagel, J.H., Roth, A.E. (Eds.), The Handbook of Experimental Economics, Volume Two. Princeton University Press, Princeton.

Kassas, B., & Palma, M. A. (2019). Self-serving biases in social norm compliance. *Journal of Economic Behavior & Organization*, *159(C)*, 388-408.

Kesten, O. (2010). School choice with consent. *Quarterly Journal of Economics, 125(3)*, 1297–1348.

Korenok, O., Millner, E., & Razzolini, L. (2017). Feelings of ownership in dictator games. *Journal of Economic Psychology, 61*, 145–151.

Li, S. (2017). Obviously strategy-proof mechanisms. *American Economic Review, 107(11)*, 3257–3287.

List, J. A., & Lucking-Reiley, D. (2002). The effects of seed money and refunds on charitable giving: Experimental evidence from a university capital campaign. *Journal of Political Economy, 110(1)*, 215-233.

Macrae, C. N., & Bodenhausen, G.V. (2000). Social cognition: Thinking categorically about others. *Annual Review of Psychololgy*, 51(1), 93–120.

Morrill, T. (2015). Making just school assignments. *Games and Economic Behavior, 92*, 18–27.

Mussweiler, T., & Ockenfels, A. (2013). Similarity increases altruistic punishment in humans. *Proceedings of the National Academy of Sciences, 110(48),* 19318-19323.

Myerson, R.B. (1981). Optimal auction design. *Mathematics of Operations Research*, 6(1), 58-73.

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.

Ockenfels, A., & Werner, P. (2014). Scale manipulation in dictator games. *Journal of Economic Behavior & Organization*, *97*, 138-142.

Ockenfels, A., & Selten, R. (2005). Impulse balance theory and feedback in first price auctions. *Games and Economic Behavior, 51(1)*, 155-170.

Paetzold, J., & Winner, H. (2016). Taking the high road? compliance with commuter tax allowances and the role of evasion spillovers. *Journal of Public Economics, 143(C)*, 1-14.

Pezanis-Christou, P. (2002). On the impact of low-balling: Experimental results in asymmetric auctions. *Game Theory, 31*, 69–89.

Rege, M., & Telle, K. (2004). The impact of social approval and framing on cooperation in public good situations. *Journal of Public Economics, 88(7)*, 1625-1644.

Riley, J.G., & Samuelson, W.F. (1981). Optimal auctions. *American Economic Review, 71(3)*, 381–392.

Rosenkranz, S., & Schmitz, P.W. (2007). Reserve prices in auctions as reference points. *Economic Journal, 117(520)*, 637–653.

Schmelzer, A. (2016). Single versus multiple randomization in matching mechanisms (SSRN Scholarly Paper No. ID 2777405). Social Science Research Network, Rochester, NY.

Shachat, J., & Wei, L. (2012). Procuring commodities: First-price sealed-bid or english auctions? *Marketing Science, 31*(2), 317–333.

Shapley, L., & Scarf, H. (1974). On cores and indivisibility. *Journal of Mathematical Economics, 1*(1), 23–37.

Spiekermann, K., & Weiss, A. (2016). Objective and subjective compliance: A norm-based explanation of 'moral wiggle room'. *Games and Economic Behavior, 96*, 170-183.

Shavell, S. (1995). Alternative dispute resolution: An economic analysis. *Journal of Legal Studies, 24(1)*, 1–28.

Skreta, V. (2015). Optimal auction design under non-commitment. *Journal of Economic Theory, 159(B)*, 854–890.

Tadelis, S. (2016). Reputation and feedback systems in online platform markets. *Annual Review of Economics, 8*(1), 321–340.

Troyan, P., Delacrétaz, D., & Kloosterman, A. (2020). Essentially stable matchings. *Games and Economic Behavior, 120,* 370–390.

Troyan, P., & Morrill, T. (2020). Obvious manipulations. *Journal of Economic Theory, 185*, 104970.

Tversky, A., & Kahneman, D. (1974). Judgment under Uuncertainty: Heuristics and biases. *Science, 185*(4157), 1124–1131.

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*(6), 466–480.

Wenzel, M. (2005). Misperceptions of social norms about tax compliance: From theory to intervention. *Journal of Economic Psychology*, *26*(6), 862-883.

Zeithammer, R. (2019). Soft Floors in Auctions. *Management Science, 65*(9), 4204–4221.

Eidesstattliche Erklärung
nach § 8 Abs. 3 der Promotionsordnung vom 17.02.2015

"Hiermit versichere ich an Eides Statt, dass ich die vorgelegte Arbeit selbstständig und ohne die Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe. Die aus anderen Quellen direkt oder indirekt übernommenen Aussagen, Daten und Konzepte sind unter Angabe der Quelle gekennzeichnet. Bei der Auswahl und Auswertung folgenden Materials haben mir die nachstehend aufgeführten Personen in der jeweils beschriebenen Weise entgeltlich/unentgeltlich (zutreffendes unterstreichen) geholfen:

Weitere Personen, neben den ggf. in der Einleitung der Arbeit aufgeführten Koautorinnen und Koautoren, waren an der inhaltlich-materiellen Erstellung der vorliegenden Arbeit nicht beteiligt. Insbesondere habe ich hierfür nicht die entgeltliche Hilfe von Vermittlungs- bzw. Beratungsdiensten in Anspruch genommen. Niemand hat von mir unmittelbar oder mittelbar geldwerte Leistungen für Arbeiten erhalten, die im Zusammenhang mit dem Inhalt der vorgelegten Dissertation stehen.

Die Arbeit wurde bisher weder im In- noch im Ausland in gleicher oder ähnlicher Form einer anderen Prüfungsbehörde vorgelegt.

Ich versichere, dass ich nach bestem Wissen die reine Wahrheit gesagt und nichts verschwiegen habe.

Ich versichere, dass die eingereichte elektronische Fassung der eingereichten Druckfassung vollständig entspricht.

Die Strafbarkeit einer falschen eidesstattlichen Versicherung ist mir bekannt, namentlich die Strafandrohung gemäß § 156 StGB bis zu drei Jahren Freiheitsstrafe oder Geldstrafe bei vorsätzlicher Begehung der Tat bzw. gemäß § 161 Abs. 1 StGB bis zu einem Jahr Freiheitsstrafe oder Geldstrafe bei fahrlässiger Begehung.

Köln, 08.01.2022

Ort, Datum

Unterschrift

Curriculum Vitae

Personal Details

Name	**Kevin Breuer**
Date/Place of Birth	16.06.1989 / Cologne, Germany
Adress	University of Cologne Department of Economics, Chair Professor Axel Ockenfels Albertus-Magnus-Platz, 50923 Cologne, Germany
E-Mail	kbreuer0@gmail.com

Education

10/2015 - present	**University of Cologne, Germany** Ph.D. in Economics
10/2012 - 09/2015	**University of Cologne, Germany** Master of Science in Economics Final Grade: 1.3 - Top three of Class Specializations: Economic Design & Behaviour, Markets and Institutions
03/2010 - 09/2012	**University of Cologne, Germany** Bachelor of Science in Economics Final Grade: 1.4 - Top ten of Class Specializations: Modern Concepts of Economic Theory, Energy Economics
09/2003 - 07/2008	**Knigin-Luise-Gymnasium, Germany** University Entrance Diploma Final Grade: 2.0 Specializations: Mathematics, Physics

Academic Positions

01/2018 - present	**Research Assistant, Department of Economics** Chair Prof. Axel Ockenfels, University of Cologne, Germany
02/2016 - 12/2017	**Research Assistant, DFG Research Unit (FOR 1371)** Design & Behavior - Economic Engineering of Firms and Markets
12/2015 - 09/2016	**Research Assistant, Department of Economics** Chair Prof. Patrick Schmitz, University of Cologne, Germany
10/2015 - 01/2016	**Ph.D. Fellow, Cologne Graduate School of Management, Economics and Social Science** Supervisor: Prof. Axel Ockenfels, University of Cologne, Germany
10/2013 - 09/2015	**Student Research Assistant, Department of Economics** Chair Prof. Axel Ockenfels, University of Cologne, Germany
08/2011 - 09/2013	**Student Research Assistant, DFG Research Unit (FOR 1371)** Design & Behavior - Economic Engineering of Firms and Markets